Developing an Automated Cytogenetic Imaging System for Detection of Dicentric Chromosomes in Biological Dosimetry

Sanaeian Pour Shirazi Z.[1], Zamani A.[1]*, Mortazavi S. M. J.[1,2], Zakeri F.[3], Dianatpour M.[4], Mosleh-Shirazi M. A.[5]

[1]Medical Physics and Medical Engineering Department, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[2]Ionizing and Non-ionizing Radiation Protection Research Center (INIRPRC), Shiraz University of Medical Sciences, Shiraz, Iran
[3]National Radiation Protection Department, Iranian Nuclear Regulatory Authority, Tehran, Iran.
[4]Genetic Department, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[5]Radiotherapy & Oncology Department, Namazi Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

ABSTRACT

Background: High-energy ionizing radiation is harmful and changes the genetic makeup of DNA, which can lead to increased risk of cancer. Thus, the exposure of radiation dose should be under control and limited. Ionizing radiation might lead to some chromosome aberrations like dicentric. There is a strong relation between the frequency of dicentric chromosome in metaphase, and the received dose.

Objective: Identifying the frequency of dicentric chromosomes is an important task especially in large-scale radiation accident for making rapid clinical decision. Given this, a large number of metaphases should be investigated for an accurate estimating of dose. It is well known that non-automated (visual) scoring of chromosome aberrations such as dicentrics requires highly trained experts. On the other hand, as thousands of cells should be scored, it is time-consuming and results in fatigue that can lead to poor concentration.

Methods: Biological dosimetry technique for assessing radiation dose is based on analyzing chromosome in peripheral blood lymphocyte. Our study describes a technique to speed up this procedure by automatically distinguishing abnormal chromosomes from normal ones. The most important feature of dicentric chromosomes is their two centromeres. Therefore, the main approach in our study is to design an automated system to identify the number and position of centromeres. We presented a method for classifying chromosomes into four groups in accordance with the number of centromeres in each chromosome.

Results: The image dataset of 311 chromosomes of normal and dicentric chromosomes is used to test the scheme. The sensitivity about 90% and specificity more than 95% for classification of chromosomes into dicentric and non-dicentric as well as accuracy more than 90% for centromere identification, are obtained.

Conclusion: If an automated system with the capability of straightening bent chromosome was applied prior to the proposed algorithm, even more accurate results would be achieved.

Keywords

Automated Scoring, Chromosome Aberrations, Biological Dosimetry, Dicentric Chromosomes, Image Processing

Introduction

Indirect dosimetry methods are applicable in poorly accessible radiation areas or when physical dosimetry does not provide sufficient information. High doses of ionizing radiation might change the genetic makeup of DNA, which can lead to increased risk of cancer. There-

*Corresponding author:
A. Zamani
Medical Physics and Medical Engineering Department, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
E-mail: zamani_a@sums.ac.ir

fore, the exposure of radiation dose should be controlled and limited. To this end, biological dosimetry is one of the most important procedures in estimating the absorbed dose. Ionizing radiation may lead to some chromosomal aberrations such as dicentric chromosomes. There is a strong relation between the frequency of dicentric chromosomes in metaphase and the received dose. Therefore, dicentric chromosome analysis (DCA) is a well-known method for biological dosimetry. This method is known as the "gold standard" in accurate individual dose estimation [1] . With this intention, a large number of metaphases (minimum of 500 metaphases or about 20000 chromosomes [2]) should be investigated for an accurate estimation of the dose. Manual scoring of dicentric chromosomes is time consuming, tedious, and requires a cytogenetic expert. Therefore, it is extremely expensive. In the present paper, a technique to facilitate this procedure by automatically distinguishing abnormal chromosomes from others is described. The most important feature of dicentric chromosomes is the presence of two centromeres. Accordingly, the main approach of the present study is to design an automated system to identify the number and position of centromeres. Centromere is the narrowest part of the chromosome where two sister chromatids are joined together. Finding the centromere position is an important feature which is essential for classifying chromosomes. This is a fundamental step for karyotyping (classification of 46 human chromosomes into a standard format called Karyogram). Four types of chromosomes classified by the position of centromeres include metacentric, submetacentric, acrocentric and telocentric [3].

The non-rigid nature of chromosomes makes it a difficult and complex task to identify centromere position. Numerous methods have been used and evaluated to extract suitable features for centromere detection. Among them, projection vectors and medial axis transformation (MAT) interested most researchers [4]. As mentioned previously, centromeres are the narrowest part of the chromosome. The centromere position can be identified by analysing the vertical and horizontal projection of the binary image of chromosomes to calculate the global minimum part of the projection [5, 6]. Unfortunately, this method is not suitable for highly bent chromosomes or acrocentric chromosomes. MAT is useful for the detection of centromeres because it gives the skeleton shape and describes shape properties such as the degree of concavity on boundaries that are noisy [7] and extract shape and density profiles to detect the lowest valley along the profile [8]. The latter method requires straight chromosomes and because it uses MAT for straightening, shape variants might produce spurious branches [9].

In the present study, we proposed an automated algorithm using a less computationally expensive technique to specify the class of chromosome based on the number of centromeres.

This paper is organized as follows: section 2 explains how the dataset is prepared, section 3 illustrates how the automated scheme works, section 4 expresses the resultant and finally section 5 ends with discussion.

Material and Methods

Image Acquisition and Database Preparation

A reliable data set with enough dicentric chromosomes for testing purposes was inaccessible. This prompted creation of an individual data set for the present study. For this purpose, fresh peripheral blood samples of healthy males and females were irradiated with 3.0 Gy Co60γ-rays at the SSDL (Secondary Standard Dosimetry Laboratory) of National Radiation Protection Department. Lymphocytes were separated from the irradiated blood and cultured in RPMI1640 containing 20% fatal calf serum for 48 h in the presence of 100 IU/ml Penicillin, 100 µg/ml Strepto-

mycin, PHA (0.2%) and Colcemid (0.05 µg /ml) (Gibco BRL) as previously described [10] . Cultured cells were treated with a KCl hypotonic solution (0.075M) at 37 °C for 20 minutes and fixed with acetic methanol (1:3). Air-dried slides were made under warm and humid conditions and stained with Giemsa. After preparation of a metaphase slide, images of well-spread metaphase were examined under microscope (Nikon Eclipse E600) and captured by (Digital camera DXM 1200F).

Approximately 80 metaphase images with dicentric aberrations were selected as the dataset. The input for the automated scheme was 311 single chromosomes, which were cropped manually.

Automated Scheme

A general scheme for automated dicentric chromosome detection is shown in Figure 1.

In the first step, pre-processing is applied to enhance image contrast. Then, the chromosomes are isolated from the background and the images are segmented. After that, suitable features for centromere localization are extracted. Chromosomes are further defined by counting the number of centromeres. Following the function of each block will be described in more details.

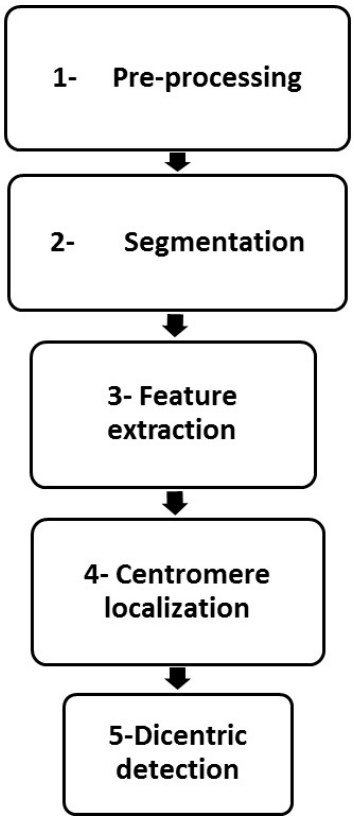

Figure 1: The algorithm for automated dicentric chromosome classification

Pre-processing

To improve the quality of chromosome images and unevenly distributed intensity [11], applying pre-processing to improve contrast and reducing inhomogeneity of chromosome images are necessary. Firstly, a median filter is applied to smooth the image, reduce salt and pepper noise, and preserve edge information. Next, adaptive histogram equalization is used to stretch the range of grey scale in histogram and increase image contrast. Figure 2 shows a sample of a main image and its histogram, and the result of adaptive histogram equalization.

Segmentation

Several methods have been studied for chromosome image segmentation [11, 12, 13, 14]. The proposed algorithm segments the chromosome images based on an Active Contour Model (ACM) [15] followed by edge detection. Active contour models use an initial framework in an attempt to minimize the energy inside and outside the contour, which moves the contour toward the outline of an object. Extensive studies have been conducted on this subject. Among them, a recent research by Zhang et al. proposed a novel (ACM) with special processing methods. By presenting a Signed Pressure Force (SPF) function (Equation 1), which adjusts the sign of pressure force inside and outside a Region of Interest (ROC); this increases both efficiency and accuracy compared to previous works.

$$SPF\left(I(x)\right)=\frac{I(x)-\frac{c1+c2}{2}}{Max\left(\left|I(x)-\frac{c1+c2}{2}\right|\right)} \qquad (1)$$

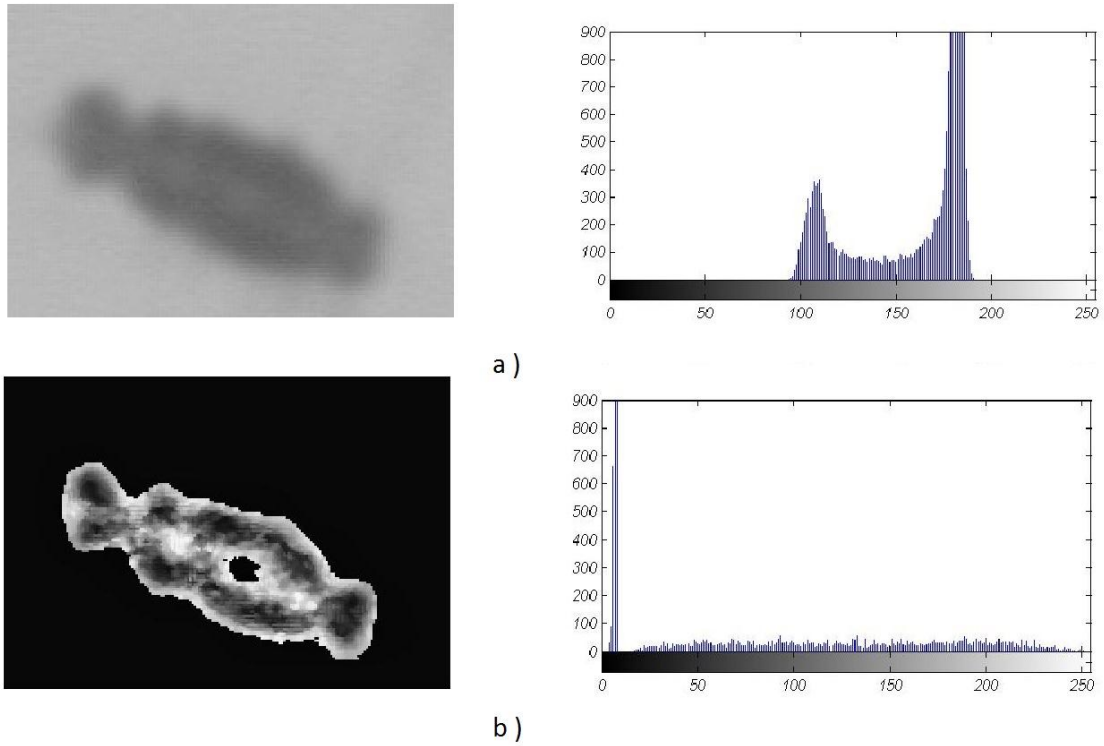

Figure 2: Image enhancement a) initial image and its histogram b) after histogram equalization

Where c1 and c2 are the average intensity of the inside and outside of the contour and I(x) is the input image. This function constrains the contour to shrink when it is outside (ROC) and expand otherwise. In the proposed algorithm, to extract the outline of the chromosome boundary, some modifications on the method proposed by Zhang et al. are applied [16]. By modifying the (SPF) function as defined in Equation 2, and σ=4 (standard deviation of Gaussian filter; Equation 3)), which has high impact on weak or blurred edges, satisfactory segmentation is achieved.

$$SPF\left(I(x)\right)=\frac{I(x)-\frac{c1+c2}{1.8}}{Max\left(\left|I(x)-\frac{c1+c2}{1.8}\right|\right)} \qquad (2)$$

$$G(X)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{x^2}{2\sigma^2}} \qquad (3)$$

This set of values is selected empirically, and provides excellent results for segmentation. To detect edges of binary images (where there is a big jump in intensity), the Canny technique, one of the most robust image processing tools for edge detection, is performed. As shown in Figure 3, these stages provide acceptable results.

Feature Extraction

Highly distinctive feature vector with the least possible length, maximize the classification performance. Due to the computational complexity in any classification procedure, efficient feature vector selection is a challenging task.

Centromeres have two unique characteristics:

- It is the thinnest region of the chromosome.
- Centromeres are significantly more concave compared to the remaining parts of the chromosome edges.

In consideration of these specific characteristics, two features are selected to identify the centromere location. The first one is the width

Figure 3: a) Segmented image based on active contour b) Canny edge detector

profile and the second one is the concavity degree of the edge. To compute these features, a two-step pre-computation is required.

Step 1) Centreline Extraction

It is more convenient to work with the chromosome images in horizontal alignment. Accordingly, the angle of the chromosome over the horizon should be initially computed. For this reason, the relative slope of the centreline of the chromosome is calculated.

Extracting the centreline; an important topological shape descriptor; is a valuable characteristic for classification, segmentation and rotation of the chromosome. In addition, the centreline is an important tool to extract features like width and density profiles. A robust technique was proposed to extract the centreline without any spurious branches in order to accurately calculate the width profile and concavity degree of the chromosome boundary and to determine the centromere position. The proposed algorithm tries to extract the centreline in two steps. First, a filling algorithm is applied to the segmented chromosome, and then an iterative thinning process is used to find the centreline. Figure 4 shows this process in a chromosome image.

One of the problems in identifying the slope of the centreline is spurious branches in the head and tail. To remove these spurious branches, 5% of the length of the centreline is truncated from the head and tail. Next a polynomial curve of degree one is fitted to the x- and y- coordinates of the centreline. After that, the slope of the straight line is computed to find the direction of the chromosome. By rotating the chromosome in the opposite angle of the slope, the chromosome is aligned horizontally. However, due to the thinning process, this method suffers from spurious branches that

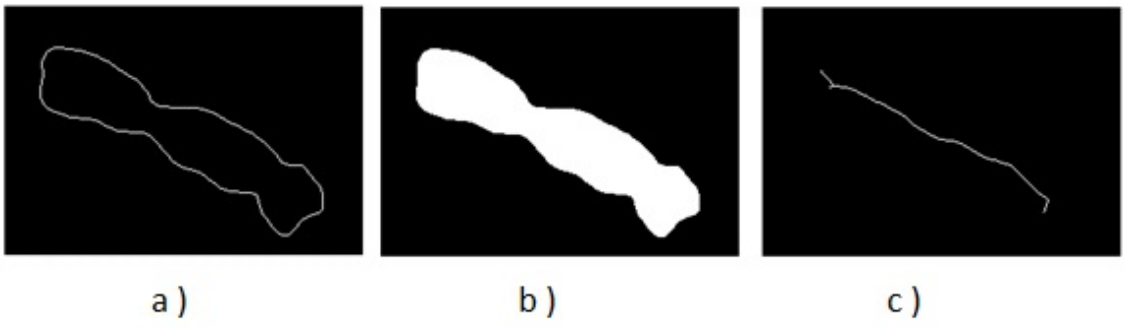

Figure 4: Extracting centreline a) segmented image, b) filling, c) thinning

might cause the polynomial curve of degree one in the wrong orientation, with the wrong slope and wrong rotation. To overcome this problem, another centreline is computed with a different method. The chromosome image of previous section is swept vertically. Then the midpoints of the lines intersecting the edges of the chromosome are chosen as the centreline. Afterwards, the slope of the centreline is computed and rotated to make the image of chromosome in the exact horizontal direction. In contrast with MAT, this technique is robust against shape variance to extract centrelines without any spurious branches.

Step 2) Edge Graph Extraction

Instead of processing a large amount of data from an image matrix, the corresponding graph of the edge is obtained by tracing the contour of image and extracting its boundary. This graph is represented by a matrix consisting of two columns (x-and y-coordinates of each pixel of the edge). This technique preserves all the information of the edge. Furthermore, a significant reduction in computational complexity can be achieved by working with a smaller amount of data from the edge, compared to larger amounts of data from the image.

According to the previous section, this graph is rotated so that the graph aligns horizontally. Figure 5 shows the result of extracting the edge graph.

Width Profile

Width profile computes the length of the perpendicular line across the centerline. It is important to align the graph horizontally before computing the width profile for accurate distance measurement. The average value of the width of graph is chosen as a threshold, thus, the graph is converted into upper and lower sub-graphs. The formula for the threshold is computed by:

$$Threshold = \frac{1}{n}.\sum_{i}^{n} = 1y_i \qquad (4)$$

Where (y_i) is the y- coordinate of each point of the graph and (n) is the total number of points in the graph.

To compute the width profile, as shown in Equation 5, the city block distance algorithm (Manhattan distance) is applied along the horizontal graph.

$$Distance = |x1 - x2| + |y1 - y2| \qquad (5)$$

Where (x1, y1) and (x2, y2) are the first and second points, respectively. First, the distance between every point of the upper or lower subgraph is computed and then the minimum distance is chosen as the actual distance.

Figure 6 shows both upper and lower subgraphs of chromosome and the resulting width profile.

Concavity Degree

As seen previously, centromeres are regions

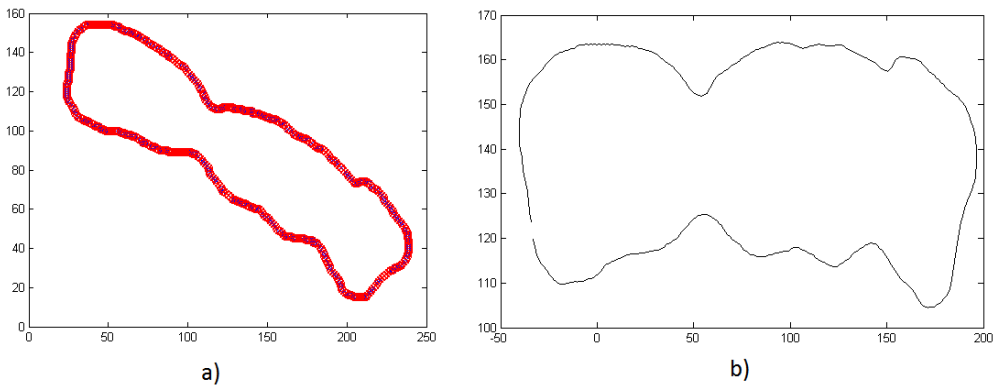

Figure 5: a) Edge graph extraction b) after rotation

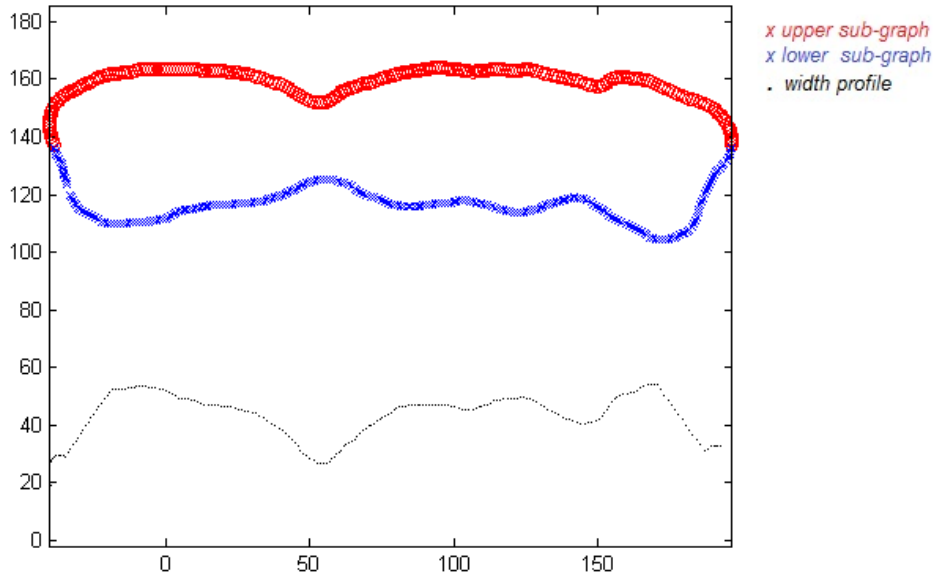

Figure 6: Upper and lower sub-graphs with width profile

with maximum concavity compared to other parts of chromosome boundary. Considering this, the proposed algorithm computes the concavity degree of the chromosome boundary. By splitting each sub-graph of the previous step into overlapped sub-sections, a second order polynomial is fitted to each part. The second derivative of each part is then calculated. The first 20% of points with the maximum derivative are labelled as the most concave points. In the next section, application of this feature will be demonstrated.

Centromere Location

Based on the obtained feature, centromere position is identified. The centromere position in the width profile is a deep valley. Due to the noisy nature of chromosome edges, searching for these valleys is a difficult procedure. To ignore weak valley, the profile is initially smoothed. Afterwards, using the same approach as the event detection of ECG signals [17], an empirical threshold is selected based on trial and error as in Equation 6. The valleys below the threshold are disregarded.

$$Threshold = STD\ (width\ profile) + 0.6 * Mean\ (width\ profile) \quad (6)$$

After that, four-point peak detection (Equation 7) is performed on the inverted width profile signal to locate the centromere position. The result of peak detection is shown in Figure 7.

$$\text{If } \begin{cases} x_2 > x_1 \\ x_2 \geq x_3 \\ x_2 > x_4 \end{cases} \text{ then x2 is peak} \quad (7)$$

As mentioned, centromere is the most concave part of chromosome. This area, as shown in Figure 8, is labelled as the most concave point on the edge graph. The combination of these two features leads to a result which is more satisfactory.

As a result, the centromere is the point where at the same length, both width profile and concavity degrees are graphically identified as centromere. Figure 9 shows both features simultaneously.

Experimental Results

The proposed algorithm can classify chromosomes into 4 spread classes, defined as those with normal chromosomes (with one centromere including metacentric chromosomes and sub-metacentric chromosomes),

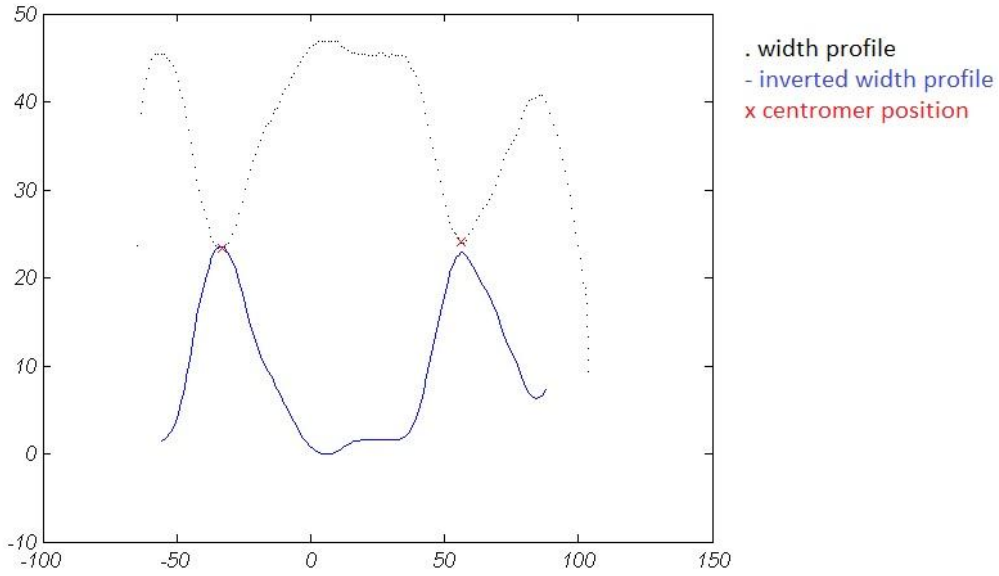

Figure 7: Centromere valleys in width profile

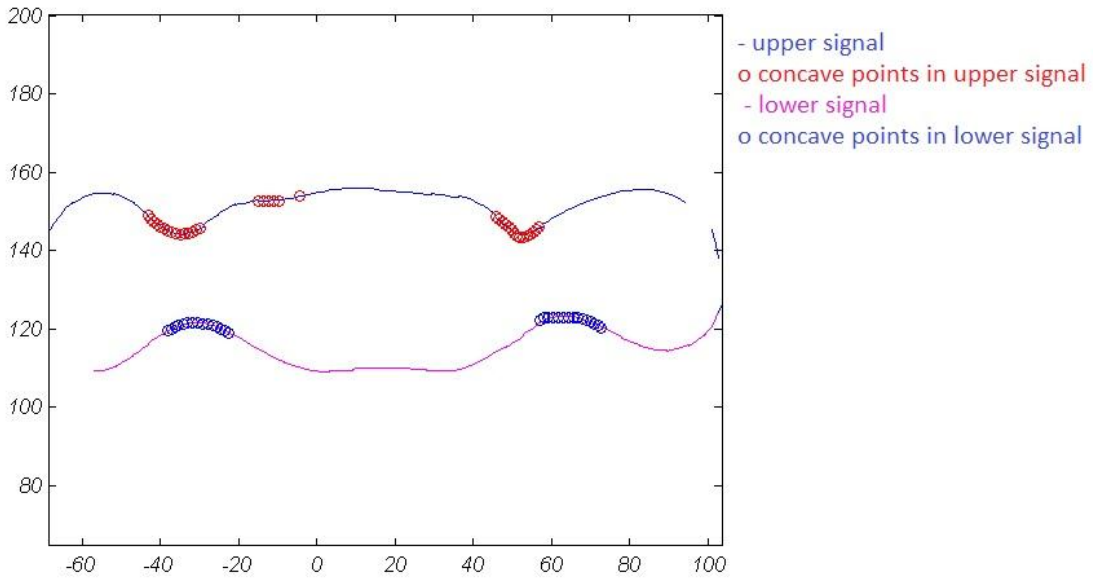

Figure 8: The most concave parts of chromosome signal

dicentric chromosomes (with 2 centromeres), thricentric chromosomes (with three centromeres) and fragmented or acrocentric chromosomes. Due to the special structure of chromosomes in the latter class, the proposed algorithm categorized them as chromosomes with no centromeres. The results of the classification of 311 chromosome images into these four classes are shown in Table 1.

Moreover, the proposed algorithm can detect the location of centromere positions. To evaluate the accuracy of centromere position, an expert human was used to compare the algorithm result with that of manual centromere segmentation. Any dislocation out of the region of interest was regarded as inaccurate centromere

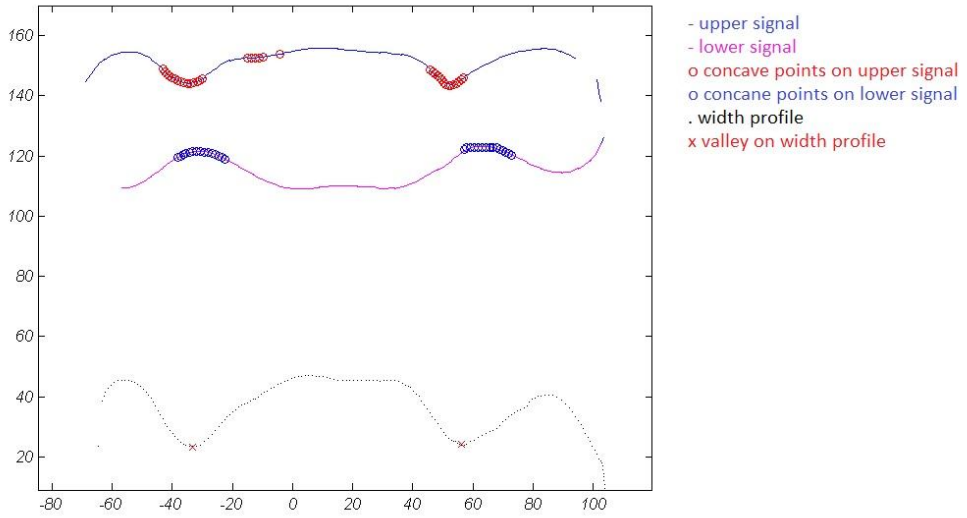

Figure 9: Width profile with centromere valleys and concave points on the chromosome signal

Table 1: Result of automated classification into 4 classes.

Class	Total No. Chromosomes	No. correct class identification	Accuracy rate of class Identification (%)
Normal (Metacentric or Submetacentri)	153	148	96.7
Dicentric	89	78	87.6
Thriccentric	5	3	60
Acrocentric or Fragment	64	60	93.8
Total	311		

position. Table 2 shows the accuracy in centromere location.

To evaluate the efficiency of the proposed algorithm in detecting normal and abnormal chromosomes, chromosomes were categorized into two distinct classes of either dicentric or non-dicentric chromosomes. By measuring two statistical indicators (sensitivity and specificity), the accuracy of this classification was computed.

Sensitivity or true positive rate (TPR), as shown in Equation 8, represents the proportion of positive cases in which the algorithm correctly identified as positive.

Specificity or true negative rate (TNR), as shown in Eq9, represents the proportion of negative cases in which the algorithm correctly identified as negative.

$$TPR=TP/(TP+FN)\times 100 \quad (4)$$

$$TNR=TN/(TN+FP)\times 100 \quad (5)$$

TP: number of chromosomes correctly classified as dicentric

TN: number of chromosomes correctly classified as non- dicentric

FP: number of chromosomes inaccurately classified as dicentric

FN: number of chromosomes inaccurately classified as non- dicentric

A sensitivity of 87.6% and specificity of 95.9% are achieved (Table 3).

Conclusion

Automated centromere detection is a critical task in karyotyping as well as detecting dicentric chromosome abnormalities in biological dosimetry. Despite the fact that much work has been done in this area, most genetic laboratories still use semi-automated systems for these routine procedures.

In the present study, an automated system for detecting centromere location was developed and tested, with fewer features (just two features), a less complicated scheme and more accurate results compared to previous research [4, 6, 18] (e.g. 94.2% versus 90.8 %)[8] or (e.g.,94.2% versus 91.3) (after removing acrocentric chromosomes from both data sets, the accuracy was computed for the most precise comparison)[19]. Additionally, if an automated system with the capability of straightening bent chromosome was applied prior to the proposed algorithm, even more accurate results would be achieved. As reported in Table 4, the results of applying the proposed algorithm to straightened chromosomes increases accuracy to 97.8 %. Fair comparison of the present method to other methods is difficult due to different preparation methods and databases utilized.

Ultimately, the input of the proposed algorithm includes single chromosomes. Nevertheless, implementing an automated system which segments all of the chromosomes of metaphase is a great improvement for the industry.

Table 2: Results of automated centromere detection

Class	Total No. Chromosomes	Total No. Centromere	No. correct Centromere position Identification	Accuracy rate of Correct centromere Position Identification (%)
Normal (metacentric or submetacentric)	153	153	148	96.7
Dicentric	89	178	164	92.1
Thricentric	5	15	13	86.6
Total	247	346	326	94.2

Table 3: Results of classification into dicentric and non-dicentric sets

Class	Total No. chromosomes	No. Correct class Identification
Dicentric	89	78
Non-dicentric	222	213

Table 4: Results of classification into dicentric and non-dicentric sets

Class	Total No. Chromosomes	Total No. Centromere	No. correct Centromere position Identification	Accuracy rate of Correct centromere Position Identification (%)
Normal (Metacentric or Submetacentri)	151	151	149	98.7
Dicentric	80	160	156	97.5
Thricentric	4	12	11	91.6
Total	235	323	316	97.8

Acknowledgment

This research was supported by Shiraz university of medical sciences. We are thankful to our colleagues Amirhossein Hadaegh in the medical genetics department, whose contribution in stimulating suggestions and encouragements, was greatly assisted in the research.

Conflict of Interest

None

References

1. Flegal FN, Devantier Y, McNamee JP, Wilkins RC. Quickscan dicentric chromosome analysis for radiation biodosimetry. *Health Phys*. 2010;**98**:276-81. doi.org/10.1097/HP.0b013e3181aba9c7. PubMed PMID: 20065694.
2. Finnon P, Lloyd D, Edwards A. Progress in automatic dicentric hunting. Chromosomal Alterations: Springer; 1994. p. 192-202.
3. Levan A, Fredga K, Sandberg AA. Nomenclature for centromeric position on chromosomes. *Hereditas*. 1964;**52**:201-20. doi.org/10.1111/j.1601-5223.1964.tb01953.x.
4. Madian N, Jayanthi K. Analysis of human chromosome classification using centromere position. *Measurement*. 2014;**47**:287-95. doi.org/10.1016/j.measurement.2013.08.033.
5. Moradi M, Setarehdan SK, Ghaffari S, editors. Automatic locating the centromere on human chromosome pictures. Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium; 2003: IEEE.
6. de Faria ER, Guliato D, de Sousa Santos JC, editors. Segmentation and centromere locating methods applied to fish chromosomes images. Brazilian Symposium on Bioinformatics; Springer; 2005. p. 181-189.
7. Mohammadi MR. Accurate localization of chromosome centromere based on concave points. *J Med Signals Sens*. 2012;**2**:88-94. PubMed PMID: 23626944. PubMed PMCID: 3632046.
8. Wang X, Zheng B, Li S, Mulvihill JJ, Liu H. A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes. *Comput Methods Programs Biomed*. 2008;**89**:33-42. doi.org/10.1016/j.cmpb.2007.10.013. PubMed PMID: 18082909.
9. Stanley R, Keller J, Caldwell C, Gader P, editors. Centromere attribute integration based chromosome polarity assignment. Proceedings of the AMIA Annual Fall Symposium: American Medical Informatics Association; 1996:
10. Zakeri F, Hirobe T. A cytogenetic approach to the effects of low levels of ionizing radiations on occupationally exposed individuals. *Eur J Radiol*. 2010;**73**:191-5. doi.org/10.1016/j.ejrad.2008.10.015. PubMed PMID: 19054641.
11. Cao H, Deng H-W, Wang Y-P. Segmentation of M-FISH images for improved classification of chromosomes with an adaptive fuzzy c-means clustering algorithm. *IEEE transactions on fuzzy systems*. 2012;**20**:1-8. doi.org/10.1109/TFUZZ.2011.2160025.
12. Tao W-B, Tian J-W, Liu J. Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. *Pattern Recognition Letters*. 2003;**24**:3069-78. doi.org/10.1016/S0167-8655(03)00166-1.
13. Grau V, Mewes AU, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans Med Imaging*. 2004;**23**:447-58. doi.org/10.1109/TMI.2004.824224. PubMed PMID:

15084070.

14. Poletti E, Zappelli F, Ruggeri A, Grisan E. A review of thresholding strategies applied to human chromosome segmentation. *Comput Methods Programs Biomed.* 2012;**108**:679-88. doi.org/10.1016/j.cmpb.2011.12.003. PubMed PMID: 22261220.
15. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *International journal of computer vision*. 1988;**1**:321-31. doi.org/10.1007/BF00133570.
16. Zhang K, Zhang L, Song H, Zhou W. Active contours with selective local or global segmentation: a new formulation and level set method. *Image and Vision computing*. 2010;**28**:668-76. doi.org/10.1016/j.imavis.2009.10.009.
17. Murthy IS, Rangaraj MR. New concepts for PVC detection. *IEEE Trans Biomed Eng.* 1979;**26**:409-16. doi.org/10.1109/TBME.1979.326420. PubMed PMID: 88412.
18. Aa AS, Samarabandua J, Knollb J, Roganb P. Automated metaphase chromosome centromere refinement using fuzzy inference systems. 6th Canadian Student Conference on Biomedical Computing and Engineering. (May 26-28, 2011).
19. Piper J, Granum E. On fully automatic feature measurement for banded chromosome classification. *Cytometry*. 1989;**10**:242-55. doi.org/10.1002/cyto.990100303. PubMed PMID: 2714109.