

A review of metaphase chromosome image selection techniques for automatic karyotype generation

Tanvi Arora¹ · Renu Dhir¹

Received: 23 November 2014 / Accepted: 11 November 2015 / Published online: 16 December 2015
© International Federation for Medical and Biological Engineering 2015

Abstract The karyotype is analyzed to detect the genetic abnormalities. It is generated by arranging the chromosomes after extracting them from the metaphase chromosome images. The chromosomes are non-rigid bodies that contain the genetic information of an individual. The metaphase chromosome image spread contains the chromosomes, but these chromosomes are not distinct bodies; they can either be individual chromosomes or be touching one another; they may be bent or even may be overlapping and thus forming a cluster of chromosomes. The extraction of chromosomes from these touching and overlapping chromosomes is a very tedious process. The segmentation of a random metaphase chromosome image may not give us correct and accurate results. Therefore, before taking up a metaphase chromosome image for analysis, it must be analyzed for the orientation of the chromosomes it contains. The various reported methods for metaphase chromosome image selection for automatic karyotype generation are compared in this paper. After analysis, it has been concluded that each metaphase chromosome image selection method has its advantages and disadvantages.

Keywords Chromosomes · Karyotype · Metaphase chromosome images · Genetic disorders

1 Introduction

A chromosome is a thread-like structure that carries the genetic information of an individual. It is made up of DNA (deoxyribonucleic acid) which is tightly packed around the specific type of proteins known as histones that support the structure of the chromosome [3]. A healthy human has 23 chromosome pairs, comprising of 22 autosomes and two sex chromosomes [15]. Since the chromosomes carry the DNA, which in turn have the instructions for the synthesis of proteins. The proteins in turn help in movement of molecules, help to break the toxins and do many other maintenance and structure building activities [14]. The mutation or a change in the gene can change the instructions for making a particular type of protein [1]. That may cause the malfunctioning of the protein or absence of the gene. This can result in a medical condition that is known as genetic disorder. One can get the genetic mutation from either one parent or both the parents; moreover, during lifetime also mutation may occur. The alteration in either the number of chromosomes or the structure of the chromosomes leads to genetic defects.

A laboratory procedure named as karyotyping is used to study the chromosomes in the cell sample. The karyotype is an organized arrangement of chromosomes that has the chromosome pairs in decreasing order of length [25]. Karyotypes are studied to find out the genetic abnormalities, and there may be some missing chromosome or an extra chromosome or a deletion of a part of a chromosome or duplication [38].

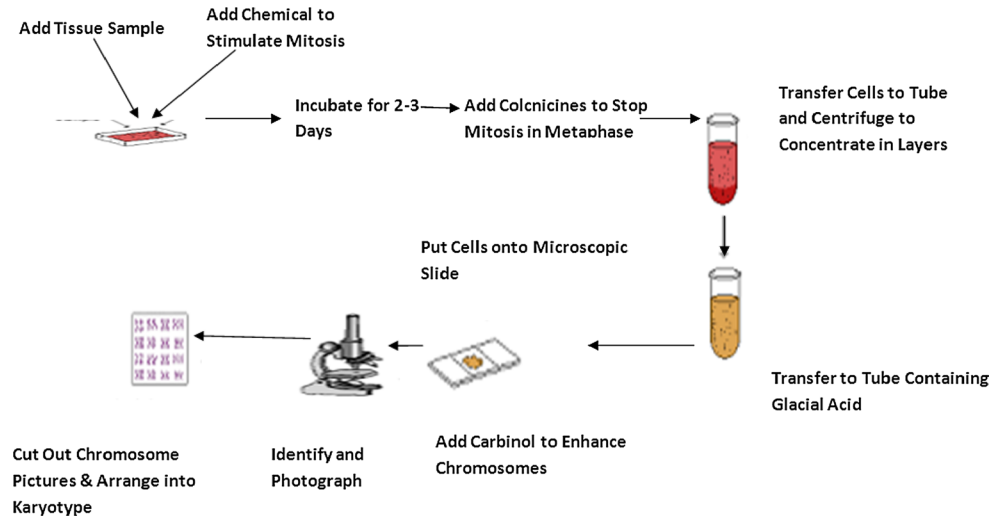
A cytogeneticist is the person who detects and interprets the chromosome abnormalities. For detecting genetic diseases, automatic scanning of specimens is quite important as it greatly reduces the work load of the cytogeneticists

✉ Tanvi Arora
tanviverma@rediffmail.com

Renu Dhir
dhirr@nitj.ac.in

¹ Department of Computer Science and Engineering,
Dr. B.R. Ambedkar National Institute of Technology,
Jalandhar, Punjab, India

Fig. 1 Metaphase chromosome extraction process



[11]. It helps in selecting the specimens that are analyzable. The cytogeneticists generally use commercially available image scanners 10x for screening and 100x for acquisition [50].

Optical microscopes are mainly used for image selection and analysis. The condenser of the microscope is responsible for the quality of the image acquired as it controls the illumination and the angle at which the light enters the front lens [37]. It has been observed that by adjusting the aperture of the condenser, we can obtain more analyzable chromosome images.

During cell division, the morphological structure of the chromosomes changes in various stages. Interphase stage has chromosomes inside the cell nucleus for DNA duplication. The next stage is mitosis, in which each chromosome is split into two identical daughter cells. The mitosis stage is a combination of many steps, out of which metaphase is of much importance, in which the daughter cells are attached at the centromere [34]. In this stage, chromosomes have the best morphological features and are studied for diagnosis purpose [17].

For the purpose of diagnosis, the metaphase chromosomes are imaged using samples of bone marrow [36], peripheral blood [51], product of conception, amniotic fluid [35], etc.

The metaphase chromosomes are extracted from the samples by the process outlined in Fig. 1. The collected sample is first incubated for some days, and during that period, it is treated with phytoagglutinins which leads to increased cell reproduction rate. Toward the end of incubation, the cell division is stopped by treating the sample with colchicines, so as to increase the count of usable metaphase chromosomes. In the second state, the sample is treated with mixture of glacial acid and carbinol so as to fix the cells at a one particular stage of cell division. After this stage, the microscopic slides are taken and the



Fig. 2 G-banded metaphase chromosome images



Fig. 3 A karyogram of G-banded chromosomes

chromosomes are spread for the purpose of observation [34].

The chromosomes obtained above are transparent bodies. So in order to analyze them, they are treated by a dye,

so that visible bands appear on the chromosomes and thus they can be distinguished based upon the band patterns. Generally, chromosomes are stained with giemsa dye, and they are known as G-banded chromosomes [28]. Figure 2 shows the meta spread image of G-banded chromosomes.

The captured images have all the 46 chromosomes randomly placed. All the metaphase chromosome images are not suitable for analysis purpose as they may contain touching or overlapping chromosomes [29]. The cytogeneticists need to check and detect the metaphase chromosome images before proceeding with karyotyping [18]. Since 4 % of the slides contain metaphase chromosomes, few of those are analyzable. Figure 3 shows the karyogram of the G-banded chromosomes.

Based upon the type of the sample, the cytogeneticists need different number of analyzable metaphase chromosome images. For bone marrow samples, they need minimum of 20 analyzable images, as a specimen will be diagnosed positive if three or more cells are found to be affected. For the blood samples, 20 analyzable images are taken, out of which 17 are used for counting the chromosomes and 3 are used for finding the structural and numerical abnormalities [50].

The first ever genetic disorder named as Philadelphia translocation was reported in 1959 by Novell and Hungerford [50]. The abnormalities related to chromosomes can be of two broad types, namely structural and numerical [28], and Philadelphia translocation is an example of structural abnormality in which there is translocation between 9th and 22nd chromosome, whereas Down's syndrome is an example of numerical abnormality in which there is an extra 21st chromosome. Various types of leukemia's are marked by the presence of a single 16th, 18th or 22nd chromosome.

The karyotyping is generally performed manually in hospitals, in which visual inspection is very laborious and time-consuming. Apart from that, it results in difference of opinion and thus incorrect or biased results [21, 30]. Therefore, there is a need for an automatic solution.

The selected metaphase chromosome images are used for segmentation of the chromosomes, and then, further each segmented chromosome is classified into one of the 22 classes or X or Y sex chromosomes [20]. Many automatic tools for segmenting and classifying the chromosomes have been developed [2, 6, 7, 9, 10, 19, 22, 26, 32, 41]. But the big challenge for all these tools is that the chromosomes are non-rigid objects and they do not always occur straightened, but they are present in different orientations, they may be bent, they may be touching one another, or they are overlapping, thus forming clusters [17].

In order to resolve the issue of touching, overlapping and bent chromosomes, the researchers have given many approximate solutions using geometric and band-based features of the chromosomes [24, 31, 33, 40, 42–45, 49]. But all these methods are based upon heuristics and do not

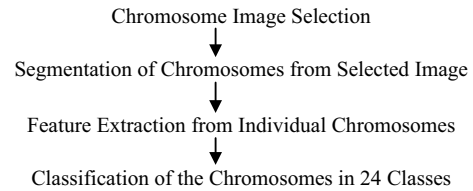


Fig. 4 Karyogram generation process

give 100 % accurate results. The biometric points of each chromosome were calculated for the identification of chromosome pairs using the parameters of band gap between each chromosome band and band lengths [8]. A re-segmentation approach has been proposed to increase the accuracy of segmentation and classification [27]. Recently, a novel imaging method of ptychography has been developed for karyotyping human chromosomes by using optical and X-ray ptychography-based methods [39].

In addition to this, the metaphase chromosome images may also contain some artifacts that are not chromosomes. So all these things make the task of automatic segmentation and classification of human chromosomes into a karyotype a tedious process. In order to get an accurate karyotype, the metaphase chromosome image that is selected for segmentation and classification of chromosomes should have more of individual chromosomes that have clear band patterns and have straight orientations and are not touching or overlapping. Figure 4 illustrates the four steps of the karyogram generation process.

To automate the process of karyotyping, hundreds of metaphase chromosome images need to be physically examined by an experienced cytogeneticist, which may require a lot of time and effort [5]. The mood of the cytogeneticist may also effect the selection of the metaphase chromosome images. So a cytogeneticist may select few images and leave the remaining images that may contain vital information, so this method of manual selection may not result in selection of the best metaphase chromosome images. In this study, the various automatic techniques for automatic metaphase chromosome image selection are studied and compared.

The paper is organized as follows: Sect. 2 contains the various methods for metaphase chromosome image selection, Sect. 3 contains the comparative analysis, Sect. 4 contains discussion and future directions, and Sect. 5 contains conclusion.

2 Methods for metaphase chromosome image selection

2.1 R Huber's method

In this approach, a three-phase detection method is used to find the metaphase chromosome image. In the

starting phase, analysis is performed on values of the filtered objects count (FOC) that have gone through band pass [16]. An analyzable metaphase chromosome image is possible if it shows sequential high values. In the middle phase, FOC is considered for horizontal distribution. If in this phase it gets consecutive high values, this indicates a possible analyzable metaphase chromosome image. In the third phase, a contour following method is applied to completely analyze the image. Thresholding is applied to the image before extraction of the features.

Following features are extracted:

- (1) Count of filtered FOC values pertaining to rectangle.
- (2) Count of objects inside the rectangle that satisfy the criteria of threshold.
- (3) Selected objects average area.
- (4) Average length of contour.
- (5) Average quotient of length of contour and area.
- (6) Ratio of squared total contour length and total area.
- (7) Summation of the product of the area of the object and the object distance from center of other objects.
- (8) Summation of all distances of objects in comparison with the center.
- (9) Ratio of summation of distance from center and summation squared distances from center.

The images are classified as metaphase images and non-metaphase images using multivariate classifier.

This method is very time-consuming and tedious. It does not consider the quality of the metaphase chromosome image being selected. It is not usable for the metaphase chromosome images that have touching and overlapping chromosomes.

2.2 Victor Gajendran's method

In this work, system is proposed that automatically counts the chromosomes in digital images. This system has two phases, namely preprocessing and counting phase. In the preprocessing phase, hysteresis thresholding segments the chromosome objects from the background. Further median filtering method is used to remove salt and pepper noise and also filling the holes of chromosomes and to smoothen the chromosome contours so that when thinning is applied, then extra branches are not created [13]. Thinning operation is performed to obtain the single pixel width skeletons of chromosomes or their clusters. After this the average width of all the skeletons is calculated. It has been observed that all the chromosomes have consistent width. So all those skeletons that are less than the average width of the skeletons are treated as noise and are not considered. Based upon the same lines, the slight connections are also removed.

The counting phase takes the noise-free and skeletonized metaphase chromosome images as input, and it labels all the skeletons present in the input metaphase chromosome image. Then, for each component, the end points and crossovers are identified. The raster scanning method is used for finding the first end point, and it is traced until next end point is reached and if a crossover is reached, then the path that first appears is taken. While tracking the skeleton all other pixels are deleted except for the crossover pixels. The count of chromosomes is incremented by one for each component, and all the above steps are repeated till all the end points are traced.

This method is just usable to count the number of chromosomes in the metaphase chromosome image. This method counts the touching and overlapping chromosomes as well. It is efficient as the error rate is just 6 %. The main drawback of the proposed method is that it does not rank the metaphase chromosome images.

2.3 Wang X's method

This method classifies the metaphase chromosome images into two categories, viz. analyzable metaphase cells and non-analyzable metaphase cells [50]. It is a five-step process. In the first step, they have taken a digital image and the image quality is enhanced using median filtering. In the second step, the thresholding is applied to remove the high gray values. Third step is region labeling to find connected components, and individual pixels are deleted. The fourth step takes the labeled components and computes the features. The five features to be computed are:

1. The count of the labeled components (N_m) is computed for each metaphase chromosome image.
2. The pixels present in each labeled component is counted ($S_i \times N_m$)
3. The circularity of each labeled component is calculated as $C_i = \frac{N_c}{N_m}$, where N_c count of pixels located in the circle and region contour and N_m stores the count of pixels inside the labeled component.
4. Each labeled components average gray value is computed as ($I_{ave} = \frac{1}{n} \sum_1^n I_i$)
5. The length between the center of the cell (x_c, y_c) and each labeled region (x_k, y_k) is calculated.

$$L_k = \sqrt{(x_c - x_k)^2 + (y_c - y_k)^2}$$

Then, in the fifth step, the computed features are passed to two machine-learning-based classifiers, namely decision trees and artificial neural networks, and the results are further optimized.

This method considers the quality of the image. But it cannot be used in images with touching and overlapping chromosomes. It is slow and does not rank the chromosome images.

2.4 Yan Wenzhong’s method

In this work, an attempt has been made to extract the count of chromosomes from a metaphase chromosome image that has overlapping chromosomes. For selecting the images, firstly they have used histogram equalization to enhance the image contrast [52]. Then, a binary image is created by using thresholding. In order to remove light and small objects, the binary image has been eroded.

Let A_k be an image that is being eroded k th time

$$A_k = A \otimes kB$$

where B represent structure element, \otimes represent erosion element. Y_k is a subset of connected components in one part of A_k if $l > k$. A_k will disappear from A_l

$$U_k = (A_{k+1} \oplus \{B\}); A_k$$

Then, $Y_k = A_k - U_k$, where \oplus denotes dilation operator

If the image has many connected components, then they are denoted as

$$Y = \bigcup_{k=1,m} Y_k$$

m represents the count of erosion operation.

The overlapping chromosomes were counted by the following method:

- Initially, it is assumed that there is only one chromosomes in the cluster.

$$\text{Max} = 1$$

- Structure element (B) of radius 3 with Euclid disk is used.
- Erode image using B
- Eight-connected component method is used to label the components, and m represents the count of chromosomes.
- The nonzero num variable is compared with max.
- Max is assigned num if num exceeds max.
- If num is less, then max go to step 3.
- If num = 0, then max is equated to number of chromosomes in the overlapping metaphase chromosome image.

The proposed method is fast, and it considers the quality of the image. It does not rank the metaphase chromosome images.

2.5 Ravi Uttamatinin’s MetaSel method

They have taken the metaphase chromosome images, and prior to analysis, they have defined the objects of the metaphase chromosome spread into four different classes. The class 1 is assigned to objects that are straight and are separate distinguishable entities, class 2 is assigned to objects that are either bent or skewed entities, class 3 is assigned to objects that are clusters of touching or overlapping entities, whereas class 4 is assigned to leftovers of the cell or artifacts [46, 47].

Their work aims at automatically choosing a high-quality metaphase chromosome spread image so as to make the task of automatic karyotyping simpler and more accurate. They have preprocessed the metaphase chromosome images using Otsu’s automatic thresholding [37] to segment the chromosome images from the background. The classification of the chromosomes has been done using width, height and its ratio, so as to categorize the objects into four classes as defined above.

$$\text{Area}_{\text{ratio}} = \frac{A_o}{A_r}$$

where A_r is count of pixels in the enclosing rectangle ($W_{\text{rect}} \times H_{\text{rect}}$) of extracted component and A_o represents pixels of extracted component. W_{rect} and H_{rect} represent width and height of rectangle. Area ratio can be used to check whether the chromosome is straight. Kernel density model-based empirical probability density function was calculated. Then, a Gaussian-based method was used to find the minimum threshold value pertaining to the ratio of area and was used to classify the chromosome into class 1. But this may also contain some cell residue or artifacts. In order to remove them, it was observed that the width of the chromosomes is consistent, so the objects classified as class 1 can be considered as cell residues or artifacts, and if their width is not consistent, then they can be assigned class to 4. If the width of the object is 1.5 times the average width of class 1 objects, then those objects are not considered.

O_w represents the objects that has width < 1 and a half times total average width. The individual objects width is defined as:

$$W_i = \frac{\text{Total number of pixels in chromosome}}{H_{\text{rect}}}$$

Average width can be calculated as

$$W_{\text{avg}} = \frac{\sum_{i \in O_w} W_i}{|O_w|}$$

Rectangle width ratio can be calculated as

$$W_{\text{rectratio}} = \frac{W_{\text{rect}}}{W_{\text{avg}}}$$

So if $W_{\text{rect, ratio}}$ for an object is less than the threshold value, then it can be classified into class 2, class 3 or class 4 categories.

To differentiate among classes, height is considered.

$$H_i = \frac{A_o}{W_{\text{avg}}}$$

The ratio between height (H_i) and height of rectangle (H_{rect}) is compared using height ratio ($H_{i, \text{ratio}}$).

$$H_{i, \text{ratio}} = \frac{H_i}{H_{\text{rect}}}$$

Using empirical density function and Gaussian model, it was found that the objects having $H_{i, \text{ratio}}$ are less than a threshold value where considered as class 4 objects.

Now in order to distinguish between class 2 and class 3 objects, one more parameter was computed, i.e., maximum width ratio ($W_{\text{max, ratio}}$).

$$W_{\text{max, ratio}} = \frac{W_{\text{max}}}{W_{\text{avg}}}$$

The empirical probability density function and Gaussian method were deployed for separating threshold for class 2 and class 3 objects. When $W_{\text{max, ratio}}$ will be greater than threshold, the object will be classified as class 3 object and the one with lesser will be classified as class 2 object.

The simple rule-based Gaussian classification techniques can rank the metaphase chromosome images depending upon the count of objects belonging to each class. The metaphase chromosome images with largest count of class 1 objects are considered as the best candidates to be chosen for further automatic segmentation and classification of chromosomes for karyotype generation.

This method considers the quality of the chromosome images, is fast, can count touching and overlapping chromosomes as well, and above all, it ranks the metaphase chromosome images. It is a computationally expensive approach.

2.6 Ravi Uttamatin's band classification method

In this work, they have classified the chromosome metaphase images into low and high band resolution considering the shape of chromosomes [48]. In the low band resolution chromosomes are small in size and are well spread and there is no touching or overlapping, so it is suitable for counting the number of chromosomes. In the case of high band resolution, the chromosomes are long, they may be bent or overlapping so these band chromosomes are used for detecting structural abnormalities. The metaphase chromosome image with low band resolution is used for counting the number of chromosomes, whereas the high band

resolution images are considered for structural abnormality checking.

In order to classify the metaphase chromosome images based upon resolution of bands, the metaphase chromosome images are preprocessed based upon gray-level adjustment and Otsu's thresholding, to separate the foreground and background objects. After segmenting the foreground and background objects, the segmented objects are rotated so that they are vertical. After this preprocessing, the parameters of the individual chromosomes are calculated such as area ratio ($\text{Area}_{\text{ratio}}$), average width (W_{avg}), width of each object [$\text{Width}(W_i)$], ratio of width and average width ($W_{\text{rect, ratio}}$), height (H_i), height ratio ($H_{i, \text{ratio}}$), maximum width ratio ($W_{\text{max, ratio}}$), length (L).

Then, based upon the following algorithm, the chromosomes are classified as low band resolution and high band resolution.

1. Image parameters such as area ratio ($\text{Area}_{\text{ratio}}$), average width (W_{avg}), width of each object [$\text{Width}(W_i)$], ratio of width and average width ($W_{\text{rect, ratio}}$), height (H_i), height ratio ($H_{i, \text{ratio}}$), maximum width ratio ($W_{\text{max, ratio}}$), length (L) are calculated.
2. If the area ratio > threshold, the objects are classified as other objects, else they are classified as straight chromosomes.
3. For straight chromosomes, W_{avg} is calculated, and then compared with $\min W_{\text{rect, ratio}}$, if it is less than threshold, then the objects are classified as small artifacts, and if that is more than the max threshold, then they are classified as large artifacts, else they are classified as individual chromosomes.
4. For other objects, $H_{i, \text{ratio}}$ is calculated, if it is less than a threshold value, then they are classified as small artifacts, else the $W_{\text{max, ratio}}$ is calculated, and if it is greater than the threshold, then the objects are classified as large artifacts, otherwise they are classified as individual chromosomes.
5. For the individual chromosomes, the length is calculated.
6. If length is greater than a minimum value of threshold, then the image is assigned as high band resolution image, else it is assigned as low band resolution image.

This method is fast and considers geometric and band features, but it does not count the number of individual or touching chromosomes.

2.7 Yuchem Qui's method

They have developed a fully automatic microscope-based image selection system which is based upon the scanning concept of time-delay integration (TDI) [36]. They have

continuously scanned the image, while the object is moving. The blur is removed by dividing the long exposure time to short exposures. The TDI-based scanning produces high-resolution images. In order to find out analyzable metaphase chromosome images, it is necessary to automatically locate region of interests (ROI).

They have applied two methods to detect ROI's, viz. online method and offline method. The online method saves the ROI's during scanning process only for those metaphase chromosome images that are suspected to be analyzable, whereas in case of offline method, each scanned image is first saved and then processed to detect, if it contains analyzable chromosomes or not.

The ROIs are segmented and labeled by region growing and labeling algorithms. The labeled regions are discarded based upon the threshold values of size and circularity. The remaining labeled regions are analyzed for average intensity, area circularity, radial length of the center. Based upon the above features of the labeled regions, the images are classified as analyzable and non-analyzable. The proposed method can find out more analyzable metaphase chromosome images and has more efficiency, and the system can directly provide the high-resolution images to the computer.

This method is fast and considers geometric and intensity-based features [23]. It does not consider touching and overlapping chromosomes, error rate is higher, and it does not rank the metaphase chromosome images.

3 Comparative analysis

The task of automatic karyotyping of human chromosomes is very challenging. The chromosomes are non-rigid bodies which are scanned using a microscope and are visible as distinct bodies only toward the end of cell division [12]. The metaphase chromosome images contain the visible chromosomes. But these chromosomes may be bent, touching or overlapping. Thus, studying the characteristic features of an individual chromosome is a tedious process. After 46 years of long research, the task of automatic chromosome segmentation and classification is still an open issue.

After reviewing the literature and work done by various researchers in the field of automatic karyotyping, it has been concluded that owing to the non-rigid nature of chromosomes, the segmentation of touching and overlapping chromosomes cannot guarantee 100 % accurate feature extraction by automated means.

So it is suggested to take up those metaphase chromosome images only for analysis or study purpose for automatic karyotyping that have least amount of bent, touching

or overlapping chromosomes, so that the chances of inaccuracies are minimized.

Most of the works for automatic segmentation and classification of human chromosomes rely on manual selection of metaphase chromosome images. The manual selection is very time-consuming process. The efficiency of the process depends upon the human behavior and the time constraints. So the search results may be biased or not up to the mark, or it may be that whole of the search space is not explored.

So there is a strong need to automate the process of metaphase chromosome image selection prior to automated segmentation and classification of the chromosomes. Few researchers have worked in this area and have come forward with methods by which they can classify the metaphase chromosome images into analyzable and non-analyzable categories, and some have tried to rank the metaphase images in order of most analyzable to least analyzable depending upon the number of individual chromosomes, bent, touching or overlapping chromosomes present in the meta spread.

To date, very less focus has been given to automate the process of automatic metaphase chromosome image selection. The automatic selection of metaphase chromosome images will be beneficial in two ways; firstly, it will reduce the time of cytogeneticist to select the metaphase chromosome image manually, and secondly, it will save the time while processing and analyzing the metaphase chromosome image for segmentation and classification purpose.

In the year 1995, Huber et al. [16] proposed a method that classifies the input image as metaphase chromosome image or a non-metaphase chromosome image considering nine geometric features, but his method has no control over quality of metaphase chromosome images, it is a slow process, it is not useful if the image contains touching, overlapping or bent chromosomes, no ranking of the metaphase chromosome images is done, and it considers only the geometric features.

In 2004, Gajendran and Rodriguez [13] proposed a method that helps to count number of chromosomes in a metaphase chromosome image, it considered three geometric features to count the chromosomes, and it was capable of counting touching and overlapping chromosomes as well. It is a slow process, and above all it does not classify the images.

In 2008, Wang et al. [50] proposed a method that considers five geometric and intensity features, to classify the metaphase chromosome image as analyzable or non-analyzable. It is not capable of handling the touching or overlapping chromosomes. It does not rank the images.

In 2009, Wenzhang [52] proposed a method that is simple and fast, and it counts chromosomes based upon connected components. It is workable in case of touching

and overlapping chromosomes, but it does not classify the image.

In 2013, Uttamatinin et al. [47] proposed a MetaSel method that can classify and rank metaphase chromosome images based upon 8 geometric features. It is workable for touching, overlapping and bent chromosomes. The only drawback of the technique is that it is computationally expensive, as it considers all the metaphase chromosome images and then ranks them.

In 2013, Uttamatinin et al. [48] proposed a band classification method that can classify the metaphase chromosome images as low band resolution and high band resolution images. The low band resolution images can be used for counting the chromosomes, whereas high band resolution images can be used for finding structural anomalies.

In 2014, Qiu et al. [36] has proposed a image selection technique based upon the scanning concept of time-delay integration. The region of interest is detected using online and offline methods. They have considered 5 geometric and intensity-based features to classify the images into either analyzable or non-analyzable. But their method cannot handle touching and overlapping chromosomes.

In this paper, an effort has been made to compare most of the proposed metaphase chromosome image selection techniques. The comparison has been done based upon 12 parameters that can compare the reported metaphase chromosome image selection techniques. The parameters considered for the comparative analysis of metaphase chromosome image selection techniques are: *number of classes* that is how many groups are being formed for the purpose of classification, *control over quality of images* it assesses whether preprocessing is being done to enhance the image quality or not, *speed* parameter judges how fast the process can execute, *does it rank images* parameter compares if the considered images are being ranked for the purpose of selection or not, *features considered* parameter compares the method based upon the type of features that are considered while classifying the images, *error rate* states the efficiency of the method, *counts number of chromosomes* parameter sees whether the proposed method counts the number of chromosomes present in the image under consideration or not, *counts number of overlapping chromosomes* parameter compares the methods if they can count the number of overlapping chromosomes, *number of features* parameter counts the number of features considered for the purpose of classification, *classifier used* sees which classification technique has been used to classify the metaphase chromosome images, *workable in case of touching and overlapping chromosome* parameter tells whether the method will work in case the input image has touching or overlapping chromosomes and *workable when the chromosomes are thin and long* parameter compares the methods for their capability to analyze thin and long chromosomes.

Every proposed method has its own set of features. In the below Table 1, all the methods are compared against a set of parameters. The comparative analysis suggests that the best method so far is the Ravi Uttamatinin's MetaSel method [47], as it ranks the metaphase chromosome images based upon the objects present, thereby making the task of karyotype generation very easy and efficient. Other methods are not much efficient as they just either counts the chromosomes or they just suggest whether the metaphase chromosome images are analyzable or not.

4 Discussion and future directions

All the methods so far proposed for automatic metaphase chromosome image selection have focused much effort on segmenting the objects from all the metaphase chromosome images, whereas only few images are to be selected for the actual analysis purpose. Therefore, the task of segmenting the objects becomes time-consuming. Time and effort are being spent on those images that have no potential to be selected as analyzable chromosome images.

Secondly, each stored image needs significant storage space; all the images are stored irrespective of the fact whether they fulfill the minimum criteria of number of objects present in the image. Therefore, a method needs to be integrated within the scanner, so that at the time of scanning only those images should be saved for future processing that satisfy minimum criteria of number of distinct objects present in the scanned image.

Thirdly, none of the reported studies have tested and validated their proposed approaches on standard datasets. So the results are not directly comparable. Each one has proposed their own approach without comparing with the existing methods.

All the methods for metaphase chromosome segmentation and classification are based upon the approximation of the segmentation results for touching and overlapping chromosomes [4]. The approximate results are not good enough to be considered for the purpose of diagnosis of a medical condition. So a method needs to be developed that gives accurate results.

So far no work has been carried out to emulate the image selection criteria of the human cytogeneticists.

5 Conclusion

Based upon the comparative analysis and the above discussion, it can be concluded that the Ravi Uttamatinin's MetaSel [47] method is the best possible approach so far as it can be used to rank the metaphase chromosome images. This method further helps in reducing the complexity of

Table 1 Comparative analysis of metaphase chromosome image selection techniques

S. no	Feature	R Huber method [16]	Victor Gajendran method [13]	X Wang method [50]	Yan Wenzhang method [52]	Ravi Uttamatinin MetaSel method [47]	Ravi Uttamatinin band classification method [48]	Yuchem Qui's method [36]
1	Number of classes	2	0	2	0	4	2	2
2	Control over quality	No	Yes	Yes	Yes	Yes	No	Yes
3	Speed	Slow	Slow	Slow	Fast	Fast	Fast	Fast
4	Does it rank images	No	No	No	No	Yes	No	No
5	Features considered	Geometric	Geometric	Geometric and intensity features	Geometric	Geometric	Geometric and intensity features	Geometric and intensity features
6	Error rate	Not specified	6 %	Not given	Not given	10 %	15 %	Not given
7	Counts number of chromosomes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Counts number of overlapping chromosomes	No	Yes	No	Yes	Yes	No	No
9	Number of features	9	3	5	0	8	8	5
10	Classifier used	Multivariate statistical	Not used	Decision trees and artificial neural networks	Not used	Rule-based Gaussian	Not given	Not given
11	Workable in case of touching and overlapping chromosomes	No	Yes	No	Yes	Yes	No	No
12	Workable when the chromosomes are thin and long	No	Yes	Yes	No	Yes	Yes	Yes

automatic segmentation and classification and make the task simpler and efficient.

The automation of the metaphase chromosome image selection process can enhance the efficiency of the automatic karyotype generation process. It can remove the reliance on the experienced cytogeneticist, thereby creating a fully automatic system for automatic karyotype generation.

References

- Abuzenadah A (2010) The causes of mutations, p 1–6
- Agam G, Member S, Dinstein H (1997) Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification. *IEEE Trans Pattern Anal Mach Intell* 19(11):1212–1222
- Alberts B (2000) Basic genetic mechanism. In: *Molecular biology of the cell*, 5th edn. Garland Science, New York, pp 191–234
- Alfredo J, Costa F, De Souza JG (2011) Image segmentation through clustering based on natural computing techniques. *Intech*, pp 57–82
- Arachchige AS (2014) Human metaphase chromosome analysis using image processing. The University of Western Ontario London, Ontario
- Carothers A, Piper J (1994) Computer-aided classification of human chromosomes: a review. *Stat Comput* 4:161–171
- Choi BSHH (2006) Automatic segmentation and classification of multiplex-fluorescence in-situ hybridization chromosome images. The University of Texas at Austin, Austin
- Devaraj S, Vijaykumar VR, Soundrarajan GR (2013) Leaf biometrics based karyotyping of G-band chromosomes. *Int J Hum Genet* 13(3):131–138
- Dhir R (2014) An efficient segmentation method for overlapping chromosome images. *Int J Comput Appl* 95(1):29–32
- El Emary IMM (2006) On the application of artificial neural networks in analyzing and classifying the human chromosomes. *J Comput Sci* 2(1):72–75
- Grisan E, Poletti E, Ruggeri A (2009) Automatic segmentation and disentangling of chromosomes in Q-Band prometaphase images. *IEEE Trans Inf Technol Biomed* 13(4):575–581
- Grisan E, Poletti E, Ruggeri A (2009) An improved segmentation of chromosomes in Q-band prometaphase images using a region based level set. In: *World congress on medical physics and biomedical engineering*, September 7–12, Munich, pp. 748–751
- Gujendran V, Rodriguez JJ (2004) Chromosome counting via digital image analysis. In: *Proceedings of international conference on image processing*, pp 2929–2932
- Harris H, Bassier B (2005) *Inside the cell*, vol 5. National Institute of General Medical Sciences, pp 1–74
- Hin Tjio J, Levan A (1925) The chromosome number of man. *Genetics* 10(6):80–85
- Huber R, Kulka U, Lörch T, Braselmann H, Bauchinger M (1995) Automated metaphase finding: an assessment of the efficiency of the METAFER2 system in a routine mutagenicity assay. *ELSEVIER Mutat Res* 334:97–102
- Jahani S, Setarehdan SK, Fatemizadeh E (2011) Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators. In: *2011 7th Iranian conference on machine vision and image processing*
- Jahani S, Setarehdan SK, Veronica M (2012) An automatic algorithm for identification and straightening images of curved human chromosomes. *Biomed Eng Appl Basis Commun* 24(6):1–9
- Janani MNS, Nandakumar R (2012) Feature extraction and pairing of G-band chromosome images using K-nearest neighbour classifier. *Int J Comput Sci Telecommun* 3(2):137–140
- Karvelis P, Likas A, Fotiadis DI (2010) Identifying touching and overlapping chromosomes using the watershed transform and gradient paths. *Pattern Recognit Lett* 31:2474–2488
- Khmelinskii A, Ventura R, Sanches J (2008) Automatic chromosome pairing using mutual information. In: *Conference on proceedings of IEEE engineering in medicine and biology society*
- Lerner B (1998) Toward a completely automatic neural-network-based human chromosome analysis. *IEEE Trans Syst Man Cybern Part B Cybern* 28(4):544–552
- Lerner B, Guterman H, Dinstein I, Romem Y (1995) Medial axis transform-based features and a neural network for human chromosome classification. *Pergamon Pattern Recognit* 28(11):1673–1683
- Li Y, Knoll JH, Wilkins RC, Flegel F, Rogan PK (2015) Automated discrimination of dicentric and monocentric chromosomes by machine learning based image processing. *BioDose* 1–2
- Lijiya A, Mumthas TK, Govindan VK (2013) Chromosome classification using M-FISH images. In: *Proceedings of international conference on advances in information technology and mobile communication chromosome*, pp 314–320
- Loganathan E, Anuja MR, Nirmala Madian (2012) Automated Identification to the centromere position and the centromere index (CI) of human chromosome in G-banded images. *Int J Adv Technol Eng Res Autom* 2(6):13–17
- Makkar A, Govindan Professor LAVK, Professor A (2014) A re-segmentation algorithm for improved M-FISH image segmentation and classification. *Int J Adv Inf Sci Technol* 32(32):87–92
- Markou C, Maramis C Automatic chromosome classification using support vector machines. pp 1–24
- Moaallem P, Karimizadeh A, Yazdchi M (2013) Using shape information and dark paths for automatic recognition of touching and overlapping chromosomes in G-band images. *Int J Image Graph Signal Process* 5(5):22–28
- Moradi M, Setarehdan SK (2006) New features for automatic classification of human chromosomes: a feasibility study. *Pattern Recognit Lett* 19:19–28
- Nair RM, Remya RS, Sabeena K (2015) Karyotyping techniques of chromosomes: a survey. *Int J Comput Trends Technol* 22(1)
- Piper J, Granum E (1989) On fully automatic feature measurement for banded chromosome classification. *Cytometry* 10:242–255
- Pravina VA (2015) Survey on techniques used for M-FISH image segmentation for classification of chromosomes. *Middle East J Sci Res* 23(8):1772–1779
- Qiu Y (2013) *Comprehensive performance evaluation and optimization of high throughput scanning microscopy for metaphase chromosome imaging*. University Of Oklahoma, Norman
- Qiu Y, Chen X, Li Y, Chen WR, Zheng B, Li S, Liu H (2013) Evaluations of auto-focusing methods under a microscopic imaging modality for metaphase chromosome image analysis. *Anal Cell Pathol* 36:37–44
- Qiu Y, Song J, Lu X, Li Y, Zheng B, Li S, Liu H (2014) Feature selection for the automated detection of metaphase chromosomes: performance comparison using a receiver operating characteristic method. *Anal Cell Pathol* 2014:1–10
- Ren HL, Li Z, Li Y, Zheng B, Li S, Chen X, Liu H (2014) The impact of the condenser on cytogenetic image quality in digital microscope system. *Anal Cell Pathol (Amst)*. 36:45–59
- Schwartzkopf WC, Bovik AC, Evans BL (2005) Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images. *IEEE Trans Med Imaging* 24(12):1593–1610

39. Shemilt L, Verbanis E, Schwenke J, Estandarte AK, Xiong G, Harder R, Parmar N, Yusuf M, Zhang F, Robinson IK (2015) Karyotyping human chromosomes by optical and X-ray ptychography methods. *Biophysj* 108:706–713
40. Somasundaram D, Kumar VRV (2014) Separation of overlapped chromosomes and pairing of similar chromosomes for karyotyping analysis. *Measurement* 48:274–281
41. Somasundaram D, Nirmala M (2010) Automatic segmentation and karyotyping of chromosomes using bio-metrics. In: International conference on emerging trends in robotics and communication technologies (INTERACT), pp 42–45
42. Somasundaram D, Palaniswami S, Vijayabhasker R, Venkatesakumar V (2014) G-band chromosome segmentation, overlapped chromosome separation and visible band calculation. *Int J Hum Genet* 14(2):73–81
43. Sreejini KS, Lijiya A, Govindan VK (2012) M-FISH karyotyping—a new approach based on watershed transform. *Int J Comput Sci Eng Inf Technol* 2(2):105–117
44. Sri Balaji V, Vidhya S (2015) A novel and maximum-likelihood segmentation algorithm for touching and overlapping human chromosome images. *ARNP J Eng Appl Sci* 10(7): 2777–2781
45. Sri Balaji V, Pragasam G, Sowmiya R, Vijayalakshmi H, Madian N (2012) Segmentation of overlapped and touching human chromosome images. *IOSR J VLSI Signal Process* 1(5):01–06
46. Uttamatani R (2013) Chromosome classification for metaphase selection. In: 13th international symposium on communications and information technologies (ISCIT), pp 464–468
47. Uttamatani R, Yuvapoositanon P, Intarapanich A, Kaewkamnerd S, Phuksaritanon R, Assawamakin A, Tongsimma S (2013) MetaSel: a metaphase selection tool using a Gaussian-based classification technique. *BMC Bioinf* 14:S13
48. Uttamatani R, Yuvapoositanon P, Intarapanich A, Kaewkamnerd S, Tongsimma S (2013) Band classification based on chromosome shapes. In: 13th international symposium on communications and information technologies (ISCIT), pp 464–468
49. Vijayan V, Remya RS, Sabeena K (2015) Survey on chromosome image analysis for abnormality detection in leukemias. *Int J Res Eng Technol* 4(4):664–669
50. Wang X, Li S, Liu H, Wood M, Chen WR, Zheng B (2008) Automated identification of analyzable metaphase chromosomes depicted on microscopic digital images. *J Biomed Inform* 41(2):264–271
51. Wang X, Zheng B, Li S, Mulvihill JJ, Wood MC, Liu H (2008) Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme. *J Biomed Inform* 42:22–31
52. Wenzhong Y (2009) A counting algorithm for overlapped chromosomes. In: 3rd international conference on bioinformatics and biomedical engineering, pp 1–3



Tanvi Arora received her B.Tech degree in Computer Science and Engineering from Punjab Technical University, Jalandhar, Punjab, India, in 2002, and M.Tech degree in Information Technology from Punjab Technical University, Jalandhar, Punjab, India, in 2007. She is currently pursuing PhD in Computer Science and Engineering from Dr. B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab,

India, and working as associate professor in the Department of Computer Science and Engineering at Baddi University of Emerging Sciences and Technology, Baddi, Himachal Pradesh, India. Her teaching and research interests include image processing, pattern recognition, machine learning, data mining and network security.



Dr. Renu Dhir received her B.Tech degree in Electrical Engineering from Guru Nanak Dev Engineering College, Ludhiana, Punjab, India, in 1983, M.Tech degree in Computer Science and Engineering from TIET, Patiala, Punjab, India, in 1997 and PhD degree in Computer Science and Engineering from Punjab University, Patiala, Punjab, India, in 2007. She is currently working

as associate professor at Dr. B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India. Her teaching and research interests include image processing, pattern recognition, machine learning and network security.