



# Preprocessing G-banded metaphase: towards the design of automated karyotyping

R. S. Remya<sup>1</sup> · S. Hariharan<sup>2</sup> · V. Keerthi<sup>1</sup> · C. Gopakumar<sup>1</sup>

Received: 15 October 2019 / Accepted: 21 November 2019 / Published online: 28 November 2019  
© Springer Nature Switzerland AG 2019

## Abstract

Genetic disorders and malignancies due to the chromosomal abnormalities are being researched in cytogenetics till date. G-banded metaphase images are analyzed, chromosomes pairs are identified and arranged into 23 classes as per ISCN ideogram features through karyotyping. This enables the cytogenetic experts to visualize and detect the chromosomal aberrations at ease. Although, the design of a fully automated karyotyping system is difficult, it eliminates the barriers of manual karyotyping. Here, we propose preprocessing techniques for G-banded metaphase images for the design of automated karyotyping system. Our method starts with a decision tree classifier that classifies the input images into analyzable and un-analyzable. Analyzable metaphase images are denoised by median filter and bilateral filter. Denoised images are enhanced using Iterative contrast limited adaptive histogram equalization and are segmented based on contour. Our method ends with an ANN classifier that classifies the segmented images into single straight, bended, touching and overlapped based on the top ten Chi square selected GLCM geometrical features.

**Keywords** Karyotyping · Iterative CLAHE · Contour · Chi square selection

## 1 Introduction

Cytogenetics is the combined study of cytology (study of cells) and genetics (study of inheritance) in which the structure and function of chromosomes are studied in detail. Chromosomes are packed but organized structure containing DNA, which carries genes. In [1–3], it is reported that humans have 23 pairs of chromosomes out of which 22 are autosomes, those responsible for structure and function of human body and the sex chromosome pair, which is responsible for the gender. Chromosomes have direct influence on human health and any changes in the number or structure of chromosomes in any cells may lead to various human disorders like mental retardation, congenital malformations, sterility, sexual abnormalities, spontaneous fetal loss, as specified in [4] or even cancer

as in [5]. Thus, cytogenetics plays an important role in the detection, diagnosis, treatment and prognosis of these human disorders due to chromosome abnormalities.

### 1.1 Karyotyping

Actual collection of chromosomes of living organisms called karyotypes, are examined by the experts through karyotyping. Human chromosome analysis or karyotyping is done manually by cytogenetic experts or physicians in the cytogenetic laboratories. For this process patient's samples from peripheral blood, amniotic fluid or bone marrow are collected and cultured. G (Giemsa) banded metaphase microscopic images are captured since at the metaphase stage of cell division, chromosomes are clearly visible. Experts identify the chromosome pairs and

✉ C. Gopakumar, gopan@ceknpy.ac.in; R. S. Remya, remya@ceknpy.ac.in; S. Hariharan, shhrcc@rediffmail.com; V. Keerthi, keerthyvasudevan.v@gmail.com | <sup>1</sup>College of Engineering Karunagappally (IHRD), Karunagappally, Kerala, India. <sup>2</sup>Regional Cancer Centre, Thiruvananthapuram, Thiruvananthapuram, Kerala, India.



ascertain the numbers. These 1–23 pairs of numbered chromosomes are arranged in a karyogram based on human ideogram published by ISCN as shown in Fig. 1. Ideograms are a schematic representation of chromosomes. They show the relative size of the chromosomes and their banding patterns. By observing the structural and numerical chromosomal aberrations from karyogram, experts diagnose various genetic disorder, malignancies and hematologic disorders. Manual karyotyping is labour intensive and time-consuming task. It also suffers from operator fatigue, human errors etc. Thus, automated chromosome classification is indeed, recommended.

### 1.2 Automated karyotyping system

As mentioned in [6], chromosomes were among the first objects to be studied using automated means, in the biological pattern recognition system due to reasons like sufficient straight forwardness, well defined to be a practical proposition, and being sufficiently monotonous. If the system could be automated then much more cases of chromosome analysis could be undertaken by a laboratory, and an increase in the output can be made even with limited staff. But ideal design and development of fully automated system is a challenging task. Various challenges at each stage of karyotyping is as shown in Fig. 2

Figure 3 shows some of such challenges in automated karyotyping. G banded metaphase images suffer from various noises like sensor noises, stain debris, Gaussian noises etc. These input images may also have unwanted structures like interphase cells. All such images, called noised images, always lead to misclassification, which

inturn, may lead to false interpretation. Such a noisy and low contrast image is shown in Fig. 3a. Even a cytogenetic expert cannot correctly identify the class of chromosomes in such cases. Overlapped chromosomes shown in Fig. 3b, is another crucial challenge in the automated classification since they are partially occluded by other chromosomes and the band information at the overlapping area, cannot be retrieved. Touching chromosomes, as shown in Fig. 3c may confuse the classifiers, as those structures may be interpreted as a single chromosome since the chromosomes are non rigid objects. Clumbed chromosome images as shown in Fig. 3d, are the unanalyzable structures that complicates the entire karyotyping process.

### 2 Related works

Automated method usually comprises of four steps: Pre-processing, Segmentation, Feature extraction and Classification. Generally, preprocessing includes algorithms for denoising and enhancement of input images. Owing to culturing, banding, staining, and imaging, image denoising and enhancement are desirable steps before feature extraction and classification. These methods improve the quality and contrast of images for efficient feature extraction and classification. Various denoising techniques using traditional smoothing and sharpening filters are used by researchers. Authors of [7] proposed a novel human chromosome enhancement algorithm based on cubic spline wavelet transform. In [8], a wavelet based algorithm using multi scale differential operators, has been applied for chromosome image enhancement. Eventhough these

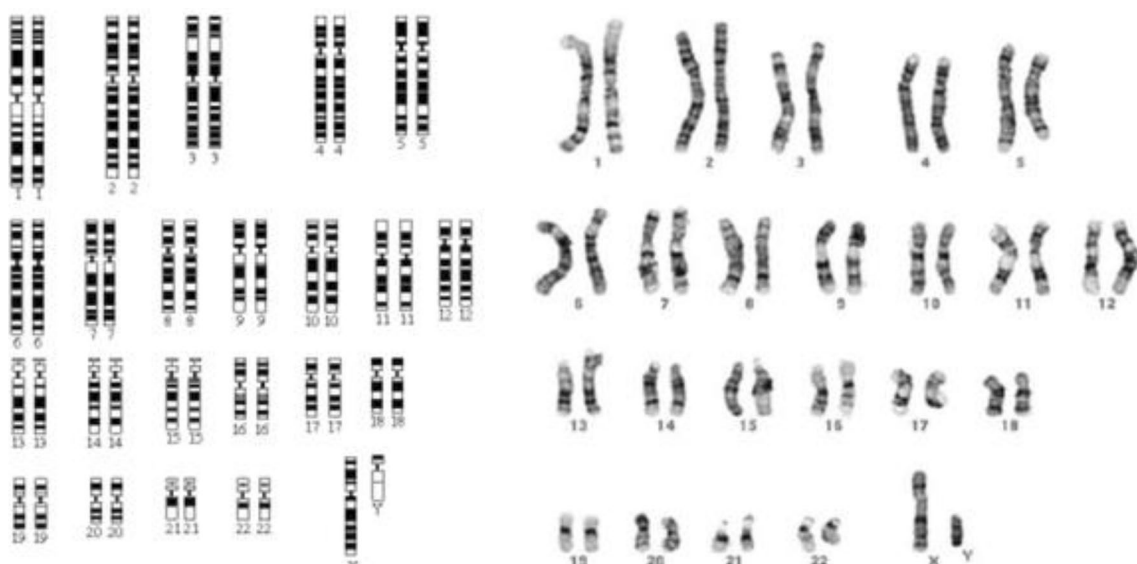
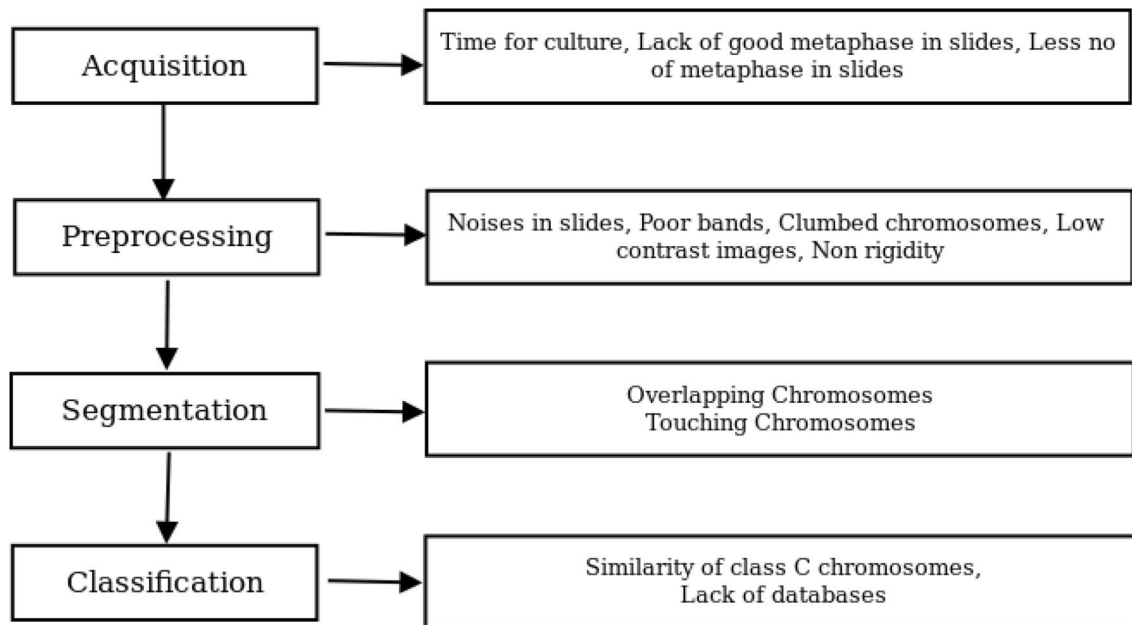
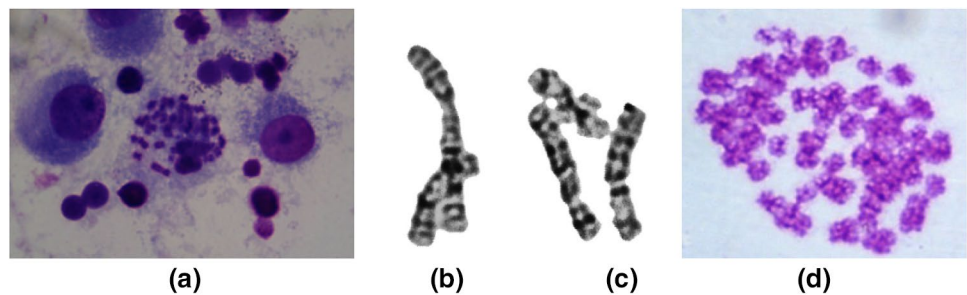


Fig. 1 ISCN human ideogram and a normal karyogram



**Fig. 2** Various challenges in each stage of karyotyping

**Fig. 3** Challenges in automated karyotyping **a** noisy and low contrast image, **b** overlapped chromosomes, **c** touching chromosomes, **d** clumped metaphase



methods improve the quality of features, due to over representation, these methods have high space complexity. In [9], an image enhancement and denoising technique based on structure self-similarity and wavelet transform coefficients has been proposed. In [10], performances of different wavelet families for image enhancement are evaluated based on their Peak Signal to Noise Ratio (PSNR) and the value of Mean Square Error (MSE). Mathematical morphology based enhancement algorithm for chromosome images has been proposed in [11]. Most of the chromosome image enhancement algorithms are reviewed in [12, 13] and found that image enhancement improves not only the display and visualization of chromosomes but also the recognition rate and the accuracy of chromosome classification. In [13], some special methods like oriented wavelets, derived from isotropic laplacian like filters, are also applied in the chromosome images for its enhancement.

Preprocessing based on histogram plays significant role in chromosome image enhancement. [14] has high

Adaptive Contrast Enhancement (ACE) technique for image enhancement. It is based on Histogram Transformation of Local Standard Deviation and uses contrast gains (CGs) for adjusting high frequency components in images. In [15], Chromosome image contrast enhancement using adaptive, iterative histogram matching is discussed. Iterative histogram matching algorithm for chromosome image enhancement based on statistical moments, is proposed in [16]. These methods increase contrast sharply and satisfactorily. The parameters have been chosen adaptively based on the input image to produce even better results and it is the major hindrance of this method. In [17], a method is proposed for the segmentation and removal of interphase cells from chromosome images using multidirectional block ranking. The efficiency of automatic karyotyping decreases with the presence of undivided, condensed mass of chromosomes called interphase cells, stain, debris and other unwanted interferences in the chromosome image. This algorithm segments and

removes these interferences and enhances the accuracy of automated karyotyping.

In segmentation, individual chromosomes are separated as foreground objects from the metaphase spread. A metaphase spread has isolated chromosomes or cluster of touching, partially occluded or overlapping chromosomes. Segmentation, Feature extraction and classification of isolated single straight chromosome is relatively easier. Region labelling, region growing, region merging, and thresholding techniques are adopted by researchers. Here, same label is assigned to all the pixels in an individual chromosome. In [18] similarity based global thresholding techniques are proposed. In [19], segmentation of chromosome images based on recursive watershed algorithm is discussed which has an issue of over segmentation. Active shape models and contour based models for segmentation were also reported [20].

Most of the currently available commercial chromosome classification systems are semi automated and requires human intervention to disentangle the touching and overlapping chromosomes in the metaphase. Another issue is that single isolated chromosomes and overlapping or touching chromosomes demand different segmentation algorithm. Most of the feature extraction and classification algorithms work well for straight chromosomes only. So, erecting bended chromosomes before feature extraction is also desirable. Automated detection of single isolated chromosomes and cluster of touching or overlapping chromosomes has been addressed in the literature. [21] proposed a system to classify the segmented chromosomes into five classes, using geometric features. Correlation-based feature selection (CFS) scheme and Classification via regression (CVR) classifier were respectively used for the feature selection and classification of the objects.

The five categories in this system are straight, overlapping, bent, touching and noise. [22] proposed a system to classify a segmented chromosome as a single chromosome or cluster of overlapping/touching chromosomes. Considering the size they were able to identify single and cluster of chromosomes, and by checking the number of end points, they were able to count the number of chromosomes in a cluster. In [23] a neural network approach is proposed for the automated identification of single chromosomes and blob of chromosomes. Significance of all these preprocessing steps in the design of automated karyotyping system is discussed in [24–26]

### 3 Proposed methods

The proposed methodology for preprocessing G-banded metaphase image, for efficient automated karyotyping, is outlined in Fig. 4 and explained in the following sessions

G-banded microscopic metaphase image collected from the cytogenetic laboratory may suffer from noise, inhomogenous illumination, low contrast etc. Some of the metaphase spread even may not be analyzable by the cytogenetic experts. Since there are sufficient number of metaphase spread from a single slide, the unanalyzable metaphase can be discarded. In this circumstance, an automated technique for classifying the metaphase as analyzable or unanalyzable is a desirable task. Thus analyzable metaphase are identified and are denoised, enhanced, segmented and post classified as single straight chromosome, bended chromosome, touching chromosomes and overlapped chromosomes. Single straight chromosomes can be directly fed into automated karyotyping system but

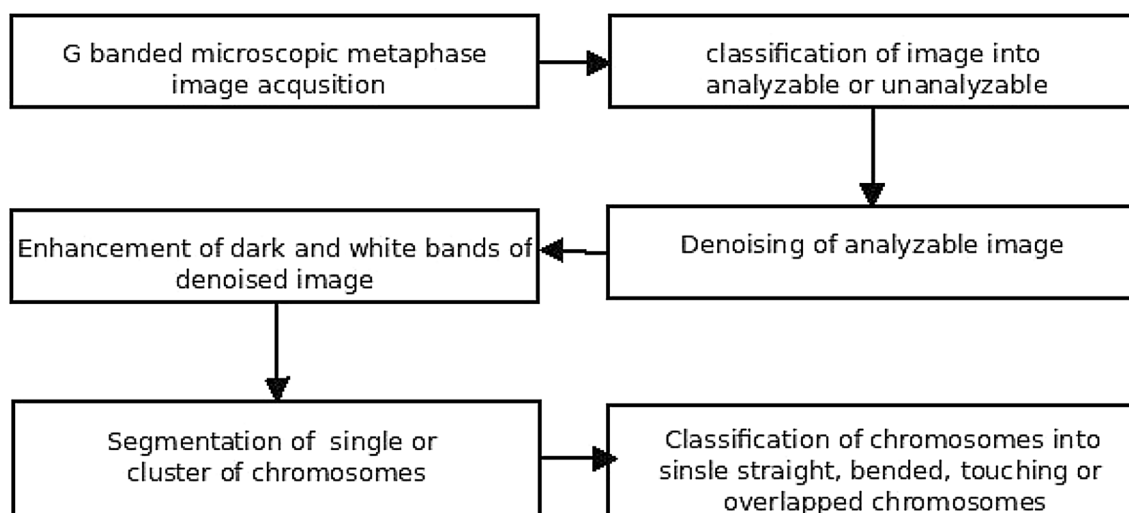


Fig. 4 Proposed methodology

the remaining class should be assigned with further geometrical correction and segmentation techniques.

### 3.1 G banded microscopic metaphase image acquisition

Giemsa stained images (G-banded) are used as the input as in Fig. 5a. G banding, or Giemsa banding is a technique used in cytogenetics to produce a visible karyotype by staining condensed chromosomes. They are then analyzed and classified based on the size and unique G-banding pattern of each chromosome class. Here, input images are captured at Regional Cancer Center, Thiruvananthapuram, Kerala, India. For this, peripheral blood from volunteers are collected. Eight drops of peripheral blood are added to 8 ml supplemented media and 80 ml freshly diluted PHA is added to this culture. This is incubated for 72 h at 37 °C and at the 69th hour 80 ml of Colchicine is added. After incubation, culture tube is centrifuged for 10 min at 800–1000 rpm. After discarding the supernatant by pipetting out the media, resuspended the cell button in 10 ml of hypotonic solution and incubated for 15–20 min at 37 °C. After this, 5 drops of fresh fixative is added. After keeping the tube at room temperature for 5–10 min, tubes are centrifuged. After discarding the supernatant and mixing the pellet thoroughly in 10 ml of fixative, the solution is kept at 4 °C overnight for fixation. After overnight fixation, again the tubes are centrifuged, supernatant is discarded and the cells are resuspended in fresh fixative. After the final centrifugation, the cells are again resuspended in a small volume of fixative approximately 0.5–1 ml, (depending on the size of the cell button) to give a slightly opaque suspension. Thus culture is harvested, slides are prepared and are banded by Trypsin, stained by Giemsa. Such slides are examined under magnification (10× of Leica Microscope) phase objective to check the cell density and spread of metaphase chromosomes. If satisfactory, they are examined under 100× oil emersion in leica DM2900 and

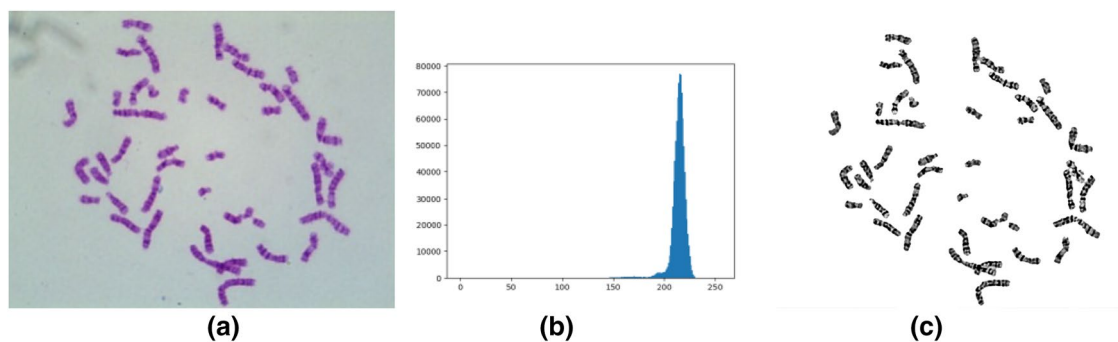
G-banded microscopic metaphase images are captured using leica DMC 2900. Sample image is shown in Fig. 5a.

### 3.2 Preprocessing G-banded metaphase image

As, G banded metaphase microscopic images acquired through cytogenetic procedure fall into two categories namely analyzable and unanalyzable, a classifier is designed to identify the analyzable images for further processing. Here, a simple decision tree is designed to classify the input images into analyzable and unanalyzable classes. For this, features are extracted from region labeled images and are used for image classification. This scheme computes five image features such as number of labelled regions, size of labelled regions, circularity of labelled region, average grey value of labelled region, radial length of each region to the cell center.

As G-banded microscopic images are susceptible to various noises, suppression of the noise from the low-quality images is desirable before the segmentation and classification of the chromosomes. So it is necessary to remove the noise and enhance the bands. Here, a traditional median filter followed by bilateral filter is applied on G-banded images for better denoising. Since the input images suffer from Gaussian noise as illustrated in Fig. 5b, a bilateral filter is proposed as it is a non-linear, edge-preserving, and noise-reducing smoothing filter for images and it replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels. Here the weights are selected based on a Gaussian distribution obtained from the input image. Separation of foreground pixels from background pixels of the input image is done by thresholding. Here the green channel of the metaphase image is Otsu thresholded as the green channel of the input image has higher intensity variation between foreground and background objects in the metaphase spread.

Since the images suffer from inhomogeneous illumination, it is essential to enhance the contrast of the images and improve the visibility of bands. For this purpose, blue



**Fig. 5** **a** Sample G-banded microscopic metaphase, **b** model of Gaussian noise in input image, **c** enhanced image using iterative CLAHE

channel of the metaphase spread is considered since the dark and white bands have comparatively good contrast. A contrast limited adaptive histogram equalization technique (CLAHE) is iteratively applied in the foreground objects of blue channel so that the dark and white bands made more clearly visible. In CLAHE, contrast of the local regions of the image or tiles, is enhanced. Each tile's contrast is enhanced to match a given histogram. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited by specifying a clipping limit to avoid amplifying any noise that might be present in the image. CLAHE overperforms on adaptive iterative histogram matching since in the latter method, the image noise if any present, may also be enhanced. As discussed in [12], the experimental result of CLAHE is shown in Table 1 in which the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM) are the measures of performance. Based on this fact, CLAHE is selected for contrast enhancement of denoised image. The clipping limit of the CLAHE algorithm for G banded metaphase image is experimentally calculated as 20 and the resultant image is shown in Fig. 5c. These denoising and contrast enhancement methods resulted in accurate segmentation, four class classification and karyotyping.

### 3.3 Segmentation and four class classification

Entire objective of karyotyping is to pair and classify 46 chromosomes in the metaphase into 23 classes. So individual chromosome should be segmented from the

metaphase. Here, contour-based segmentation is proposed which yields single or cluster of chromosomes. For chromosome contour extraction, the binary image of the chromosome is convolved with the kernel. Convolved images shows only the boundary pixel with high intensity and this information is used to segment the chromosomes from the CLAHE enhanced blue channel. For this minimum area rectangle enclosing these contours are considered. As chromosomes are non-rigid objects, they are present in different orientations in the metaphase spread. To correct the orientation of the chromosomes, the angle of inclination of these minimum area rectangle is found out and are rotated to align the chromosomes vertically. Such segmented as shown in Fig. 6.

Here, for the feature selection, Chi square technique is applied to identify combined top 10 prominent geometrical features and GLCM features. Selected features are shown in the Table 2.

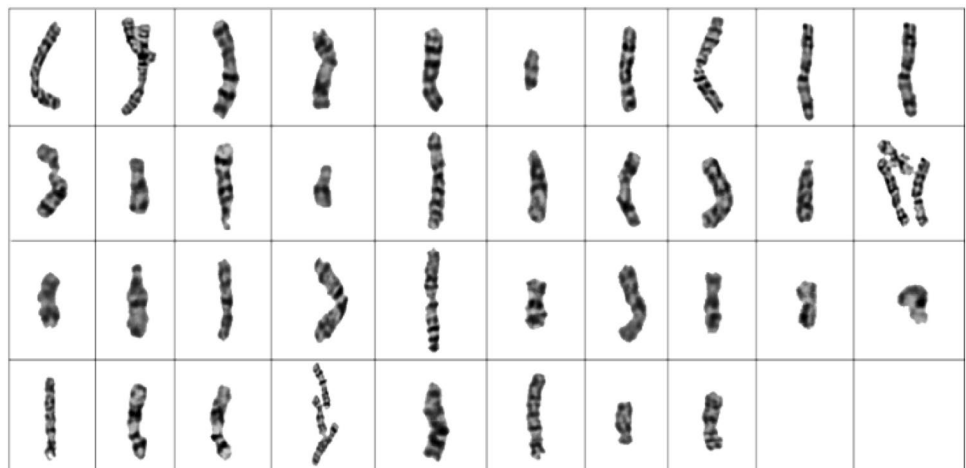
Further, a neural network is designed for four-class classification in which these top 10 features of the segmented chromosomes are fed to 10 input layer neurons to classify the segmented objects into four categories namely, straight single (Fig. 7a), bended (Fig. 7b), touching (Fig. 7c) and overlapped (Fig. 7d). These four classification determines whether the chromosomes should process further or not. In karyograms the single chromosomes are always aligned vertically, so there is no further processing for single straight chromosomes. If the chromosome is bended one, then it should be straightened in order to arrange it into the karyogram. Also in the case of touching and overlapping chromosomes each chromosome image should be separated to arrange them in the form of 23 pairs of chromosomes in karyogram.

The pretrained model is tested by the dataset of 36 chromosomes out of which 23,9,2,2 are the single straight, bended, touching, overlapped, respectively and an accuracy of 91.7% is obtained. Analysis of the models with

**Table 1** Experimental results of CLAHE

Method	PSNR	SSIM
CLAHE	24.5843	0.9703

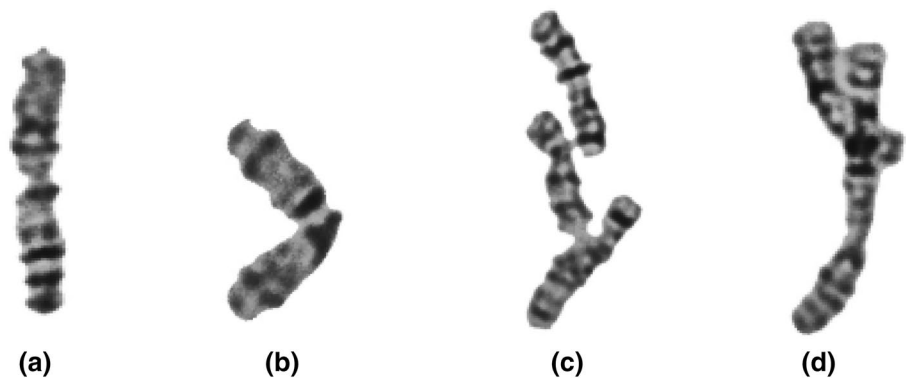
**Fig. 6** Segmented image



**Table 2** Top 10 prominent geometrical and GLCM features selected using Chi square technique

Geometrical features	GLCM features	Top 10 Chi square selected features
Length	Mean	Geometrical
Area	Variance	Area
Convex area	Standard deviation	Convex area
Perimeter	Kurtosis	Eccentricity
Equidiameter	Skewness	Minor axis length
Major axis length	Entropy	Orientation
Minor axis length	Contrast	Solidity
Solidity	Sum of squares and variance	GLCM
Eccentricity	Cluster shade	Mean
Extent	Cluster prominence	Kurtosis
Euler number	Sum of entropy	Entropy
Circularity	Difference of entropy	Length
Orientation	Sum of average	

**Fig. 7** Classification results **a** single straight, **b** bended, **c** touching, **d** overlapped



**Table 3** Four class classification efficiency

Method	Accuracy (%)
Geometrical features	87.1
GLCM features	80.6
Top 10 Chi square selected features	91.7

Geometrical and GLCM features separately and combined is shown in Table 3.

#### 4 Discussion and conclusion

In this paper, preprocessing of G-banded metaphase microscopic image for efficient karyotyping is discussed. Analyzable and unanalyzable images can be identified by using a decision tree classifier using features extracted from region labelled images. After denoising and enhancement of input image, contour based segmentation is proposed that yields both single

chromosomes and cluster of touching or overlapping chromosomes. A four class classification of segmented parts as single straight, bended, touching or overlapped chromosome is proposed, in which top 10 Chi square selected features are used for classification. It is found that the four class classification is having 91.7% accuracy and specific post processing methods and classification techniques can be applied for these classes, for karyotyping. In future, better contrast enhancement techniques and feature selection techniques can be used for improving the accuracy of the classifier. This work can be extended with a five class classifier that includes one class explicitly for classifying interphase cells so that objects belong to that class can be directly eliminated before karyotyping.

**Acknowledgements** Authors are grateful to Engineering and Technology Programme (ETP) of KSCSTE (Kerala State Council for Science, Technology and Environment) (Grant No. ETP/18/2016/KSCSTE) for the financial support and RCC (Regional Cancer Center, Thiruvananthapuram) for the technical support.

## Compliance with ethical standards

**Conflict of interest** The author(s) declare that they have no conflict of interest.

## References

- Therman E, Susman M (1993) Human chromosomes: structure, behavior and effects, 3rd edn. Springer, New York
- Tjio JH, Levan A (1956) The chromosome numbers of man. *Hereditas* 42:1–6. <https://doi.org/10.1111/j.1601-5223.1956.tb03010.x>
- Trask BJ (2002) Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet* 3:769–778
- Shah VC, Murthy DK, Murthy SK (1990) Cytogenetic studies in a population suspected to have chromosomal abnormalities. *Indian J Pediatr* 57:235–243
- Pui C-H, Crist WM, Look AT (1990) Biology and clinical significance of cytogenetic abnormalities in childhood acute lymphoblastic leukemia. *Blood* 76(8):1449–1463
- Britto AP, Ravindran G (2007) A review of cytogenetics and its automation. *J Med Sci* 7:1–18
- Wu Q, Castleman KR (1998) Wavelet-based enhancement of human Chromosome images. In: International conference of the IEEE engineering in medicine and biology society, pp 963–966
- Wang Y-P, Wu Q, Castleman KR et al (2003) Chromosome image enhancement using multiscale differential operators. *IEEE Trans Med Imaging* 22(5):685–693
- Feng J, Yonghua X, Guowen X, Xiong N et al (2010) Image enhancement and denoise based on structure self-similarity and wavelet transform coefficients. In: International conference on mechanic automation and control engineering, pp 6335–6340
- Dubey S, Tiwari D, Singh OP, Dixit A (2013) Performance evaluation of different wavelet families for chromosome image denoising and enhancement. *IOSR J Eng* 3:50–56. <https://doi.org/10.9790/3021-03315056>
- Yan W (2009) Mathematical morphology based enhancement for chromosome images. In: International conference on bioinformatics and biomedical engineering, pp 1–3
- Arsa DMS, Jati G, Santoso A et al (2017) Comparison of image enhancement methods for chromosome karyotype image enhancement. *J Comput Sci Eng* 10(1):50–58
- Yan W (2011) Enhancement methods for chromosome images. In: International conference on electrical and control engineering, pp 3024–3026
- Chang D-C, Wu W-R (1998) Image contrast enhancement based on a histogram transformation of local standard deviation. *IEEE Trans Med Imaging* 17(4):518–531
- Ehsani SP, Mousavi HS, Khalaj BH (2011) Chromosome image contrast enhancement using adaptive, iterative histogram matching. In: 7th Iranian conference on machine vision and image processing, pp 1–5
- Ehsani SP, Mousavi HS, Khalaj BH (2012) Iterative histogram matching algorithm for chromosome image enhancement based on statistical moments. In: 9th IEEE international symposium on biomedical imaging (ISBI), pp 214–217
- Rajaraman S, Vaidyanathan SG, Chokkalingam A (2013) Segmentation and removal of interphase cells from chromosome images using multidirectional block ranking. *Int J Bio-Sci Bio-Technol* 5(3):79–92
- Agam G, Dinstein I (1997) Geometrical separation of partially overlapping nonrigid objects applied to automatic chromosome classification. *IEEE Trans Pattern Anal Mach Intell* 19(11):1212–1222
- Karvelis P, Fotiadis DI, Syrrou MV, Georgiou I, Greece (2005) Segmentation of chromosome images based on a recursive watershed transform. In: The 3rd European Medical and Biological Engineering Conference, vol 11, pp 1727–1983
- Albert PB, Ravindran G (2005) A review of deformable curves from the perspective of chromosome image segmentation. *J Med Sci* 5:363–370
- Arora T, Dhir R (2016) Correlation-based feature selection and classification via regression of segmented chromosomes using geometric features. *Med Biol Eng Comput* 55:1–13
- Sahar S, Setarehdan K, Fatemizadeh E (2011) Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators. In: IEEE Iranian conference on machine vision and image processing
- Rahimi Y, Amirfattahi R, Ghaderi R (2008) Design of a neural network classifier for separation of images with one chromosome from images with several chromosomes. In: Third international conference on broadband communications, information technology biomedical applications, pp 186–190
- Munot MV (2018) Development of computerized systems for automated chromosome analysis: current status and future prospects. *Int J Adv Res Comput Sci*. <https://doi.org/10.26483/ijarcs.v9i1.5436>
- Wang X, Zheng B, Wood M et al (2005) Development and evaluation of automated systems for detection and classification of banded chromosomes: current status and future perspectives. *J Phys D Appl Phys* 38:2536–2542
- Wang X, Zheng B, Li S et al (2009) Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme. *J Biomed Inform* 42(1):22–31

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.