



Research article

Automated identification of single and clustered chromosomes for metaphase image analysis

E-Fong Kao^{a,*}, Ya-Ju Hsieh^a, Chien-Chih Ke^a, Wan-Chi Lin^b, Fang-Yu Ou Yang^b^a Department of Medical Imaging and Radiological Sciences, College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan^b Isotope Application Division, Institute of Nuclear Energy Research, Taoyuan, Taiwan

ARTICLE INFO

Keywords:Computer-aided detection
Chromosome image
Chromosome cluster
Image feature analysis

ABSTRACT

Background: Chromosome analysis is laborious and time-consuming. Automated methods can significantly increase the efficiency of chromosome analysis. For the automated analysis of chromosome images, single and clustered chromosomes must be identified. Herein, we propose a feature-based method for distinguishing between single chromosomes and clustered chromosome.

Method: The proposed method comprises three main steps. In the first step, chromosome objects are segmented from metaphase chromosome images in advance. In the second step, seven features are extracted from each segmented object, i.e., the normalized area, area/boundary ratio, side branch index, exhaustive thresholding index, normalized minimum width, minimum concave angle, and maximum boundary shift. Finally, the segmented objects are classified as a single chromosome or chromosome cluster using a combination of the seven features.

Results: In total, 43,391 segmented objects, including 39,892 single chromosomes and 3,499 chromosome clusters, are used to evaluate the proposed method. The results show that the proposed method achieves an accuracy of 98.92% by combining the seven features using support vector machine.

Conclusions: The proposed method is highly effective in distinguishing between single and clustered chromosomes and can be used as a preprocessing procedure for automated chromosome image analysis.

1. Introduction

Chromosome analyses, such as chromosome number counting, karyotyping [1] and aberration scoring [2,3], are laborious and time-consuming. Automated methods can significantly increase the efficiency of chromosome analysis. In recent years, many methods [4] have been developed to analyze chromosome images automatically. Most methods used to automatically analyze chromosomes are based on a single chromosome. However, chromosome clusters comprising many single chromosomes typically appear in chromosome images, which necessitates a different algorithm for analyzing the chromosome clusters. Hence, single and clustered chromosomes must be identified to perform automated chromosome image analysis.

Studies have been conducted pertaining to methods by which single chromosomes can be distinguished from clustered chromosomes. Some of the relevant studies are briefly introduced below.

* Corresponding author. Department of Medical Imaging and Radiological Sciences, College of Health Sciences, Kaohsiung Medical University, Kaohsiung, 807, Taiwan.

E-mail address: tonykao@kmu.edu.tw (E.-F. Kao).

<https://doi.org/10.1016/j.heliyon.2023.e16408>

Received 10 April 2023; Received in revised form 12 May 2023; Accepted 16 May 2023

Available online 20 May 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Rahimi et al. [5] proposed a neural network classifier with nine features to separate single chromosomes from clustered chromosomes. They used 178 images (109 single chromosomes and 69 chromosome clusters) to evaluate the proposed method. The accuracy of the proposed method was reported to be 73%.

Yan et al. [6] proposed an automatic chromosome-counting algorithm. In this method, the area was used as a criterion to distinguish between single and clustered chromosomes. A segmented object with an area greater than twice the average area was considered a chromosome cluster.

Jahani et al. [7] proposed a method for identifying overlapping and touching chromosomes using morphological operators. The skeleton was extracted by applying a thinning procedure to each segmented object; subsequently, the end points were extracted from the skeleton. Any segmented object with more than two end points was considered a chromosome cluster, and the proposed method was evaluated using 904 segmented objects (776 single chromosomes and 128 clustered chromosomes). The identification accuracies of the clustered and single chromosomes were 96% and 99%, respectively.

Moallem et al. [8] proposed a method similar to that proposed by Jahani et al. [7], which uses the number of end points in the skeleton to identify chromosome clusters. This method was evaluated using 1930 segmented objects (1695 single chromosomes and 235 clustered chromosomes), and the detection accuracies of the clustered and single chromosomes were 86.4% and 98.5%, respectively.

Uttamatani et al. [9] proposed a rule-based classification method for metaphase selection. In this method, the segmented objects were classified into four classes, and the third class was characterized as a chromosome cluster. The classification accuracy of chromosome clusters was 89.44%.

Minaee et al. [10] proposed a geometry-based approach for chromosome segmentation. In the method, three geometric indicators were used to detect chromosome clusters: the ratio of the minor axis length to the major axis length of the surrounding ellipse (surrounding ellipse method), the ratio of the area in the original image to that in its convex hull (convex hull method), and the number of end points in the skeleton for each segmented object.

Kubola et al. [11] proposed a technique for counting chromosome numbers. This method, which is similar to that proposed by Jahani et al. [7], uses the endpoints and intersection points of a skeletonized chromosome image to identify chromosome clusters. The method was evaluated using 300 segmented chromosome images (100 single chromosomes, 100 clusters with two chromosomes, and

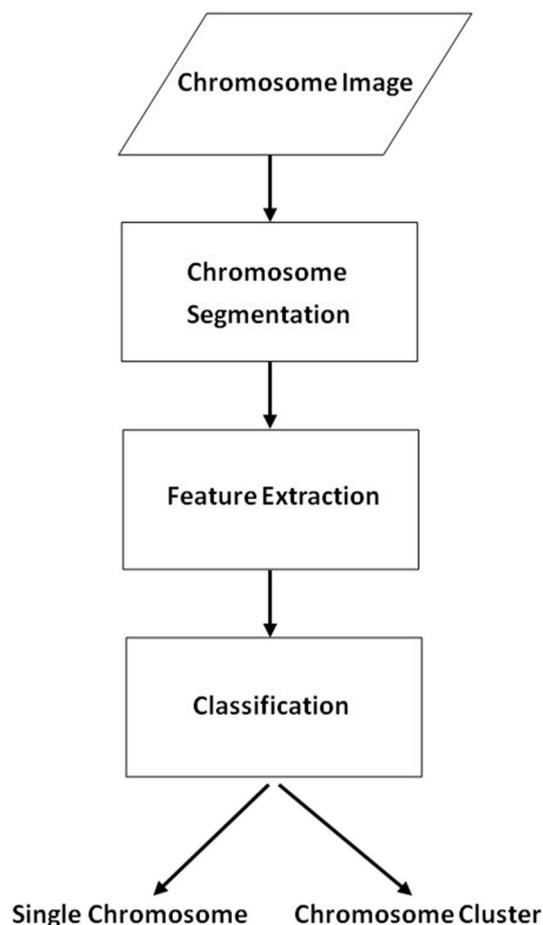


Fig. 1. Overall scheme of the proposed method.

100 clusters with more than two chromosomes). The detection accuracies were 100%, 100%, and 79.12% for single chromosomes, clusters with two chromosomes, and clusters with more than two chromosomes, respectively.

Yilmaz et al. [12] proposed an improved chromosome segmentation method. This method is similar to that proposed by Minaee et al. [10] for detecting chromosome clusters, which includes the convex hull method, the skeletonization method, and a comparison of area and number of objects after a segmented object is eroded.

Arora et al. [13] proposed an approach for chromosome segmentation using region-based active contours. In this method, segmented objects were classified as single or clustered based on their features, including length, area, and circularity.

Lin et al. [14] proposed a framework for identifying chromosome clusters using 11 geometric features and machine learning algorithms. The method was evaluated using 633 samples (528 chromosome instances and 105 chromosome clusters) and demonstrated a classification accuracy of 96.21%. In addition, Lin et al. [15] proposed a method for chromosome cluster-type identification. This method can classify chromosome clusters into touching, overlapping, and touching-overlapping chromosome clusters.

Although many relevant methods have been proposed, how to increase the classification performance of single and clustered chromosomes is still a worthwhile issue. Herein, we propose a novel method that uses seven features to identify single and clustered chromosomes. In our method, some of the features are designed based on concepts inspired by previous studies, whereas others are developed based on our observations of the characteristics of the junction between chromosomes. Five classifiers are used and compared to combine the seven features. The discriminative performances of the seven features and their combination are evaluated using large datasets.

2. Materials and methods

2.1. Image database

The proposed method was evaluated using 1038 metaphase chromosome images obtained from the Institute of Nuclear Energy Research in Taiwan. All chromosome images were of the color JPEG format and were obtained using a microscope (Zeiss, Imager Z2) equipped with an objective lens providing $64\times$ magnification and a scanning system (Metasystems, Metafer4) with an image size of 1360×1024 pixels. A color JPEG image comprises red (R), green (G), and blue (B) components. Each pixel value (R, G, and B) in the color JPEG images was converted into a grayscale value between 0 and 255 for further analysis. Higher and lower pixel values corresponded to the background (bright regions) and foreground (dark regions) in the images, respectively.

2.2. Overall scheme of proposed method

Fig. 1 shows the overall scheme of the proposed method for distinguishing between single and clustered chromosomes. The method comprises three steps. In the first step, chromosome objects are segmented from the metaphase chromosome images in advance. In the second step, seven features are extracted from each segmented object. Based on the extracted features, the segmented objects are classified as either a single chromosome or a chromosome cluster in the final step. The procedures and definitions of these features are described in detail below.

2.3. Chromosome segmentation

A chromosome image and its histogram in terms of the grayscale values are shown in Fig. 2(a) and (b), respectively. The histogram

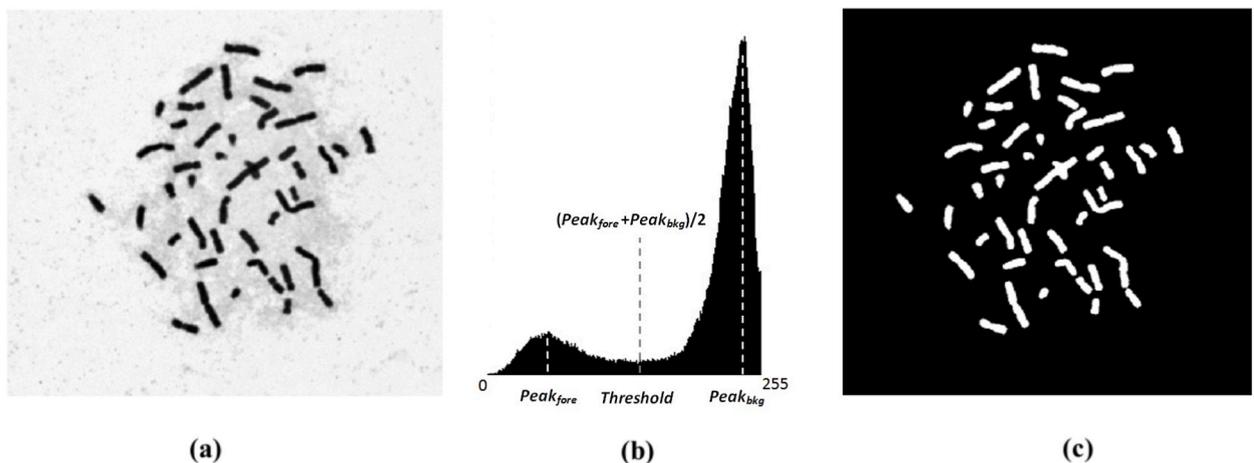


Fig. 2. (a) Metaphase chromosome image; (b) histogram of Fig. 2(a) and threshold used to segment the chromosomes; and (c) segmented objects corresponding to Fig. 2(a).

contains two major clusters: one with a lower value corresponding to the chromosomes or foreground, and the other with a higher value corresponding to the background. The threshold between the two clusters must be determined in advance to segment the chromosomes. In this study, the threshold was determined by the mean value of the two clusters, as shown in Fig. 2(b). For the chromosomes, a grayscale value with the maximum number $Peak_{fore}$ was searched between 0 and 100. Subsequently, $Peak_{bkg}$ was searched between 150 and 255 for the background. The threshold was calculated as $(Peak_{fore} + Peak_{bkg})/2$. A binary image of the chromosome candidates was obtained by segmenting pixels with values less than the threshold, as shown in Fig. 2(c). Subsequently, the segmented objects, including single and clustered chromosomes, were classified based on their features.

2.4. Features for distinguishing between single and clustered chromosomes

After the chromosomes were segmented, each segmented object was rotated toward the vertical position along its long axis for feature extraction, as shown in Fig. 3(a)–3(c). In each metaphase chromosome image, the segmented objects were organized based on their areas. The middle-sized segmented object was selected and used for normalization. Fig. 4 shows selected chromosome with a median size for normalization in a metaphase chromosome image. The seven features used in this study were designed based on the characteristics that differed between the single and clustered chromosomes. Because the correlation between those features is low, their combination can compensate for one another to improve the discrimination performance. The definitions of the seven features extracted from the segmented object are presented next.

2.4.1. Normalized area

The size of segmented objects is an important feature for distinguishing between single and clustered chromosomes. The area of a chromosome cluster is typically larger than that of a single chromosome, as shown in Fig. 5(a)–5(d). In this study, the area of each segmented object was normalized to the area of the middle-sized chromosome (Fig. 4) in each metaphase image. The normalized area A_n is formulated using Eq. (1) as follows:

$$A_n = A_i / A_{mid} \quad (1)$$

where A_i is the area of the segmented object and A_{mid} is the area of the middle-sized chromosome. For a single chromosome, the normalized area A_n is approximately 1.0. By contrast, the normalized area of a chromosome cluster deviates significantly from 1.0.

2.4.2. Area/boundary ratio

For a chromosome cluster, some portions of the boundary vanish in the overlapping region, as shown in Fig. 6(a)–6(c). To detect this phenomenon, the area/boundary ratio $R_{a/b}$ can be used, which can be formulated using Eq. (2) as follows:

$$R_{a/b} = A_i / N_{boundary} \quad (2)$$

where A_i is the number of pixels in the segmented object area and $N_{boundary}$ is the number of pixels in the segmented object boundary. The $R_{a/b}$ for a chromosome cluster is larger than that for a single chromosome owing to the fewer boundary pixels of the former.

2.4.3. Side branch index

Based on our observations of overlapping chromosomes, when one chromosome is rotated vertically along its long axis, the other chromosomes are represented as side branches, as shown in Fig. 7(a) and (b). This information is valuable for detecting overlapping chromosomes. Hence, we used it to define a side branch index to detect overlapping chromosomes. To determine the index, each segmented object is rotated toward the vertical position in advance, as shown in Fig. 3, and the maximum width W_{max} is obtained by searching the widths along the long axis, as shown in Fig. 7(b). Subsequently, the side branch index W_{branch} is formulated using Eq. (3) as follows:

$$W_{branch} = W_{max} / W_{mid} \quad (3)$$

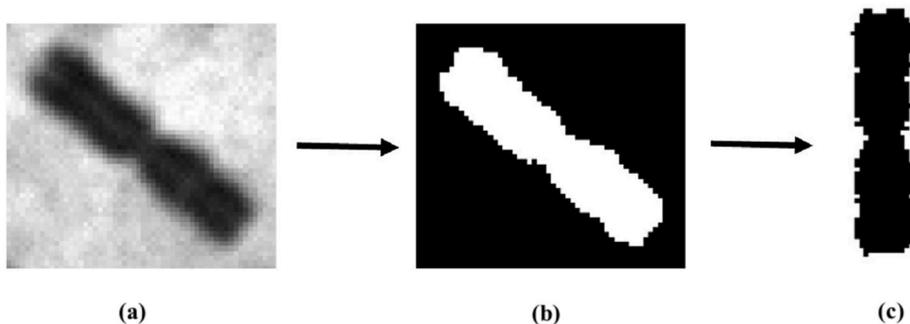


Fig. 3. Segmented object rotated toward vertical position for feature extraction.

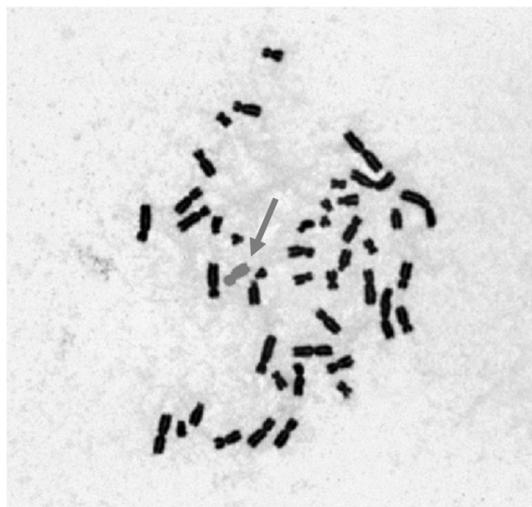


Fig. 4. Middle-sized chromosome (indicated by the arrow) used for normalization in metaphase image.

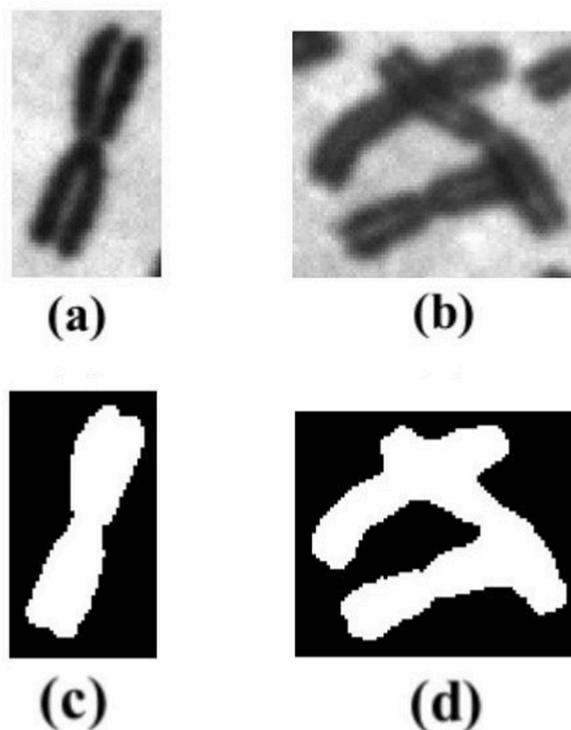


Fig. 5. Size comparison between single chromosome (left) and chromosome cluster (right).

where W_{max} is the maximum width of the segmented object and W_{mid} is the maximum width of the middle-sized chromosome (Fig. 4) in each metaphase image. W_{branch} refers to the maximum width of each segmented object normalized to the maximum width of the middle-sized chromosome. For a single chromosome, the value of W_{branch} is approximately 1.0. By contrast, the W_{branch} value for overlapping chromosomes with a “side branch” is significantly higher than 1.0.

2.4.4. Exhaustive thresholding index

For a large portion of the touching chromosomes, the boundary between the chromosomes is clearly visible, as shown in Fig. 8(a). Clustered chromosomes with clear boundaries can be further separated via exhaustive thresholding, as shown in Fig. 8(c). In exhaustive thresholding, the threshold value is decreased at each iteration to determine whether the segmented object is divided into

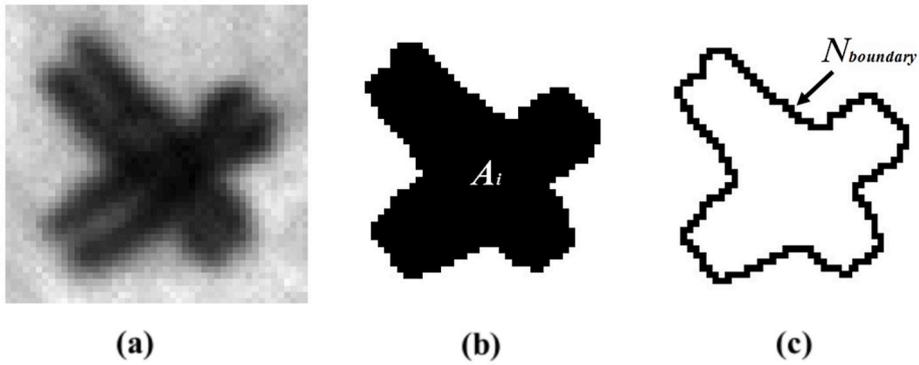


Fig. 6. (a) Overlapping chromosomes and area/boundary ratio determined by (b) number of area pixels A_i divided by (c) number of boundary pixels $N_{boundary}$.

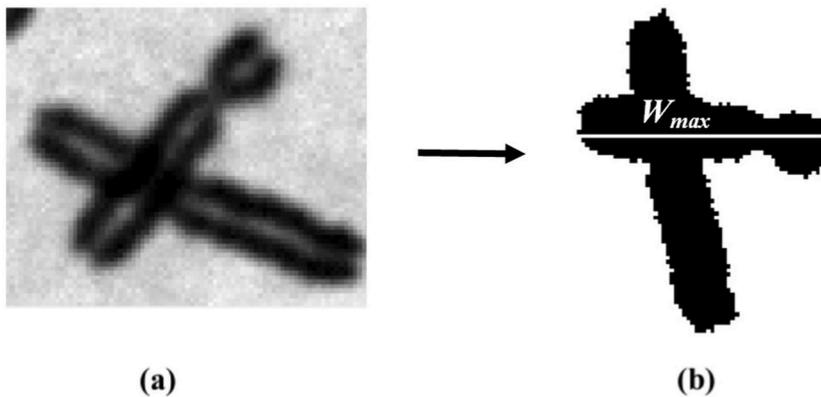


Fig. 7. (a) Chromosome cluster segmented and rotated for determining (b) the maximum width W_{max} in the vertical position.

more than one segment, as shown in Fig. 8(b) and (c). In this study, the minimum threshold value used to separate the potential single chromosome was calculated using formula $0.8 \times Peak_{fore} + 0.2 \times Peak_{bkg}$, as demonstrated similarly in Fig. 2(b). If a segmented object can be divided into many pieces via exhaustive thresholding, then it can be considered a chromosome cluster. However, some single chromosomes can be separated into many pieces via exhaustive thresholding, as shown in Fig. 8(d)–8(f). In this study, we defined an exhaustive thresholding index to overcome this problem. For a single chromosome divided into several pieces, the distance between the pieces on the chromosome is small. In contrast to the inner distance, the distance between touching chromosomes is relatively large. Based on this concept, the distance D_{mass} between the centers of mass of the divided pieces was used to distinguish between single and clustered chromosome, as shown in Fig. 8(g). The exhaustive thresholding index I_{ET} is defined as shown in Eq. (4):

$$I_{ET} = D_{mass} / W_{mid} \quad (4)$$

where D_{mass} is the distance between the mass centers of the divided pieces and W_{mid} is the maximum width of the middle-sized chromosome described above. If the final number of pieces is 1 (no further division), then the value of D_{mass} is set to zero. If the number of divided pieces is greater than two, then the value of D_{mass} is the average of all distances between the mass centers. The exhaustive thresholding index I_{ET} indicates that the distance D_{mass} is normalized to the width of the middle-sized chromosome in each metaphase image. For a single chromosome, the value of I_{ET} is less than 1.0. By contrast, the I_{ET} value exceeds 1.0 for chromosome clusters with divided pieces.

2.4.5. Normalized minimum width

Touching chromosomes without a clear boundary cannot be detected via exhaustive thresholding (as described above). Other features must be designed to detect clustered chromosomes in which the chromosomes merge. Hence, the junction characteristics of the touching chromosomes were used to design the following features.

When chromosomes are in contact without a clear boundary, the junction between chromosomes typically appears narrow, as shown in Fig. 9(a). The minimum width of a segmented object can be used to detect junctions between chromosomes. To achieve this, each segmented object was rotated toward the vertical position, and the minimum width W_{min} was determined by searching along the long axis, as shown in Fig. 9(b). The minimum width W_{min} was subsequently normalized to the middle-sized chromosome 's width

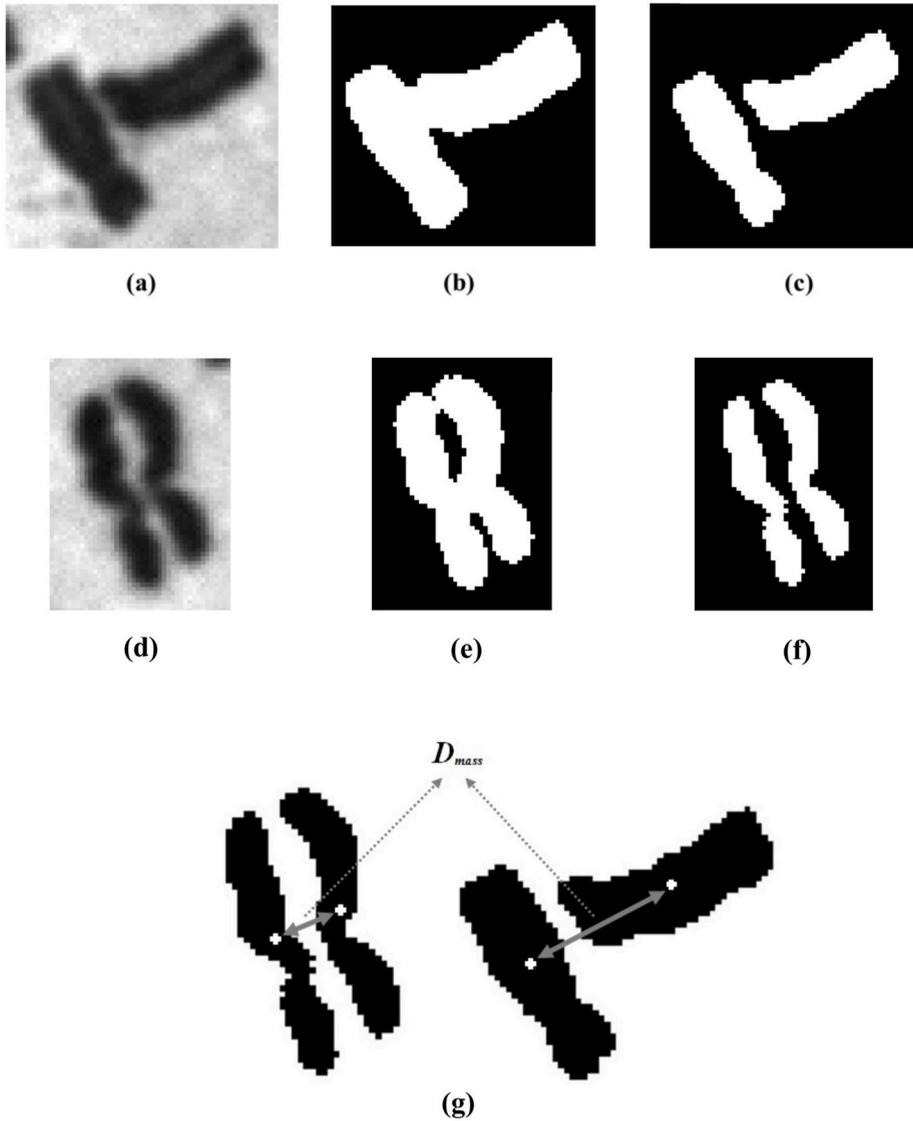


Fig. 8. (a) Touching chromosomes with visible boundary and (b) the corresponding segmentation and (c) exhaustive thresholding result; (d) single chromosome with an internal fissure and (e) its segmentation and (f) exhaustive thresholding result; (g) distance D_{mass} between centers of mass of the divided pieces.

(W_{mid}) to obtain the normalized minimum width W_n , which is expressed by Eq. (5) as follows:

$$W_n = W_{min} / W_{mid} \quad (5)$$

The normalized minimum width W_n can provide information regarding the narrowest section of a segmented object.

2.4.6. Minimum concave angle

In addition to the junction between the chromosomes in a cluster, as shown in Fig. 10(a), the centromere [16] is the narrowest section of a single chromosome, as shown in Fig. 10(c). A concave angle near the minimum width was used to distinguish the junction from the centromere. The nearby angle for a centromere is typically blunt (obtuse angle), as shown in Fig. 10(d), and the angle near the junction is typically an acute angle, as shown in Fig. 10(b). These concave angles provide more information regarding the minimum width. We defined an index, i.e., the minimum concave angle θ_{min} , which was determined by selecting the minimum angle from the left and right concave angles, by Eq. (6) as follows:

$$\theta_{min} = \min(\theta_{left}, \theta_{right}) \quad (6)$$

The minimum concave angle θ_{min} and normalized minimum width W_n described above can be used to detect narrow junctions

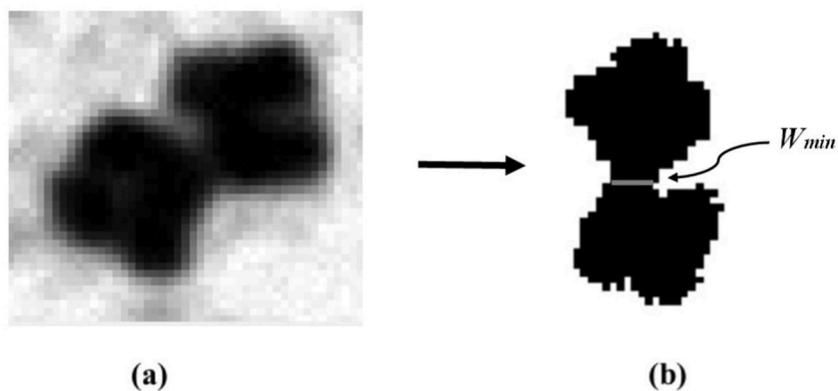


Fig. 9. (a) Chromosome cluster segmented and rotated for determining (b) the minimum width W_{min} in the vertical position.

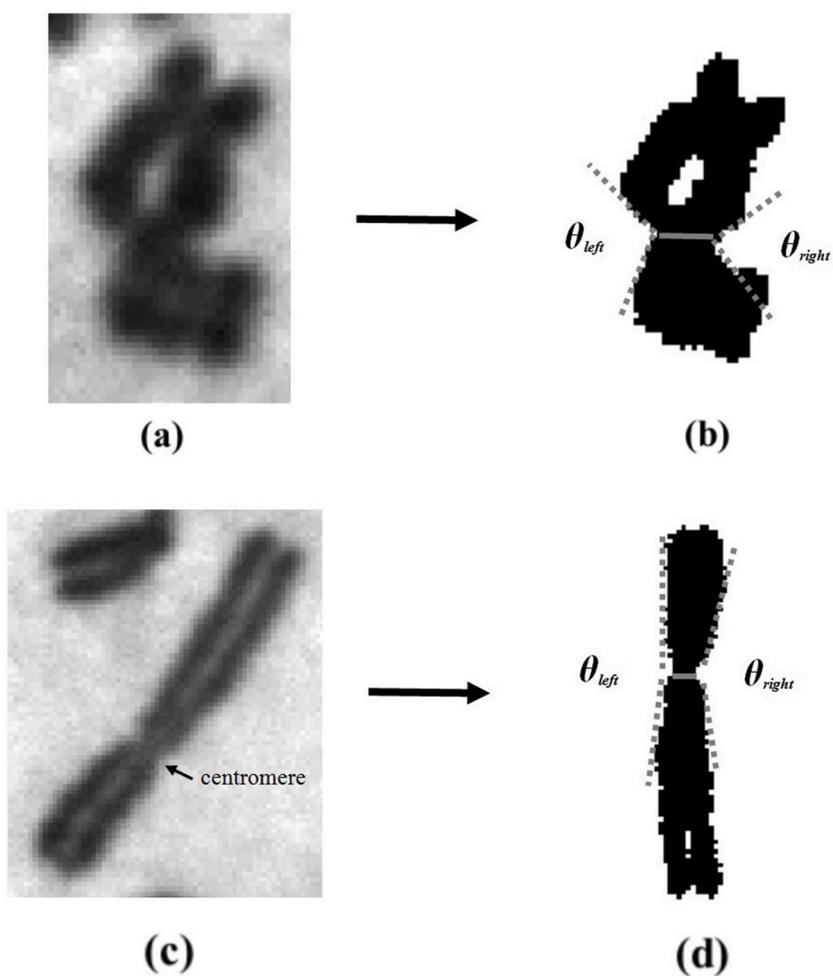


Fig. 10. Narrowest sections corresponding to (a) the junction between touching chromosomes and (c) the centromere of a single chromosome; the concave angles near the minimum width for a (b) chromosome cluster and (d) single chromosome.

between the touching chromosomes.

2.4.7. Maximum boundary shift

A distance shift was observed along the boundary in the junction between chromosomes, as shown in Fig. 11(a)–11(c). The

boundary was smooth for a single chromosome. By contrast, the boundary of a chromosome cluster showed an abrupt displacement near the junction. The shifting distance can be a useful indicator for detecting junctions. Hence, we defined another index, i.e., the maximum boundary shift S_n . To obtain the index, each segmented object was rotated vertically, and the maximum shifting distances of the left and right boundaries were determined, as shown in Fig. 11(c). The maximum S_{max} was selected from the left and right maximum boundary shifts using Eq. (7).

$$S_{max} = \max(S_{max}^{left}, S_{max}^{right}) \quad (7)$$

Subsequently, the maximum shift S_{max} was normalized to the middle-sized chromosome's width (W_{mid}) to obtain the maximum boundary shift S_n using Eq. (8) as follows:

$$S_n = S_{max}/W_{mid} \quad (8)$$

A segmented object is considered a chromosome cluster if its S_n value is high.

2.5. Classification

After feature extraction, all the segmented objects were classified as either single or clustered chromosomes based on the features described above. In this study, five classifiers were used to classify the segmented objects: the decision tree (DT), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM), and multi-layer perceptron (MLP). All classifications were performed using a machine learning tool (scikit-learn) [17].

2.6. Evaluation method

All segmented objects in the 1038 chromosome images, as described in Section 2.3, were classified as single or clustered chromosomes by two experts. In total, 43,391 segmented objects were divided into 39,892 single chromosomes and 3,499 clustered chromosome. The classification results were used as the gold standard for evaluating the proposed method.

The dataset was randomly separated into training and testing data sets. A training dataset was used to train each classifier and create the model. After the model was established, the test set was sent to the model to obtain the classification accuracy. The performance of each classifier was evaluated via five-fold cross-validation.

3. Results

In this study, seven features were designed and used to classify the segmented objects into single and clustered chromosomes. The effectiveness of each feature was evaluated individually. Fig. 12(a)–12(g) show the distributions of the single and clustered chromosomes in terms of the feature values. The mean and standard deviation of each feature for the single and clustered chromosomes are summarized in Table 1. The values of A_n , $R_{a/b}$, W_{branch} , I_{ET} , W_n , θ_{min} , and S_n were 0.99 ± 0.42 and 2.29 ± 0.94 , 3.52 ± 0.79 and 4.15 ± 0.58 , 1.16 ± 0.24 and 2.23 ± 0.87 , 0.02 ± 0.15 and 1.42 ± 1.07 , 0.71 ± 0.17 and 0.50 ± 0.27 , 125.33 ± 28.00 and 83.01 ± 39.20 , and 0.20 ± 0.18 and 1.19 ± 0.78 for the single and clustered chromosomes, respectively. To some degree, all features can distinguish between single and clustered chromosomes.

To combine the seven features for classification, five classifiers (DT, LDA, QDA, SVM, and MLP) were used and compared. Table 2 presents a comparison of the classification performances of the five classifiers. The results show that the SVM with seven features demonstrated the best classification performance, with the highest accuracy of 98.92%. After training the SVM classifier, a hyperplane for separating the single and clustered chromosomes was obtained. The distance between each segmented object and the separating hyperplane is shown in Fig. 13. Single and clustered chromosomes can be separated successfully using an SVM with a combination of the seven features.

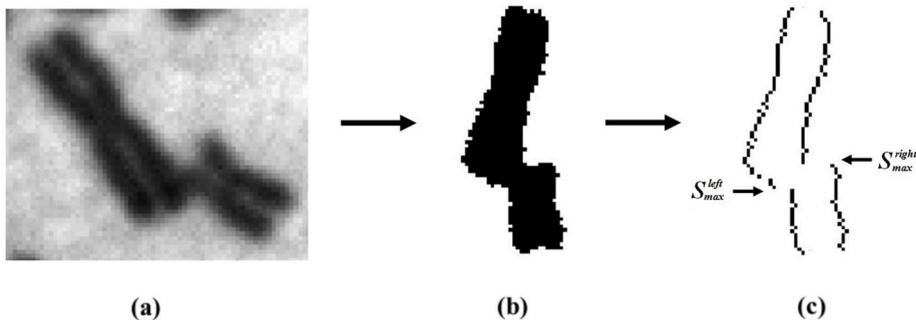


Fig. 11. (a) A chromosome cluster (b) segmented and rotated for determining (c) the maximum boundary shift in the vertical position.

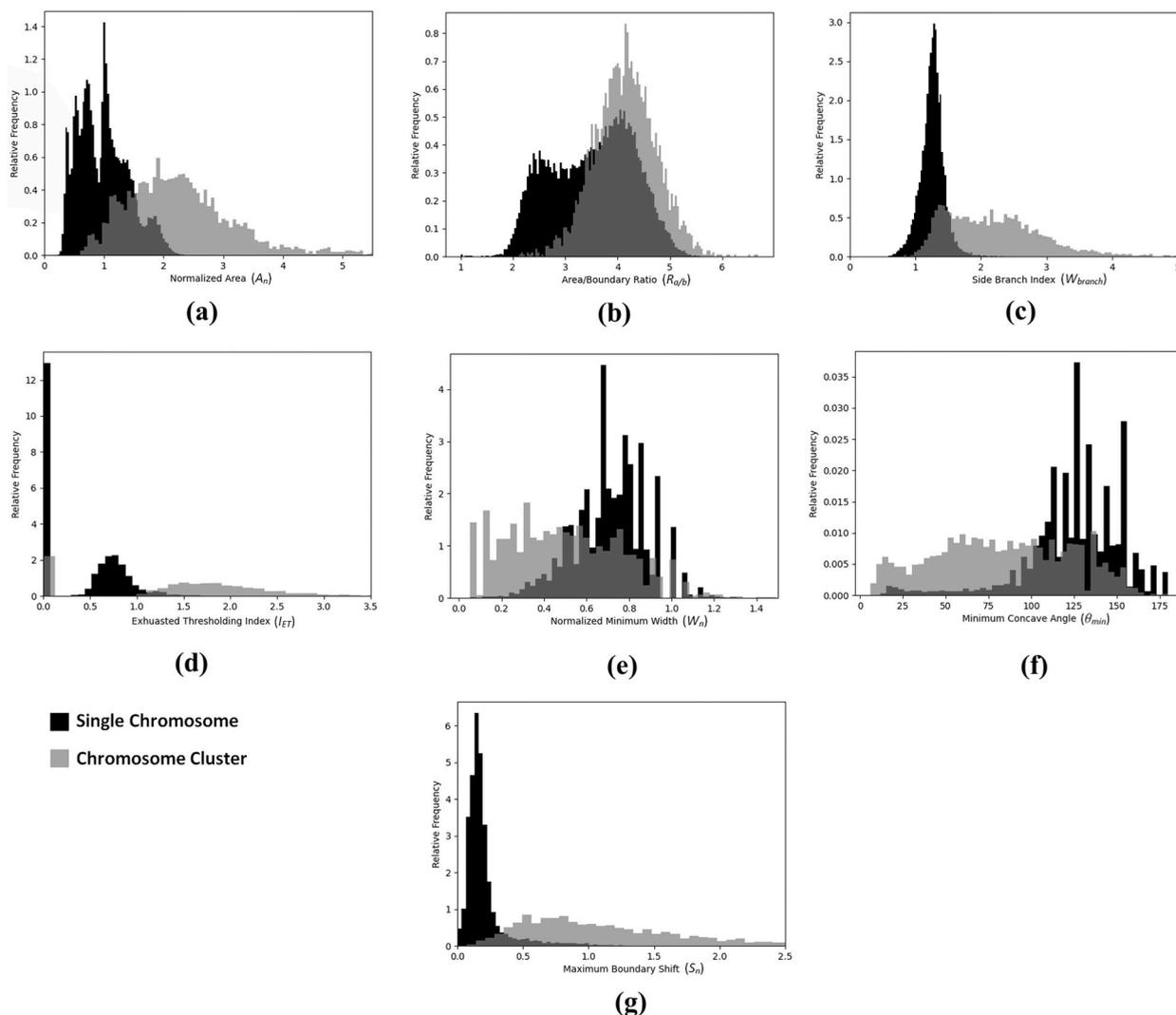


Fig. 12. Distributions of single and clustered chromosomes in terms of the features: (a) Normalized area, (b) area/boundary ratio, (c) side branch index, (d) exhaustive thresholding index, (e) normalized minimum width, (f) minimum concave angle, and (g) maximum boundary shift.

Table 1

Mean and standard deviation of the features for single and clustered chromosomes.

| Feature | Single Chromosome | Chromosome Cluster |
|--|--------------------|---------------------|
| Normalized Area (A_n) | 0.99 ± 0.42 | $2.29 \pm 0.94^*$ |
| Area/Boundary Ratio ($R_{a/b}$) | 3.52 ± 0.79 | $4.15 \pm 0.58^*$ |
| Side Branch Index (W_{branch}) | 1.16 ± 0.24 | $2.23 \pm 0.87^*$ |
| Exhaustive Thresholding Index (I_{ET}) | 0.02 ± 0.15 | $1.42 \pm 1.07^*$ |
| Normalized Minimum Width (W_n) | 0.71 ± 0.17 | $0.50 \pm 0.27^*$ |
| Minimum Concave Angle (θ_{min}) | 125.33 ± 28.00 | $83.01 \pm 39.20^*$ |
| Maximum Boundary Shift (S_n) | 0.20 ± 0.18 | $1.19 \pm 0.78^*$ |

* Significant difference with single chromosome ($p < 0.05$).

4. Discussion

In image feature analysis, the discriminating capability of the features significantly affects the classification results. In this study, we designed seven features based on the characteristics of a single chromosome and chromosome cluster. The usefulness of each feature was analyzed individually as follows: The results of the normalized area A_n indicate that the chromosome cluster was larger than a single chromosome. This shows that size is an important feature for distinguishing chromosome clusters from single

Table 2
Classification results of the classifiers.

| Classifier | Classification Accuracy (%) |
|---------------------------------------|-----------------------------|
| Decision Tree (DT) | 98.23 ± 0.16 |
| Linear Discriminant Analysis (LDA) | 97.90 ± 0.07 |
| Quadratic Discriminant Analysis (QDA) | 96.08 ± 0.18 |
| Support Vector Machine (SVM) | 98.92 ± 0.06 |
| Multi-layer Perceptron (MLP) | 98.68 ± 0.11 |

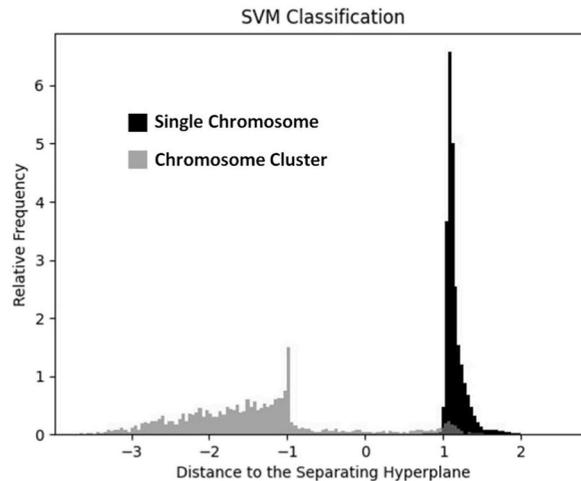


Fig. 13. Distributions of single and clustered chromosomes in terms of the distance to the separating hyperplane of SVM classifier.

chromosomes. The $R_{a/b}$ was larger for the chromosome clusters because of the smaller boundary pixels due to the disappearance of the boundary in the overlapping region. The results of the side branch index W_{branch} show that the value for a single chromosome was approximately 1.0, whereas the value for the chromosome clusters deviated significantly from 1.0. This implies that widths (side branches) of chromosome clusters are typically larger than those of single chromosomes. For the exhaustive thresholding index I_{ET} , the results show that the distance between the divided pieces inside a single chromosome was small, whereas it was relatively large for the divided chromosomes from the clusters. Based on the normalized minimum width W_n and minimum concave angle θ_{min} , the results show that the junction between chromosomes indicated a narrow width and an acute angle. By contrast, for the narrowest region (the centromere) of a single chromosome, the nearby angle was large. To detect abrupt displacements near the junction, the maximum boundary shift S_n was determined; the result shows that the S_n value for the chromosome cluster was higher than that for a single chromosome. This implies that the boundary of a single chromosome is smooth, whereas that of a chromosome cluster typically exhibits significant shift along the boundary. In summary, all the designed features can be used to distinguish between single and clustered chromosomes.

To combine the seven features for distinguishing clustered chromosome from single chromosome, we used five classifiers and compared their classification performances. The results show that the SVM performed the best, with an accuracy of 98.92%. In addition, the MLP and DT exhibited competitive results. This implies that the MLP and DT are also suitable classifiers for performing classification using the feature space comprising the seven features. The comparisons between the proposed and related methods are presented in Table 3. The proposed method was evaluated using a large data set and showed competitive performance compared with the related methods.

In this study, size- or length-related features were normalized to the size or width of the middle-sized chromosome in each metaphase image. Chromosome clusters with excessively large or small feature values can be detected more easily compared with the middle-sized chromosome. For example, the values of A_n and W_{branch} are approximately 1.0 for single chromosomes, and the segmented objects can be regarded as chromosome clusters if the feature values deviate significantly from 1.0. In addition, metaphase chromosome images can be acquired using microscope systems equipped with objective lens of different magnifications or scanning systems with different image sizes. This poses problems to size-related features, which can be solved through normalization. Hence, by adopting normalization, the proposed method can be readily applied to metaphase chromosome images obtained at different resolutions or magnifications.

To overcome the limitations of the proposed method, misclassified cases were analyzed. Fig. 14 shows the primary types of single chromosomes misclassified as chromosome clusters, including highly bent chromosome (Fig. 14(a) and (b)), double-stranded chromosomes with long arms (Fig. 14(c)), and long chromosomes with middle fissures (Fig. 14(d)). For the highly bent and long-arm single chromosomes, the maximum width W_{max} was overly large, which resulted in an extremely high value of the side branch index W_{branch} .

Table 3
Comparison with related studies.

| Method | Data set | | Accuracy |
|------------------------|----------|----------------------------|---|
| | Single | Cluster | |
| Rahimi et al. [5] | 109 | 69 | 73% |
| Jahani et al. [7] | 776 | 128 | 99% for single 96% for cluster |
| Moallem et al. [8] | 1,695 | 235 | 98.5% for single 86.4% for cluster |
| Uttamatinin et al. [9] | 6,409 | 994 | 89.44% for cluster |
| Kubola et al. [11] | 100 | 100 (n = 2) 100 (n > 2) | 100% for single 100% for cluster (n = 2) 79.12% for cluster (n > 2) |
| Lin et al. [14] | 528 | 105 | 96.21% |
| Proposed Method | 39,892 | 3,499 | 98.92% |

n: number of chromosomes in a cluster.

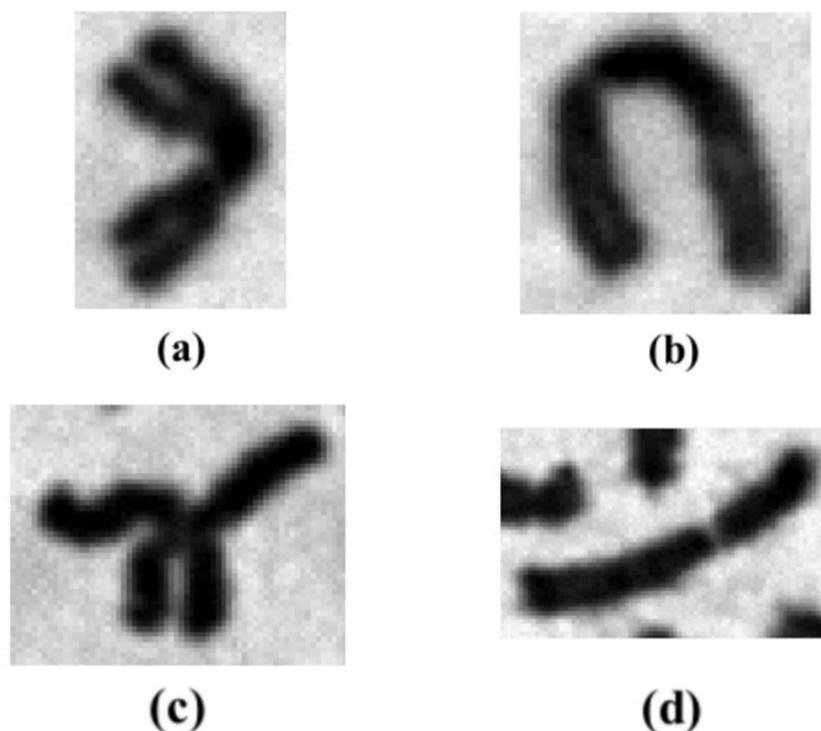


Fig. 14. Single chromosome misclassified as chromosome cluster by the proposed method.

Therefore, these types of single chromosomes were classified as clustered chromosomes. For a long single chromosome with a middle fissure, the segmented object was divided into two pieces via exhaustive thresholding. In this study, the distance between pieces was used to verify whether the pieces belonged to a single chromosome, based on the criterion that the distance is small. However, the distance between the mass centers of the pieces for a long single chromosome is relatively large, and these pieces are considered separate chromosomes. This causes the segmented object of a long chromosome to be misclassified as a cluster during exhaustive thresholding.

By contrast, Fig. 15 shows the primary types of chromosome clusters misclassified as a single chromosome, including small chromosomes located in close proximity to each other (Fig. 15(a) and (b)) and chromosomes overlapping in a straight line (Fig. 15(c) and (d)). As described above, the distance between the divided pieces was used to determine whether the pieces belonged to a single chromosome or separate chromosomes for a segmented object in the exhaustive thresholding. For a segmented object in which small chromosomes are located in close proximity to each other, the distance between chromosomes separated by exhaustive thresholding is small. This causes the segmented object of the cluster to be misclassified as a single chromosome. In addition, the segmented object of a cluster, in which the chromosomes overlap in a straight line, is extremely similar to that of a single chromosome. If the area is not sufficiently large, then the segmented object of a cluster will be misclassified as a single chromosome because of the similar values of

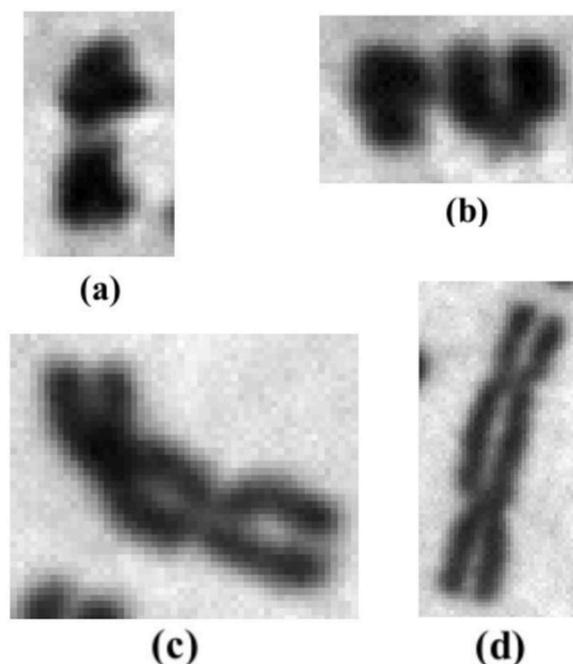


Fig. 15. Chromosome cluster misclassified as single chromosome by the proposed method.

the other features compared with those of a single chromosome.

Although the proposed method exhibits satisfactory performance, some aspects can be considered for further improvement. Highly bent single chromosomes and double-stranded chromosomes with long arms exhibit features such as "side branches" and have high W_{branch} values. The features of bent and long-arm single chromosomes must be designed appropriately to distinguish them from chromosome clusters. For the exhaustive thresholding index, we used only the distance between the divided pieces to distinguish between single and clustered chromosomes. This is insufficient for small chromosomes located in close proximity to each other and long single chromosomes with middle fissures. Other features of the divided pieces must be designed appropriately to accommodate these situations. However, the detection of clustered chromosomes that overlap in a straight line appears to be the most challenging task in the future.

5. Conclusions

The proposed method is highly effective in distinguishing between single and clustered chromosomes; thus, it can be used as a preprocessing procedure for automated chromosome image analysis.

Author contribution statement

E-Fong Kao: Conceived and designed the experiments; Wrote the paper.
 Ya-Ju Hsieh, Chien-Chih Ke: Analyzed and interpreted the data.
 Wan-Chi Lin, Fang-Yu Ou Yang: Contributed reagents, materials, analysis tools or data.

Funding statement

This work did not receive any specific funding or grant.

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C. O'Connor, Karyotyping for chromosomal abnormalities, *Nature Edu.* 1 (1) (2008) 27.
- [2] H.M. Swartz, B.B. Williams, A.B. Flood, Overview of the principles and practice of biodosimetry, *Radiat. Environ. Biophys.* 53 (2014) 221–232.
- [3] H. Romm, E. Ainsbury, S. Barnard, L. Barrios, J.F. Barquintero, C. Beinke, M. Deperas, E. Gregoire, A. Koivistoinen, C. Lindholm, J. Moquet, U. Oestreicher, R. Puig, K. Rothkamm, S. Sommer, H. Thierens, V. Vandersickel, A. Viral, A. Wojcik, Automatic scoring of dicentric chromosomes as a tool in large scale radiation accidents, *Mutat. Res.* 756 (2013) 174–183.
- [4] M.V. Munot, Development of computerized systems for automated chromosome analysis: current status and future prospects, *Int. J. Adv. Res. Comput. Sci.* 9 (1) (2018) 782–791.
- [5] Y. Rahimi, R. Amirfattahi, R. Ghaderi, 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*, 2008, pp. 187–190.
- [6] W. Yan, S. Shen, An automatic counting algorithm for chromosomes, in: *2nd International Conference on Bioinformatics and Biomedical Engineering*, 2008, pp. 2492–2494.
- [7] S. Jahani, S.K. Setarehdan, E. Fatemizadeh, Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators, in: *Iranian Conference on Machine Vision and Image Processing*, 2011, pp. 1–4.
- [8] P. Moallem, A. Karimizadeh, M. Yazdchi, 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) (2013) 22–28.
- [9] R. Uttamantanin, P. Yuvapoositanon, A. Intarapanich, S. Kaewkamnerd, R. Phuksaritanon, A. Assawamakin, S. Tongsimma, MetaSel: a metaphase selection tool using a Gaussian-based classification technique, *BMC Bioinf.* 14 (16) (2013) S13.
- [10] S. Minaee, M. Fotouhi, B.H. Khalaj, A geometric approach to fully automatic chromosome segmentation, in: *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2014, pp. 1–6.
- [11] K. Kubola, P. Wayalun, Automatic determination of the g-band chromosomes number based on geometric features, in: *15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2018, pp. 1–5.
- [12] I.C. Yilmaz, J. Yang, E. Altinsoy, L. Zhou, An improved segmentation for raw g-band chromosome images, in: *5th International Conference on Systems and Informatics (ICSAI)*, 2018, pp. 944–950.
- [13] T. Arora, R. Dhir R, A novel approach for segmentation of human metaphase chromosome images using region based active contours, *Int. Arab J. Inf. Technol.* 16 (1) (2019) 132–137.
- [14] C. Lin, A. Yin, Q. Wu, H. Chen, L. Guo, G. Zhao, X. Fan, H. Luo, H. Tang, Chromosome cluster identification framework based on geometric features and machine learning algorithms, in: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 2357–2363.
- [15] C. Lin, G. Zhao, A. Yin, Z. Yang, L. Guo, H. Chen, L. Zhao, S. Li, H. Luo, Z. Ma, A novel chromosome cluster types identification method using ResNeXt WSL model, *Med. Image Anal.* 69 (2021), 101943.
- [16] N. Madian, K.B. Jayanthi, Analysis of human chromosome classification using centromere position, *Measurement* 47 (2014) 287–295.
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.