Int J Hum Genet, 18(3): 228-237 (2018)
DOI: 10.31901/24566330.2018/18.3.704

# Structural Similarity and Probabilistic Neural Network Based Human G-Band Chromosomes Classification

**D. Somasundaram[1*], N.Kumaresan[2], Vinodhini Subramanian[3] and S. Sacikala[3]**

*[1]Department of Electrical Engineering, Indian institute of Science, Bangalore, India*
*[2]Department of ECE, Anna University, Regional office, Coimbatore, India*
*[3]Center for Medical Genetics, Chennai, India*
*E-mail: [*]<somgce@gmail.com>*

**KEYWORDS** Birth Defects. Centromere. Chromosomes. Cytogenetic Analysis. Similarity Index. Support Vector Machine

**ABSTRACT** Chromosome classification has a vital role to play in achieving karyotype. Generally, human chromosomes consists of 46 chromosomes. Karyotyping is the important clinical procedure for screening and diagnosing genetic disorders and cancer. Manual karyotyping is a labor-insensitive and time-consuming task hence, developing automated computer-assisted systems have gained importance. In past decade, chromosome length and centromere positions were considered for classification. In chromosome analysis it is essential to segment the object of interest from the background. This object often consists of two or more chromosomes, either touching, overlapping or multiple overlapping with each other. In this paper, dataset of 1000 touching chromosomes, 1000 overlapping and 500 multiple overlapping chromosomes are taken for the study. SVM and Probabilistic Neural Network (PNN)-based classification is carried out and compared with other classifiers. PNN provides ninety-seven percent of classification accuracy and proves to be better compared to other classifiers.

## INTRODUCTION

Karyotyping is performed with new technologies using different dying process. In biotechnology, research is carried out to find the new additives for better analysis of chromosomes to attain karyotype. In another end, newer image processing tools and algorithms are developed to obtain the perfect karyotyping. In these developments, chromosome and its characterization play a major role to find out the disorders through karyotype. Chromosome classification has a key role for attaining karyotyping. Many parameters are used to classify the human chromosomes. Few parameters are standard to obtain the perfect classification and karyotyping. Based on the chromosome length and centromere position. Granlund (1976) initialized a work in chromosomes based on integrated density profiles which classifies the chromosomes as touching, overlaping, bend, and heavily bend categories using Fourier descriptors. Initially the method provides seventy-five percent accuracy in classification. Lerner (1998) has investigated the application of neural networks (NN's) for automatic analysis of chromosome images. NN's have been applied to perform all major stages of automatic human chromosome analysis, namely feature extraction, image segmentation, and classification. By applying MLP based hierarchical classification strategies to a well-explored chromosome database, a classification performance of 83.6 percent was achieved. This is higher than ever published on this database and an improvement of more than ten percent in the error rate. Therefore, chromosome analysis based on the NN-based techniques that are developed in this research leads towards a completely automatic human chromosome analysis. Stanley et al. (1998) presented a technique for similarity matching for chromosome identification based on the specific classes, banding pattern and centromere index using neural network. Prototype chromosome images can be quantitatively synthesized from a subspace to objectively represent the chromosome images of a given type or population, and the transformation in the subspace can be utilized as the extracted feature measurements for classification purposes. In particular the formation of three well-known subspaces, namely the ones derived from Principal Component Analysis 19 (PCA), Fisher's Linear Discriminant Analysis, and the Discrete Cosine Transform (DCT) are examined. Experimental results show that previously unseen prototype chromosome image of high vi-

sual quality can be synthesized using the proposed subspace-based method, and that PCA and the DCT significantly outperform the well-known benchmark technique of weighted density distribution functions in classifying 2-D chromosome images. Cho (2000) presented a method for automated chromosome karyotyping using a two-layer classification platform. The hypotheses is that by selecting most effective feature sets and adaptively optimizing classifiers for the different groups of chromosomes with similar image characteristics reduces the complexity of the automated karyotyping scheme and improve its performance and robustness. So, an image database involving chromosomes is assembled and implemented with a genetic algorithm to optimize the topology of multi-feature based Artificial Neural Networks (ANN). In the first layer of the scheme, a single ANN was employed to classify 24 chromosomes into seven classes. In the second layer, seven ANNs were adaptively optimized for seven classes to identify individual chromosomes. The scheme was optimized and evaluated using a "training-testing-validation" method. In the first layer, the classification accuracy of the validation dataset was 92.9 percent. In the second layer, classification accuracy of seven ANNs ranged from 67.5 percent to 97.5 percent, in which six ANNs achieved accuracy above 93.7 percent. The maximum difference in classification accuracy between the testing and validation datasets is <1.7 percent. The study demonstrates that this new scheme achieves higher and robust performance in classifying chromosomes. Classification is not based on the chromosome length. Conroy et al. (2000) presented a study for classification based on neural network, singular value decomposition and Fisher discriminant analysis. In this, Hidden markov model is used. Mousavi et al. (2000) introduced a multi-class feature analysis for chromosomes based on the centromere position. In this, chromosomes are initially segmented using fuzzy algorithm and gradient method where the heteromorphic chromosome are classified. Karvelis et al. (2009) presented a technique for classification method of M-Fish chromosome images using region Bayes classifier and vector median filtering. This method outperforms compared to pixel by pixel classifier. The algorithm was tested over 183 six channel chromosome M-Fish images and its accuracy is more than ten percent compared to existing methods and artifacts are removed using binarization method. Wang et al. (2009) developed automated classification scheme for chromosomes using multi-feature based two-layer Artificial Neural Networks. This scheme is validated with an accuracy of ninety-three percent. This method is tested over 6900 chromosomes and classified into seven classes. The proposed method used different groups of chromosomes with similar image characteristics. This study considered and tested for metaphase chromosomes with normal samples. Ming and Tian (2010) explained the concept of pattern extraction and classification based on multi-scale wavelets Bi-spline. In this, pale-path algorithm is introduced for the segmentation of touching and overlapping chromosomes. Middle point algorithm is used to detect the medial axis in chromosomes and the method used for evaluation are average gray profile, gradient profile and shape profile. Weighted density distribution algorithm is used as chromosome classifier. Recent works for chromosome classification using wavelet neural network is discussed by Oskouei and Shanbehzadeh (2010), pixel based analysis is carried out using Band analysis discussed by Somasundaram (2016), VLSI based karyotyping methodology proposed by Abid et al. (2017), and Kusakci et al. (2017) discussed a method for autonomous human chromosome system using CSVMT. Limitations in the processing of multicolor chromosomes are discussed by Lee et al. (2001).

**Objectives**

Chromosomes are classified based on shape descriptors, chromosome length, and centromere index and band patterns. The features considered are, size and banding pattern. Size is one of the parameter to differentiate two normal chromosomes. Banding pattern is dark and bright bands formed on chromosome that depends on the type of dye used for staining. There are different staining procedures to form the bands.

In this paper, SVM and PNN based classification method for human chromosomes are discussed. For the classification, parameters like centromere position, p-arm, q-arm and total length of chromosomes are considered. Additionally, structural similarity is used to identify each chromosome.

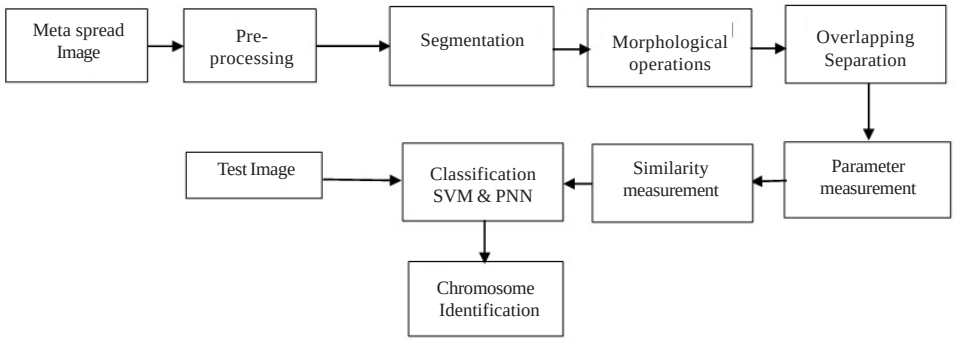


**Fig. 1. Proposed method**

## METHODOLOGY

Over all proposed method for chromosome analysis is shown in Figure 1. Initially, metaspread chromosome image is taken for analysis. Preprocessing steps are carried out for the metaspread chromosome images. In this, median filter is applied to reduce the noise and morphological operations are carried out for the image. The morphological operations include erosion, dilation, opening and closing operations. Various segmentation algorithms are carried out to separate the chromosomes from metaspread chromosomes. The methodology uses MOGAC method to segment the chromosomes from metaspread. After obtaining individual chromosomes, few chromosomes are perfectly separated but few chromosomes have the problem of overlapping and touching chromosomes. Hypothesis analysis is used for separating overlapping chromosomes to single chromosomes. After successful separation of chromosomes, the classification of these chromosomes is necessary to obtain karyotyping.

This can be done using parameter analysis, like measuring of length of the chromosome, centromere position and area of chromosomes. These parameters are major and mandatory process to categorize chromosomes. Still some challenges are there to identify the chromosomes. For the better identification of chromosomes, Riesz transform is used for additional support. Previously, methods for morphological operations and parameter evaluations are well explained in the literature. In this section, profile extraction, centromere index calculation, length of chromosome measurement and structural similarity are explained.

First, the scheme computes the size and the length of each chromosome by counting the total number of pixels comprising the chromosome and the number of pixels along the detected medial axis, respectively. Second, the scheme computes two feature profiles for each individual chromosome. Third, the scheme extracts the features of these profiles and adaptively selects effective ones for different types of chromosomes. The two computed feature profiles are described below:

### Shape Profile

A perpendicular line with weighted width is taken vertically across the medial axis of the image and a profile is drawn and formulation is given below:

$$\sum_{i=1}^{n}(g_i \times d_i^2) / \sum_{i=1}^{n} d_i^2, \quad (1)$$

Where, $g_i$ is the gray value and $d_i$ is Euclidean distance. Shape profile is obtained by summing the product of every gray value with Euclidean distance corresponding to the medial axis by its distance sum. A sample for the shape profile of the chromosome is shown in Figure 2a.

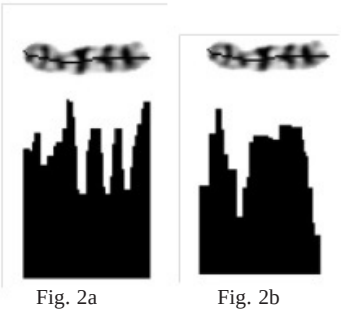

Fig. 2a Fig. 2b

**Fig. 2a, Fig. 2b. Shape and density profile for chromosome 10**

**Density Profile**

Perpendicular lines are drawn across the medial axis and corresponding average gray value of every perpendicular line gives density profile and is given as:

$$D(x)= [\sum_{i=1}^{n} g_i(x)] / n, \quad (2)$$

Where, $g_i$ is the gray value and n is pixel count in every perpendicular line. The density profile is shown in Figure 2b. Filtering is performed to reduce impulse and noises in the profile. Filter used here is median filter.

**Centromere Index Identification**

The centromere index is computed as the ratio of the length of a shorter arm to the total length of the chromosome. Therefore, the length of the p-arm and the total length are calculated. The centromeric index is defined as:

$$\text{Centromere index} = \frac{\text{Length of the short arm (p arm)}}{\text{Total length of chromosome}} \times 100 \quad (3)$$

Then the chromosome length (p-arm and q-arm) and centromere index is calculated by using the medial axis transformation. The length and centromere index values for some of the chromosomes are shown in Table 1.

**Table 1: Results for length and centromere index**

| *Chromosome number* | *p-arm length* | *q-arm length* | *Total length* | *Centromere index* |
|---|---|---|---|---|
| 2 | 176 | 51 | 227 | 77.53 |
| 5 | 121 | 12 | 133 | 90.97 |
| 8 | 125 | 8 | 132 | 93.98 |
| 10 | 73 | 37 | 110 | 66.36 |
| 11 | 93 | 19 | 112 | 83.03 |

**Chromosome Length**

Medial axis is obtained for all chromosomes and the length is calculated. The images are normalized to remove the different scale problem in the image and the length values are compared. Standardized length formulation is given as:

$$\text{Len}= \frac{L}{\bar{L}} \quad (4)$$

Where, L is the length of the chromosome and $\bar{L}$ is the average length of the metaphase.

**Extraction of PC and GM**

In chromosome images, image intensity has sharp changes and directly considered as features. The PC (Phase Congruency) attained in that points and the phase is maximum in that Fourier modules. These features are biologically believable model as described in [16-17]. This PC is a dimensionless measurement for individual chromosome structure. Each structure is divided into two parts, one is even symmetric and another one is odd symmetric. This can be represented as:

$$[e_n(x), o_n(x)] = [g(x) * M_n^e, g(g) * M_n^o] \quad (5)$$

$M^e_n$ - is even symmetric, $M^o_n$ - is odd symmetric

Local amplitude on scale is described as,

$$A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2} \quad (6)$$

For PC calculation

$$\text{Let, } F(x) = \sum_n e_n(x) \text{ and } H(x) = \sum_n o_n(x) \quad (7)$$

$$\text{Phase Congruency } (PC)(x) = \frac{E(x)}{\varepsilon + \sum_n A_n(x)}$$

$$\text{Where, } E(x) = \sqrt{F^2(x) + H^2(x)} \quad (8)$$

Transfer function of a log-Gabor filter in the frequency domain is

$$G(\omega) = exp\left(-\left(log(\omega/\omega_0)\right)^2 / 2\sigma_r^2\right) \quad (9)$$

2-D Gabor function is,

$$G_2(\omega, \theta_j) = exp\left(-\frac{\left(log\left(\frac{\omega}{\omega_0}\right)\right)^2}{2\sigma_r^2}\right) . exp\left(-\frac{(\theta - \theta_j)^2}{2\sigma_\theta^2}\right) \quad (10)$$

Modulating ω0 and θj and convolving G2 with the 2D image, $[e_n, \theta_j(X)^2, O_n, \theta_j(X)^2]$

The local amplitude and local energy on scale n and orientation θj is,

$$A_n, \theta_j(x) = \sqrt{e_n, \theta_j(X)^2 + O_n, \theta_j(X)^2} \quad (11)$$

$$E_{\theta_j}(X) = \sqrt{F_{\theta_j}(X)^2 + H_{\theta_j}(X)^2} \quad (12)$$

Where, $F_{e1}(X) = \Sigma_n e_n \theta_j(X)$ $H_{\theta j}(X) = \Sigma_n O_n \theta_j(X)$

For Chromosome Image Phase Congruency is,

$$PC_{2D}(X) = \frac{\sum_j E_{\theta_j}(X)}{\varepsilon + \sum_n \sum_j A_n, \theta_j(X)} \quad (13)$$

Gradient Magnitude of chromosome is,

$$GM = \sqrt{G_x^2 + G_y^2} \quad (14)$$

**Riesz Transform based Similarity Measurement**

The spatial representation of the Riesz kernel is,

$$\left(R_x(X), R_y(X)\right) = \left(\frac{x}{2\pi|X|^3}, \frac{y}{2\pi|X|^3}\right), \quad (15)$$

$$X = (x, y)\epsilon\mathbb{R}^2$$

Transfer function in the Fourier domain is,

$$\left(H_u(u), H_v(u)\right) = \left(-i\frac{u}{|u|}, -i\frac{y}{|u|}\right),$$

$$u = (u, v)\epsilon\mathbb{R}^2 \quad (16)$$

Chromosome image f(x), the monogenic signal is defined as the combination of f and its Riesz transform is,

$$f_M(X) = \left(f(X), R_x\{f\}(X), R_y\{f\}(X)\right) \quad (17)$$

$$f_M(X) = \left(f, R_x * f, R_y * f\right) \quad (18)$$

Local orientation of each chromosome is calculated using equation 19,

$$\theta = arctan\frac{R_y\{f\}}{R_x\{f\}}, \theta\epsilon[0, \pi) \quad (19)$$

Local phase and amplitude can be calculated using,

$$|(h_1 * f_\theta)(0)| = \left|-\frac{1}{\pi}\int_{t\epsilon\mathbb{R}} \frac{f(tcos\theta, tsin\theta)}{t} dt\right| (20)$$

$$\varphi = atan2\left(\sqrt{R_x^2\{f\} + R_y^2\{f\}}, f\right), \varphi\epsilon[0, \pi) \quad (21)$$

Local amplitude of a chromosome image is,

$$A = |f_M(X)| = \sqrt{f^2 + R_x^2\{f\} + R_y^2\{f\}} \quad (22)$$

Similarity locations of each chromosome are calculated using equation 23,

$$d_i(x, y) = \frac{2f_i(x, y).\, g_i(x, y) + c}{f_i^2(x, y) + g_i^2(x, y) + c} \quad (23)$$

$$D_i = \frac{\sum\sum d_i(x, y).\, M(x, y)}{\sum\sum M(x, y)}$$

Similarity matrix is calculated using,

Similarity index = $\Pi_{i=1}^{5}\ D_i$

$$f(x) = \beta_1\left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x-\beta_3)}}\right) + \beta_4 x + \beta_5 \quad (24)$$

Similarity measurement is taken for all possible combinations of chromosomes. Chromosomes are classified based on the basic parameters using two kinds of classifiers:

1. Support Vector Machine (SVM)
2. Probabilistic Neural network (PNN)

Classifiers are explained in the following section.

## Support Vector Machine Based Chromosome Classification

Let *{pi. qi}* be a set of N data Vectors with $p_i \epsilon R_d$, $q_i \epsilon (+1, -1)$, i=1,...N.

$p_i$ is the i[th] data vectors belongs to a binary class $q_i$. The objective of training 2v-SVM is to find the hyper-plane that separates two classes with the widest margins, that is,

$$w(p) + b = 0 \quad (25)$$

Subject to

$$q_i(w\varphi(p) + b) \geq (\rho - \theta_i), \theta_i \geq 0 \quad (26)$$

To minimize

$$\frac{1}{2}\|w\|^2 - \Sigma_i C_i(v\rho - \theta_i) \quad (27)$$

Where, *p* margin is position and *φ(p)* is error factor. is the mapping function of feature spaces where mapping of data space is done and gives a decision function which need not be linear for any training set data.$C_i$ *(vp-*$\theta_i$*)* is the cost function. W is normal vector, b the bias, and $\theta_i$ is slack variable for classification errors, *v* is the error parameter computed by $v_1$ and $v_-$ and $v_-$ training the positive and negative classes are done respectively.

$$v = \frac{2v_+v_-}{v_+ + v_-}, 0 < v_+ < 1\ and\ 0 < v_- < 1 \quad (28)$$

Error penalty $C_i$ is calculated as,

$$C = \begin{cases} C_+, & if \quad p_i = +1, \\ C_-, & if \quad q_i = -1, \end{cases} \quad (29)$$

Where,

$$C_+ = \left[n_+\left(1 + \frac{v_+}{v_-}\right)\right]^{-1} \quad (30)$$

$$C_- = \left[n_-\left(1 + \frac{v_-}{v_+}\right)\right]^{-1} \quad (31)$$

And $n_+$ and $n_-$ are the number of training points for the positive and negative classes, respectively. 2v-SVM training can thus be formulated as,

$$max_{(\propto_i)}\left\{-\frac{1}{2}\Sigma_{i,j}\propto_i\propto_j\ q_iq_jK(p_ip_j)\right\} \quad (32)$$

Where $0 < \propto_i < C_i$, $\Sigma_i \propto_i q_i = 0$, $\Sigma_i \propto_i > v$,

$i, j\ \varepsilon\ 1,...N$, $\propto_i$, $\propto_i j$ are the Lagrange multipliers and kernel function is,

$$K(p_i, p_j) = \varphi(p_i)\varphi(p_j) \quad (33)$$

Decomposition of Image I,

$$[I_{Ai}, I_{Hi}, I_{Vi}, I_{Di}] = RDWT(I) \quad (34)$$

Where, *i* = 1, 2, 3 represents the level of decomposition and A, H, V and D represent the approximation, horizontal, vertical and diagonal bands respectively. Approximation and detailed bands at each decomposition level are used to compute the quality factor of the bands.

Let QA, QH, QV, and QD be the quality factor for the approximation, horizontal, vertical and diagonal bands respectively. The quality factor for each band is:

$$Q_A = \sum_{i=1}^{s} \sum_{k,j=1}^{n} I_{Ai}(k,l), \quad (35)$$

$$Q_H = \sum_{i=1}^{s} \sum_{k,j=1}^{n} I_{Hi}(k,l), \quad (36)$$

$$Q_V = \sum_{i=1}^{s} \sum_{k,j=1}^{n} I_{Vi}(k,l), \quad (37)$$

$$Q_D = \sum_{i=1}^{s} \sum_{k,j=1}^{n} I_{Di}(k,l), \quad (38)$$

Further, these quality factors are combined to compute the quality score $Q$ of image $I$,

$$Q = \frac{m_A Q_A + m_H Q_H + m_V Q_V + m_D Q_D}{m_A + m_H + m_V + m_D} \quad (39)$$

Where, mA, mH, mV, and mD are the weight factors computed using the following equations:

$$m_A = \sum_{i=1}^{s} \frac{1}{1 + \sum_{k,j=1}^{n} \nabla(I_{Ai}(k,l))}, \quad (40)$$

$$m_H = \sum_{i=1}^{s} \frac{1}{1 + \sum_{k,j=1}^{n} \nabla(I_{Hi}(k,l))}, \quad (41)$$

$$m_V = \sum_{i=1}^{s} \frac{1}{1 + \sum_{k,j=1}^{n} \nabla(I_{Vi}(k,l))}, \quad (42)$$

$$m_D = \sum_{i=1}^{s} \frac{1}{1 + \sum_{k,j=1}^{n} \nabla(I_{Di}(k,l))}, \quad (43)$$

Where, i represent the level of decomposition and $\nabla$ represent the gradient operation. The weight factors ensure proper weight assignment to all bands depending on the information contained in each of the bands.

## Feature Extraction and Input Data for SVM

In the chromosome classification, feature extraction is obtained using statistical features such as chromosome length, centromere index and similarity index. The advantage of these features is that they reflect the correlation among pixels within an image. Chromosome length is extracted from each image using morphological operations. Features are extracted from the images and given as input to the SVM linear classifier. In this, 20 samples for each chromosome are taken, for 23 chromosomes overall 460 chromosomes are considered for normal male chromosomes, and 230 female normal chromosomes are taken. For abnormality, in each category 10 male and female chromosomes are taken, in total 460 chromosomes are taken for training. In SVM, Training set of 690 normal chromosomes and 460 abnormal chromosomes are given as input to the SVM.

## Probabilistic Neural Network Based Chromosome Classification

In this, chromosome length, perimeter and centromere position are considered as features for classification. Based on the length, centromere position and perimeter, chromosomes are classified into 12 classes, which are used to train the PNN network. This PNN network is designed with these input parameters to segment and separate the overlapped chromosomes and also used to classify the class of separated chromosomes with its statistical values. In the PNN architecture each class is given to input layer. The significant advantages of the PNN classifier are its speed and simplicity of the training process. The training process is simply to select a kernel function and its smoothing parameter. There is no need for weight adaptation as in other neural networks. Similarity measurement of male chromosomes as shown in Table 1. Human chromosomes are classified into seven groups based on the centromere position and chromosome length as shown in Table 3. Based on the classification, 12 classes are selected to train the PNN network. Twelve (12) classes of chromosome and its parameters are shown in Ta-

**Table 2: Similarity measurement of male chromosomes (1 to 8 chromosomes)**

| *Chromosome* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.00215 | 0.0015 | 0.00205 | 0.00603 | 0.00102 | 0.00875 | 0.00138 |
| 2 | 0.00215 | 1 | 0.00849 | 0.00285 | 0.00256 | 0.0022 | 0.00715 | 0.00198 |
| 3 | 0.0015 | 0.00849 | 1 | 0.00279 | 0.00663 | 0.00195 | 0.00786 | 0.00766 |
| 4 | 0.00205 | 0.00285 | 0.0279 | 1 | 0.00501 | 0.00132 | 0.00953 | 0.00770 |
| 5 | 0.00603 | 0.00256 | 0.00663 | 0.00501 | 1 | 0.00326 | 0.0024 | 0.0012 |
| 6 | 0.00102 | 0.0022 | 0.00195 | 0.00132 | 0.00326 | 1 | 0.00404 | 0.0013 |
| 7 | 0.00875 | 0.00715 | 0.00786 | 0.00953 | 0.0024 | 0.00404 | 1 | 0.0014 |
| 8 | 0.00138 | 0.00198 | 0.00766 | 0.00770 | 0.0012 | 0.0013 | 0.0014 | 1 |

**Table 3: Chromosome classification based on centromere position**

| *Chromosome number* | *p-arm length (in pixel)* | *q-arm length (in pixel)* | *Group of chromosome* | *Type of chromosome* |
|---|---|---|---|---|
| CH 1A | 77 | **77** | A | Metacentric |
| CH 1B | 82 | 101 | A | Acrocentric |
| CH 2A | 63 | 91 | A | Acrocentric |
| CH 2B | 79 | 60 | A | Acrocentric |
| CH 3A | 62 | 84 | A | Acrocentric |
| CH 3B | 51 | 81 | A | Acrocentric |
| CH 4A | 33 | 80 | B | Acrocentric |
| CH 4B | 30 | 37 | B | Sub-metacentric |
| CH 5A | 89 | 43 | B | Acrocentric |
| CH 6A | 60 | 38 | C | Acrocentric |
| CH 7A | 48 | 32 | C | Acrocentric |
| CH 7B | 44 | 37 | C | Sub-metacentric |
| CH 8A | 33 | 60 | C | Acrocentric |
| CH 8B | 28 | 47 | C | Acrocentric |

ble 4. After training the PNN network, overlapped chromosomes are given as input to the network. Segmentation of the chromosome occurs based on the concave and convex points and interesting points are identified. Centromere point and chromosome length are calculated based on the shape and density profile. P-arm, q-arm length and centromere index is taken as feature and to improve the classification, perimeter is considered additionally.

Overlapped chromosome is shown in Figure 3. In this, input class is having two chromosomes. Initially, the overlapping needs to be separated to classify each chromosome.

## RESULTS AND DISCUSSION

Linear Kernel Support Vector Machine is used for the classification. In this, normal and abnormal classification is analyzed using the length of the chromosomes. Above classification is carried out in MATLAB environment and a Graphical user interface is created for this classification as shown in Figure 4. This GUI contains two phases; Training phase and Test phase. In training phase, each male and female dataset were given to train the neural network. In the testing phase of chromosome, random selections from the unknown chromosome images are considered. Similarity generated input for each chromosome is given in Table 2 for entire 46 chromosomes. In this, identical chromosomes similarity is unity and others have different combinations as listed. Four kinds of chromosomes are considered such as metacentric, sub-metacentric, acrocentric, subacrocentric chromosomes. These chromosomes are classified based on the centromere position of each chromosomes.

In overlapped chromosome, input class is having two chromosomes. Initially, the overlapping needs to be separated to classify each chromosome as shown in Figure 5.

Each concave and convex points in each separated chromosome is shown in Figure 5.

From each concave and convex points, possible separation lines are drawn over the chro-

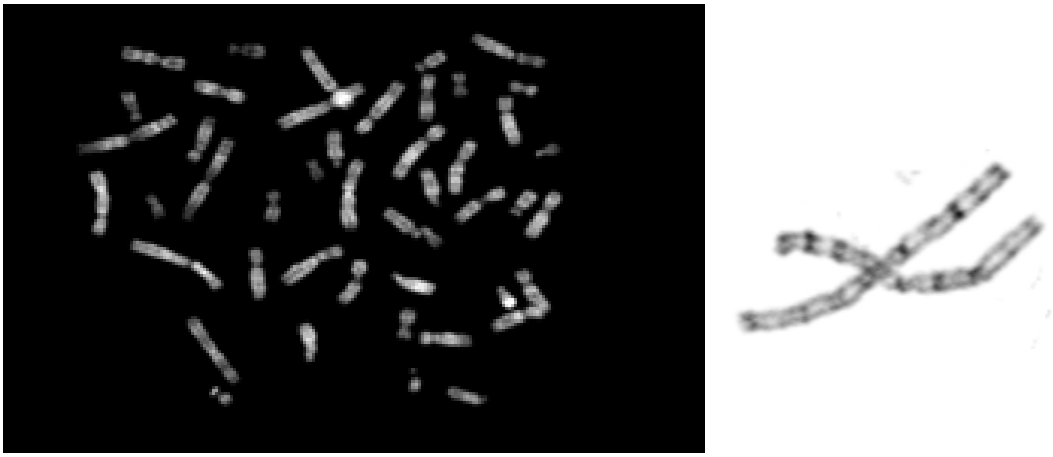

**Fig. 3. Metaspread and overlapped chromosome image**


**Fig. 4. Simulations GUI for classification**

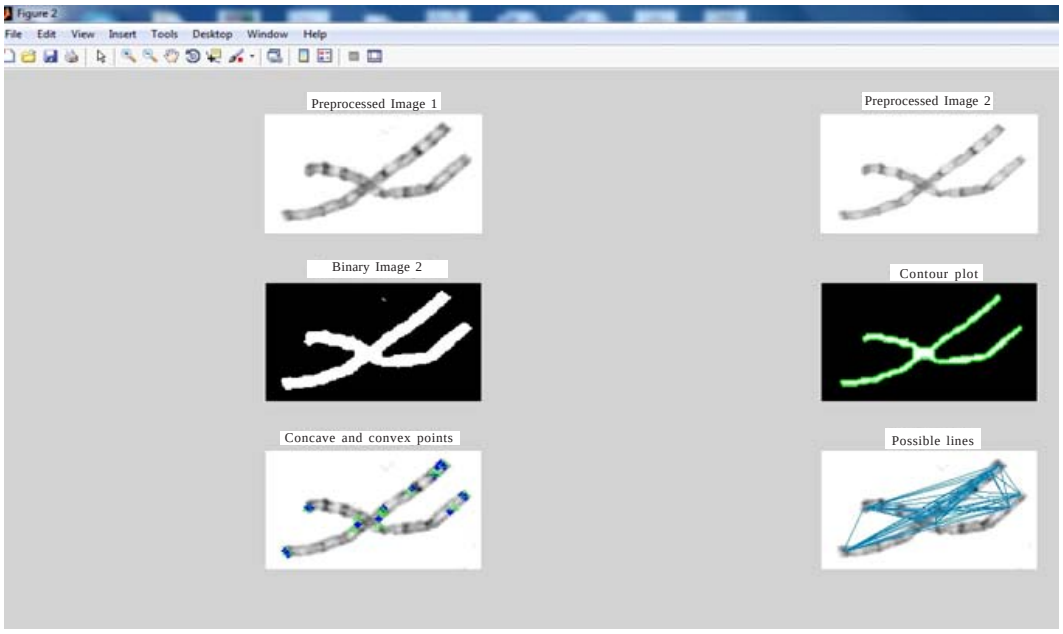


**Fig. 5. Overlap separation**

**Table 4: Chromosome classes with parameters**

| *Chromosome classes* | *Length (pixels)* | *Perimeter (pixels)* |
|---|---|---|
| Class 1a | 168 | 343 |
| Class 1b | 142 | 317 |
| Class 2a | 146 | 304 |
| Class 2b | 139 | 298 |
| Class 3a | 128 | 267 |
| Class 3b | 121 | 255 |
| Class 4 a | 104 | 241 |
| Class 4b | 113 | 213 |
| Class 5a | 85 | 215 |
| Class 5b | 83 | 203 |

Overlapped chromosome belongs to class 3. Thus, the PNN network classified each chromosome with its parameter values such as length and perimeter. This PNN Classification provides ninety-seven percent accuracy in classification of human chromosomes. Overall comparison is shown in Table 5.

Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD) and Support Vector Machine (SVM) methods are compared with Probabilistic Neural Network (PNN). Among these, SVM and PNN provide better results for normal

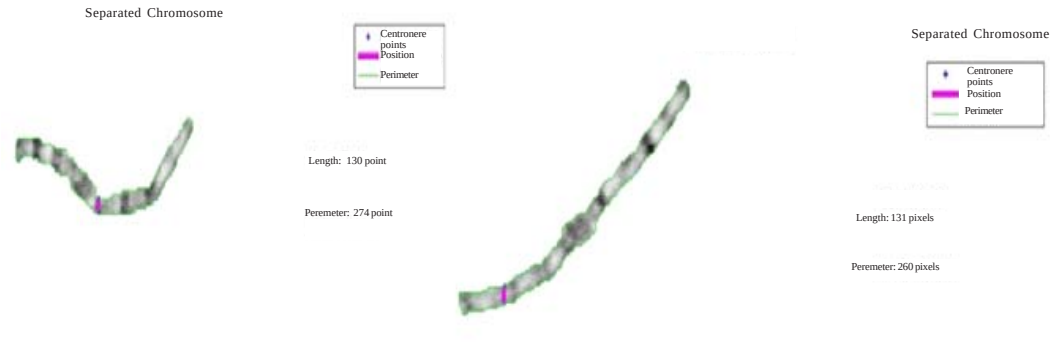


**Fig. 6. Separated chromosome**

mosomes. Based on this, the separation of the overlapped chromosome is carried out. The extracted chromosomes are shown in Figure 6. chromosomes and PNN outperforms for abnormal chromosomes.

**Table 5: Chromosome classifier comparisons**

| *Classifiers* | *Accuracy (%)* | | *Error (%)* | |
|---|---|---|---|---|
| | *Normal chromosomes* | *Abnormal chromosomes* | *Normal* | *Abnormal* |
| PCA | 87 | 81 | 13 | 19 |
| DWT-KPCA | 90 | 85 | 10 | 15 |
| SVD | 94 | 89 | 6 | 11 |
| SVM | 96 | 92 | 4 | 8 |
| PNN | 97 | 93 | 3 | 7 |

## CONCLUSION

Chromosome classification methods such as Principal Component Analysis, Discrete Wavelet Transform Principal Component Analysis, Support Vector Machine based classification and Probabilistic Neural Network (PNN) based classification are analyzed. In this, PNN provides better classification accuracy. The classification results from the Riesz transform based structural similarity index and SVM were better than other neural networks and fuzzy logic algorithms. This method outperforms when chromosome images are distorted in nature. Many methods fail when uncertainty or inadequate information is involved. This proposed method overcomes the low-level identification problem and constitute successful information compaction. Proposed method brings the more efficient classification of chromosome images. This method provides ninety-seven percent of classification accuracy among the normal and abnormal chromosomes.

## RECOMMENDATIONS

The proposed classification method is best suited for normal and abnormal chromosome classifications. This study further extended towards the highly degraded chromosomes to increase the classification accuracy.

## ACKNOWLEDGEMENTS

The researchers would like to thank DST-SERB for supporting this project: NPDF101198.

## REFERENCES

Abid F, Hamami L, Badache F, Derdour H 2017. A system on chip for automatic karyotyping system. *Computers and Electrical Engineering,* 64: 1-14.

Cho JM 2000. Chromosome classification using back propagation neural networks. *IEEE Engineering in Medicine and Biology Magazine,* 19(1): 28-33.

Conroy JM, Kolda TG, O'Leary DP, O'Leary TJ 2000. Chromosome identification using hidden Markov models: Comparison with neural networks, singular value decomposition, principal components analysis, and Fisher discriminant analysis. *Laboratory Investigation,* 80(11): 1629-1641.

Granlund GH 1976. Identification of human chromosomes by using integrated density profiles. *IEEE Transactions on Biomedical Engineering,* 23(3): 182-192.

Karvelis PS, Fotiadis DI, Tsalikakis DG, Georgiou IA 2009. Enhancement of multi-channel chromosome

classification using a region-based classifier and vector median filtering. *IEEE Transaction on Information Technology in Biomedicine,* 13(4): 561-570.

Kusakci AO, Ayvaz B, Karakaya E 2017. Towards an autonomous human chromosome classification system using competitive support vector machines teams (CSVMT). *Expert Systems with Applications,* 86: 224-234.

Lee C, Gisselsson D, Jin C, Nordgren A, Ferguson DO, Blennow E, Fletcher JA, Morton CA 2001. Limitations of chromosome classification by multi-color karyotyping. *The American Journal of Human Genetics,* 68(4): 1043-1047.

Lerner B 1998. Toward a completely automatic neural-network-based human chromosome analysis. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics,* 28(4): 544-552.

Ming D, Tian J 2010. Automatic pattern extraction and classification for chromosome images. *Journal of Infrared Millimeter and Terahz Waves,* 31(7): 866-877.

Mousavi P, Ward RK, Lansdorpt PM, Fels SS 2000. Multi-feature analysis and classification of human chromosome images using centromere segmentation algorithms. *Proceedings of the IEEE International Conference on Image Processing,* 1: 152-155.

Oskouei BC, Shanbehzadeh J 2010. Chromosome Classification Based on Wavelet Neural Network. *International Conference on Digital Image Computing: Techniques and Applications,* 1-3 December, Sydney, NSW Australia. US: IEEE, pp. 605-610.

Somasundaram D 2016. Pixel sensible local band analysis in microscopic chromosome images using CSPA. *Cytology and Genetics,* 50(1): 42-46.

Stanley RJ, Keller JM, Gader P et al. 1998. Data-driven homologue matching for chromosome identification. *IEEE Transactions on Medical Imaging,* 17(3): 451-462.

Wang X, Zheng B, Li S, Mulvihill JJ, Wood MC, Liu H 2009. Automated classification of metaphase chromosomes: Optimization of an adaptive computerized scheme. *Journal of Biomedical Informatics,* 42(1): 22-31.

**Paper received for publication on February 2018**
**Paper accepted for publication on July 2018**