CHAPTER 2

LITERATURE SURVEY

This chapter reviews the publications related to chromosomes analysis. The first section is related to the segmentation and the separation of overlapping and touching chromosomes. The second section reviews the work done on parameter identification for the classification of chromosomes. The third section discusses the work on the straightening of human chromosomes and the final section is about the classification of chromosomes with various classifiers.

2.1 SEGMENTATION

Lucas van Vliet et al (1990) proposed a semi automated system for karyotyping process of metaspread chromosomes. The system performs segmentation and classification of overlapping and touching chromosomes by their band information. Optimal density based thresholding is used to find the dark band regions. The regions are then subjected to Laplace filtering and second order derivative filtering to detect the hills and valleys in the image. The obtained hill points are labelled and every set of connected points form the candidate of a band.

Liang Ji (1994) presented an automated segmentation of overlapping and touching chromosomes by eliminating interphase nuclei, stain debris, and noise. A rule based approach is performed for the segmentation process. The success rate of the algorithm is 90-95%. The
technique is presented to predict chromosome number which helps in controlling the splitting process.

Hongchi Shi et al (1998) discussed a parallel algorithm for splitting the touching chromosome images. Local features for finding the potential paths are obtained and the distance measure of the possible paths is calculated. Practical mesh algorithm is used to find the shortest potential path between any two vertices with suitable distance measure used for chromosome segmentation. The local features are cut points, skeleton points, junction points and ravine points. Cut points are obtained using curvature function with possible concave points. Thinning algorithm is used to obtain the skeleton of the image. Junction points are the point which have at least three branches. The ravine points are the points on the image which have the value less than their neighborhood pixel points which form the pale path on chromosome images. The author claims that the practical mesh algorithm is a cost effective method.

Graham Charters & Jim Graham (2002) proposed a trainable shape model where the shape properties are used to get the geometric information. This addresses the overlapping problems in chromosome. The model is trainable and yields a good probability estimation. Classification along with shape model provides an improved performance in resolving the overlap problems in chromosomes.

Cristina Urdiales Garcia et al (2003) discussed a new procedure for karyotyping. Chromosomes under non uniform resolution are taken for a better segmentation. Segmentation is done using foveal polygon which is a hierarchical structure. The segmented output deals with the overlapping and touching regions. A post processing is done which calculates the longitudinal axis and uses banding pattern to separate the overlapped chromosome images. The separated chromosomes are identified using their curvature function.
Separated chromosomes are characterized into seven groups using a normalized size. The feature vectors are obtained from the chromosome shape.

Petros Karvelis et al (2005) discussed the segmentation of touching chromosomes. Watershed transform (WT) is used for the initial segmentation. Further, each segmented area is subjected to WT until no segmented area is obtained. The path with high concave points are only considered for the final segmented output. 940 chromosome images are taken with 396 touching and 29 overlapping images. The accuracy of the algorithm is 92% for touching images.

Wacharapong Srisang et al (2006) proposed a method using computational geometry to separate the overlapping chromosomes. Initially, thresholding is done and the input images are converted to binary image. An image contour is thus obtained. The interesting points are obtained using a curvature function. High concave points are considered for the separation process. Separation is done using voronoi diagram by identifying all possible cut points obtained from the contour line of the overlapping chromosome images. The target cut points which indicate the overlapping region are obtained by Delaunay triangulation. The algorithm is tested on 35 chromosome images. 28 out of 35 overlapping images are separated correctly with an accuracy of 80%.

Hyohoon Choi et al (2006) proposed a decomposition method which includes information about the maximum likelihood frame work of the chromosome clusters. The method is evaluated using multiple hypothesis which includes geometric information, chromosome size and pixel classification. The Hypothesis with a maximum likelihood is considered as the best decomposition for the cluster. The accuracy of the method is about 90%.
Enrico Grisan et al (2007) proposed a space variant thresholding for resolving overlapping and touching problems in chromosomes and this method is successful even in the presence of hyper or hypo fluorescent region in an image. The proposed method is tested with 30 images with 1380 chromosomes and helps in resolving 92% of adjacencies and 90% of overlaps.

Prabhu Britto Albert & Ravindran (2007) discussed a Discrete Cosine Transform (DCT) based Gradient Vector Flow (GVF) active contour for the segmentation of chromosome images. DCT has a good energy compaction which helps better in describing the image property. Therefore, DCT along with GVF gives a better segmentation result.

Xiangzhi Baia et al (2009) proposed a new algorithm which includes identifying the concave points and performing ellipse fitting for splitting the touching cells. The algorithm includes contour preprocessing which helps in smoothing out the false contours and fluctuations in the contour to find the concave points. The concave points divide the cell into two individual segments and the ellipse fitting helps in showing the different segments of the contour into a single cell.

Yan Wenzhong & Feng Xiaohui (2010) proposed the watershed segmentation for overlapping problems. The watershed segmentation has oversegmentation issues. The overlapped chromosome images are subjected to various pre-treatment to eliminate the over segmentation issues. The distance transformed image is obtained and given as an input for the watershed segmentation and the algorithm is proved successful. When chromosomes are overlapped together very tightly, the segmentation result is not satisfied.

Mousami Munot et al (2011) proposed a modified snake algorithm along with the greedy approach based on the computational geometry of
boundary pixel to separate the touching chromosomes from the metaspread image. The accuracy of the algorithm with two touching chromosomes is 100% and with 3 and 4 touching chromosomes, it is 95%.

Enea Poletti et al (2012) presented a comparative study on eleven thresholding methods for automated identification of blobs that help in representing the chromosomes in singles or in clusters. Thresholding is the initial step in the disentanglement of chromosomes. From the comparative study, it is seen that adaptive thresholding and region based level set methods give the best performance.

Mukul Joshi et al (2012) proposed an automated method to separate the overlapped chromosome images by automatically detecting the cut points using computational geometry. The algorithm is tested on synthesized images of LK1 data set with 100% accuracy for one and two overlaps and 88% for three and four overlaps.

Hui Wangn et al (2012) proposed a novel method for splitting the clumps. The method includes three steps i) shape classification ii) identifying the pair of points for clump split and iii) joining the selected points. Shape classification is done using the shape features and support vector machine. The splitting line is done based on minimizing the image energy. The proposed method performs well in accurate clump splitting.

Yongqiang Zhao et al (2013) presented an automated method for the chromosome analysis which combines the pixel based geometric information and decision based pairing information. The segmentation is based on the geometric and the spectral information of chromosome clusters. This is followed by a graph based theoretical pairing method to identify the chromosome pairs.
Wenzhong Yan & Lei Bai (2013) presented a maximum likelihood function to segment the overlapping and touching chromosome images. A decomposition method is also used for segmentation. A hypothesis that has the maximum likelihood value will be considered as the best decomposition of the cluster. The segmented accuracy of two and three chromosome cluster is 90%.

Mousami Munot et al (2013) proposed an automated segmentation method by identifying the cut points on the overlap region, based on the computational geometry of the boundary pixels. The method is applied to M-FISH images and the algorithm is also tested on 40 real images. The overall accuracy is about 98% in identifying the cut points in two, three and four overlapping chromosomes.

Siddharth & Tripathi (2013) presented a method for separation of overlapping and touching chromosome images using an artificial neural network. The method initially finds the blobs in the metaspread chromosome images with overlap and the training algorithm is used for separating the overlapping and touching regions. The obtained results are compared with the healthy ones to detect the chromosomal abnormalities.

Devaraj Somasundaram & Vijay Kumar (2014) proposed an automated method for the separation of overlapped chromosome images and homologue chromosome identification by centromere position. An automated geometry separation algorithm is proposed for separating the overlapped chromosome images. Identification of cut points and cut lines help to obtain the overlap region. The Hypothesis verification is done for separating the chromosome images with 94% accuracy.

Tanvi & Renu Dhir (2014) presented a novel approach for the segmentation of overlapped chromosome images using computational
geometry. The method initially finds the contour of the images then the cut points are tracked on the overlapping region. A specific number of cut points are selected using computational geometry for separating the chromosomes. The algorithm accuracy is about 87.4%.


2.2 PARAMETER IDENTIFICATION

Goesta Granlund (1976) compared four different types of descriptors such as curves, non-uniformly sampled curves, Fourier descriptors, and distribution functions to identify the human chromosomes using integrated density profile. These profiles help us to group the homologue chromosomes. A limited test set of 5 cells with 230 chromosomes are considered, among which only 192 chromosomes are used for analysing. The recognition rates for the above mentioned descriptors are 82.9%, 85%, 77.6%, and 90.1%. When all the 230 chromosomes are considered, the recognition rate will be 69.1%, 70.1%, 64.7% and 75.2%

Frans Groen et al (1989) described various techniques for automated chromosome analysis. A piecewise linear approximation method is used for determining the axis of chromosomes. Two methods are compared for identifying the centromere of a chromosome. The first method deals with identifying the closest pair opposite contour points. The second method plots the profile for the width of chromosomes. The profile conveys information about the distance between the borders perpendicular to the medial axis. The first method yields a better result over the second method with an accuracy of 85% and 68% for Leyden data base and 93% and 76% for Copenhagen data.
Local laplacian band descriptors are used to identify bands in the chromosomes.

Jim Piper & Erik Granum (1989) introduced global shape features for identifying centromere position in chromosomes. Piecewise linear approximation method is used for chromosome axis determination. The profile value of the chromosome axis is obtained by summing the values of the points along the transverse line which is perpendicular to the tangent of the axis. A shape profile is identified to locate the centromere of a chromosome. Global band descriptors are used to identify bands in chromosomes. 16 different features are measured using an automatic feature selection algorithm and helps in developing an automated Karyotype operation.

Ronald Stanley et al (1998) presented an image analysis technique which describes the numerical abnormalities progressed from the structural abnormalities. Chromosomes are classified using the neural network. The parameters for the classification are banding pattern and CI which help in homologue chromosome matching. Homologue matching algorithm is better than the transportation algorithm for obtaining the density profile and the band representation contour.

Parvin Mousavi et al (2002) introduced an iterative fuzzy algorithm and a gradient descent method for centromere identification of human chromosome images. Florescence in situ hybridization chromosome images are used for the analysis. The centromere region pixels are assigned with fuzzy membership values. The error function is minimized using the iterative process. The method is tested on chromosome 22. Homologue chromosome pairs are identified and classified based on intensity value of the centromere position and their morphological differences. The algorithm proved to be
better in segmenting the centromere position and helps in homologue chromosome classification with a multi feature analysis.

Enea Poletti et al (2008) proposed an automated classification system based on medial axis of chromosomes, chromosome polarization, and feature preprocessing. Classification is done using neural network. 119 karyotype images are taken for the study among which 70 karyotype images are used for polarization and training and 49 for testing. A correct chromosome polarization is achieved for 92.3% of cases. The proposed classification system achieved an accuracy of about 95.6%.

Jau Hong Kao et al (2008) proposed an approach to find centromere by determining medial axis of chromosomes and profile matching. Medial axis is obtained by a simple cross section analysis. Subsequence matching technique is used for chromosome classification using the band profile obtained from medial axis. The experimental result shows that the proposed approach helps in automatically determining the medial axis of chromosome and has a better classification accuracy.

Rahmadi Trimananda (2010) discussed a simple and powerful technique for identifying the chromosomes. A pattern vector for one of the chromatid is considered and it is compared with the chromosome prototype. The relative error is calculated for identifying the object that matches with the prototype.

Mousami Munot et al (2012) proposed an efficient method to find the propinquity within the chromosomes for automated homologue chromosome identification. A novel nearness factor is defined for chromosome pair identification based on the banding pattern similarity measure. 50 images are taken for the analysis which belong to group A (Chromosome pair 1-3) and C (Chromosome pair 6-12) and an accuracy of
100% is achieved for the chromosomes belonging to group A and 97% accuracy for group C chromosomes.

Akila Subasinghe Arachchige et al (2013) proposed a multistage algorithm which includes discrete curve evolution, functional approximation of curve segments, gradient vector active contours and support vector machine for an accurate chromosome classification by detecting the centromere. The algorithm also includes Laplacian thickness measurement which incorporates contour and intensity information to obtain the accurate centromere position. The method helps in the separation of sister chromatids and measurement of width profile. The method is tested on 226 human metaspread chromosome images and proved better compared to centerline based method.

Somasundaram Devaraj et al (2013) proposed a LEAF algorithm to obtain the centromere position by identifying the biometric points of each chromosome. The parameter other than centromere is the band gap between each chromosome band and band length of each chromosome pairs.

2.3 STRAIGHTENING

Mehrsan Javan Roshtkhari & Seyed Kamaledin Setarehdan (2008) presented an effective algorithm for straightening the highly curved chromosome images using vertical and horizontal projection vectors with various rotation angles. The algorithm is performed on binary image which is obtained using thresholding. Rotation score is defined by amplitude of the peak values of the horizontal projection. The rotated images help in finding the bending axis and considered as the bending region of a chromosome. By the rotation, the artificial straightening of the curved chromosomes are obtained. Density profile is obtained for chromosomes before and after
performing the straightening operation. The proposed approach helped in achieving a good quantitative analysis.

Sahar Jahani & Seyed Kamaledin Setarehdan (2012) proposed an automated method to identify the parameter for classifying the chromosome into 23 pairs. The centromere position is the narrowest portion in the image. A linearly varying gray level mask is calculated for each chromosome image and then it is multiplied by its binary version to achieve the global minimum which is the centromere position. The algorithm is tested on 54 highly curved chromosomes and an average error rate of 4.2 pixels for centromere identification and 5.8 pixels for calculating the length of the chromosomes are obtained which is an acceptable error rate according to the skilled operator.

Devaraj Somasundaram & Vijay Kumar (2014) proposed a novel projective straightening algorithm for straightening the chromosomes to find the length of the chromosomes. The centerline of the chromosome is obtained using stentiford thinning method and then the straightening is performed. The proposed method straightens the curved chromosomes closely up to 178 degree. The parameters like straighten angle, length of the line and area are considered for further operation to identify the individual chromosomes.

2.4 CLASSIFICATION

Phil Errington & Jim Graham (1993) proposed an automated classification approach for metaspread chromosome images using multilayer perceptron neural network. The classifier inputs considered are the size of a chromosome, centromere index and banding profile. Three types of datasets considered for the work are Copenhagen, Edinburgh, Philadelphia. The classification error rate is compared between Network classifier, Parametric classifier using weighted density distribution (WDD) function and Parametric
classifier using a local band. The classification error rate is less using a
network classifier compared with the other two classifiers.

Walter Sweeney et al (1994) described a probabilistic neural
network for the classification of human chromosome images. 30 different
types of features are taken for classification. Copenhagen, Edinburgh, and
Philadelphia database images are used for the classification process. The
experimental results show that the recognition rate of the network is better
compared to a maximum likelihood and back propagation neural network
 techniques.

Jim Piper (1995) proposed the genetic algorithm for solving the
optimization problem in automated chromosome classification. The algorithm
proved to be better with a minimum error rate compared to many of the best
classification methods. Banding pattern is taken as one more constraint which
helped in further minimizing the error rate.

Gunter Ritter et al (1995) introduced a statistical model to derive
ML classifier for the correct classification of chromosomes. The model
includes elliptically symmetric feature vectors. Experimental results show that
the error rate is less than 2% and the chromosome images considered for the
study is Copenhagen data set.

Boaz Lerner et al (1998) proposed a new approach called
Classification-Driven Partially Occluded Object Segmentation (CPOOS)
method which determine the partial occlusion in the images. This method is
very robust to obtain a better classification output without any post processing
process. The chromosome analysis is done by considering two geometrical
features (length and Centromere index) and 64 density profile features.
Initially all the isolated images are classified into the respective classes and
then CPOOS method is applied to the cluster image. The method involves
binarization using K-means clustering of algebraic moment representation of the image pixel. A common problem with the threshold value selection is eliminated by using this process. Then the touching chromosome images are identified by their size. Concave points are selected along the curvature and these points are called as potential cut points. The potential cut points are taken to draw the separation lines in the touching chromosome images. Multilayer perceptron (MLP) trained back propagation classification algorithm is used to verify the hypothesis. The classifier helps in selecting the correct separation line. The accuracy of the method is 90%.

Boaz Lerner (1998) investigated the automated chromosome image analysis using neural networks. The separation of chromosome images are obtained by lines connecting the cut points based on hypothesis study. Cut points are the points related to the concave regions. MLP neural network is used to verify the hypothesis and helps in segmentation of chromosomes with separation lines. The segmentation process is also done using a classification driven approach without any shape based information and yields promising results. With this neural network implementation, Sammon’s mapping is used where features are extracted by the principal component analysis to minimize the dimensionality rate, allowing a high classification accuracy of 83.6%.

John Conroy et al (2000) proposed four automated approaches for identifying chromosomes. They are singular value decomposition, principal component analysis, fisher discriminant analysis and hidden Markov models. These approaches are tested on G band chromosome images of well known database (Philadelphia, Edinburgh, and Copenhagen). All the four approaches are then compared with neural network classifier. Comparing all the approaches, hidden Markov model proves better with classification accuracy of about 97% and with 95% even when telomeres of the chromosome disappear and also when a small portion of the chromosome is inverted. The
model yields better classification accuracy for both normal and abnormal G banded chromosomes images.

Jong Man Tho (2000) described the back propagation neural network which helps to overcome the nonlinearity problem associated with the chromosome images. The input features taken for the study are relative length of the chromosomes, centromere position and density profile of the G band chromosome image. The limitations of the proposed work are the computation time taken by the network and the centromere position of the telocentric chromosome.

Zhenzhen Kou (2002) proposed a molecular cytogenetic method for detecting the chromosomal imbalance. The method proposed is the comparative genomic hybridization which helps in automated karyotyping. This is proved to be better than DAPI images which need enhancement. The author gives a simple feature extraction process that deals with density profile of the chromosomes. Support vector machine (SVM) is then used for the classification process. SVM is considered as a better classifier because it takes only limited training samples. The success rate of the algorithm is more than 90%.

Guimar Zes et al (2003) proposed a method based on the shape of the chromosome for the classification of human chromosomes. The shapes of the human chromosomes are converted to signatures and decomposition process is carried out using wavelet packet transform. Chromosomes are classified into six groups associating wavelet packet coefficients with corresponding chromosome shape signatures. The six groups are assigned with alphabets like A,B,C,D,E,F and G.
Delshadpour (2003) proposed a technique for automated classification of human chromosomes using MLP. The MLP helps in reducing the dimensional rate for better classification. The output of the neurons is reduced to an order of $n \to \log_2 \{n\}$ and this helps in reducing the dimensionality of the network, required number of training data, training time and generalization of the network error. Copenhagen database images are taken for the analysis. 304 images are taken for the study and the classification accuracy is 88%. The training time of the proposed MLP is five times better than the standard MLP.

Pravesh Biyani et al (2005) proposed a joint optimization solution for classification and pairing of human chromosomes. The problem is addressed using optimal three dimensional assignment of maximum likelihood function with objective function. Two different problems arise with this optimization process. The problems are estimation of posterior probability for classifying the chromosomes into their respective classes and three dimensional assignments which suffer NP. The method gives a better classification result. This is obtained using the statistical features of the chromosomes. Lagrangian-type relaxation method is used to overcome the NP problem to achieve the optimal solution.

Wade Schwartzkopf (2005) proposed a method for chromosome segmentation and classification by multispectral information in M-FISH chromosome images. Probabilistic model of M-FISH is used for both classification and segmentation processes. The method is proved to be robust for chromosomal identification. It also helps in detecting various anomalies related to chromosomes. The work is better than the conventional M-FISH
classification where the segmentation information is not considered for processing chromosomal images.

Qiang Wu et al (2005) proposed an automated classification for prototyping chromosomal images using subspace based approach. Chromosomal prototype helps in representing chromosome images for a given class which are synthesized quantitatively from the subspace. The features for chromosome classification are obtained from the subspace using transformation coefficients. The subspaces are derived using Principal Component Analysis (PCA), Discrete Cosine Transform (DCT) and Fisher’s linear discriminant analysis. Prototype of two dimensional images, classification of images and chromosome profiles are obtained using these subspaces. The proposed method proves good in prototyping chromosome images and subspace derived using PCA and DCT performs well compared to WDD technique.

Petros Karvelis et al (2006) proposed an automated method for classifying M-Fish chromosomal images. Morphological watershed segmentation is applied on intensity gradient image which helps in decomposing the image into a set of homogeneous regions. Bayes classifier is applied for classification of each region. The overall accuracy is about 89% and the technique is better than few other methods using watershed basins.

Ibrahiem El Emary (2006) proposed a software tool for analysing chromosomal images. The work is about filter the noise present in the images and enhance the quality of the image. After filtering, segmentation is performed on overlapping and touching chromosomal images. Classification is done further by computing the length of the chromosome. Length is given as input to the multi-layer neural network with back-propagation algorithm.
Length of chromosomal images are obtained by thinning process. The quality of the image is improved by 91.7%.

César Martínez et al (2007) proposed an Elman network which is a partial recurrent connectionist model which helps in analysing banding patterns and iterative contextual classification algorithm is used for better classification. Copenhagen dataset images are taken for analysis. This method gives a better performance compared to other neural network approaches.

Petros Karvelis & Dimitrios Fotiadis (2008) proposed a region based decorrelation stretch transform on decomposed multichannel images with homogeneous regions for enhancing the classification ratio. 15 images are taken for the analysis. A Bayes pixel by pixel classifier is used and an improvement of 12.28% is achieved.

Xingwei Wang et al (2008) presented a comparative analysis of two classifiers which include decision tree and Artificial Neural Network (ANN). The two classifiers help in detecting the analysable and unanalysable cells. Decision tree takes input features of individual chromosomes and ANN takes the input features from metaphase cells. The method is tested on 170 images. The processes adopted are image filtering, thresholding and labelling few set of features are also extracted for better classification. The receiver operating characteristic curve for the complete database is 0.93.

Petros Karvelis et al (2008) proposed a multichannel watershed transform to perform segmentation which helps in obtaining the similar spatial and spectral characteristic regions. Later classification is done using Bayes classifier. The inputs for the proposed method are M-FISH chromosomal images. The basic steps of the method are calculating the gradient magnitude of the chromosome image, marking the homogenous
regions using watershed transform and combining the similar adjacent regions. The overall accuracy is about 82.4%.

Hyohoon Choi et al (2008) proposed expectation maximization (EM) algorithm with feature normalization method and also an unsupervised non parametric classification method is introduced for analysing M-FISH chromosomal images. EM algorithm with feature normalization helps in improving the pixel classification accuracy by about 20%.

Artem Khmelinskii et al (2008) described a strategy for pairing the homologous chromosomes with mutual information. This information helps in discriminating the textural difference between the chromosomal images. 836 chromosomal images are taken for analysis and supervised non linear classifier is employed for identifying the homologous chromosomes.

Yaser Rahimi et al (2008) described the chromosome classification using multi-layer feedforward perceptron neural network by extracting features like surface of the chromosomes, boundary pixels of the chromosomes and calculating the six momentums. The overall accuracy of the work is about 73%.

Benoit Legrand et al (2008) developed a Dynamic Time Warping (DTW) classifier for classifying chromosomes based on length and density profile. DTW helps in classifying chromosomes with various elongations. The classification accuracy is about 81%. DTW requires only a small number of datasets compared to Bayesian classifiers and neural networks. DTW helps in detecting the chromosomal abnormality better than other methods.

Ganesh Vaidyanathan et al (2008) discussed a non linear decision boundary for analysing the two different classes of chromosomes namely 18
and 19. Optimal boundary points of the two classes are obtained using decision boundary by performing the sampling on two dimensional feature spaces. The classification accuracy of the method is 94%.

Petros Karvelis et al (2009) proposed the Bayes classifier to improve the classification accuracy compared to pixel by pixel classifier. Image filtering is performed by vector median filter. 183 chromosomal images are taken and the improved classification accuracy is 9.99% compared with pixel by pixel classifier. This work helps in identifying the abnormalities related with deoxyribonucleic acid.

Petros Karvelis et al (2009) developed a classification model for M-FISH images using the Support Vector Machine (SVM). Multichannel watershed segmentation is applied for image segmentation and SVM is used for pixel classification. The classification accuracy is improved by 10.16% compared to Bayesian classifier.

Sunthorn Rungruangbaiyok & Pornchai Phukpattaranont (2010) presented a probabilistic neural network classifier to classify the chromosomes into their respective (24) classes. The features considered for the classification of chromosomes are area, perimeter, band area and profile and singular value decomposition. Probabilistic neural network is used to group the chromosomal images in two steps with six groups in the first step and 24 groups in the second step. Density profile is sampled at 10, 30, 50 and 80 samples respectively and classification results for female and male are 68.19% at 30 samples and 61.3% at 50 samples respectively.

Delie Ming & Jinwen Tian (2010) presented the pale path algorithm to segment the touching and overlapping chromosomes. Middle point algorithm is employed for extracting the medial axis of the
chromosomes. Multiscale wavelet B spline algorithm is then used to enhance 
the banding pattern of the chromosomes which helps in extracting the shape 
and gradient profiles. The weighted density distribution algorithm is used to 
calculate the chromosomal pattern and classify the chromosomes to their 
respective classes using multilayer classifier.

Enea Poletti et al (2012) proposed an improved method to estimate 
the medial axis of the chromosomes. This is used to extract the features for 
classification of the chromosomes. Estimation of polarization is done after 
marking medial axis which helps in estimating the orientation of the 
chromosomes. Normalization techniques and feature rescaling are the results 
of polarization. They help in reducing the intra class variance with an increase 
in inter class variance. A novel class reassignment algorithm is used along 
with the neural network classifier to increase the probability of classification 
rate. A total of 5474 chromosomes are considered for classification and an 
accuracy of 94% is achieved.

Wenzhong Yan & Lei Bai (2013) proposed automated 
chromosomal classification by ANN and wavelet transform. The work is a 
comparative study of these two algorithms. The Copenhagen and Genzyme 
dataset of M-FISH chromosome images are used in the classification process.

The algorithms which are applied on the chromosomal images for 
segmentation, parameter identification and classification are summarized in 
Table 2.1.
<table>
<thead>
<tr>
<th>Author</th>
<th>Segmentation Method</th>
<th>Accuracy (in %)</th>
<th>Remarks</th>
</tr>
</thead>
</table>
| Gady Agam (1997)       | Hypothesis Based Approach   | -              | • Worked on overlapping and touching contour images  
• Concave/ convex points and separation lines are identified  
• Hypothesis based analysis is performed |
| Petros Karvelis et al (2005) | Watershed Transform (WT) | 92             | • Worked mainly on touching chromosomes  
• Segmentation done by concave points  
• 940 images considered (340 touching and 29 overlaps) |
• Separation done using voronoi diagram with possible cut points obtained by DT  
• 28 out of 35 overlaps are separated correctly |
90 for overlaps. | • Worked on overlapping and touching chromosomes  
• 30 metaspread images with 1380 chromosomes are considered for analysis  
• Segmentation done based on Hypothesis tree |
Table 2.1 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Segmentation Method</th>
<th>Accuracy (in %)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrico Grisan et al (2009)</td>
<td></td>
<td>90</td>
<td>162 image from 117 cells with 6683 chromosome are taken for analysis</td>
</tr>
<tr>
<td>Petros Karvelis et al (2010)</td>
<td>Watershed Transform</td>
<td>90.6 for touching and 80.4 for overlapping</td>
<td>• Worked on overlapping and touching M-FISH chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 183 M-FISH images are taken for analysis</td>
</tr>
<tr>
<td>Mousami Munot et al (2011)</td>
<td>Modified Snake Algorithm with Greedy Approach</td>
<td>100 for 2 touch 95 for 3 &amp; 4 touch</td>
<td>Worked only on touching chromosomes</td>
</tr>
<tr>
<td>Mukul Joshi et al (2012)</td>
<td>Computational Geometry</td>
<td>100 for 1 &amp; 2 overlaps and 88 for 3 and 4 overlaps.</td>
<td>• Worked only on overlapping chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Tested on synthesized images of LK1 data set</td>
</tr>
<tr>
<td>Mousami Munot et al (2013)</td>
<td>Computational Geometry</td>
<td>98 (only cut point identification)</td>
<td>• M-FISH images are considered for analysis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Worked only on overlapping chromosomes</td>
</tr>
<tr>
<td>Tanvi &amp; Renu Dhir (2014)</td>
<td>Computational Geometry</td>
<td>87.4</td>
<td>• Worked only on overlapping chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Cut points are tracked along contour for separation of overlaps</td>
</tr>
</tbody>
</table>
Table 2.1 (Continued)

Segmentation and Separation of Overlapping and Touching Chromosome Images

<table>
<thead>
<tr>
<th>Author</th>
<th>Segmentation Method</th>
<th>Accuracy (in %)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devaraj Somasundaram (2014)</td>
<td>Geometry Separation Algorithm</td>
<td>94</td>
<td>• Worked only on overlapping chromosomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Cut points and cut lines are identified and hypothesis verification done for separation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Homologue chromosome identification is done by centromere position</td>
</tr>
</tbody>
</table>

Parameter Identification

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Accuracy (in %)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goesta Granlund (1976)</td>
<td>Centromere using Integrated Density Profile</td>
<td>70</td>
<td>• Four different types of descriptors are obtained using integrated density profile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 230 chromosomes are considered, the recognition rate for above descriptors are 69.1%, 70.1%, 64.7% and 75.2%</td>
</tr>
<tr>
<td>Frans Groen et al (1989)</td>
<td>Centromere using Piecewise Linear Approximation Method</td>
<td>85 and 68 for Leyden data base and 93 and 76 for Copenhagen data base</td>
<td>• Closest pair of opposite contour points are identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Profile for the width of chromosomes is plotted</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Medial axis is obtained for identification of centromere</td>
</tr>
<tr>
<td>Author</td>
<td>Method</td>
<td>Accuracy (in %)</td>
<td>Remarks</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Xingwei Wang et al (2008)</td>
<td>Centromere and Polarity Assignment using Multi Stage Rule Based Algorithm</td>
<td>91.4 for centromere 97.4 for polarity assignment</td>
<td>• Three feature profile are extracted&lt;br&gt;• Medial axis of chromosome is obtained using thinning algorithm</td>
</tr>
<tr>
<td>Mousami Munot et al (2012)</td>
<td>Banding Pattern Similarity Measure</td>
<td>100 for group A 97 for group C chromosomes</td>
<td>• Nearness factor is calculated for banding pattern similarity measure&lt;br&gt;• 50 images of group A and C chromosomes are taken for the analysis</td>
</tr>
</tbody>
</table>

### Classification

<table>
<thead>
<tr>
<th>Author</th>
<th>Classifier</th>
<th>Accuracy</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boaz Lerner et al (1998)</td>
<td>MLP Trained Back Propagation Classification Algorithm</td>
<td>90</td>
<td>• Classification-Driven Partially Occluded Object Segmentation is performed&lt;br&gt;• Two geometrical features (length and Centromere index) are considered for classification</td>
</tr>
<tr>
<td>Author</td>
<td>Classifier</td>
<td>Accuracy</td>
<td>Remarks</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| John Conroy et al  | Singular Value Decomposition, Principal Component Analysis, Fisher Discriminant Analysis and Hidden Markov Models | 97       | • Tested on G band chromosome images of well known database (Philadelphia, Edinburgh, and Copenhagen)  
• Markov model proves better for classification |
| Zhenzhen Kou        | Support Vector Machine                                                      | 90       | • Simple feature extraction process done with density profile and use normalized profile as feature vector |
| Delshadpour         | Multi Layer Perceptron Neural Network                                        | 88       | • Output neurons are reduced to an order of $n \to \log_2 \{n\}$ helps in reducing dimensionality of the network, required number of training data, training time and generalization of network error.  
• Copenhagen database images are taken for the analysis.  
• 304 images are taken for the study |
<p>| Petros Karvelis et al | Bayes Classifier                                                          | 89       | Morphological watershed segmentation is applied on intensity gradient image which helps in decomposing the image into a set of homogeneous regions |
| Xingwei Wang et al  | Decision Tree and Artificial Neural Network (ANN)                           | 86       | Tested on 170 images and process adapted are image filtering, thresholding and labeling algorithms |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Classifier</th>
<th>Accuracy</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petros Karvelis et al (2008)</td>
<td>Bayes Classifier</td>
<td>82.4</td>
<td>Spatial and spectral characteristic regions are taken as input for classifier</td>
</tr>
<tr>
<td>Yaser Rahimi et al (2008)</td>
<td>Multi-Layer Feedforward Perceptron Neural Network</td>
<td>73</td>
<td>Features for classification are surface of chromosome, boundary pixel of chromosomes, and six momentums</td>
</tr>
<tr>
<td>Benoit Legrand et al (2008)</td>
<td>Dynamic Time Warping</td>
<td>81</td>
<td>Features for classification are length and density profile</td>
</tr>
<tr>
<td>Sunthorn Rungruangbaiyok &amp; Pornchai Phukpattaranont (2010)</td>
<td>Probabilistic Neural Network</td>
<td>68.19 for female samples 61.3 for male samples</td>
<td>Features considered for chromosome classification are area, perimeter, band area, profile and singular value decomposition</td>
</tr>
</tbody>
</table>
| Enea Poletti et al (2012)      | Neural Network Classifier                       | 94       | - 5474 chromosomes are considered for the classification  
- Medial axis and polarization estimation is done for feature extraction.  
- Novel class reassignment algorithm is used along with the neural network classifier to increase the probability of classification rate |
2.5 SUMMARY

From the literature survey it is understood that various chromosomal abnormalities are detected by analysing the images. Various algorithms are proposed by researchers for chromosomal analysis. The most important steps in chromosomal analysis are segmentation and separation of overlapping and touching chromosomes and parameter identification for classification of chromosomes. Many segmentation algorithms are applied either on overlapping or touching chromosomes. Real time chromosomal images are very rarely used. Parameter identification is a unique process for every chromosome and depends on the structure of the chromosome.

The proposed work applies a common algorithm for overlapping and touching chromosomes. Identification of parameters is done for curved, bent and other irregular structures in chromosomes.