Indian J.Sci.Res. 17(2): 154-157, 2018 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

CHROMOSOMAL KARYOTYPE AND ABNORMALITY DETECTION USING EFFICIENT CLASSIFIER ALGORITHM

[1]Kishore. T, [2]Harish Kumar. K, [3]Vanitha. L
[1,2,3] Department of Electronics and Communication Engineering,Prathyusha Engineering College, Chennai

Abstract -The main objective of this system is to categorize the human chromosomes and determines the chromosomal disorders automatically without human intervention. Microscopic chromosome image is acquired, processed and features are extracted and k-means algorithm is used for categorization.A human cell has 23 pairs of chromosomes, accounting to a total of 46 per cell. 50 images for each pair chromosomes are considered from standard database and the featureslength, width, area, entropy, standard deviation are extracted from the image. For each pair 30 chromosomes are used for training the Support Vector Machine (SVM) and 20 chromosomes are used for testing. Karyotyping is the identification, classification and presentation of 24 classes (Chromosome labeled from 1-22, X chromosome and Y chromosome) into a solitary picture.Any deviation from the normal karyotype, in terms of chromosome number or structure is known as chromosomal abnormality. The presence of 3 numbers in chromosome 21 is called as Trisomy 21 or Down Syndrome.

Keywords–Karyotype, S V M ,Chromosome abnormality, Down Syndrome

I.Introduction

Chromosomes are complex structures found in a cell nucleus, mainly the "packages" that contain the DNA. They contain thousands of genes, whichmanage our physical and medical individualities, such as hair colour, blood type and proneness to disease.The chromosomes appear as thin, thread-like structures when observed under a microscope. The chromosomes have a long arm and short arms which are partitioned by a primary constriction called centromere. The short arm is considered as '***p'*** and the long arm as '***q'***.A normal human cell contains 22 pairs of chromosomes (labeled as chromosomes 1-22) and one pair of sex chromosomes; females have two X chromosomes, while males have one X and one Y chromosome. Thus a normal human cell contains 46 chromosomes.

A Chromosome band is described as a section of a chromosome,which shows relatively darker or lighter stain as compared to the adjacent sections of the same chromosome.

All the twenty three pairs of chromosomes have an explicit band pattern.Figure 1shows the Karyotype image of a female chromosome. Figure 2 shows the Karyotype image of a male chromosome.

Figure 1: Karyotype image of a Female Chromosome

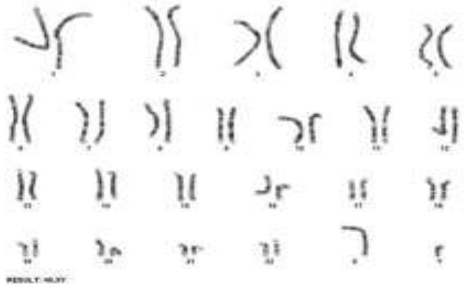

Figure 2: Karyotype image of Male Chromosome

Human chromosome analysisis an indispensable task in cytogenetic, particularly in prenatal screening, genetic syndrome diagnosis and cancer pathology research.One of the objectives of chromosome analysis is the establishment of karyotype, which is used to examine the features of chromosomes and to predict several genetic disorders.Problemsencounteredbythe technicians in examiningthehuman chromosomes are: (1) Counting the number of chromosomes (taking into consideration ofcoinciding and touching chromosomes), (2) In order to get accurate results, this process might be repeated several times with various types of cells. Thus,mutually the time and effort to accomplish these tasks are relatively long.

II.Methodology

Figure 3 illustrates the block diagram of Pattern Recognition System. The main objective of pattern recognition is the categorizing of patterns and sub patterns in an image. A pattern recognition system includes:Subsystem to describe pattern class, Subsystem to extract designated features and Subsystem for classification known as classifier.The classifier used in this work is Support Vector Machine (SVM) algorithm.

[1]Corresponding Author

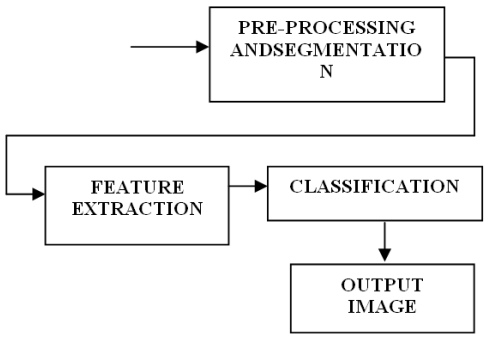

Figure 3: Block diagram of Pattern Recognition System

The different modules of this work are:

1. Pre-processing

2. Feature extraction

3. SVM for classification

4. Identification of chromosome abnormalities

A. Pre-Processing

The main purpose of pre-processing is to eliminatenoise from the image. This is mandatory for the reliable extraction of features as feature extraction algorithms givespoor results in the existence of a noisy background.

B. FeatureExtraction

The three main approaches for feature extraction and classification based on the type of features are Statistical approach, Syntactic or structural approach and Spectral approach. In statistical approach, pattern is defined by a set of statistically extracted features characterized as vector in multidimensional feature space. The statistical features might be based on first-order, second-order, or higher-order statistics of gray level of an image. In case of syntactic approach, texture is described by texture primitives, which are spatially organized according to placement rules to produce a complete pattern. A systematic relationship is drawn between the structural design and syntax of the language.In spectral method, textures are described by spatial frequencies and are calculated by autocorrelation function of a texture.

The features are extracted by defining the medial line of a chromosome by applying the medial axis transformation (MAT). Thinning algorithm is used that iteratively removesthe edge points of a region subject to the constraints that deletion of these points does not remove end points, does not break connections.The different features extracted from chromosomes are relativelength, relative area, centromericindex(C.I.), mean, standard deviation, entropy, contrast, correlation and variance

A. Relativelength:

The length of everychromosome is determined bycounting the number of pixels in the medial line. Therelativelengthofthei-thchromosome(l_{ri})canbeacquiredbynormalizingthemedial axislengthusingthefollowingequation.

$$l_{ri} = \frac{l_i}{l_t} \qquad (1)$$

where,

l_i(i=1,2,…,24) – lengthof i-thchromosome

l_t -totallengthof all46Chromosomesof onecell.

B.Relativearea:

The relative area of the$i$-thchromosome(A_{ri})canbeobtainedby countingthepixelsof thechromosomebodyandbynormalizingthearea usingthefollowingequation.

$$A_{ri} = \frac{A_i}{A_t} \qquad (2)$$

where

A_i(i=1,2,…,24) – areaof$i$-thchromosome

A_t-total areaof all46 chromosomesofonecell.

C.Centromericindex(C.I.):

$$\text{C.I}=\frac{\text{short arm length}}{\text{whole length of medial axis}}$$

whole length of medial axis --- (3)

D. Mean:

$$\mu_p = \frac{1}{n^2}\sum_{r=0}^{n-1}\sum_{s=0}^{n-1} P_{r,s} \text{--- (4)}$$

where,

$P_{r,s}$ is pixel at location (r,s).

E. Standard deviation:

$$\sigma_p = [\ \frac{1}{n^2}\sum_{r=0}^{n-1}\sum_{s=0}^{n-1}[\ P_{r,s} - \mu_p\ ]^2]^{1/2} \text{--- (5)}$$

F. Entropy:

It is a measure of randomness or degree of disorder in a system.

$$e = -\sum_{b=0}^{L-1} p(b) log_2\, p(b) \text{--- (6)}$$

where,

$p(b) \approx N(b)/n^2$ for $\{0 \le b \le L\text{-}1\}$,

where,

L-number of different values which pixels can adopt

N(b) = number of pixels of amplitude (b) in the pixel

window of size 'n×n'.

The co-occurrence matrix personifies the spatial interrelationships of the gray tones in an image. The values of the co-occurrence matrix elements existing relative frequencies with which two neighboring pixels areparted by a distance d and at an angle θ appear on the image.One of which has gray level i and j and their joint likelihood of occurrence is given by Pi,j.

G. Contrast:

Contrast is defined as

$$\sum_{n=0}^{N_g} n^2 \quad (7)$$

where $n=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{i,j}\cdot|i-j|$ --- (8)

Ng = number of gray levels

H. Correlation:

Correlation is an amount of gray-level linear dependency of the image.Correlation feature is defined as

$$\text{Correlation} = \frac{\sum_i \sum_j (i,j) P_{i,j} - \mu^2}{\sigma^2} \quad \text{--- (9)}$$

where μ and σ are the mean deviation and standard deviation of the co-occurrence matrix respectively.

I. Variance:

Variance is the measure of how much the gray level is varying from the mean.

$$\text{Variance} = \sum_i \sum_j (i,j) P_{i,j} - \mu^2 \quad \text{--- (10)}$$

A. Chromosome categorization

Chromosome categorization is prepared using Support Vector Machine (SVM). SVM is a discriminative classifier defined by a separating hyperplane. The features are given as training data (supervised learning), and the algorithm outputs an optimal hyperplane which categorizes the testing dataset. In two dimentionalspaces this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

B. Chromosome abnormality recognition

Chromosome abnormality usually occurs when there is a mistake in cell partition. Any divergence from the usual karyotype(physical composition), in terms of chromosome number or structure, is known as a chromosome abnormality. The chromosomal abnormalities are categorized as: Numerical abnormalities and Structural abnormalities.

C. Numerical Abnormalities:

When either one pair of chromosome is missing from a couple (monosomy) or has more than two chromosomes of a couple (trisomy), it is called Numerical Abnormality. The set of abnormalities associated with exact symptoms is known as Syndrome. An example of numerical abnormalities is Down syndrome which is due to three copies of chromosome 21, rather than two. Turner Syndrome is an example of monosomy, where the individual - in this case a female - is born with only one sex chromosome(X) instead of normal two X chromosome.

D. Structural Abnormalities:

When the chromosome's structure is changed it is called as Structural Abnormalities. This can take numerous forms as deletions, duplications, translocations inversions, rings.

This paper, deals only with Numerical abnormality. The output of the neural network is used to find out the Numerical abnormality. If the number of chromosomes in a group is less than or greater than the required number, then numerical abnormality is reported.

III.Implementation And Results

The input to the feature removal algorithm is the chromosome images. The outline vectors (features) extracted from the images is fed as input to the SVM. Large database are necessary for the classifier to execute the classification correctly. In this system 30 images per chromosome number are used for training and 20 images for testing. The Classification accuracy and error rate is obtained by using the succeeding formula:

$$\text{Accuracy} = \frac{\text{number of correctly classifiedchromosomes}}{\text{total number of chromosomes}} \quad \text{--- (11)}$$

$$\text{Error Rate} = \frac{\text{number of misclassifiedchromosomes}}{\text{total number of chromosomes}} \quad (12)$$

Table 1: Result of SVM classification

ALGORITHM	CLASSIFICATION ACCURACY	ERROR RATE
SVM	91 %	9 %

Chromosome Abnormality Detection:

The result of the Second stage classifier is analysed for irregular number of chromosomes. Turner's syndrome which shows unnatural female chromosomes is shown in Figure 5. The output of this stage is displayed in Table4 which shows only one X

chromosome instead of two X chromosomes. Figure 4 shows the image for Down syndrome with 3 chromosomes in chromosome 21.

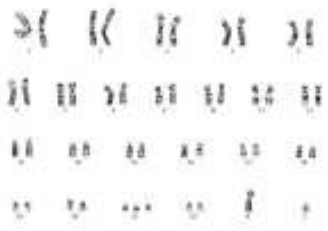

Figure 4:Down syndrome – Three chromosomes in Chromosome 21

Tools for Experiment:

In this study, MATLAB is used as a programming language for feature extraction, chromosome tabulation and to find the chromosomal abnormalities.

Table 2. Chromosome Abnormality diagnosis

Class Number	Number of Chromosomes	Class Number	Number of Chromosomes
1	2	13	2
2	2	14	2
3	2	15	2
4	2	16	2
5	2	17	2
6	2	18	2
7	2	19	2
8	2	20	2
9	2	21	3
10	2	22	2
11	2	X	1
12	2	Y	1
Total Number of Chromosomes : 47			
Diagnosis : Patient is suffering fromDownSyndrome			

IV.Conclusion And Future Scope

The classifier SVM algorithm is used to categorize the chromosome data and the system has the ability to find the numerical abnormalities of the chromosomes. This work focuses with the diagnosis of numerical abnormality of chromosomes. This work can be continued to diagnose various anatomical abnormality of chromosomes in forthcoming years.

References

[1] Guisong Liu and Xiaobin Wang, " An Integrated Intrusion Detection System by using Multiple Neural Networks", IEEE,pp22-28,2008

[2] Syed Zahid Hassan and BrijeshVerma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", Seventh International Conference on Intelligent Systems Design and Applications,pp.503-507,2007

[3] J.Cho, S.Y.Ryu, S.H.Woo,"A Study for the Hierarchial Artificial Neural Network Model for Giemsa stained Human Chromosome Classification", IEEE conference,pp.4588-4591,2004.

[4] Ibrahiem M. M. El Emary," On the application of Artificial Neural Networks in Analyzing and classifying the Human Chromosomes", Journal of Computer science 2 (1): 72-75, 2006,ISSN 1549-3636, 2006 Science Publications

[5] B.D.Singh, "Genetics", Kalyani Publishers,2005.

[6] Earl Gose, Richard Johnson-baugh, Steve Jost, "Pattern recognition and image analysis", Prentice Hall of India Private Limited, New Delhi,2000

[7] LaureneFausett, Fundamentals of Neural Networks, Pearson Education,2007

[8] Simon Haykin, Neural Networks – A comprehensive foundation, Pearson Education,2001

[9] Robert C. Gonzalez, Richard E.Woods, Steven L.Eddins, Digital Image Processing using MATLAB, Pearson Education, 2005

[10] Duane Hanselman, Bruce Littlefield, Mastering MATLAB 7, Pearson Education,2008.

[11] http://www.genome.gov/

[12] http://www.ncbi.nlm.nih.gov

[13] www.genetics.org

[14] www.ias.ac.in/jgenet

[15] www.vivo.colostate.edu/hbooks/genetics/medgen