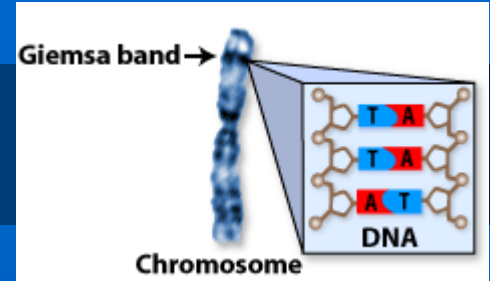# BCUAKS (Biomedical Cairo University Automated Karyotyping System)

Chromosomes Classification Based on Neural Networks, Fuzzy Rule Based, and Template Matching Classifiers
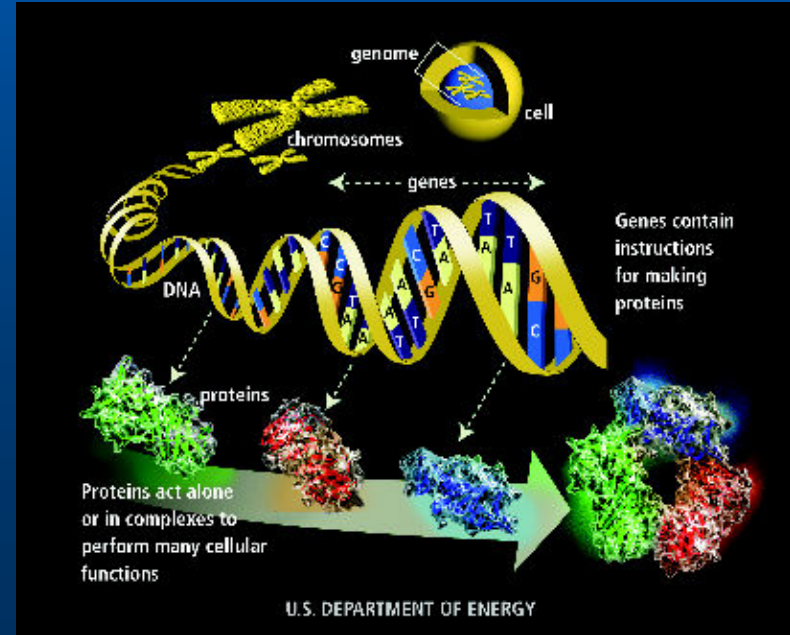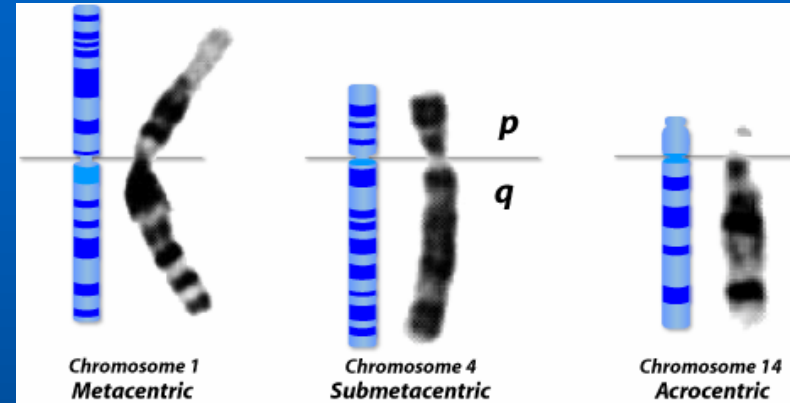
Introduction

Giemsa band → DNA
Chromosome

Chromosome?

•The chromosome is a very long DNA molecule that contains many genes and associated proteins, that carry information of an organism.

•The number of chromosomes in human cells is 46 with 22 autosomal (non sex) pairs and 2 sex chromsosmes – 2 X chromosomes for the female or an X and an Y for the male.

Chromosome 1
Metacentric

Chromosome 4
Submetacentric

Chromosome 14
Acrocentric

genome

cell

chromosomes

genes

DNA

Genes contain instructions for making proteins

proteins

Proteins act alone or in complexes to perform many cellular functions

U.S. DEPARTMENT OF ENERGY
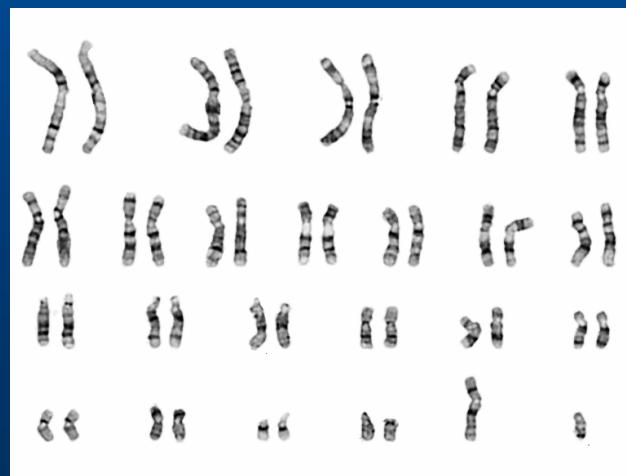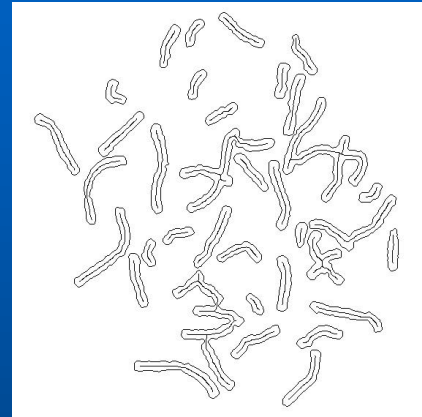
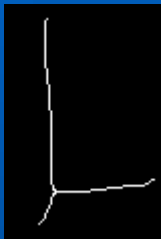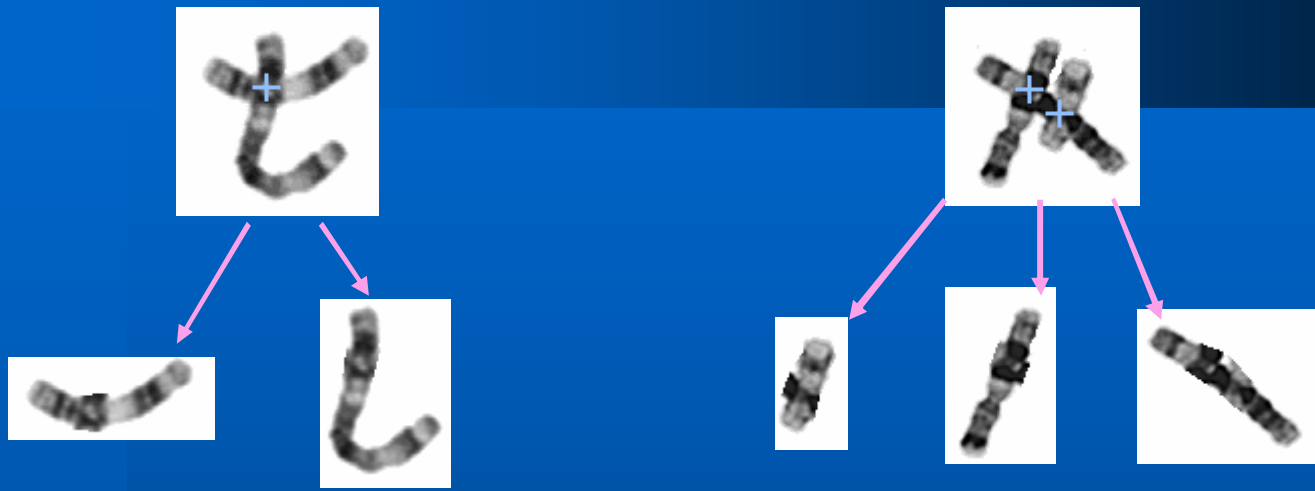# Karyotyping Problem Definition

Problem?

Time?
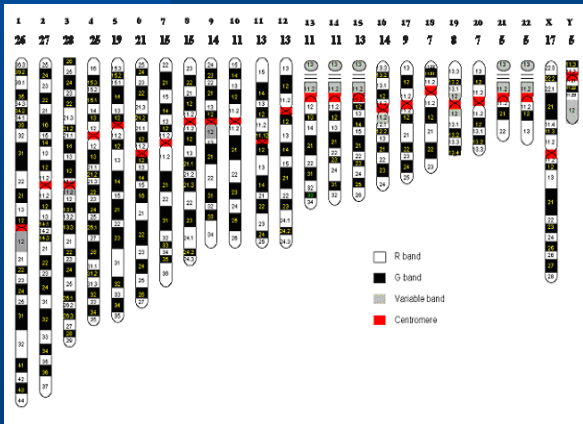
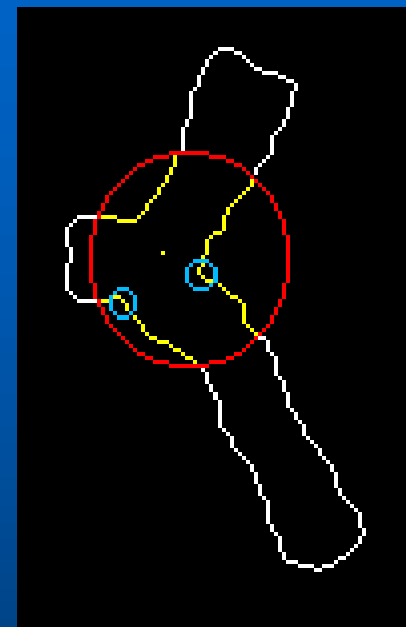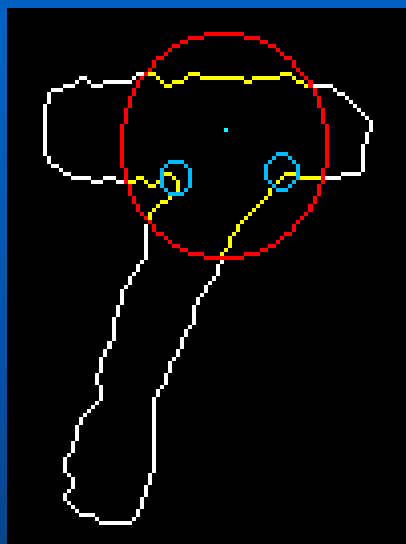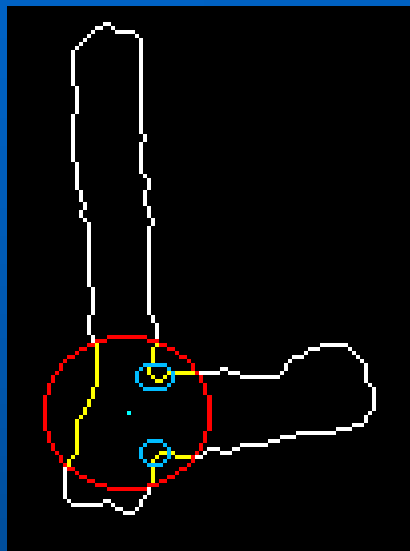# Image Separation

- **Bifurcation examples**

- **Cross-Over examples**

CO

BF

Cut Locations

BF

CO
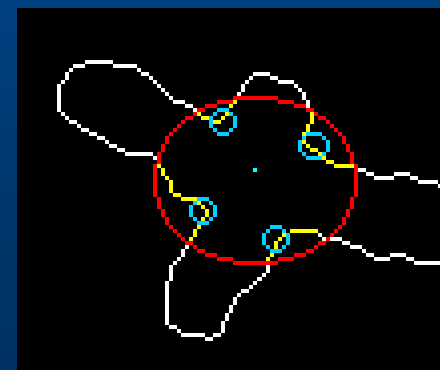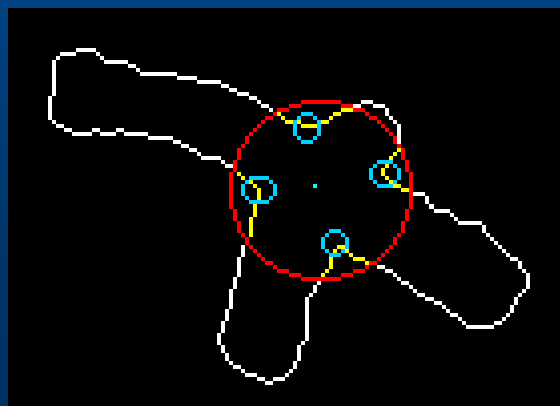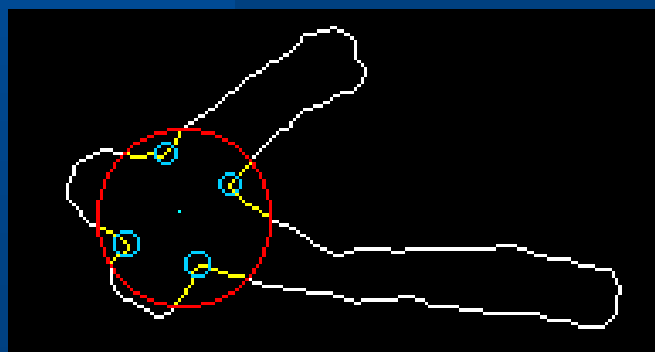
Centromere Localization

- **Centromere is first located on the chromosome's contour as the two points (far apart in the sequence by a certain threshold) having the minimum distance between each others.**

- **Then the centomere is located on the medial axis as the intersection point between the medial axis and the line joining the 2 centromeric points of the contour.**

Chromosome 1
Metacentric

Chromosome 4
Submetacentric

Chromosome 14
Acrocentric
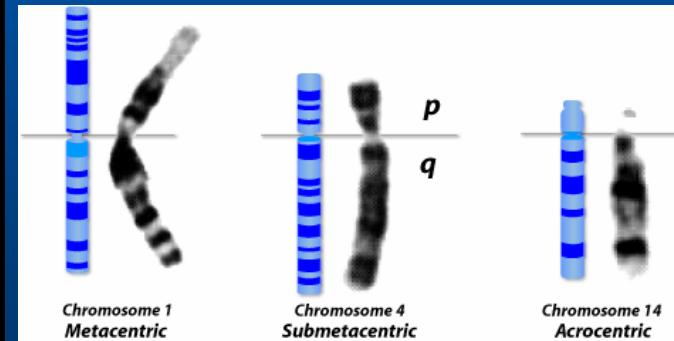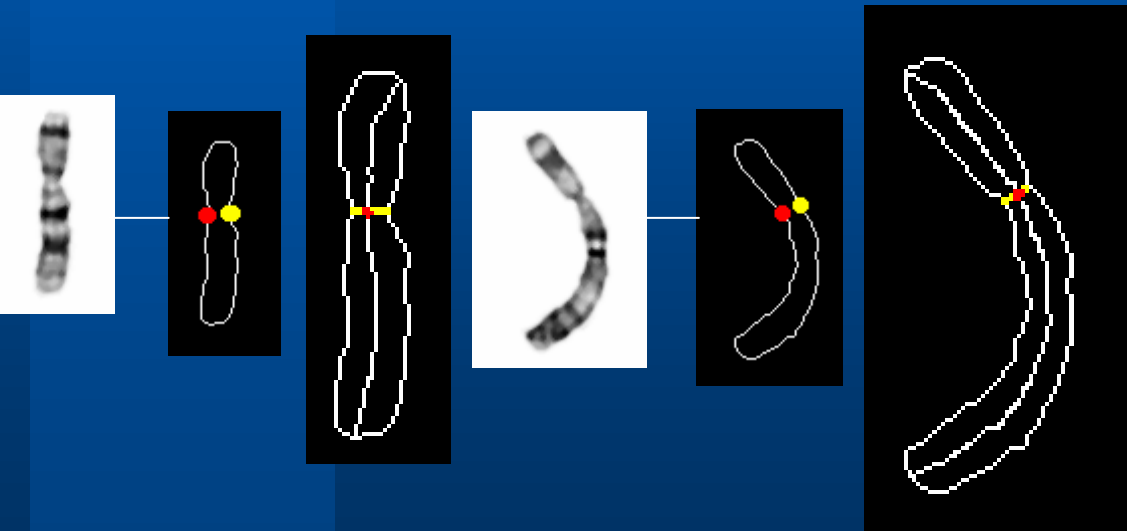
Features Extraction

Geometric and grayscale features were extracted.
The following categories of features were then extracted:

1) G-Banding grayscale profile features:
These are 12 fixed thickness grayscale average features.

2) Global chromosome's features:
a) Medial axis length (Normalized)
b) Contour length (Normalized)
c) Area (Normalized)
d) Mass (Normalized)

3) Chromosome's centromeric features:
a) Medial axis length ratio of P/Q (Normalized)
b) Contour length P/Q ratio (Normalized)
c) Area P/Q ratio (Normalized)
d) Mass P/Q ratio (Normalized)

4) Number of real bands:
Number of maxima & minima in profile.

5) Real gray level banding:
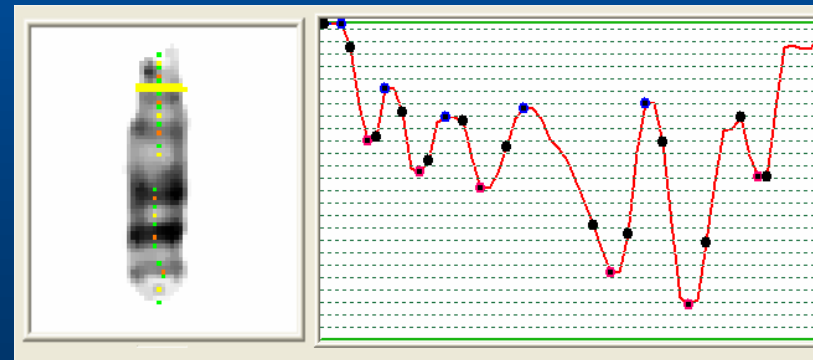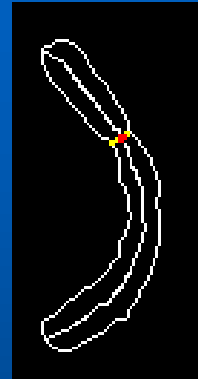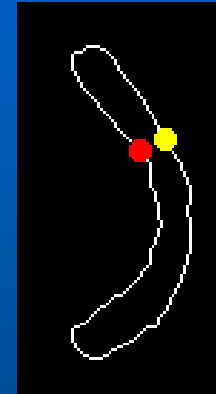These are the average graylevel of the real band.

6) Distances between centers of bands:
These are the distances between the bands centers
 (along medial axis).

7) Real bands thickness:
These are the real banding thickness (along medial axis).

Proposed Matching Methods

A. Neural Networks

Chromosomes features Input Layer — Hidden Layer — Chromosome Class Output Layer
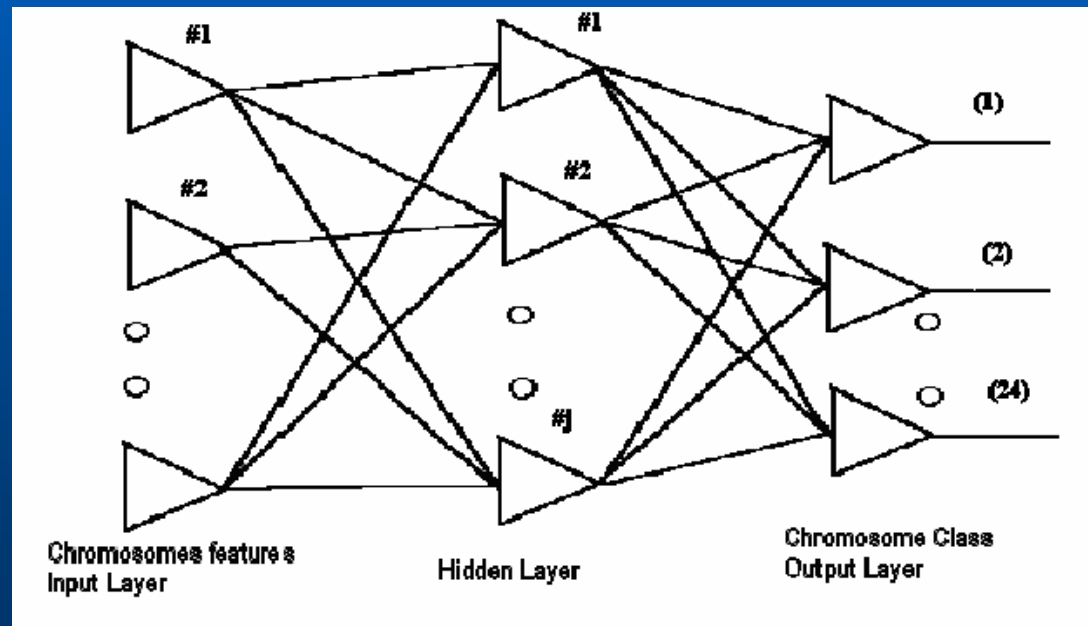
B.   *Fuzzy Rule Based*

eatures vectors were first fuzzified using its defined membership functions of 11 fuzzy sets each as shown.

he training set of 593 vectors for 24 classes of hromosomes were used to derive the fuzzy rules. 79 vectors are used to test the system.

utput fuzzy sets are shown.

uzzy rules were first generated, validated, nd then were used to evaluate the system via n inference mechanism using *SUP MIN* ompositional rule of inference
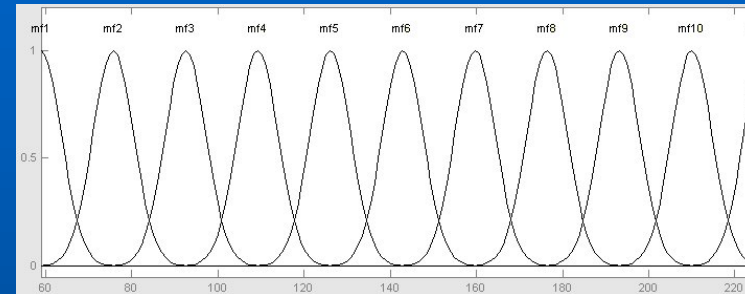
Input fuzzy sets

Output fuzzy sets

C. Template Matching Method

24 template chromosomes types are from chromosome 1-22, X, and Y chromosomes.

Features were **normalized and stretched** to get an overall average template of **fixed length** for every type (**42 bands, 108 features**).
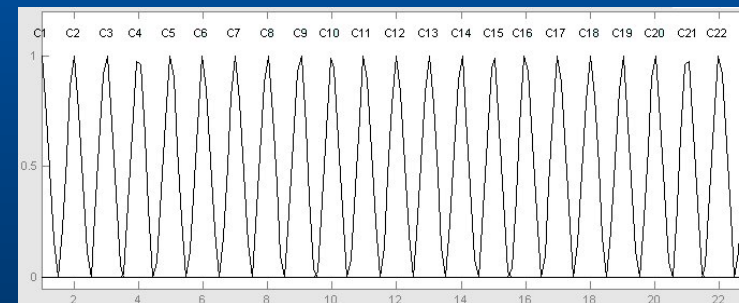
We have a set of reference patterns (**templates**) and we have to decide which one of these reference patterns an unknown one (**test pattern**) matches best.

A defined simple measure for fixed length features is the **Euclidean** distance cost measure.

Euclidean cost measures for all features after stretching to fixed size were calculated.

12 cost functions (contour-length, medial-axis length, area, mass, contour length-centromeric-index, medial-axis-length-centromeric index, mass-centromeric-index, banding-numbers, 12-fixed-length-banding-profile, real bands grayscale, and real-band-thicknesses) **were weighted** using **neural network** for maximum recognition, where we trained a net taking the inputs as the costs and the output is either one if correctly classified and zero if not.

In classification, first we get the costs of different features of this chromosome with either the templates or the whole references, then entering these costs to the net, the one having highest output will be classified to that chromosome average template or reference.
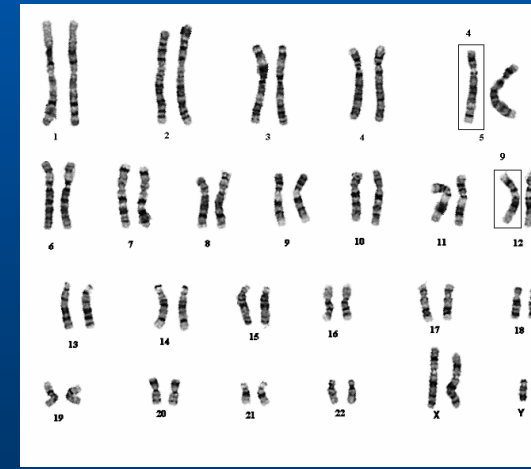
Matching Flowchart
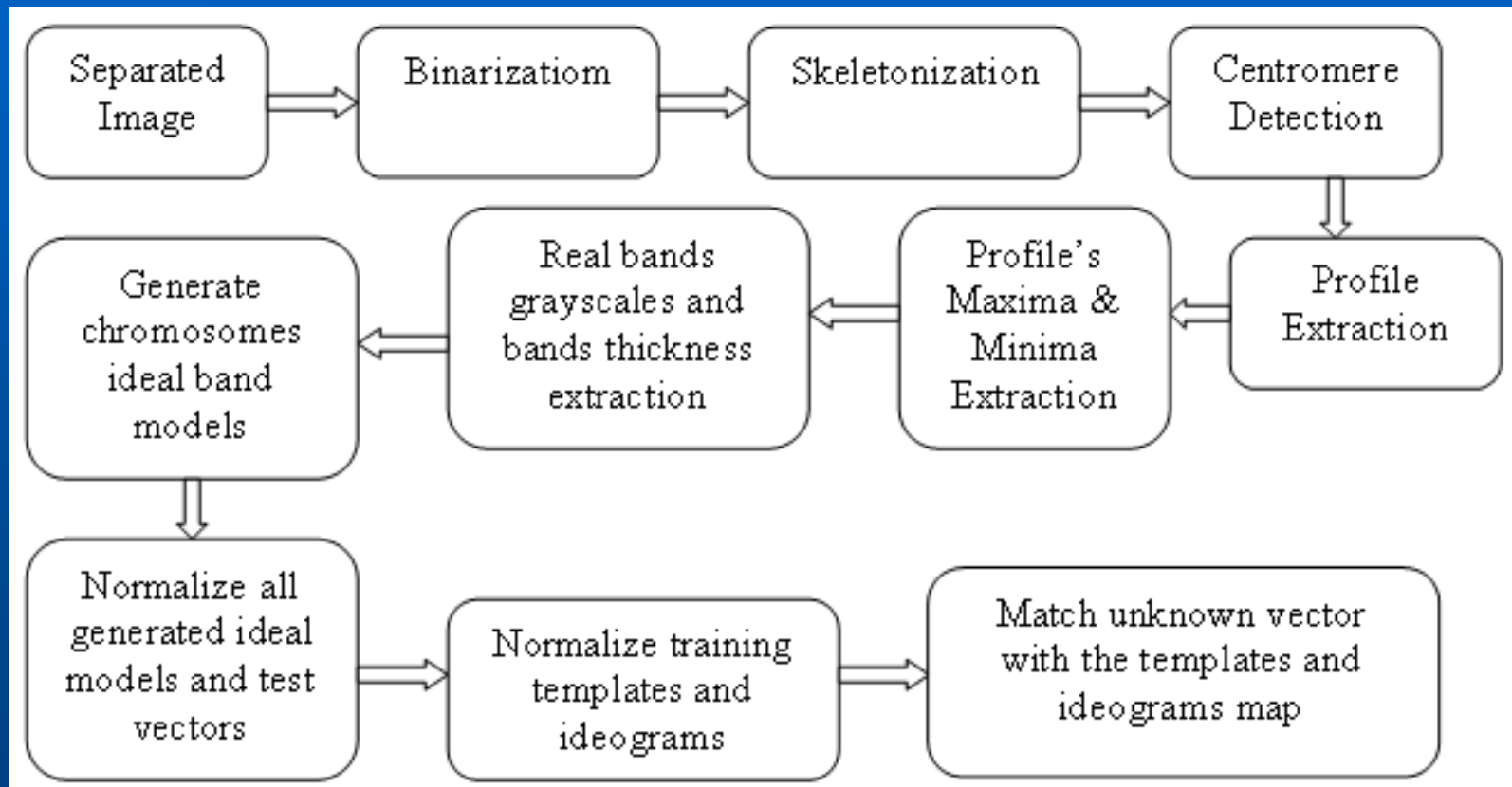
Results

Classification Method	Training %	Testing %
Fuzzy Rule Based Classification	100	93.54
Neural Networks Classification	98.81	94.76
Matching with 24 average templates	96.9	95.96
Matching with all reference vectors	97.65	96.89
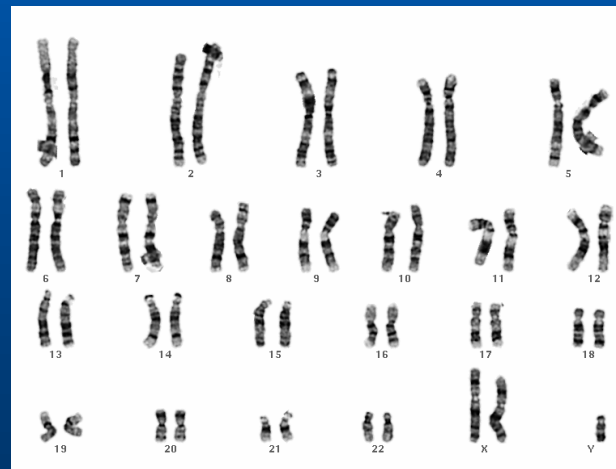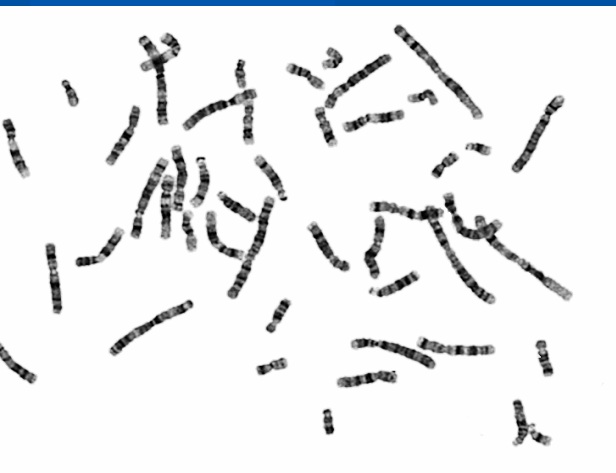
Conclusions

A new features extraction and classification algorithms for real G-banded chromosomes images is proposed. Real bands grayleves features and real band thicknesses features as well as real number of bands (hills and valleys) were automatically extracted for every chromosome together with the shape and fixed banding features.

Classification results were found to be more than 93% and were compared for different classification methods. The matching comparison results of the three classification methods using the whole features categories suggests the usage of template matching method with all the references since it has the highest classification rate among other methods (96.89%).

Real banding graylevels and thickness will be suitable to analyze chromosome deletions and translocations via   bands registration to diagnose structural abnormalities.

Acknowledgments

- **This work was partially supported via a Cairo University grant for chromosomes separation and classification (2002). We are grateful to Professor of Clinical Pathology, Dr. Mervat El-Ansari at Cairo Medical School for her valuable suggestions and data preparation of this work.**

Thank You