CrossMark

# A survey of neural network based automated systems for human chromosome classification

**Faroudja Abid**[1] · **Latifa Hamami**[1]

**Abstract** Chromosome classification and karyotype establishment are important procedures for genetic diseases diagnosis. Various computer-aided systems have been developed to automate this tedious and time consuming task, which is performed manually in most cytogenetic laboratories. This paper provides a comprehensive review of past and recent research in the area of automatic chromosome classification systems. We start by reviewing methods for feature extraction, followed by a neural network based chromosome classifiers survey. We sum-up various techniques and methods in this area of research and discuss important issues and outcomes within each study for both chromosome feature extraction and classification. Although the ANN based chromosome classifiers are the main topic of this survey, a number of classifiers based on other algorithms are exposed to give an overall idea about additional techniques employed in chromosome classification.

**Keywords** Classification · Human chromosome · Neural Network

## 1 Introduction

Cytogenetic concerns the cell activity study, focusing on the chromosome analysis, which is an important procedure for genetic diseases diagnosis at an early stage. The genetic diseases are a consequence of chromosomal abnormalities (Natarajan 2002) that can be structural or numerical. Traditional methods for chromosome karyotyping process are performed manually in most cytogenetic laboratories; this process is tedious and time consuming, therefore expensive (Lerner 1998). The computer-aided systems for chromosome classification are needed to automate the chromosome analysis and help cytologists to perform this task.

Since the 1980s several researches have focused on automatic chromosome classification and numerous computer-aided systems (Lerner 1998; Moradi and Setarehdan 2005; Lerner

✉ Faroudja Abid
abidfaroudja@gmail.com

1 Département d'électronique, Ecole Nationale Polytechnique, 10 Avenue Hacen Badi, BP 182, El Harrach, Alger, Algeria

and Lawrence Neil 2001; Popescu et al. 1999; Poletti et al. 2008; Carothers and Piper 1994; Wang et al. 2005) have been developed to automate this process for abnormalities and genetic diseases detection (Wang et al. 2010; Catherine et al. 1998; Mohammadi 2012). Automatic chromosome classification system includes mainly three principal parts: Image processing, feature extraction and classification. The image processing step begins by improving the quality of the metaphase images (images taken at the third stage of cell division) in order to remove the noise factors, for better readability that facilitates the analysis. After that the segmentation (procedure that separates the chromosomes from the background of the image) and the skeletonization of the chromosome images are carried out for chromosome feature (geometrical and/or band pattern based features) extraction. The last step is the chromosome classification process; the purpose of this step is to assign each chromosome to its corresponding class. According to the research reported in the field of computer-aided systems for chromosome classification; the artificial neural networks (ANN) are the widely used methods, due to their parallelism that increases efficiency and reduces the computational complexity and time.

This paper presents an overview of computer aided systems for human chromosome classification. We mainly give a survey of the approaches reported in the literature regarding chromosome feature extraction and ANNs based chromosome classifiers; the first aspect of automation, i.e. chromosome images processing, is not in the scope of this paper. The rest of this paper is organized as follows: Sect. 2 briefly presents the human chromosome abnormalities. Sect. 3 undertakes a review of the previous works related to feature extraction, whereas Sect. 4 deals with works exposing ANN based systems for chromosome classification. In Sect. 5 we give an overview of a number of additional classifiers based on other algorithms apart from the ANN. A discussion regarding the important issues within the techniques exposed in Sects. 3 and 4 is given in Sect. 6. Finally, Sect. 7 concludes this paper.

## 2 Human chromosome abnormalities

Chromosomes that carry the genetic information are structures found in the nucleus of cells; they appear in cell division phases. A normal human cell contains 46 chromosomes arranged by karyotyping into 22 pairs of autosomes (classes 1–22) and two sex chromosomes (gonosomes either XY (male) or XX (female)). The correlation between human diseases and chromosome abnormalities has been recognized since 1956, when Tjio and Levan (1956) discovered that the number of human chromosomes was 46. To identify chromosomes different staining methods have been developed (Estandarte 2012). For example, the Giemsa stained metaphase images where the chromosomes appear as a succession of dark and light bands. This technique is called G-banding; it was the most commonly staining method used in karyotyping. Additional staining techniques, such as Fluorescence In-Situ Hybridization (FISH), the Comparative Genomic Hybridization (CGH) and Multiplex-FISH (M-FISH) (Siepcher et al. 1996) are employed for color karyotyping. The M-FISH technique uses five colors and an extra fluorophore, DAPI (4-6-Diamidino-2-phenylindole) to stain chromosomes with different colors; each chromosome class appears with a distinct color. This allows deletions and translocations detection. The CGH are three colors (red, green and blue) images that are employed for detecting chromosomal imbalances.

The chromosomal abnormalities are the result of an accident either during meiosis or during mitosis, which can involve one or more chromosomes. They are a cause of genetic diseases, thus analyzing chromosomes morphology is an important procedure for cancer
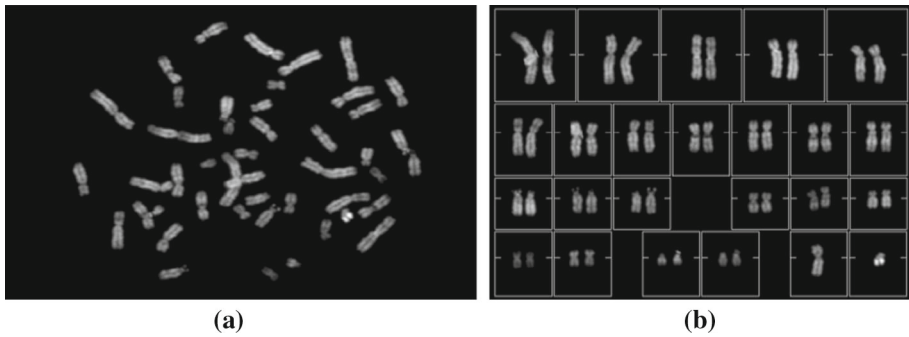
**Fig. 1** chromosomes cell and its corresponding karyotype **a** Metaphase cell, **b** Human karyotype (Bioimlab)

and genetic diseases diagnosis at an early stage. The chromosomes analysis result is the karyotype, which consists of the arrangement of the chromosomes in a standard format that helps in abnormalities detection and diagnosis procedure. Figure 1 shows chromosomes cell and its corresponding karyotype.

There are two main types of chromosome abnormalities, namely, numerical abnormalities that include either missing chromosome (monosomy) or extra chromosome (trisomy). The trisomy is the presence of three chromosomes of a pair instead of two; it affects some pairs of autosomes (13, 18 and 21) and the sex chromosomes. The structural abnormalities are changes that affect the structure of the chromosomes. They take several forms, namely, deletions, insertions, inversions and translocations. They are two types of translocation, reciprocal and Robertsonian translocation (Catherine et al. 1998).

Both structural and numerical abnormalities have been associated with human health. As stated in Wang et al. (2005) the chromosome aberrations are present in perhaps 20 % of all conceptions, in 50 % of early spontaneous abortions, in 10 % of mentally retarded individuals and in many cancers.

We summarize in Table 1 the common types of chromosome abnormalities and their descriptions.

## 3 Feature extraction

A reliable system for chromosome classification is mainly related to the feature extraction; a good features vector allows an appropriate classification. This section undertakes a review of previous works related to feature extraction. Diverse features have been used to describe chromosomes; the mainly used features are geometrical and chromosome banding pattern based features, which are represented by a density profile (DP). Density profile value is computed as the mean intensity of pixels along the line perpendicular to the chromosome axis. Figure 2 depicts an individual chromosome with its medial axis (the approximation of the central curve of the chromosome) and density profile curve. The most important and widely used geometrical features are the chromosome length (L) (usually the total length is determined by the number of pixels along the medial axis) and the centromeric index (CI), which is the ratio of the length of the short arm (p) to the total chromosome length. The short arm (p) and the long arm (q) are joined by the centromere that is the thinnest region of the chromosome as shown in Fig. 2.

**Table 1** The common types of chromosome abnormalities

| Chromosome abnormalities | | | Structural abnormalities | |
|---|---|---|---|---|
| Numerical abnormalities | | | | |
| Type of abnormality | Description | | Type of abnormality | Description |
| Polyploidy | Presence of extra chromosomes in each pair. Triploid (presence of 3 chromosomes in each pair) Tetraploid (presence of 4 chromosomes in each pair) | | Deletions | Missing or deletion of chromosomal material |
| Aneuploidy | Monosomy Missing of one chromosome e.g. Turner syndrome (sex chromosome missing) | | Insertions | Presence of additional chromosomal material |
| | Trisomy Extra chromosome in one pair of autosomes or extra chromosome in the sex chromosomes pair. pair 21 (Down syndrome) pair 13 (Patau syndrome) pair 18 (Edward syndrome) Sex chromosomes pair (Klineflter syndrome) | | Inversions | In this case a portion of the chromosome has broken off, turned upside down and reattached, therefore the genetic material is inverted |
| | | | Translocations | Chromosomal material has moved from one chromosome to another, they are two types of translocation, reciprocal and Robertsonian (fusion of two chromosomes by their centromere) |
| | | | Isochromosomes | Chromosomes with identical arms |
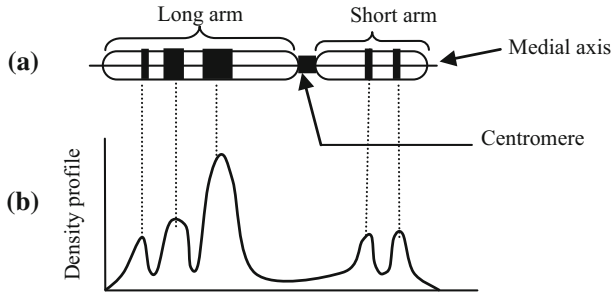
**Fig. 2** Individual chromosome with its medial axis **(a)** Density Profile curve **(b)**

The centromere location is an important process for chromosome classification; according to the centromere position, the chromosomes are classified into three groups. The metacentric group, where the centromere is in the middle (the q-arm and the p-arm have the same length (chromosomes 1, 3, 18, 21, 22)). The submetacentric group; the chromosomes of this group have unequal length (chromosomes 15, 16, 17, 23, 24). The acrocentric group includes chromosomes where the centromere is located near the end.

Several methods have been reported in the literature for automatic centromere location (Lerner et al. 1995; Moradi et al. 2003a, b; Madian and Jayanthi 2014; Wang et al. 2008a, b). In Wang et al. (2008a, b) the authors developed an automatic computerized scheme for the centromere location and chromosome polarity assignment. They used 2287 chromosomes extracted from 50 bone narrow metaphase cells, including 26 normal and 24 abnormal (leukemia) cells. The thinning algorithm is used to identify the medial axis; and three features profiles are extracted, namely shape profile, density profile and banding profile. The accuracy of 91.4 % is obtained for centromere location and 97.4 % for polarity assignment for all groups of chromosomes. Authors affirmed that compared to the literature their scheme has lower accuracy only for centromere identification in small size group of chromosomes (19, 20 and 22), which are either overlapped or bent.

Another approach for centromere identification, which is based on vertical projection vectors (the pixels of each column of the binarized chromosome image) and horizontal projection vectors (the pixels of each line of the binarized chromosome image) is proposed by Moradi et al. (2003a, b). The centromere is regarded as the total minimum in the horizontal projection vector. Authors applied their approach on 87 chromosomes; the centromere location recognition rate is about 96 %, however this algorithm presents a principal limitation in the case of strongly curved chromosomes. The same authors proposed another method (Moradi et al. 2003a, b) based on the automatic localization of the key points (landmarks) in the skeleton of the chromosomes. These landmarks are used for the parameter extraction.

In most reported methods regarding the centromere identification and other feature extraction, the chromosome's medial axis is first estimated. This estimation is based on the MAT (Medial Axis Transform) which is the mainly applied technique to detect the medial axis (Mohammadi 2012; Lerner et al. 1995; Moradi et al. 2003a, b; Madian and Jayanthi 2014; Ming and Tian 2010; Khan et al. 2012), since it preserves the shape of the chromosome. In the study exposed in Lerner et al. (1995) the MAT has been used for feature extraction; two approaches were used to MAT computation based on the binary image of the chromosomes. The first approach is based on the skeletonization (Xia 1989) using the fire front's propagation and extinction. The second one is the piecewise linear PWL approximation that is more suitable for the bent chromosomes. In this approach, the second order moments of

the chromosome gray levels is used to calculate the orientation of the principal axis; the columns perpendicular to the principal axis are considered, and the points on the middle of each column is found. A linear extension of the axis is applied for both approaches to obtain a continuous medial axis. The authors concluded that the medial axis extracted using the skeleton approach is finer than the PWL based one.

Recently, Madian and Jayanthi (2014) addressed a comparison study between two approaches for the centromere position determination, namely the MAT approach and the projection vector approach. The authors concluded their study by the fact that the MAT approach is adapted for only the straight chromosomes and the performance of this algorithm decreases if the chromosomes are curved. The data sets used in this study include55 chromosomes. The centromere identification rate is about 90 % for the projection vector and about 78 % when the MAT is applied. Authors claimed that projection vector proves better than MAT in identifying the centromere position for bent chromosomes. However the two algorithms present limitations when high curved chromosomes are considered. The centromere location is also used in chromosome polarity assignment studied in Poletti et al. (2012), Gregor and Granum (1991) and Wang et al. (2008a, b).

The work exposed in Gregor and Granum (1991) discusses a markov network model, which is used for centromere finding, where only band pattern information is considered. The band pattern features are also used as unique features for chromosome classification in numerous works. For example, chromosomes shape has been used in the method proposed in Uttamatanin et al. (2013a, b) for classification based band resolution. This study shows that the numerical abnormality is well detected by the low resolution bands whereas high resolution bands are appropriate for structural abnormality. In Eskiizmirliler et al. (1996) the band patterns of chromosomes are represented by applying wavelet transform technique on the gray level profiles of chromosomes.

The feature extraction scheme that utilizes the active contour and B-spline curve modelling for the parameters extraction is developed in the method proposed by Theerapattanakul et al. (2005). The contour of the chromosomes is extracted initially by using active contour technique; this technique is also used for the extraction of the chromosome's medial axis, and then density profiles computation. Technical B-spline is then applied to the chromosome in order to extract the centromere. An enhanced standard Laplacian thickness measurement algorithm was discussed in Arachchige et al. (2013) for centromere location. Authors in Loganathan et al. (2013) developed an algorithm that helps in calculating chromosome's length by curve fitting method based on skeletonization.

The fuzzy logic is also considered in centromere detection, the work reported in Mousavi et al. (2002) deals with an iterative fuzzy segmentation algorithm for centromere segmentation. The intensity of the segmented centromere are used to classify the homologs chromosome 22 into two classes (maternal and paternal) the data base used in this study includes 12 images, which are stained using DAPI and Fluorescein isothiocyanate-centromere (FITC), this probe has a green appearance and it allows mainly the chromosome centromere visualization.

Accordingly much research is done for automatic and accurate feature extraction, mainly centromere finding, which is an important parameter for chromosome classification. Table 2, summarizes the important issues (applied techniques, features, data sets, and accuracies) within a set of studies considered in this paper, regarding feature extraction methods; particularly methods related to centromere identification since it is the widely used feature for classification.

**Table 2** Summary of some studies regarding feature extraction methods exposed in this paper

| References | Methods | Features searched | Data sets #chromosomes | Accuracy % |
|---|---|---|---|---|
| Madian and Jayanthi (2014) | MAT and PV | Centromere | 55 | 78 % (MAT) |
| | | | | 90 % (PV) |
| Wang et al. (2008a, b) | MAT | Centromere and polarization | 2287 | 91.4 % (centromere identification) |
| | | | | 97.4 % (polarization) |
| Moradi et al. (2003a, b) (1) | PV | Centromere | 87 | 96 % |
| Moradi et al. (2003a, b) (2) | Automatic localization of the key points in the chromosomes skeleton (*landmarks*) | Centromere | 219 (groupe E) | NA (not available) |
| | | Length | | |
| Theerapattanakul et al. (2005) | MAT active contour and B-spline | Centromere | NA | NA |
| | | Length Density profile | | |
| Lerner et al. (1995) | MAT | Geometrical features and density profile | 622 | 98 % for MAT |
| | PWL | | 5classes (2, 4, 13, 19 et X). | 97 % for PWL |
| Gregor and Granum (1991) | Markov network model | Centromere | 6985 from 179 cells | 94 % correctly estimated centromere |

**Table 3** summary of some studies regarding ANN based chromosome classification systems exposed in this paper

| References | Methods | Data sets composition | Accuracy (%) | |
|---|---|---|---|---|
| Esküzmirliler et al. (1996) | Hybrid ANN Kohonen and MLP | 178 normal metaphase images 8818 chromosomes | Maximum error rates 33.1 % | |
| Lerner et al. (1995) | MLP | 150 cells including 622 single chromosomes | Probability of correct test set classification 98 % | |
| Graham et al. (1992) | Combined MLP | 8106 (Copenhagen data sets) 4569 (Edinburgh data sets) 5817 (Philadelphia data sets) | Single MLP 6.9 % 19.3 25.6 | Compound MLP 6.2 % 17.1 % 23 % |
| Wang et al. (2009) | Two-layer classifier with eight ANNs (combined ANNs) | 6900 chromosomes | 86.7 % | |
| Uttamatanin et al. (2013a, b) (1) | A rule based classification (RBC) | 192 images | 99.42 %. for the first class | |
| | | 7817 segmented objects | 90.67 % for the class 2 89.44 % for the class3 93.25 % for the class4 | |
| Uttamatanin et al. (2013a, b) (2) | RBC of band-resolution chromosome images | 68 images 822 straight chromosomes 1012 chromosomes including artifactual | 73.08 % for low and band resolution group 95.24 % for high band resolution group | |
| Poletti et al. (2012) | A two-layer ANN | 119 Q-band images 5474 chromosome | 94 % | |
| Ming and Tian (2010) | A multistage classifier (multilayer and Bayesian classifier) | NA | 85.6 % for the multistage 72.3 % for Bayesian classifier | |

**Table 3** continued

| References | Methods | Data sets composition | Accuracy (%) |
| --- | --- | --- | --- |
| Rungruangbaiyok and Phukpattaranont (2010) | Probabilistic ANN | 60 images | 68.18 % female cell images |
| | | 2760 chromosomes | 61.30 % for male cell images |
| Oskouei and Shanbehzadeh (2010) | Wavelet NN (WNN) | 450 chromosomes | 94 % |
| Ruan (2000) | Fuzzy Hopfield neural network CFHN | 100 chromosomes in each | Identification error rate 0 % |
| | | 24 chromosome classes | Un-identification error rate of 3.33 % |
| Badawi et al. (2003) | 1-A Multi-layer neural network (MLNN) | G-banded images | 94.76 % for MLNN |
| | 2-A fuzzy rule based classifier | | 93.54 % for fuzzy rule based |
| | 3-Template matching | | 96.98 % for template matching classifier |

## 4 Neural network based chromosome classification

Substantial studies have investigated automatic chromosome classification. The artificial neural network based approach for chromosome classification has been greatly used. A multiplicity of ANNs are included, principally the MLP (Multi-Layer Perceptron) based system (Popescu et al. 1999; Lerner et al. 1995; Moradi et al. 2003a, b). In Lerner et al. (1995), authors proposed the MLP based classifier of the 5 classes of the chromosomes. In this study a features selection mechanism was applied. The later is based on the Knock-out algorithm to select the most dominant features. Chromosomes of five classes (2, 4, 13 19 and X) that are extracted from 150 cells including 622 single chromosomes of the amniotic fluid are considered. From the MAT of each chromosome, 66 features (64 DP + L + CI) are extracted. Authors used several combinations of the features for testing their MLP based classifier. According to this study the CI and L features are more suitable for classification than the DP features. A probability of correct test set classification of 98 % is reported by the authors. A combination of ANNs has been also used in several works (Wang et al. 2009; Graham et al. 1992; Errington and Graham 1993). In Wang et al. (2009) a multti-step classifier that includes two decision layers with eight ANNs was discussed. Wang et al. applied a genetic algorithm (GA) to find the optimal topology of the ANN including optimal features set as input neurons and the number of hidden neurons. Authors claimed that their classifier achieves 86.7 % classification accuracy and a small data set is needed to train each ANN. In Wang et al. (2008a, b), the same authors developed an automated scheme for metaphase chromosome cells identification and classification. The chromosome cells are classified into analysable and un-analysable cells using two classifiers, namely decision tree (DT) classifier and an ANN classifier with three-layer feed-forward topology. According to the reported results in this work, the scheme can correctly identify more than 90 % analyzable banded chromosome cells and almost 85 % of un-analyzed cells are eliminated.

Another type of ANN such as a probabilistic ANN has been considered for chromosome classification (Sweeney et al. 1993; Rungruangbaiyok and Phukpattaranont 2010; Poletti et al. 2011). In Rungruangbaiyok and Phukpattaranont (2010) a probabilistic neural network with two layers was adopted for the classification process. This later was divided into two steps. In the first step, the chromosomes were classified into six groups, whereas in the second step, the chromosomes of the six groups were classified into 24 classes. The feature vector components used in this study consisted of: chromosomes area, length, band's area, maximum value of singular value matrix from SVD (singular value decomposition) and density profile. These features were extracted from 60 images including 2760 chromosomes. The classification rate was about 68.18 and 61.30 % for female and male, respectively. According to the authors, probabilistic neural network offers faster training time compared to back-propagation neural network (a processing time of 3.50s is reported for the probabilistic NN).

A Wavelet neural network was proposed for chromosome classification in Oskouei and Shanbehzadeh (2010). In this work, a classifier that combines the wavelet into neural network (WNN) is developed. The WNN performs classification of chromosomes in group E (16, 17, 18). The features vector includes length, centromere index, normalized average gray density (NAGD) position, intensity and width (extracted from density profile). The data sets used in this study include 450 chromosomes (150 for each class of the group E). A comparison of the performances of the WNN against the ANN was addressed in this work. It is reported that the WNN improves the training and accuracy compared to ANN classifier, due to the nonlinear characteristic of the WNN that fits the nonlinearity of chromosome structure. The best classification accuracy was about 94 and 89 % for the WNN and ANN, respectively.

A hybrid ANN (HANNE) structure that combines a supervised ANN (MLP) for numerical abnormalities detection and unsupervised ANN (Kohonen) for structural abnormalities detection was discussed in Eskiizmirliler et al. (1996). The second stage of this hybrid ANN consists of 24 ANNs, which are assigned to each class of chromosomes. In this study Eskiizmirliler et al. applied two algorithms that they developed, namely an automatic object detection algorithm (AODA) and object skeletonization algorithm (OSA) for chromosomes detection and feature vector components calculation, respectively. The feature vector components are length, centromere, gray level profile values, area and the band patterns of chromosomes. Regarding the data sets used in this work, 178 metaphase images including 8818 chromosomes are considered. Authors compared the classification performance of their hybrid classifier (HANNE) with the rule based classifier (RBC). They found that the maximum error rates were about 33.1 and 36.7 % for the HANNE and RBC respectively.

A chromosome classifier based on a fuzzy Hopfield neural network was developed in Ruan (2000), the exposed CFHN (Classifier Fuzzy Hopfield Neural network) is a Neurocomputing and fuzzy computing model. Authors claimed that for test data including 100 chromosomes in each 24 chromosome classes, their CFHN has an identification error rate of zero and the un-identification error rate of 3.33 %. The used features include chromosome's length, centromere index and density profile.

In Khan et al. (2012) a multistage classifier for chromosome recognition and classification is discussed. The classifier includes a multilayer NN and a Bayesian classifier, which is a statistical classifier that predicts the class probabilities. In the first stage the chromosomes are classified into nine groups using geometric features, two geometrical parameters are used, namely length and centromere index extracted after medial axis estimation using the midpoint algorithm. In the second stage, chromosomes of each group carried out in the first stage are classified by a Bayesian classifier. In this stage, the band profile features including average gray profile, gradient profile and shape profile are employed. The multistage classifier accuracy rate is about 85.6 % for geometric classification (first stage) and 72.3 % for Bayesian classifier (second stage). Authors compared the accuracy of their multistage classifier and a single Bayesian classifier (that gives an accuracy of 60.7 %); they found that the multistage classifier has better effects than single Bayesian one.

A comparison between three automatic classifiers for G-banded images, namely a multi-layer neural network, a fuzzy rule based classifier and a template matching classifier was addressed in the study exposed in Badawi et al. (2003). The features vector used in this work includes both geometrical and banding grayscale profile features. The classification results that Badawi et al. reported are 94.76, 93.54 and 96.98 % for the multi-layer NN, fuzzy rule based classifier and template matching classifier, respectively.

In Poletti et al. (2012) authors proposed a modular scheme for automatic classification. This scheme includes feature extraction based on the medial axis, which is estimated by using the modified vessel-tracking algorithm that was developed by the authors. The features vector includes chromosome length, 64 density profile, 64 contour function, area and perimeter. Note that the two last geometrical features are not based on the medial axis. The features are extracted from 119 Q-band chromosome images including 5474 individual chromosomes (the image data sets used in this work is publicly downloadable (Bioimlab)). In this scheme the chromosome polarity assignment is performed before the classification step. The reported performance of the system is about 94 % with a two-layer ANN classifier. A comparison with the best methods reported in the literature is also made in this work.

A rule based classification of chromosome images for good metaphase spreads selection is discussed in Uttamatanin et al. (2013a, b). Three parameters have been defined, namely width ratio, height ratio and area ratio. The chromosomes are classified into four classes.

The first class denotes the straight individual chromosomes; the second class for the bended individual chromosomes and the third and fourth classes denote the overlapping chromosomes and artifacts, respectively. Authors found that the classification accuracies are about 99.42, 90.67, 89.44 and 93.25 % for the first class, the second class, the third class (overlapping chromosomes) and the fourth class, respectively.

In the work exposed in Stanley et al. (1998), the feed-forward neural network is used for determining primary chromosome and candidate chromosomes that are used later with the matching algorithm to identify the chromosomes within metaphase spreads. The maximum recognition rate was about 89.1 % for 55 metaphase spreads. Stanley et al. used the transportation algorithm (TP) (Tso and Graham 1991) as benchmark to compare their homologue matching algorithm. In Tso and Graham (1991), the karyotyping is considered as a transportation problem where the chromosome classes are considered as the source and the chromosomes as the destinations. A Backpropagation neural network (Cho 2000) and a Single-layer (Narayanana et al. 2004) are also considered for chromosome classification purpose. It is noteworthy to precise that even though a multiplicity of ANNs are employed in chromosome classification, the most used one is the MLP based ANN either as a single stage or a combination of MLPs. We give in Table 3 a summary of some studies considered in this paper regarding the ANN based chromosome classification systems; in term of methods (type of ANN considered), data sets used, and outcomes.

## 5 Additional types of chromosomes classifiers

Although, the ANN based chromosome classifiers are the main topic of this survey, we expose in this section a number of chromosome classifiers based on other algorithms apart from the ANN, such as the support vector machine (SVM), Bayesian and fuzzy logic based classifiers. We also present a number of works, regarding the classification of chromosomes with different stains, mainly the M-FISH.

The SVM algorithm, which is based on statistical learning theory (Vapnik 1999), has been used in chromosome classification in several studies (Kou and Zhang 2002; Lerner and Lawrence Neil 2001; Markou et al. 2012; Arachchige et al. 2013). This type of classifier is new compared to its predecessor classifiers, i.e., ANN based classifiers and Bayesian classifiers. The study reported in Kou and Zhang (2002) exposes an approach for CGH human metaphase karyotyping. The proposed approach is based on the SVM algorithm. Kou et al. used the integrated density profiles as features. According to this study the SVM classifier has good generalization ability and requires a limited sample for training. A success rate of 90 % is reported.

A comparison between four classification techniques, namely a SVM, a neural network (two layers Perceptron NN), a Bayesian NN (BNN) and a Naive Bayesian classifier (NBC) has been addressed in Lerner and Lawrence Neil (2001). Lerner et al. applied these four techniques to FISH signals extracted from 400 FISH images. The reported classification accuracies are 98.1, 98.8, 98.4, 94.0 % for NN, BNN, SVM and NBC respectively. Authors claimed that the BNN is the most accurate followed by the NN and SVM with a comparable accuracies; the least accurate technique is the NBC (Lerner 2004). Authors also reported the accuracy of 94.6 % for the linear classifier, which is less than the BNN, NN, SVM but it is better than the NBC. In Markou et al. (2012) a two-stage classification strategy based on SVM is exposed, the first stage, called one context-independent SVM classifier, is used for each chromosome class. The results of the classifiers of the first stage are used as inputs

for the second stage, which represents a context-aware classification for post classification. An error rate of 6.68 % was reported for 99 Giemsa-stained cell images. The fuzzy logic is also adopted in the chromosome classification filed (Choi et al. 2006; Sjahputera and Keller 1999).

In Choi et al. (2006) an unsupervised classifier based on fuzzy logic for M-FISH images is exposed. Choi et al. (2006) compared the achieved classification rate using the fuzzy logic classifier against a maximum likelihood (ML) classifier used in Sampat et al. (2002). Rate about 88 % for the fuzzy classifier and 89.8 % for the ML classifier are reported by the authors.

The ML is also employed in Schwartzkopf et al. (2005), in this study, Schwartzkopf et al. proposed a chromosome identification system based on a combination of chromosome segmentation and classification using a ML and multispectral information in M-FISH chromosome images. The multicolor chromosome images are also used for experiments in the studies reported in Lijiya et al. (2013) and Wang (2004). The Bayesian classifier is adopted for pixel by pixel classification of the M-FISH images in Lijiya et al. (2013); the intensity and size of the chromosomes are used as features. The proposed classifier was tested on five M-FISH images; an average classification accuracy of 91.86 % was reported.

## 6 Discussion

The most important and widely used features in chromosome classification are chromosome length, the centromere index (CI) which is calculated after centromere position identification and the chromosome banding pattern based features, which are represented by the density profile.

The chromosome features, mainly size, shape and banding pattern are derived from the medial axis. Hence the first step to extract these features is the chromosome's medial axis estimation. As shown in most references tabulated in Table 2, the medial axis is mainly extracted by using the MAT technique that is widely used, since it preserves the shape of the chromosome. Medial axis estimation is particularly useful for centromere location, chromosome length estimation (determined by the number of pixel along the medial axis) and DP computation.

It is noticeable that besides the MAT technique, which was the subject of several papers, other techniques have recently been applied to centromere identification, such as the projection vector (PV). This technique gives better accuracy for identifying the centromere position compared to the MAT technique, mainly for bent chromosomes. According to Madian and Jayanthi (2014) accuracy about 78 % for the MAT against 90 % for PV is observed. In Moradi et al. (2003a, b) authors observed almost 96 % when PV is applied. Additional techniques have been considered such as the piecewise linear approximation (PWL) and markov network model. An accuracy of 97 % is obtained with the PWL, while in case of markov network model about 94 % correctly estimated centromere is observed as reported in Gregor and Granum (1991).

Regarding the ANN based chromosome classifiers that are summarized in Table 3, the MLP based classifier has been widely adopted in the past reported works (Popescu et al. 1999; Lerner et al. 1995; Moradi et al. 2003a, b), whereas a multiplicity of ANNs are investigated in recent research. The classification rate with the MLP neural network varied from 78 to 98 % that is the maximum rate as shown in Table 3. In case of other types of ANN, accuracy

rates varied from 61.30 to 96.98 %. For example, rates about 68.18, 94.76 and 94 % for the probabilistic ANN, the multi-layer NN and the Wavelet NN, respectively were reported.

It is worth noting that even though the classification rate obtained for the ANNs besides the MLP is acceptable and comparable to MLP, this latter is the most suitable and the widely used ANN for chromosome classification. It gives the maximum rate, which is about 98 % as reported in Lerner et al. (1995).

This comparison is indicative because of the different chromosome data sets and images that have been used in each study.

Beside the neural network based chromosome classifiers, other types of classifiers have been reported, such as the SVM, Bayesian and fuzzy logic based classifiers. It is noteworthy to precise that these classifiers are mainly employed for multicolor chromosome images, such as the M-FISH images. The SVM classifier is new compared to the ANN based classifier. The comparison exposed in Lerner and Lawrence Neil (2001) shows that the SVM and ANN give comparable classification accuracies.

## 7 Conclusions

Substantial methods have been proposed to automate chromosome feature extraction and classification for abnormalities detection. In this paper we attempted to summarize the progress made up to date, in this field. We surveyed several methods and techniques developed in this area of research that still represent an open issue. The important outcomes within each method are resumed in this survey.

According to the reviewed studies, the artificial neural network based approach for chromosome classification has been widely used, since they increase efficiency and reduce the computational complexity and time due to their parallelism. A multiplicity of ANNs are considered; the most used and suitable one is the MLP based system either as a single stage or a combination of MLPs. Other types of ANN are also investigated, namely a probabilistic ANN, wavelet neural network WNN, Backpropagation ANN and a single-layer.

Apart from the neural network based chromosome classifiers, other types of classifiers have been investigated, such as the SVM based classifier which is new compared to the ANN based one. Nevertheless the SVM and ANN give almost comparable classification accuracies.

Regarding the method used for feature extraction, the MAT is the most used technique since it preserves the shape of the chromosome; it has been applied in most methods to detect the medial axis. Much research is done for automatic and accurate feature extraction, mainly centromere finding, which is an important parameter for chromosome classification.

We consider that the given survey helps researchers to have an overview of the state-of-the-art in automatic chromosome feature extraction and neural network based classification systems, in order to improve such systems for efficient abnormalities detection and genetic diseases diagnosis.

## References

Arachchige AS, Samarabandu J, Knoll JHM, Rogan PK (2013) Intensity integrated laplacian-based thickness measurement for detecting human metaphase chromosome centromere location. IEEE Trans Biomed Eng 60(7):2005–2013

Badawi AM, Hasan K, Aly E, Messi RA (2003) Chromosomes classification based on neural network fuzzy rule based and template matching classifiers. In: Proceedings of circuits and systems, IEEE 46th midwest symposium on, vol 1, pp 383–387

Bioimlab http://bioimlab.dei.unipd.it

Carothers A, Piper J (1994) Computer-aided classification of human chromosomes: a review. Stat Comput 4:161–171

Catherine TF, Gilchrist JM, Pericak-Vance MA, Speer MC (1998) Using neural networks as an aid in the determination of disease status: comparison of clinical diagnosis to neural-network predictions in a pedigree with autosomal dominant limb-girdle muscular dystrophy. Am J Hum Genet 62:941–949

Cho JM (2000) Choromosome classification using backpropagation neural network. In: Proceedings of IEEE engineering in medicine and biology, pp 28–33

Choi H, Castlman KR, Bovik AC (2006) Segmentation and fuzzy logic classification of M-FISH chromosomes images. In: Proceedings of image processing, 2006 IEEE international conference, pp 69–72

Errington PA, Graham J (1993) Classification of chromosomes using a combination of neural networks. In: Proceedings of IEEE international conference on neural networks, pp 1236–1241

Eskiizmirliler S, Erkmen AM, Basaran F, Beksaç MS, Ĉakar AN (1996) Ahybrid intelligent diagnostic system based on neural networks and image analysis technique in the field of automated cytogenetics. In: Proceedings of international conference on image processing, pp 315–318

Estandarte AKC (2012) A review of the different staining techniques for human metaphase chromosomes. Master thesis, University of London

Graham J, Errington P, Jennings A (1992) A neural network chromosome classification. J Radiat Res 33:250–257

Gregor J, Granum E (1991) Finding chromosome centromere using band pattern information. Biol Med 21:55–67

Khan S, DSouza A, Sanches J, Ventura R (2012) Geometric correction of deformed chromosomes for automatic karyotyping. In: Proceedings of the 34th annual international conference of the IEEE EMBS, pp 4438–4441

Kou Z, Ji L, Zhang X (2002) Karyotyping of comparative genomic hybridization human metaphases by using support vector machines. Cytometry 47:17–23

Lerner B (1998) Towards a completely automatic neural-network-based and human chromosome analysis. IEEE Trans Syst 28(4):544–552

Lerner B (2004) Bayesian fuorescence in situ hybridisation signal classification. Artif Intell Med 30:301–316

Lerner B, Lawrence Neil D (2001) A comparison of state-of-the-art classification techniques with application to cytogenetics. Neural Comput Appl 10:39–47

Lerner B, Guterman H, Dinstein I, Romem Y (1995) Medial axis transform-based features and a neural network for human chromosome classification. Pattern Recognit 28(11):1673–1683

Lijiya A, Mumthas TK, Govindan VK (2013) Chromosome classification using M-FISH images. In: Proceedings of international conference on advances in information technology and mobile communication, pp 314–320

Loganathan E, MR Anuja, Madian N (2013) Analysis of human chromosome images for the identification of centromere position and length. In: Proceedings of IEEE point-of-care healthcare technologies (PHT), pp 314–317

Madian N, Jayanthi KB (2014) Analysis of human chromosome classification using centromere position. Measurement 47:287–295

Markou C, Maramis C, Delopoulos A, Daiou C, Lambropoulos A (2012) Automatic chromosome classification using support vector machines. iConceptPress, Hong Kong

Ming D, Tian J (2010) Automatic pattern extraction and classification for chromosome images. J Infrared Milli terahz Waves 31:866–877

Mohammadi MR (2012) Accurate localization of chromosome centromere based on concave points. J Med Signals Sens 2(2):88–94

Moradi M, Setarehdan SK (2005) New features for automatic classification of human chromosomes: a feasibility study. Pattern Recognit Lett 27:19–28

Moradi M, Setarehdan SK, Ghaffari SR (2003a) Automatic landmark detection on chromosomes' images for feature extraction purposes. In: Proceedings of the 3rd international symposium on image and signal processing and analysis, pp 567–570

Moradi M, Setarehdan SK, Ghaffari SR (2003b) Automatic locating the centromere on human chromosome pictures. In: Proceedings of 16th IEEE symposium on computer-based medical systems CBMS, pp 56–61

Mousavi P, Ward RK, Fels SS, Sameti M, Peter M (2002) Lansdorp feature analysis and centromere segmentation of human chromosome images using an iterative fuzzy algorithm. IEEE Trans Biomed Eng 49(4):336–371

Narayanana A, Keedwella EC, Gamalielssonb J, Tatinenia S (2004) Single-layer artificial neural networks for gene expression analysis. Neurocomputing 61:217–240

Natarajan AT (2002) Chromosome aberrations: past, present and future. Mutat Res Elsevier Sci 504:3–16

Oskouei BC, Shanbehzadeh J (2010) Chromosome classification based on wavelet neural network. In: Proceedings of the digital image computing: techniques and applications, pp 605–610

Poletti E, Rugger A, Grisan E (2008) Automatic classification of chromosomes in Q-band images. In: Proceedings of 30rd IEEE annual international conference of the EMBS, pp 1911–1914

Poletti E, Rugger A, Grisan E (2011) An improved classification scheme for chromosomes with missing data. In: Proceedings of the 33rd IEEE annual international conference of the EMBS, pp 5072–5075

Poletti E, Grisan E, Rugger A (2012) A modular framework for the automatic classification of chromosomes in Q-band images. Elsevier Comput Methods Progr Biomed 105:120–130

Popescu M, Gader P, Keller J, Klein C, Stanley J, Caldwell C (1999) Automatic karyotyping chromosomes of metaphase cells with overlapping. Elsevier Comput Biol Med 29:61–82

Ruan X (2000) A classifier with the fuzzy hopfield network for human chromosomes. In: Proceedings of the IEEE 3rd word congress on intelligent control and automation, pp 1159–1164

Rungruangbaiyok S, Phukpattaranont P (2010) Chromosome image classification using a two-step probabilistic neural network. J Sci Technol 32(3):255–262

Sampat MP, Castleman KR, Bovik AC (2002) Pixel-by-pixel classification of MFISH images. In: Proceedings of the 24th annual conference and the annual fall meeting of the biomedical engineering society EMBS/BMES conference, pp 999–1000

Schwartzkopf WC, Bovik AC, Evans BL (2005) Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images. IEEE Trans Med Imaging. 12:1593–610

Sjahputera O, Keller JM (1999) Evolution of a fuzzy rule-based system for automatic chromosome recognition. In: Proceedings of the IEEE international fuzzy systems conference, pp 129–134

Siepcher MR, Ballard SG, Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet 12:368–375

Stanley R, Keller JM, Gader P, Caldwell CW (1998) Data-driven homologue matching for chromosome identification. IEEE Trans Med Imaging 17(3):451–462

Sweeney WP, Musavi MT, Guidi JN (1993) Probabilistic neural network as chromosome classifier. In: Proceedings of the international joint conference on neural networks, pp 935–938

Theerapattanakul J, Plodpai J, Pintavirooj C, Lertprasert P (2005) Feature extraction for automatic chromosome analysis. In: Proceedings of international conference on electricalengineering/electronics, computer, telecommunications andinformation technology (ECTI-CON), pp 399–402

Tjio JH, Levan A (1956) The chromosome number of man. Hereditas 42:1–6

Tso M, Graham J (1991) An efficient transportation algorithm for automatic chromosome karyotyping. Pattern Recognit Lett 12(1991):117–126

Uttamatanin R, Intarapanich A, Kaewkamnerd S, Yuvapoositanon P, Tongsima S (2013a) Chromosome classification for metaphase selection. In: 13th international symposium on communications and information technologies (ISCIT), pp 464–468

Uttamatanin R, Intarapanich A, Kaewkamnerd S, Yuvapoositanon P, Tongsima S (2013b) Band classification based on chromosome shapes. In: Proceedings of the biomedical engineering international conference (BMEiCON), pp 1–5

Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999

Wang YP (2004) M-fish image registration and classification. In: Proceedings of the IEEE international symposium on biomedical imaging: nano to macro, pp 57–60

Wang X, Zheng B, Wood M, Li S, Chen W, Liu H (2005) Development and evaluation of automated systems for detection and classification of banded chromosomes: current status and future perspective. J Phys D Appl Phys 38:2536–2542

Wang X, Zheng B, Li S, Mulvihill JJ, Liu H (2008a) A rule-based computer scheme for centromere identification. Comput Methods Progr Biomed 89:33–42

Wang X, Li S, Liu H, Wood MC, Chen WR, Zheng B (2008b) Automated identification of analyzable metaphase chromosomes depicted on microscopic digital images. J Biomed Inf 41:264–271

Wang X, Zheng B, Li S, Mulvihill JJ, Wood MC, Liu H (2009) Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme. J Biomed Inf 42:22–31

Wang X, Zheng B, Li S, Mulvihill JJ, Chen X, Liu H (2010) Automated identification of abnormal metaphase chromosome cells for the detection of chronic myeloid leukemia using microscopic images. J Biomed Opt 15(4):1–12

Xia Y (1989) Skeletonization via the realization of the fire front's propagation and extinction in digital binary shapes. IEEE Trans Pattern Anal Mach Intell 2(10):1076–1087