

# Algorithms for Chromosome Classification\*

Wenzhong Yan, Lei Bai

Department of Computer, North China Institute of Science and Technology, Beijing, China

Email: yanwenzhong@ncist.edu.cn

Received 2013

## ABSTRACT

Automated chromosome classification has been an important pattern recognition problem for decades. In order to improve the performance of automated chromosome classification, artificial intelligence and machine learning methods have been widely used in the computer-assisted chromosome detection and classification systems. This paper is focused on these algorithms, especially on artificial neural network (ANN) and wavelet transform algorithms. The principle and the realization of these algorithms are analyzed. Results of these algorithms are compared and discussed.

**Keywords:** Chromosome; Classification; ANN; Wavelet; M-FISH

## 1. Introduction

Chromosomes are genetic information carriers and chromosome analysis constitutes an important procedure in clinical and cancer cytogenetics studies. Chromosome karyotyping refers to the classification and subsequently a formatted display of the chromosomes found in a cell spread. A karyotype is required to assign each chromosome to one of 24 classes (22 autosomes and two sex chromosomes). **Figure 1** shows a sample result of the karyotype. Since karyotyping is a time consuming procedure, computer-based classifiers have been proposed. Most of these classifiers make use of an intuitive transformation of the chromosome image density distributions into a set of features to be used by some sort of statistical discriminator. These types of classifiers have not shown high performance results [1,2].

In order to improve the performance of automated chromosome classification, artificial intelligence and machine learning methods have been widely used in area. Among them, artificial neural networks (ANN) and wavelet transform algorithms are the most popular tools. This paper is focused on these algorithms. The principle and the realization of these algorithms are analyzed. Results of these algorithms are compared and discussed.

## 2. Artificial Neural Network Based Algorithms

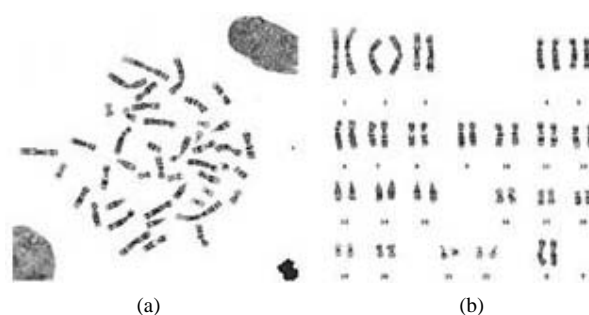
### 1) Basic Theory

Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological

nervous systems. The basic processing elements of neural networks are called artificial neurons, or simply neurons or nodes. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm [3].

### 2) Classification Algorithms Based on ANN

Backpropagation training method is commonly used to train ANNs. In multi-layer feed-forward ANNs, the number of output neurons is often fixed (from 1 to 24), but the number of input neurons, hidden neurons, steepness of the activation function, learning rate, momentum term, number of learning iterations and upper bound of training error are all programmable. Determining these training or optimization parameters is important for the performance



**Figure 1.** (a) A metaphase cell spread; (b) A karyotype of the chromosomes in (a).

\*This research is supported by the Fundamental Research Funds for the Central Universities (2011A010).

and robustness of an ANN used in chromosome classification [4].

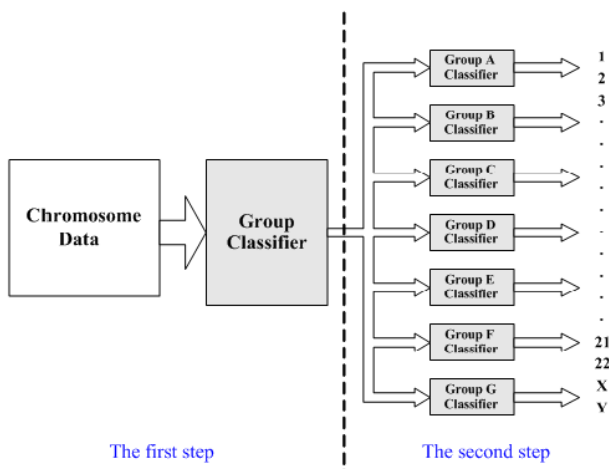
In order to improve the performance of traditional multilayer ANNs, a number of other more sophisticated neural networks have been proposed and tested in this area.

A hierarchical multi-layer neural network with an error back-propagation training algorithm has been adopted for the automatic classification of Giemsa-stained human chromosomes. Firstly, chromosomes data is classified into 7 major groups based on their morphological features such as relative length, relative area, centromeric index, and 80 density profiles. Then each 7 major groups are classified into 24 subgroups using each group classifier. **Figure 2** shows the two steps of chromosome classification. The classification error decreased by using two steps of classification and the classification error was 5.9% [5].

A fuzzy Hopfield neural network is a combination model of neuro and fuzzy computing. Its main difference from the traditional ANN is that it holds fuzzy clustering capability and learning mechanism of acquiring knowledge about the targets (human chromosomes) from the noisy training samples. It develops a Classifier with the Fuzzy Hopfield Network (CFHN) to identify each observed human chromosome and assign it to one of the 24 human chromosome classes. In a test involving 100 human chromosomes, the fuzzy Hopfield neural network produces a very low unidentification rate of 3.33% [6].

### 3. Wavelet Transforms Based Algorithms

Some researchers have set out to explore the use of wavelet-based band pattern descriptors for chromosome classification. Compared with other methods, wavelet transforms use different basis functions that lead to the desirable property of characterizing and localizing signal



**Figure 2. Architecture of the hierarchical multi-layer neural network.**

features simultaneously in both the space and transform domains. Furthermore, they offer a means of signal representation that facilitates multi-resolution analysis [7].

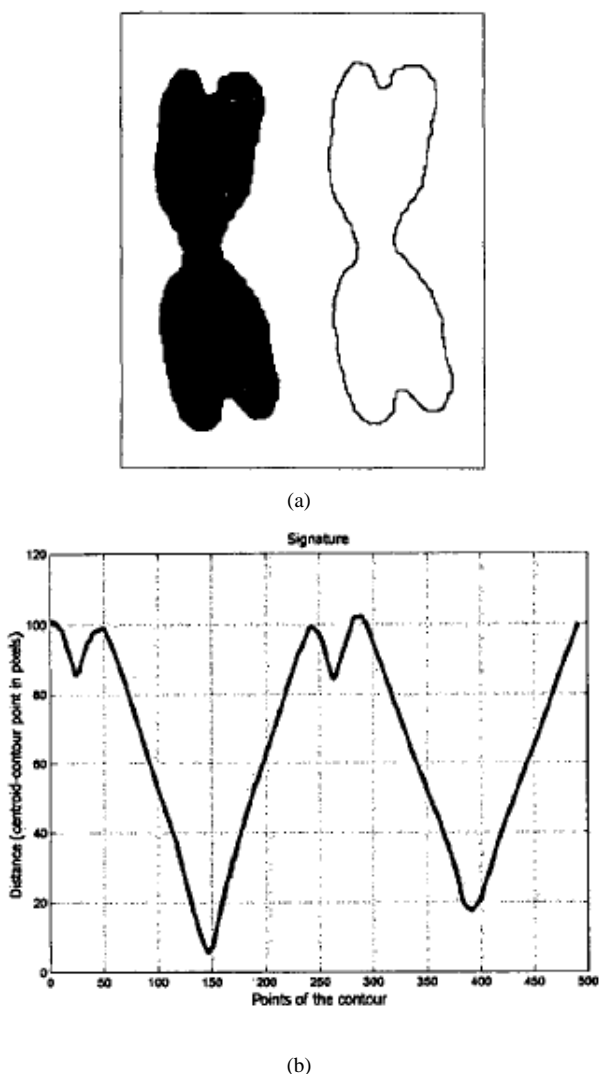
An expanded basis function system that allows high resolution decomposition of a signal is called the wavelet packet transform. These are computed by iterating not only down the lowpass scaling function branch of Mallat’s Discrete Wavelet Transform (DWT ) algorithm tree, but also down the highpass wavelet branch [7]. Thus, the wavelet packet transform offers more basis functions for signal analysis than the wavelet transform. Compared with the wavelet transform, the wavelet packet transform uses only full-width basis functions. They are all orthogonal, and all but the first have zero area. These are intuitively more satisfying weighting functions than the ever-narrowing wavelet basis functions.

In one study, researchers describe their recent study to employ wavelet packets as basis function sets to compute chromosome band pattern features. During the study, they evaluated a total of 28 wavelet packet basis function sets, including the well known Haar and Daubechies’4 and Daubechies’6 wavelet packets. They conducted experiments on two benchmark chromosome datasets and compare the experimental results with the results of the currently best performing Weighted Density Distribution (WDD) method. **Table 1** summarizes the experimental results. For the sake of clarity and page limit, this table includes only those of the well-known Haar, Daubechies’4 (D4), Daubechies’6 (D6), and the best-performing wavelet packet (BPWP), in comparison to the WDD results [8].

Another research proposes a method for chromosome classification based on the chromosome shape analysis. This approach is based on wavelet packet transform (WPT) and best basis algorithm (BBA). First, the chromosome image is preprocessed and binarized. Second, the contour of the chromosome is detected and the signature of the contour is produced. **Figure 3** shows a binarized chromosome, its contour and its signature of the contour. Third, the signature is decomposed by the wavelet packet transform and its best basis is found. Finally, the coefficients of the best tree of wavelet packet transform correspondent to the signature of chromosomes are compared in order to classify the chromosomes. The results obtained show that the proposed method provides an effective chromosome classification based on WPT and BBA of the shape signature of the chromosomes [9].

**Table 1. Summary of the chromosome classification experimental results based on the two data sets**

|                | WDD   | BPWP  | Haar  | D4    | D6    |
|----------------|-------|-------|-------|-------|-------|
| Copenhagen set | 96.8% | 96.2% | 95.7% | 95.2% | 94.0% |
| Genzyme set    | 85.3% | 84.6% | 83.8% | 79.9% | 80.0% |



**Figure 3.** (a) A binarized chromosome and its contour; (b) The signature of the contour in (a).

#### 4. Other Algorithms

Besides the ANN and wavelet transforms based algorithms, there are also other algorithms used for chromosome classification.

One research is interested in classification of chromosomes from either complete or incomplete cells. Researchers investigate globally optimal algorithms for automated classification and pairing of human chromosomes. Even in cases where the cell data are incomplete as often encountered in practice, they can still formulate the problem as a transportation problem, and hence find the globally optimal solution in polynomial time. In addition, a technique of homologue pairing via maximum-weight graph matching is proposed. It obtains the globally optimal solution by forming all homologue pairs simultaneously under a maximum likelihood criterion, rather than finding one pair at a time as in existing heuristic algo-

gorithms. After the optimal homologue pairing, chromosome classification can also be done by maximum-weight graph matching. This new graph theoretical approach to chromosome pairing and classification is more robust than the transportation algorithm [10].

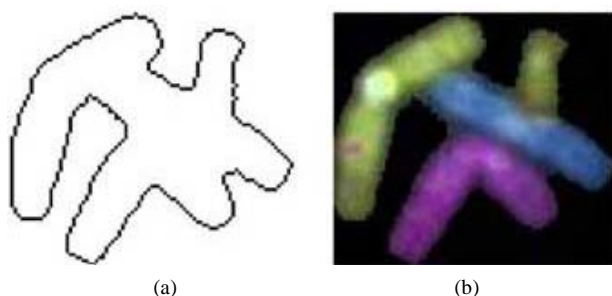
Traditional chromosome imaging has been limited to grayscale images. In the mid-1990s, a new technique for staining chromosomes was introduced. It produced an image in which each chromosome type appeared as a distinct color [11]. This multispectral staining technique is called multiplex fluorescence in-situ hybridization, or MFISH, which made analysis of chromosome images easier, not only for visual inspection of the images by humans, but also for computer analysis of the images. M-FISH uses five color dyes that attach to various chromosomes differently to produce a multispectral image, and a sixth dye that attaches to all chromosomes to produce a grayscale image. Thus, it is possible to envision new and improved methods for the location, segmentation and classification of chromosome images by exploiting the color information in M-FISH images.

One study addresses the topics of segmentation and classification of MFISH chromosome images. It introduces a probabilistic model of M-FISH chromosomes that allows for simultaneous segmentation and classification. The additional information provided by multiple spectra in chromosome images makes it feasible to distinguish chromosomes that overlap and touch within clusters. **Figure 4** shows the comparison of two types of cluster information. Thus, researchers develop a joint segmentation-classification algorithm that optimizes probabilistic information obtained from the multispectral chromosome pixels, and enables the decomposition of overlapping and touching chromosomes, and moreover, provides estimates of confidence in the chromosome segmentation-classification [12].

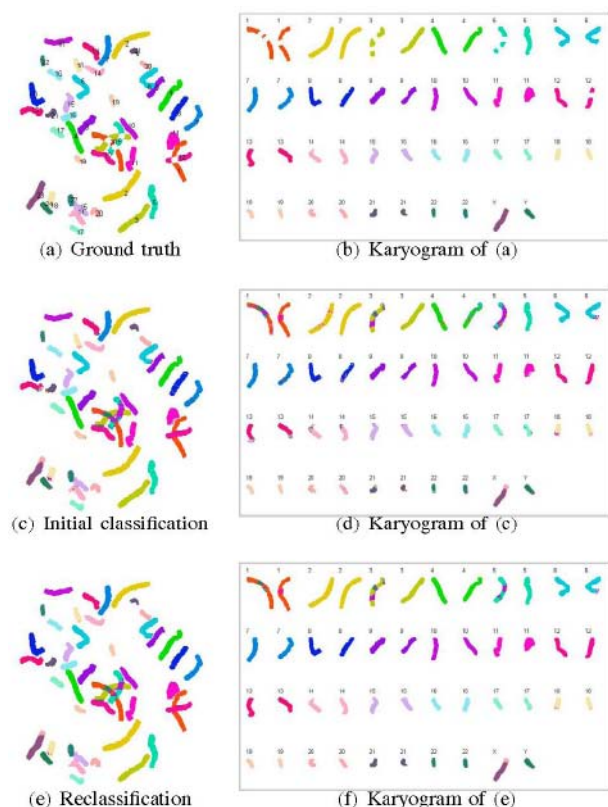
Another study presents a new segmentation method between chromosomes and background and a novel unsupervised classification method based on a fuzzy logic classifier specifically designed for M-FISH images. Utilizing the chromosome boundaries, the initial classification results improved significantly after the prior adjusted reclassification while keeping the translocations intact. **Figure 5** shows the fuzzy logic classification and prior adjusted reclassification. This study also presents a new segmentation method that combines both spectral and edge information. Ten M-FISH images from a publicly available database were used to test our methods. The segmentation accuracy was more than 98% on average [13].

#### 5. Discussion and Conclusion

The problem of automated chromosome classification has been investigated in many studies. A large number of



**Figure 4. Comparison of two types of cluster information. (a) Boundary of cluster; (b) Multispectral information in cluster.**



**Figure 5. Fuzzy logic classification and prior adjusted reclassification.**

novel techniques have been investigated by a number of research groups around the world. In this paper we reviewed some typical algorithms, such as ANN and wavelet transform algorithms etc. We analyzed the principle and the realization of these algorithms and also discussed the results of these algorithms.

## REFERENCES

- [1] M. Zardoshti-Kermani and A. Afshordi, "Classification of Chromosomes Using Higher-Order Neural Networks".
- [2] O. Sjahputera and J. M. Keller, "Evolution of a Fuzzy Rule-Based System for Automatic Chromosome Recognition," *IEEE International Fuzzy System Conference Proceedings*, 1999, pp. 129-134.
- [3] P. H. Sydenham and R. thorn, "Handbook of Measuring System Design," John Wiley & Sons, Ltd., 2005. <http://dx.doi.org/10.1002/0471497398>
- [4] J. Cho, "Chromosome Classification Using Backpropagation Neural Networks," *IEEE Engineering in Medicine and Biology Magazine*, Vol. 19, 2000, pp. 28-33. <http://dx.doi.org/10.1109/51.816241>
- [5] J. Cho, S. Y. Ryu and S. H. Woo, "A Study for the Hierarchical Artificial Neural Network Model for Giemsa-Stained Human Chromosome Classification," *Proceeding of the 26th Annual International Conference of the IEEE EMBS*, 2004, pp. 4588-4591.
- [6] X. Ruan, "A Classifier with the Fuzzy Hopfield Network for Human Chromosomes, Intelligent Control and Automation," *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, Vol. 2, 2000, pp. 1159-1164.
- [7] C. S. Burrus, R. A. Gopinath and H. Guo, "Introduction to Wavelets and Wavelet Transforms," Prentice-Hall, Englewood Cliffs, NJ, 1997.
- [8] Q. Wu and K. R. Castleman, "Automated Chromosome Classification Using Wavelet-Based Band Pattern Descriptors," *13th IEEE Symposium on Computer-Based Medical Systems*, 2000, pp. 189-194.
- [9] L. V. Guimaraes, J. A. Schuck and A. Elbern, "Chromosome Classification for Karyotype Composing Applying Shape Representation on Wavelet Packet Transform," *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, 2003, pp. 941-943.
- [10] X. L. Wu, P. Biyani and S. Dumitrescu, "Globally Optimal Classification and Pairing of Human Chromosomes," *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 2004, pp. 2789-2792.
- [11] M. R. Speicher, S. G. Ballard and D. C. Ward, "Karyotyping Human Chromosomes by Combinatorial Multicolor FISH," *Nature Genetics*, Vol. 12, 1996, pp. 368-375. <http://dx.doi.org/10.1038/ng0496-368>
- [12] C. S. Wade, C. B. Alan and L. E. Brian, "Maximum-Likelihood Techniques for Joint Segmentation-Classification of Multispectral Chromosome Images," *IEEE Transaction on Medical Imaging*, Vol. 24, No. 12, 2005, pp. 1593-1610. <http://dx.doi.org/10.1109/TMI.2005.859207>
- [13] H. Choi, K. R. Castleman and A. C. Bovik, "Segmentation and Fuzzy-Logic Classification of M-FISH Chromosome Images," *IEEE International Conference on Image Processing*, 2006, pp. 69-72.