VOL.2 NO.2September 2013 - ISSN 2233 – 1859

Operations Research Society in Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing

Available online: www.scjournal.com.ba

Denver Groups Classification of Human Chromosomes Using CANN Teams

Sadina Gagula-Palalic, Mehmet Can

International University of Sarajevo, Faculty of Engineering and Natural Sciences, HrasnickaCesta 15, Ilidža 71210 Sarajevo, Bosnia and Herzegovina

Article Info

Article history:
Received 17 Sep.2013
Received in revised form 17 Oct 2013

Keywords:
Human chromosomes, classification, Neural Networks, CANNT

Abstract

Unbanded human chromosome can be classified into seven Denver Groups (A-G) based on their lengths and the ratio of the length of the shorter arm to the whole length of the chromosome, which is called the centromere index (CI). In this article, the novel artificial neural network committee machines technique (CANNT) developed earlier, is applied to the Denver Groups and the correct classification rate in Denver Groups Classification of Human Chromosomes raised from 96%, to a level of 98%.

1. INTRODUCTION

In 1956 Tjio and Levan, using the improved cell culturing and staining technique, discovered that the number of human chromosomes is 46 (Tjio and Levan, 1956). From this time on, the research on chromosomal abnormalities, as a cause of diseases, became one of the main branches of the molecular biology.

Disorder in human chromosomes is a powerful indicator in diagnosis of leukemia, skin and breast cancers, and other geneticdiseases. Clinical laboratories routinely performed researches to identify chromosome abnormalities, and provide medical doctors the diagnostic results and help them decide therapeutic treatments for patients.

The most prominent difficulty in chromosome analysis is the absence of clear microscopic chromosome images. The variation of cell culturing conditions, chromosome staining, and microscope illumination make finding analyzable chromosomes in a genetics clinical laboratories very difficult. For human experts, identification and classification of chromosomes is a tedious and time-consuming task. The human error also introduces variation and affects the accuracy of the diagnostics made by physicians.
The development of computer-assisted metaphase finding and karyotyping systems, slowed down by the noisy cell images.

2. HUMAN CHROMOSOMES

Since Waldeyer in 1898 coined the term chromosome (Vermaand Babu, 1995), it is known that chromosomes resides within a cell's nucleus, and contains the person's deoxyribonucleic acid (DNA). Each chromosome is made up of a single extremely long DNA molecule. Using cells cultured from fetal lung tissue, Tjio and Levan, demonstrated that human cells contain 46 chromosomes as they appear during cell division or mitosis. A healthy human cell nucleus includes 44 autosomes and 2 sex chromosomes: X and Y.

The test cells used for chromosome imaging and analysis are taken mostly from blood sample, amniotic fluid, and bone marrow. These test samples are cultured overnight in a mitotic arresting agent. Then cells are processed with hypotonic solutions to increase cell volume. This procedure spreads the chromosomes apart.

The methanol-acetic acid is used to fix them for analyses. The fixed cells are dropped onto a standard glass microscope slide and allowed to dry. If karyotyping and classification are going to be performed using banded chromosomes, the slide is then subjected to a staining process. Staining makes clear the distinctive reproducible patterns of bands along chromosomes. These bands permit accurate identification of chromosomes and recognition of abnormalities.

2.1 CLASSIFICATION OF HUMAN CHROMOSOMES

In order to improve the performance of automated chromosome classification including recognition of disordered chromosomes, artificial intelligence and machine learning methods have been widely used in the computer-assisted chromosome detection and classification systems (Gagula-Palalicand Can 2012/2012). Among them, ANN is the most popular tool owing to its capability of modeling the human brain decision making process to recognize objects based on incomplete or partial information, as well as its simple topographic structure and easier training process (Mitchell, 1997).

Early studies also indicated that ANN performance could achieve comparable results compared with that obtained by simpler statistical methods (Sweeney, 1993). A large number of different feature based and pixel value distribution based ANN have been tested and evaluated in classification of banded chromosomes, which include supervised multi-layer neural networks (Delshadpour,2003, and Wu, et al., 1990); Hopfield network (Ruan,2000), and unsupervised architecture of self-organizing nonlinear maps (Lerner et al., 1996), SOFM (Kyan et al., 1999) and mutual information maximization based training method (Mousavi et al., 1999).

There are a huge number of researches to replace technicians in the cytogenetic labs with software and computers. Some of them use image processing techniques for segmentation of human cells photographs in metaphase, and artificial neural networks in chromosome classification and pairing.

Artificial intelligence and machine learning methods have been widely used techniques to improve the performance of the computer-assisted chromosome detection and classification systems. Because of their capability to recognize objects based on incomplete or partial information, Artificial Neural Network (ANN) is the most popular tool. Its architecture is simple and training process is simple process (Mitchell, 1997 and Haykin, 2009). A large number of different ANNs have been tested in classification of human chromosomes, which include supervised neural network architecture. Multi-layer neural networks are studied in (Delshadpour,2003, Wu, et al., 1990, Lu, andYa 1989, Erington and Graham, 1993,ElEmari , 2006, Wang et al.,2009, Can, and Gagula-Palalic, 2012) and Hopfield network in (Ruan, 2000); fuzzy neural techniques in (Ruspini, 1973a, Ruspini, 1973b, Ramstein et al., 1992, Keller et al., 1995, Sjahputera and Keller, 1999, Sarosa et al., 2000); and unsupervised architecture of nonlinear maps (Lerner et al., 1996), self organizing feature maps (Kyan et al., 1999)and mutual information maximization based training method (Mousavi et al., 1999).

In chromosome classification and pairing, back-propagation training method is used to train ANNs. In multi-layer feed-forward ANNs, the number of output neurons is equal to the number of human chromosome types. The number of input neurons is equal to the dimension of the input data, which is the number of features used for classification. Often, principal components replace real features to reduce the dimension of the input data, and hence the computation cost. The number of hidden layers, number of hidden neurons, steepness of the activation function, learning rate, and momentum factor, number of learning iterations and upper bound of training error are chosen by the user experimentally.

While the proper choice of these parameters is important for the performance and robustness of an ANN used in chromosome classification (Cho, 2000), studies indicated that ANN performance was slightly lower than that obtained using simpler statistical methods (Granum and Thomason, 1990, Sweeney sndMusavi, 1993, Conroy et al., 2000). Unnecessary complexity of the ANN architecture and overtraining of ANNs dramatically reduce the robustness of the ANN in chromosome classification. One study (Mitchell, 1997), using multi-layer perception based ANN, obtained 0% error rate in the training data set but 24.2% error rate in the testing data set. To increase ANN performance, another study showed that by reducing the complexity of an ANN, its testing accuracy can be increased from 75.8% to 88.3% (Delshadpour,2003).

One of the other more sophisticated neural networks proposed and tested in this area is a fuzzy Hopfield neural network. It holds fuzzy clustering capability and learning mechanism of acquiring knowledge about the human chromosomes from noisy inputs. In a test involving 100 human chromosomes Ruan (2000) succeeded to achieve a very high identification rate of 96.67%.

Recently Palalic and Can (2013) developed a novel committee of neural network machines, competing artificial neural network teams technique (CANNT) which over scores almost all previous human chromosome classifiers.

2.2 CLASSIFICATION OF UNBANDED CHROMOSOMES

When the chromosomes are not banded, they can be classified into seven Denver Groups (A-G) (H. C. S.

Group, 1960) as seen in Table1. Denver Group classification is mainly based on:
(1) the length or size of each chromosome and
(2) the ratio of the length of the shorter arm to the whole length of the chromosome, which is called the centromere index (CI).

Table 1: The classification of chromosomes based on Denver Group classification

Chromosome Class	Denver Group
#1-#3	Group A
#4-#5	Group B
#6-#12,X	Group C
#13-#15	Group D
#16-#18	Group E
#19-#20	Group F
#21-#22,Y	Group G

In this article, the competing artificial neural network teams (CANNT) method Supplemented by a nearest neighbor technique will be used to perform the Denver Group classification of a given set of unbanded human chromosomes.

3. DATA DESCRIPTION

The data used in this work is taken from Copenhagen data base. We omitted gray level features, and only keep (1) the length of each chromosome and (2) the ratio of the length of the shorter arm to the whole length of the chromosome, which is called the centromere index (CI).

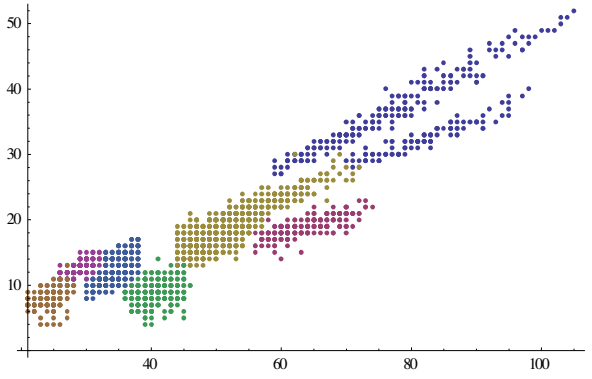

Figure 1.The distribution of 2200 human chromosomes into seven Denver Group classes from A, to G.

4.COMPETING ARTIFICIAL NEURAL NETWORK (CANNT)

Architecture of ANN

We represent the network consisting of 2 inputs x[i], i=1, 2, 12 neurons in the hidden layer and one neuron in the output layer as shown in the Fig 1. A special organized committee of 42 simple perceptrons is used to improve the rate of correct classification of 7 types of unbanded human chromosomes. Each of these simple perceptrons is trained to distinguish between two types of chromosomes. These multilayer perceptrons use Back-Propagation algorithm.

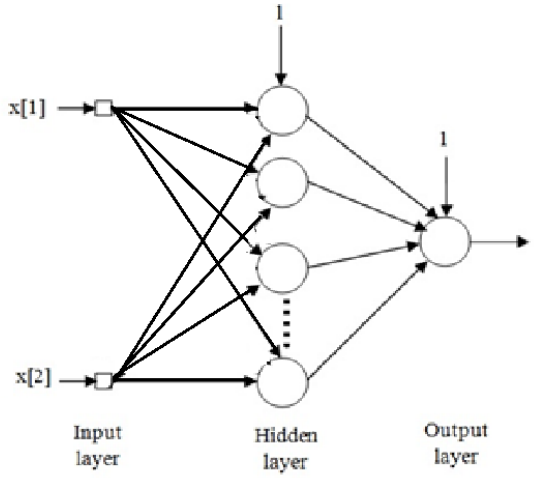

Fig 2: Neural network architecture for a simple multilayer perceptron

Assembling votes

Let $P(i,j)$ be the simple perceptron which is trained to distinguish chromosomes of type i, and of type j, and let

$$T(s) = \{P(i,s), P(s,j) \mid i = 1 \dots 7, j = 1 \dots 7\}, s = 1 \dots 7$$

be the seven teams of perceptrons.

When a new data of type x enters the network, the perceptrons

$$P(i,x), \quad i = 1 \dots 7$$

of the team $T(x)$ creates mostly an output 1, while the perceptrons

$$P(x,j), \quad j = 1 \dots 7$$

of the same team $T(x)$ creates mostly an output -1. The perceptrons of other teams also creates outputs either 1, or -1. But since the other teams are not trained to distinguish chromosomes of type x from other chromosome types, their consensus will be weaker than the consensus of team

T(x). So we expect that the team T(x) will be the winner of the competition.
For completeness, the dummy perceptrons $P(j,j), j = 1,2,\ldots,7$ which always give output 0 are added. When 7×7 perceptrons are arranged as a 7×7 grid, the votes of teams appear in crosses:

0	-1	1	1	1	1	-1
-1	0	1	-1	1	-1	1
-1	-1	0	-1	-1	-1	-1
-1	-1	1	0	-1	-1	-1
-1	-1	1	1	0	-1	-1
-1	1	1	1	1	0	1
1	1	1	-1	-1	-1	0

Fig. 3: Example of the decision matrix. The team T(3)is the winner of the competition. The nearest competitor is T(4).

The score of each team is its distance to its consensus. In Figure 3, the score of the team T(3) is zero, while the score of nearest competitor T(4) is four. The team with smallest score is the winner of the competition, and the new chromosome data entered, belongs to the chromosome type of winners label.

Another representation of the winner team can be visualized attaching gray levels to the team members proportional to their scores as seen in Figure 4:

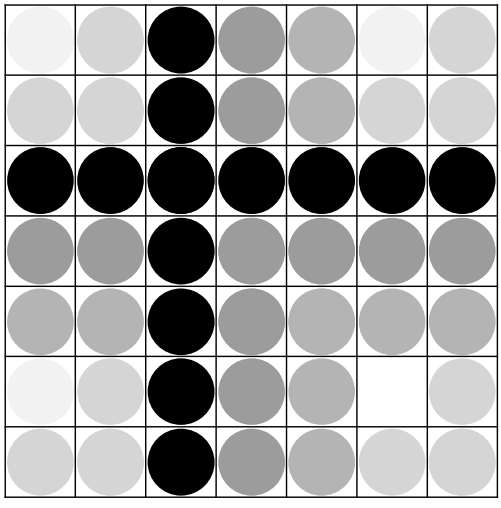

Fig 4: Competing teams. The darkest cross is the one which consist of 3rd row and 3rd column that wins the competition. The nearest competitor to team 3 is team 4.

5. RESULTS

During the training of 462 simple multilayer perceptrons, it is possible to complete training with zero error. But this leads to overtraining that causes lower rates in testing. From each chromosome type 50 random samples are chosen for training. The same numbers of random samples are also chosen for validation and testing. We have seen that it is possible to go over 97% correct classification rates with this special committee of perceptrons [Table 2].

Table 2. Correct classification rates during training and testing. Using a validation data set, the overtraining is prevented.

Denver group	Correct Classification Rate (%)		
	Training	Validating	Testing
A	98	100	100
B	100	100	98
C	96	98	92
D	100	98	100
E	100	96	94
F	92	96	100
G	94	98	88
Average	97.14	98.00	96.29

6. CONCLUSION

In this study we presented a special organized committee of 42 simple perceptrons used to improve the rate of correct classification of 7Denver types of unbandedhuman chromosomes. Each of these simple perceptrons is trained to distinguish between two types of chromosomes. When a new data is entered, the votes of these 42 simple perceptrons and additional 7 dummy perceptrons create a decision matrix of the size 7×7. By a special assembling of these votes we get a higher rate of correct classification of 7 Denver types of human chromosomes, with an average of 98.00% correct classification when tested on Copenhagen Chromosome Dataset.

REFERENCES

Can, M., and S. Gagula-Palalic (2012) Application of Ensemble Machines of Neural Networks to Chromosome Classification, SEJSC, Vol 1, No. 231-35.

Cho J.M. (2000) Chromosome Classification using back-propagation neural networks, IEEE Engineering Medicine and Biology, January/February 2000, pp. 28-34

Conroy J.M., Kolda T.G., O'Leary D.P., O'Leary T.J. (2000) Chromosome Identification using Hidden Markov Models: Comparison with Neural Networks, Singular Value Decomposition, Principal Components Analysis and Fisher Discriminant Analysis, Laboratory Investigation, Vol. 80, No. 11, pp. 1629-1641

Delshadpour, S, (2003) Reduced size multi-layer perceptron neural network for human chromosome classification, Proceedings of the 25th Annual

International Conference of the IEEE (Engineering in Medicine and Biology Society).

El Emari I. M. (2006) On the application of artificial neural networks in analyzing and classifying the human chromosomes, Journal of Computer Science Vol. 2, No.1, pp. 72-75

Erington P.A., Graham J. (1993) Classification of Chromosomes using a combination of neural networks, IEEE International Conference on Neural Networks, 28 Mar – 1 Apr, San Francisco, Vol. 3, p.p. 1236-1241

Granum E., Thomason M. G. (1990) Automatically inferred Markov Network Models for Classification of Chromosomal Band Pattern Structure, Cytometry, Vol. 11, pp. 26-39

Tjio J. H., and A. Levan (1956) The chromosome number in man, Hereditas, vol. 42, pp. 1-6.

Verma R. S., and A. Babu (1995) Human Chromosomes, Principles and Techniques, 2nd Ed. New York: McGraw-Hill.

Gagula-Palalic S.G, and M. Can (2012) Automatic Segmentation of Human Chromosomes, SEJSC, Vol 1, No. 2, pp. 80-83.

Gagula-Palalic S, and M. Can (2012) Extracting Gray Level Profiles of Human Chromosomes by Curve Fitting, SEJSC, Vol 1, No. 2, pp. 66-71

Gagula-Palalic S, and M. Can (2013) An Organized Committee of Artificial Neural Networks in the Classification of Human Chromosomes International Journal of Computer Applications, Vol. 80, No. 8, pp. 38-41

Haykin S. (2009) Neural Networks and Machine Learning, Pearson, Third edition, pp. 153-205

Keller, J. M., P. Gader, O. Sjahputera, and C. W. Caldwell (1995) A fuzzy logic rule based system for chromosome recognition, presented at Proceedings of the Eighth IEEE Symposium on Computer-Based Medical Systems.

Kyan, M. J., L. Guan, M. R. Amison, and C. J. Cogswell, (1999)Feature extraction of chromosomes from 3D confocal microscope images, presented at 1999 International Conference on Image Processing.

Lerner, B., H. Guterman, M. Aladjem, and I. Dinstein (1996) Feature extraction by neural network nonlinear mapping for pattern classification," presented at Proceedings of the 13th International Conference on Pattern Recognition.

Lu, Y., and Y. Ya (1989) An expert system for banded chromosomes recognition, presented at Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Mitchell, T. M. (1997) Machine Learning. Boston MA: WCB McGraw-Hill.

Mousavi, P., P. K. Ward, and P. M. Lansdorp (1999) Feature analysis and classification of chromosome 16 homologs using fluorescence microscopy image," presented at IEEE Can J Elec. & Comp Eng..

Ramstein, G., M. Bernadet, A. Kangoud, and D. Barba (1992) A rule-based image analysis system for chromosome classification, presented at Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Ruan, X. (2000) A classifier with the fuzzy Hopfield network for human chromosomes, Intelligent Control and Automation," presented at Proceedings of the 3rd World Congress on Intelligent Control and Automation.

Ruspini, E. (1973a) New experimental results in fuzzy clustering. Inf. Sci., 6, 273–284.

Ruspini, E. (1973b) A Fast Method for Probabilistic and Fuzzy Cluster Analysis using Association Measures. Proceedings of the 6th International Conference on System Sciences, Hawaii, pp. 56–58.

Sarosa M., Ahmad A.S., Riyanto B., Noer A.S. (2007) Optimization of Neuro-Fuzzy System using genetic algorithm for chromosome classification, ITB J. ICT, Vol 1 C, No. 1, pp. 56-69

Sjahputera O., and J. M. Keller, (1999) Evolution of a fuzzy rule-based system for automatic chromosome recognition, presented at 1999 IEEE International Fuzzy Systems Conference Proceedings.

Sweeney W. P., and M. T. Musavi, (1993) Application of neural networks for chromosome classification," presented at Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,

Wang X., Zheng B., Li S., Mulvihill J.J., Wood M.C., Liu H. (2009) Automated classification of metaphase chromosomes: Optimization of an adaptive computerized scheme, Journal of Biomedical Informatics 42, pp. 22–31

Wu, Q, P. Suetens, and A. Oosterlinck, (1990), Chromosome classification using a multi-layer perceptron neural net, presented at Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society.