INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Classification Of Chromosomes Using Feed Forward Neural Network Back Propagation Algorithm

Rajalakshmi.M
PG scholar, Sri Balaji Chockalingam Engineering College
Arni,Thiruvannamalai (d.t), India

Boopathi.S
Assistant professor,Department Of ECE , Sri Balaji Chockalingam Engineering College
Arni,Thiruvannamalai(d.t), India

Abstract:

Karyotyping is a common method in cytogenetic. Automatic classification of the chromosomes within the microscopic images is the first step in designing an automatic karyotyping system. This is a difficult task especially if the chromosome is highly curved within the image. The main aim of this paper was to define a new group of features for better representation and classification of chromosomes. this paper proposes classification & analysis of human chromosomes which includes the following steps i)we use image processing utilities and filter to remove noise .ii)the filtered image is then entered into segmentation algorithm to segment the image .iii)then the segments enter into two tracks for classifying chromosomes. the first one depends on image processing for measuring the length of chromosomes where the second one deals with initiating the feed forward neural network which is trained by means of back propagation algorithm. By using feed forward neural network and back propagation algorithm, width, position and the average intensity of chromosome was determined. back propagation algorithm achieves high accuracy with minimum training time, which makes it suitable for real-time chromosome classification in the laboratory.in our paper ,segmentation is done by using image processing and classification is done by using feed forward neural network and back propagation algorithm.

1.Introduction

The nucleus of the cell in the human body contains 23 pairs of homologous chromosomes. In each pair,one chromosome is inherited from mother known as "maternal homolog" and the other chromosome is inherited from the father known as "paternal homolog". Cancer is believed to be related to specific chromosome abnormalities. In order to study the characteristic of cancerous cell,it is essential to classify chromosome into paternal and maternal homologous classes so as to have the ability to analyze separate homologs. Fig shows the structure of a metaphase chromosome consisting of duplicate sister chromatids .the central part of the chromosome is known as centromere .telomere form the ends of the P & Q arms. Telomere length are believed to have an important role in cell life span and the development of cancerous cell.

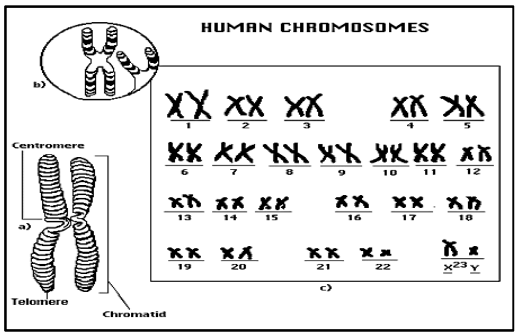

Figure 1: Structure Of Human Chromosome

2.Previous Research

MULTICOLOR fluorescence in situ hybridization(M-FISH) is a combinatorial labeling technique that is developed for the analysis of human chromosomes The technique has been used for the characterization of chromosomal translocations, to search for cryptic rearrangements, and to study mutagenesis, tumors, and radiobiology . In this technology, chromosomes are labeled with five dyes and a DNA stain known as 4'-6-diamidino-2-phenylindole (DAPI) that attaches to DNA and labels all chromosomes. A fluorescent microscope that is equipped with a filter wheel is used to capture the chro-mosome images. Each dye is visible in a particular wavelength and can be captured by the use of a specific filter. Therefore, M-FISH signals can be obtained as multispectral or multichannel images, in which a chromosome was stained to be visible (signed as "1") or not visible (signed as "0"). For a number n, the number of Boolean combination is 2n.

Hence, five spectra are sufficient to distinguish the 24 classes of chromosomes in human genome.

In addition to that, DAPI is used to counterstain each chromosome such that all of the chromosomes are visible in a DAPI channel. By simultaneously viewing six different channel images, pixel-wise classification of human chromosome is possible. This technique is also called color karyotyping in cytogenetics .

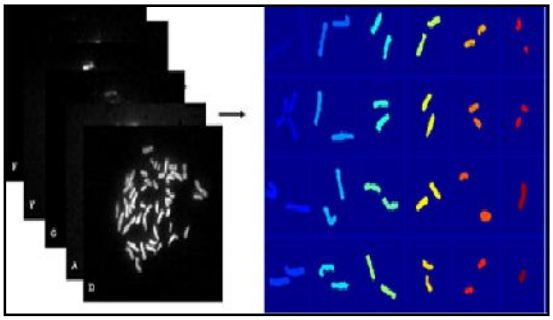

Figure 2:Twenty-four classes of chromosomes are classified from the five channel

spectral images; each class of chromosome is displayed with a different pseudo color. This pixel-wise classification technique is called color karyotype.

Fig 2 shows an example of M-FISH images of a male cell, where 22 autosomes and both sex chromosomes are classified from a five-channel spectral image data and are displayed by the use of 24 pseudocolors. For a normal cell, each chromosome should be painted with the same color. Otherwise, it indicates that chromosomal abnormalities might exist, which are associated with certain genetic diseases and cancers. The detection of chromosomal abnormalities depends on ac-curate pixel-wise classification techniques. Even though many attempts have been made to automate image analysis procedure, the reliability of the diagnosis technique has not reached the level for clinical use due to a number of factors that include nonhomogeneity of staining, variations of intensity levels within and between image sets, and emission spectral overlaps between fluorophores. The sizes of the misclassified regions are often larger than the actual chromosomal rearrangements or lost, which often lead to incorrect interpretation by cytogeneticists. To improve the detection of chromosomal abnormalities for clinical diagnosis, accurate segmentation and classification algorithms have to be developed.

In this paper, feed forward neural network & back propagation algorithm was applied to the classification of images by considering intensity inhomogeneity, which often exist in the images. Different from the previous AFCM algorithm that is proposed for MRI

image analysis, we proposed an improved back propagation method, which yields better background compensation and results in improved chromosome segmentation and classification.

3.Proposed Software

The process of human chromosome classification executed by our proposed software passes through two main stages given by the following:

3.1.Image Processing

This phase is concerned with analyzing the image which is taken from microscope via digital camera and examines this image if it is in RGB form or not. If it was in that form, we should convert it into gray scale image to make it easier to be handled by matlab .after that, we apply an advanced filtering technique to remove the noise which exists in the image. we create our own filter because all types of prefabricated filters are used to remove creation types of noise. But in the treated images, the object which we considered it as noise and must be removed. Several of built in filters use the value of an output pixel that is determined by the median of the neighborhood pixels rather than the mean ,but in our image we want to separate some pixels from its neighborhood. Our filtering algorithm removed 91.7% of the noise in the image .after that the segmentation process begins with cutting the image into the most likely segments which usually contain one chromosome. However some critical segments could contain the overlapping or touching chromosomes

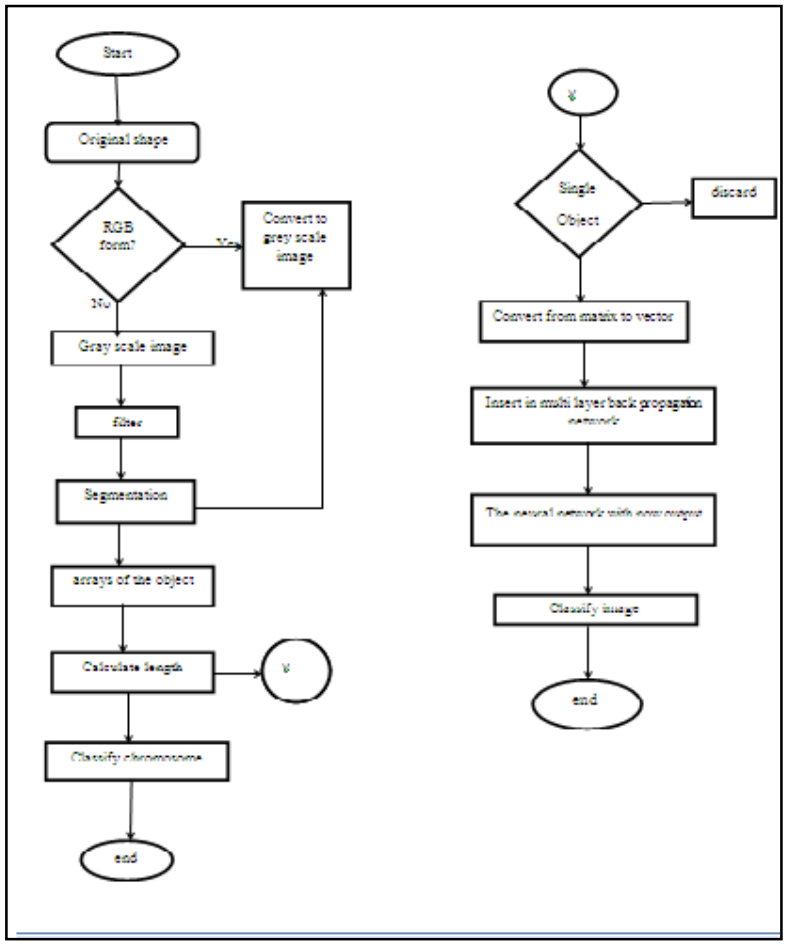

Figure 1: Flow chart of our proposed method

the neural network will be trained to distinguish them and rearrange them into pairs .each pair must have two identical chromosome.if there is any extra or lost or defected chromosome, the neural network will show it through its decision . also the network will separate the noise which the filter algorithm could not remove still in the image after segmentation.

3.2 .Feed Forward Neural Network

The basic architecture consists of three types of neuron layers: input, hidden, and output. In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple layers of units, but no feedback connections are present. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process

such that the network will evolve to a stable state in which these activations do not change anymore.

In other applications, the changes of the activation values of the output neurons are significant; such that the dynamical behavior constitutes the output of the network.A feed-forward network has a layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer. The Ni inputs are fed into the first layer of Nh;1 hidden units. The input units are merely 'fan-out' units; no processing takes place in these units. The activation of a hidden unit is a function Fi of the weighted inputs plus a bias, as given in in eq

$$Y_k(t+1)=f_k(s_k(t))=f_k(\sum w_{jk}(t)y_j(t)+\Theta_k(t)$$

The output of the hidden units is distributed over the next layer of Nh;2 hidden units, until the last layer of hidden units, of which the outputs are fed into a layer of No output units .

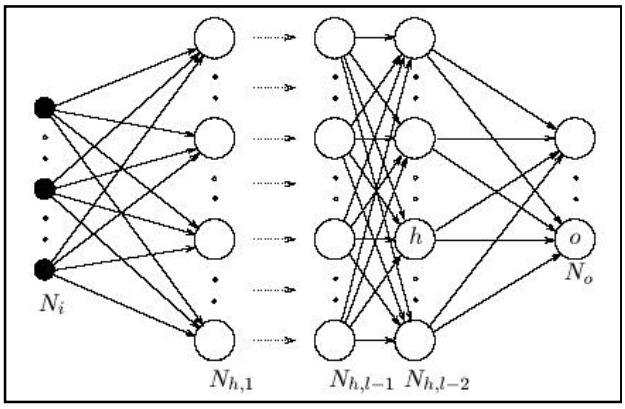

Figure 2: Architecture Of Feed Forward Neural Network

Although back propagation can be applied to networks with any number of layers, just as for networks with binary units it has been shown that only one layer of hidden units success to approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are non-linear. In most applications a feed-forward network with a single layer of hidden units is used with a sigmoid activation function for the units.

3.3.Back Propagation Algorithm

Backpropagation, is an abbreviation for "backward propagation of errors", is a common method of training neural network. Back propagation algorithm is the most popular

approach to implement learning in neural network. This algorithm is a multilayer network using a weight adjustment based on the sigmoid function. This method of training is an example of a unsupervised learning where the target function is known.
activation

It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, networks making up the training set. It is most useful for feed- (networks that have no feedback, or simply, that have no connections that loop). Backpropagation requires that the activation method used by theartificial neurons (or "nodes") be differentiable..

3.4.Algorithm For Feed Forward Neural Network & Back Propagation Algorithm

One of the problems with the traditional back propagation algorithm is the predetermination of the number of neurons in the hidden layer within a network. To overcome this problem the construction algorithm for feed forward networks may be used, which constructs the network during training. Thus we can have an optimal number of neurons in the hidden layer to attain a satisfactory level of efficiency for a particular problem. Besides applying the early stopping method of training using cross-validation we can also train the network in a relatively short estimation period (training period). In the construction algorithm proposed by Rudy Setiono and Huan Liu they have defined the stopping condition of the training by classifying all the input patterns. It means that while the efficiency is 100%, the training will stop. But in most cases with the benchmarking classification problems 100% efficiency may not be achieved. This is why we used a new algorithm for pattern classification that defines the stopping condition by the acceptance of efficiency level. Another consideration we have made that the desired or acceptable efficiency on the test sets may not be achieved even though the mean square error on training set is minimum. These considerations encouraged us to propose an algorithm that will combine the learning rule of backpropagation algorithm to update weights of the network and the construction algorithm to construct the network dynamically and also consider the efficiency factor as a determinant of the training process. the backpropagation learning algorithm can be divided into two phases: propagation and weight update.

3.4.1.Phase 1:Propagation

Each propagation involves the following steps:

1.Forward propagation of a training pattern's input the neural network in order to generate the propagation's output activations.ssss

2.Backward propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

3.4.2.Phase 2: Weight Update

For each weight-synapse follow the following steps:

1.Multiply its output delta and input activation to get the gradient of the weight.

2.Bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight.

This ratio influences the speed and quality of learning; it is called the *learning rate*. The sign of the gradient of a weight indicates where the error is increasing, this is why the weight must be updated in the opposite direction.

Repeat phase 1 and 2 until the performance of the network is satisfactory.

The following steps are followed to build and train a network

- Create an initial neural network with number of hidden unit h = 1. Set all the initial weights of the network randomly within a certain range.
- Train the network on training set by using a training algorithm for a certain number of epochs that minimizes the error function.
- If the error function av x on validation set is acceptable and, at this position, the network classifies desired number of patterns on test set that leads the efficiency E to be acceptable then stop.
- Add one hidden unit to hidden layer. Randomly initialize the weights of the arcs connecting this new hidden unit with input nodes and output unit(s). Set h = h + 1 and go to step 2.

For back propagation algorithm the weight adjustment is:

For the output-layer weights:

$W^o_{kjnew} = W^o_{kjold} + \eta \delta^o_{pk} i_{pj}$ .....(4.1)

Where $\delta^o_{pk} = \delta_{pk} f^o_k(net^o_{pk})$

For hidden layers

$$W^{h}_{jinew}=w^{h}_{jiold} + \eta\delta^{h}_{pj}x_{pi}.....(4.2)$$

Where

$$\delta^{h}_{pj}=f^{h}_{j}(net^{h}_{pj})\sum\delta^{o}_{pk}w^{o}_{kj}$$

Where, k indicates the kth output unit, j indicates the jth hidden unit; i indicates the ith input node, p is the input vector, Η is the learning rate, δ is the error term, xpi is the input value to the i, fok(netopk) is the output function of the kth output unit, fhj(nethpj) is the output function of j connected to k.The error function is usually defined as the mean-squared-errors

$$E_k=dk(n)-yk(n).....(4.3)$$

$$\Xi(n)=1/2\sum ek2(n).....(4.4)$$

$$\Xi a_v=1/N\sum\xi(n).....(4.5)$$

Where, k denotes kth output unit, n denotes the nth iteration, C is the number of output units, N is the total number of patterns, k d denotes the desired output from k, ky denotes the actual output of neuron k, ke denotes the error term for kth output unit.

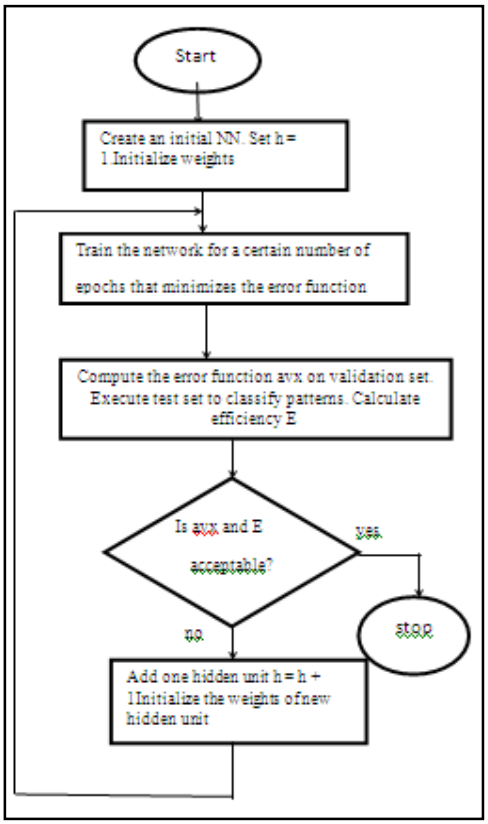

Figure 4: Flow Chart Of Back Propagation Algorithm

4.Result

The proposed algorithm is now applied to the images in the data set and the location of the Centromere and the length of each chromosome were extracted. Next, all chromosome images were considered by a cytogenetic expert and the location of the Centromere and the length of each chromosome were manually identified which are considered as the gold standard. For performance evaluation of the proposed algorithm the Euclidian distances between the experts defined Centromeres and the automatically identified ones were calculated in terms of the number of pixels for all chromosomes in the data set. Similarly, the differences between the expert defined chromosome lengths with those automatically identified by the proposed algorithm were calculated for each chromosome in terms of the number of pixels and the differences are considered as the algorithm's error.

Centromere location	Mean value of the absolute error	4.2 (pixels)
	Standard deviation of the absolute error	3.2
	Mean value of the normalized error	0.043
	Standard deviation of the normalized error	0.027
Chromosome length	Mean value of the absolute error	5.8
	Standard deviation of the absolute error	3.7

Table 1:Comparison results of the automatically defined Centromere locations and chromosome lengths to those manually identified by the expert

The mean value and the standard deviation of the absolute values of these errors are then calculated which are shown in Table I. In addition, since due to the different sizes of the chromosomes the absolute errors in Centromere locations have different meanings, the mean and standard deviation of the normalized errors are also calculated by dividing each absolute error to the expert defied chromosome lengths and shown in table I. Using these methods prior to the proposed algorithm the two important parameters of location of the Centromere and the length of the chromosome can be automaticallyidentified for any straight, curved or most important ones used for chromosome identification and automatic Karyotyping

5.Reference

1. Rajesh parekh,jihoon yang and vasant honavar"constructive neural network learning algorith ms for pattern classification",IEEE transactions on neural network,vol.11,no.2,mar 2000.
2. Ryu,seung,m.choo,m.jong and h.woo,seung, a study for the feature selection to identify human chromosome based on artificial neural network
3. Gonzalez,Rafael c.,2002 digital image processing IInd edition.
4. Cho,j.m., 2000.chromosome classification using back propagation neural network.IEEE engineering in medicine and biology.
5. "Elements of Artificial neural network" by Krishnan Meharota&SanjayRanga,MITpress