IOP Conf. Series: Materials Science and Engineering **402** (2018) 012194 doi:10.1088/1757-899X/402/1/012194

Literature survey of chromosomes classification and anomaly detection using machine learning algorithms

Nimitha N [1], Arun C[2] , Puvaneswari AS[3] , Paninila B[4], Pavithra VP[5], Pavithra B[6,*]

[1]Assistant Professor, Department of ECE, R.M.K College of Engineering & Technology, Chennai

[2]Professor, Department of ECE, R.M.K College of Engineering & Technology, Chennai

[3,4,5,6]Student, Department of ECE, R.M.K College of Engineering & Technology, Chennai

*Corresponding Author: pavithrabala9@gmail.com

Abstract Detection of chromosomal anomaly is done to prevent diseases at early stage. Karyotyping is the oldest manual method of detecting chromosomal abnormalities by dividing the chromosomes in laboratories. Reviews on karyotyping and previous other classification reviews on classification state that classifications were not extremely accurate. Some of them needed operator's interaction in the identification and separation of overlapping or touching chromosomes. They also didnot work properly for acrocentric, slanted, curved, banded chromosomes. Some works only for particular chromosomal anomaly. In our paper we are proposing chromosomal anomaly detection through various classification techniques to reach out the best accuracy.

Key Words: (Bilateral filter; Fuzzy C means; Feature extraction; Segmentation; Classification; Neural network)

1. Introduction

Many deadly diseases like Williams Syndrome, Tuner Syndrome, Down Syndrome, Giedion syndrome-Miller Dieker Syndrome cannot be cured at final stage. In order to prevent and identify these diseases earlier, chromosomes anomaly detection is done. It includes detecting chromosomal mutations like deletion, duplication by the classification of image.

Classification is generally done to extract the information classes. Decision tree, nearest neighbor, support vector machine, boosted tree, linear, random forest, neural network are various image classification methods made used. Most important among them was neural network. Because they process different records at different record times, and learn comparing classification of records with actual classification of records. Errors of initial record is fed back into the network and they modify network algorithm for different further iterations.

2. Review on chromosome classification:

Petros S. Karvelis, Dimitrios I. Fotiadis, Alexandros Tzallas and Ioannis Georgiou(2007) presented a paper on Multichannel Image Segmentation (MIS) & Region Bayes Classification

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

(RBC). Chromosome analysis can be made better but it produces misclassification errors which might be misunderstood as chromosome abnormalities. But the overall accuracy was only 83.59%.

Hyohoon Choi, Alan C. Bovik, and Kenneth R. Castleman (2008) did classification based on Expectation Maximization (EM) Algorithm. Accuracy of pixel classification is improved by 20 percentage after EM normalization. But classification accuracy will be low.

C.K. Neocleous, K.H. Nicolaides, K.C. Neokleous and C.N. Schizas (2010) classified using Artificial Neural Network structures for early identification in chromosomally abnormal foetuses. But it is true only for Down Syndrome.

Sahar Jahani, S. Kamaledin Setarehdan, and Emadedin Fatemizadeh (2011) used Automatic Karyotyping System for classification. It is simple yet effective algorithm. It is used to identify any cluster of overlapping or touching chromosomes. But the identification and separation of overlapping/touching chromosomes need operator's intervention.

Hongbao Cao, Hong-Wen Deng, and Yu-Ping Wang (2012) used Fuzzy C-Means Algorithm. Due to minimal classification and segmentation error, which will contribute to improved diagnosis of genetic diseases and cancers. But it classified only unbanded chromosomes.

Jagath Samarabandu Akila subasinghe arachchige, Peter K. Rogan, , Joan H. M. Knoll (2013) presented classification on Intensity Integrated Laplacian-Based Thickness Measurement. Standard Laplacian thickness measurement algorithm is used in the incorporation of contour information and intensity information to obtain greater accuracy of centromere location. But standard Laplacian thickness measurement algorithm done to imply contour information and intensity information to obtain more accurate centromere location.

B. Dhivyapriya and Dr.V.Vijaya Baskar (2014) presented classification on Relative Length Calculation using LabVIEW. Computational complexity is reduced by this proposed method. This algorithm holds good for every human chromosomal images with minimal error. It is still a complex and time consuming task which needs experienced operator or a cytogenetic expert to intervene.

Nirmala Madian, K.B. Jayanthi, and Dr.S.Suresh (2015) used Contour Based Segmentation for classification. This work helps in the automated identification of overlap zone which helps in segmentation and disentangling of chromosomes without manual intervention. This becomes a complex task when more than one overlap occurs in the chromosomes mainly due to the non rigid nature of the chromosomes.

V. Keerthi, R. S. Remya and K. Sabeena (2016) used GLM (Gray Level Mask). This technique gives accurate results for most of the chromosome classes. But this algorithm performance is poor for curved chromosome.

M. Neethu Sathyan, R. S. Remya and K. Sabeena (2016) performed classification on Gray Level Co-occurence Matrix (GLCM) & ANN Classifier. Gray Level Co-occurence Matrix (GLCM) & ANN Classifier Gray Level Co-occurence Matrix (GLCM) & ANN Classifier individual chromosome. But overall classification accuracy is **75%.**

3. Methods:

3.1 Artificial Neural Network:

An artificial neural network is an intelligence network depending on size and functions majorly related to biological terms. It learns by itself .There are various types of neural network shown as follows.

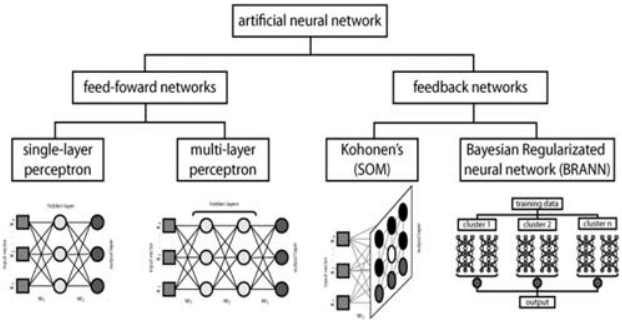


Figure 1. Classification of Artificial Neural Network

3.2 Feed-Backward Network:

When loops are introduced in the network signals will be travelling in both the direction in Feed-Backward Networks. These networks are so much complicated and they are much powerful. Computations derived from earlier input are fed back into the network, which provides a type of memory. Networks used for feedback are dynamic and their state will be continuously changing till they reach equilibrium point. Until the input changes they remain at the equilibrium point and a new equilibrium needs to be found.

Types :

1. Kohonen’s neural network
2. Bayesian Regularizated Neural Network(BRANN)

3.2.1 Kohonen's networks:

Kohonen's networks is a type of self-organizing neural network. Self-organizing capacity provides new possibilities like adapting to previously unknown input data. This method is used in human brains in which no patterns are defined. These patterns take shape during learning process, which is then later combined with normal work.

Kohonen's networks is a whole group of nets which uses self-organizing and competitive type learning method. First signal should be set up on the net's inputs and later winning neuron should be chosen. Scheme of rivalry and then modifications at later stage of synaptic wages will have different forms. Based on rivalry there are many sub-types, which themselves differ by precise self-organizing algorithm.

3.2.2 Bayesian regularized neural network:

To verify the best model of Neural Networks, the model was tested with upto five-neuron architectures. The analysis says that the BRANN training algorithm gives better performance than the Levenberg and other algorithm. It will reduce or eliminate the usage of wider crossvalidation. In Bayesian regularized (BRANN) models, regularization techniques includes imposing some prior distributions on the architect parameters. The function includes

$$F=\beta ED(D|w,M)+\alpha EW(w|M)$$

Where EW(w|M), is sum squares of weights, M is the Neural Networks architecture model in statistical jargon, α and β are core function parameters ,(known as regularization function or hyper-parameters which take only the positive values) that have to be estimated properly . The right hand side of equation which has , αEW, is known as weight decay and α, is its coefficient, supports small values for **w** and thus decreases the tendency of a architecture to overfit.

3.3 Feed-Forward Network:

The basic model has three types of layers. They are input layer , hidden layer , and output layer. In feed-forward neural networks, the direction of signal flow from input to output layers only , exactly in a feed forward direction. The processing of data can enlarge over many layers of units, but no extra feedback terms are present. Recurrent networks holds feedback connections.

Contradictory to feed forward networks, the dynamic properties of feed forward neural network are more important. In many cases, the activation values of neural networks terms engage in with a relaxation procedure such that neural network can evolve to the stable state condition, in which most of these activation function do not change anywhere. In some other applications, the changes of these activation functions of these output neurons layers are very significant; thus the dynamic behavior corresponds to the output of the neural network.

A feed-forward network consists of three layered structure. Where each layer consists of functions which will receive their input from corresponding terms, from the neural layer which is directly below and transfer their output to equivalent functions in a layer which is directly above the respective unit. There are no specified connections between the layer. The Ni inputs units are fed into the first neural layer of the Nh;1 hidden function. The input units are mostly the 'fan-out' units; no further processing takes place in such units. The activation principle of a hidden unit is function Fi of the evaulated inputs plus a bias, is given as equation as follows:

$$Yk(t+1)=fk(sk(t))=fk(\sum wjk(t)yj(t)+\Theta k(t)$$

The output obtained from the hidden units is distributed evenly over the next layers of Nh;2 hidden functions, until the last neural layer of the hidden functions, of which the output extracted are fed into the neural layer of No output terms .

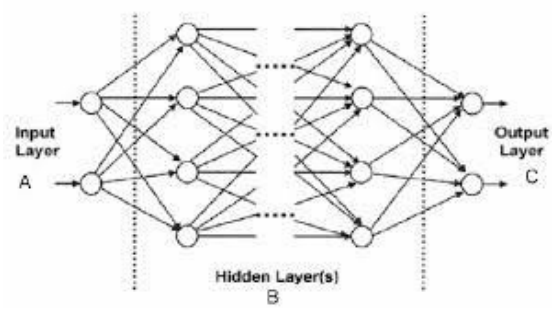


Figure 2. Networks of Artificial Neural Network

Backpropagation is applied to the neural network whenever the data is lost in the network.So ,when the data is lost ,it works back-froth algorithm and try to tertieve the data as soon as possible .Thus ,feedforward algorithm gives the better result of all.

Types :

1. Single layer perceptron
2. Multi-layer perceptron

3.3.1. Single Layer Perceptron:

The uncomplicated kind of artificial neural network is the single layer perceptron neural network, which consists only of single layer of input , output and hidden units.Thus ,it moves only in the forward direction and produces the best result

The single layer input is multi-dimensional.function $x = (X_1, X_2, .., X_n)$. Input nodes (or functions) are connected (technically fully) to one node (or multiple functions) in the next neural layer. A node in the next neural layer takes a evaluated sum of all the inputs.

Evaluated input = $\sum_i w_i I_i$

3.3.2. Multilayer perceptron:

The classification undertaken by us is done under multi-layered perception. It provides suitable solutions for various complex problems. The benefits of neural network is done by extracting the features of the images.

Neural network architecture has three layers of neuron networks which includes input layer , hidden layer, output layer. In Feed Forward Network (FFN) the direction of signal flow is from the input layer to the output layer. The direction of the network should be strictly followed. In addition to feed forward network the back propagation are also important.

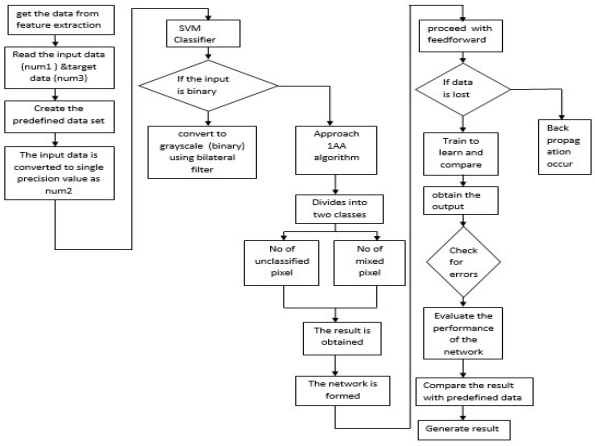


Figure 3. Classification flowchart using multilayered feed-forward network

FFN is a layered structure.First it receives the data from the feature extraction and stores. Now the some predefined data set will be given to obtain the output .Then enters the SVM classifier ,generally the SVM classifier checks whether the input data is binary or not. If it is binary ,it will proceeds with 1AA algorithm or else it will convert to grayscale and proceeds. The 1AA algorithm divides the data into unclassified and mixed pixels.Then the result is stored and continue with feed forward algorithm as explained earlier.The obtained result will now check with the predefined data set and identify the type of disease.

4. PROCEDURE

The chromosomes anamoly detection is done here. The sample of 24 chromosomal images are taken from MFISH data base.the output of each stage is stored into next stage.

- First step is preprocessing , where noise is eliminated using bilateral filter and the image is converted to grayscale image since everything is digitized.
- Second step is he segmentation, where the image gets segmented using Fuzzy-C algorithm. Segmentation is done to get the detailed study of the image.
- Third step is the feature extraction ,where the features are extracted individually using Wavelet Transformation and helps to analyse the image with sensitivity ,specificity, dimension etc.
- Fourth step is classification. The image gets classified using Artificial Neural Network. With helps of predefined data set the disease are identified.

Table 1. Disease identified with accuracy

Disease	Accuracy
Williams Syndrome	98.3%
Tuner Syndrome	98%
Down Syndrome	97%
Giedion syndrome-Miller Dieker Syndrome	93%

5. CONCLUSION

In this paper we conclude that classification can be best obtained by using multi-layered feedforward network. It works for acrocentric, slanted, curved, banded chromosomes also. It can detect various chromosomal abnormalities like Williams Syndrome, Tuner Syndrome, Down Syndrome, Giedion syndrome-Miller Dieker Syndrome. The accuracy obtained is about 97% which is far better than other accuracies obtained so far.

6. REFERENCES

[1] Sreejini K S; Lijiya A ;V K Govindan "M-FISH Karyotyping - A New Approach Based On Watershed Transform" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), **Vol.2,** No.2, April 2012.

[2] Vinya Vijayan; Remya R S;Sabeena K "Survey On Chromosome Image Analysis For Abnormality Detection In Leukemias" International Journal of Research in Engineering and Technology , **Vol 04**, Issue: 04 ,April 2015 .

[3] Neethu Sathyan M ; Remya ; Sebina " Automated Karyotyping Of Metaphase Chromosome images based on texture features" International conference of Scientific Research(IJSR),**Vol 2**,Issue12 ,December 2015.

[4] Revathy M Nair, Remya R S, Sabeena K "Karyotyping Techniques of Chromosomes: A Survey" International Journal of Computer Trends and Technology (IJCTT) ,**Vol 22**, Number 1,April 2015.

[5] Mehmet Can " Denver Groups Classification of Human chromosomes using CANN Teams" International Journal Of Research in Engineering and technology,August 2014.

[6] C.K.Neoclous; K.H.Nioclaides; N.K.Neokleous "Artificial neural networks for non-invasive chromosomal " Internal journal of research in in Engineering and Technology , **Vol.3** , No.3,April 2014.