International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.114-117
ISSN: 2278-2419

An Approach for Breast Cancer Classification using Neural Networks

S.Vijaya[1], D. Gladis[2]
[1] Research Scholar, PG and Research Department of Computer Science, Presidency College.
[2] Associate Professor, PG and Research Department of Computer Science, Presidency College.
E mail: vijayapeterraj@gmail.com, svijaya2774@gmail.com

Abstract-Breast Cancer,an increasing predominant death causing disease among women has become a social concern. Early detection and efficient treatment helps to reduce the breastcancerrisk.AdaptiveResonanceTheory(ART1),anunsuper vised neural network has become an efficient tool in the classification of breast cancer as Benign(non dangerous tumour) or Malignant (dangerous tumour). 400 instances were pre processed to convert real data into binary data and the classification was carried out using ART1 network. The results of the classified data and the physician diagnosed data were compared and the standard performance measures accuracy, sensitivity and specificity were computed. The results show that the simulation results are analogous to the clinical results.

Keywords: Breast Cancer, WBCD, ART1.

I. INTRODUCTION

Breast cancer is the second largest cancer that takes the life of human being. The root cancer is the disorderly development of cells in an organ of human body. Despite the fact that this affects both genders women are the most affected victims. A computer aided early diagnosis of Breast Cancer can reduce the mortality rate of victims.Breast cancer cells that develop in the tissues of breast grow and divide by itself to form new cells and accumulate to form a mass of tissues called a lump or tumour. Cancer cells infect the nearby breast tissue, make their way to lymph nodes and have a pathway to other parts of the body[1]. In the different stages of cancer, the early is stage I and advanced is stage IV, where cancer cells spread to other parts of the body especially liver. Breast cancer is mostly due to genetic anomaly. 5 – 10 % of cancer are due to inheritance from parents, 90% is due to generic abnormalities due to aging and lifestyle[1]. Researchers using various neural network models have obtained different levels of accuracy . Marcano *et al* using Artificial Metaplasticity Multilayer Perceptron (AMMLP) has obtained an accuracy of 99.26%.[3] Fawzi *et al* using Resilient Back Propagation Network has obtained an accuracy of 98.73%[5], R.R. Janghel *et al* using Linear Vector Quantization has obtained 95.82% of accuracy and 97.92% using Back propagation algorithm [2]. Seema *et al* using Adaptive Resonance Theory neural network(ART2) network model obtained an accuracy of 82.64%[4].

II. ADAPTIVE RESONANCE THEORY

Adaptive Resonance Theory, an unsupervised learning network is capable of forming clusters of arbitrary succession by self organization. It groups unconventionally and learns new clusters if necessary. The input patterns given to the network, selects suitable cluster. The cluster unit learns by adjusting the weights. The vigilance parameter manages the degree of similarity on same cluster. ART overcome the plasticity stability dilemma problem. The self organizing structure of network allows self-directed recognition and learning.The basic ART architecture consists of 3 neurons, F1 the input processing layer,F2 the clustering layer and reset mechanism that controls the degree of similarity. The F1 layer consists of an input portion for processing input, an interface portion that combines F2 and F1's input portion, for comparing the similarity of input signal with the weight vector for the cluster selected for learning. To control the degree of similarity between the connection interface portion units and cluster layer units, the bottom up weights connect interface portion to cluster layer and top down Weights connect cluster unit to interface portion. The highest net input in the cluster unit is the one selected to learn the input pattern. The interface unit integrates the data from input portion and cluster units.Based on the Similarity between input vector and top down weight vector, the reset mechanism unit decides which cluster unit may be allowed to learn. When the cluster unit is not allowed to learn and inhibited, a new cluster is selected. ART has plasticity and stability. New categories can be formed when the environment doesn't match any of the stored pattern. Unless there is sufficient similarity the environment cannot change the stored pattern.

III. METHODOLOGY

ART1 network[Figure 1], an unsupervised neural network consists of a comparison layer, a recognition layer, vigilance parameter and a reset module. When the set of weights of single neuron matches closer to input vector, the comparison layer takes an input vector and transfer it to the recognition layer. Each neuron in recognition layer outputs a negative signal, proportional to that neurons quality of match to the input vector, to each of other neurons and inhibits the output. This field allows each neuron to represent a category to which input vectors are classified. After classifying the input vector reset module compares the strength of match of recognition field to vigilance parameter. Training commences once the vigilance threshold is reached otherwise the recognition neuron is inhibited until a new input vector is applied. Until the vigilance parameter is satisfied by a recognition match the recognition neurons are disabled one by one by the reset function. If there is no recognition match an uncommitted neuron is committed and adjusted towards the matching input vector.[12]

The data for the classification of Breast cancer is obtained from University of Wisconsin, UCI machine learning repository[10].

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.114-117
ISSN: 2278-2419

The features that causes cancer are taken as input parameters for classification of breast cancer are listed[7][Table 1].

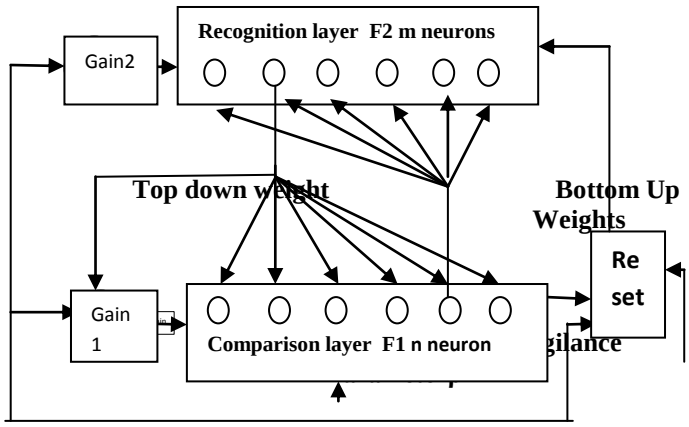

Figure1 ART1 network

Table 1. input parameters

No	Attribute	Domain
1	Clump thickness	1 – 10
2	Uniformity of cell size	1 – 10
3	Uniformity of cell shape	1 – 10
4	Marginal adhesion	1 – 10
5	Single epithelial cell size	1 – 10
6	Bare nuclei	1 – 10
7	Bland chromatin	1 – 10
8	Normal nucleoli	1 – 10
9	Mitosis	1 – 10
10	Class	2for benign and 4 for malignant

The missing attributes if any are replaced by median value method[7]The median value for n features is calculated by[6]

$$\text{median} = \frac{\text{size}(n+1)}{2}$$

where n is the number of input features.

The data normalization, an essential preprocessing step improve the performance of neural networks. Each attribute of data should be normalized in the digital form and used as input to neural network. The min max normalization technique[6] is used to normalize the data which is given by

$$v'''''''' := \frac{v - \min_A}{\max_A - \min_A}\left(new\_max_A - new\_min_A\right) + new\_min_A$$

Where

A is an instance of the input data,

v' is the new value in the required range,

min and max are the minimum and maximum value of the instances.

Algorithm:

Step by step clustering procedure

The input vector is I_H, where H represents the features of an instance. The vigilance parameter ρ. The output clusters are grouped according to the similarity determined by ρ. Initialize G1=0 & G2=0. Set nodes in F1 layer & F2 layer to zero.

Assign values to G1 and G2

G1 = 1 if $I_H \neq 0$ and output from F2= 0; 0 otherwise

G2 = 1 if $I_H \neq 0$; 0 otherwise

Step 1: Initialization

Initialize bottom up weights W(t) and top down weights V(t),

Initialize vigilance parameter ρ ($0.3 < \rho < 0.7$)

Step 2: repeat the steps 3 to 8 for all input vectors I_H, presented to F1 layer.

Step 3: choose input pattern vector

Step 4 : compute input Y for each node in F2

Step 5: select winning neuron k from step 4

If an indecision tie is noted then perform vigilance test

else go to step 6.

Step 6: compute activation in F1

Step 7: calculate similarity between activation in F1 and input.

Step 8: test similarity with vigilance parameter.

If similarity is true

Temporarily disable node k.

Update top down weights

Update bottom up weights

Update weight matrix w(t) and

v(t) for next input vector

If done with all input patterns then stop

the process otherwise repeat step 3 to 8 for next input pattern.

IV. PERFORMANCE EVALUATION

The data for the classification of Breast cancer(WBCD) is obtained from University of Wisconsin, UCI machine learning repository[10]. Out of 699 instances 400 instances are taken as inputs where each instance depicts nine attributes as clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nuclei and mitosis. 400 instances are trained in ART1 network and classified as '0'(Benign) and '1'(Malignant)The performance evaluation for 400 instances are carried out by determining the accuracy, sensitivity & specificity where accuracy is the standard measurement of performance for classification, sensitivity and specificity considered are supplementary measurements[9] . The values are represented using a confusion matrix.[Figure 2]

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Similarly the Sensitivity and specificity is calculated by

$$\text{Sensitivity(true positive rate)} = \frac{TP}{(TP + FN)}$$

$$\text{specificity(true negative rate)} = \frac{TN}{(TN + FP)}$$

where TP is positive cases that are correctly classified as positive, TN is negative cases that are correctly classified as negative, FP is negative cases that are wrongly classified as positive, FN is positive cases that are wrongly classified as negative.The vigilance parameter considered for training are $0.3 \leq \rho \leq 0.7$ and the confusion matrix are drawn for the instances of 100, 200, 300 and 400. Performance evaluation

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.114-117
ISSN: 2278-2419

parameters for the neural networks accuracy, specificity and sensitivity are calculated.The accuracy of this ART 1 neural network in classifying the breast cancer as benign and malignant is 70% approximately .For 100 instances the accuracy is 66% and for the instances 200, 300 and 400 the accuracy remains as 70% for the vigilance parameter 0.3. The performance evaluation parameters are calculated for other vigilance parameters. The accuracy remains 70% approximately for all the vigilance parameters [Table 2,3,4,5].

		Condition (as determined by "Gold Standard")		
		Condition positive	Condition negative	
Test out come	Test outcome positive	True positive TP =	False positive FP =	Positive predictive value TP/(TP + FP)
	Test outcome negative	False negative FN =	True negative TN =	Negative predictive value TN/(FN + TN)
		Sensitivity TP/ (TP + FN)	Specificity TN/ (FP + TN)	Accuracy TP + TN/ (TP+TN+FP+FN)

Table 2 performance for 100 inputs

Vigi Para	No of inputs = 100						
	TP	TN	FP	FN	Acc	Sens	Spec
0.3	37	29	20	14	**0.66**	0.73	0.60
0.5	45	28	21	06	**0.73**	0.88	0.57
0.7	51	21	28	0	**0.72**	1.00	0.43

Table 3 Performance for 200 inputs

Vigi Para	No of inputs = 200						
	TP	TN	FP	FN	Acc	Sens	Spec
0.3	80	61	41	18	**0.70**	0.82	0.60
0.5	90	58	43	09	**0.74**	0.90	0.57
0.7	97	47	55	01	**0.72**	0.99	0.46

Table 4 performance for 300 inputs

Vigi para	No of inputs =300						
	TP	TN	FP	FN	Acc	Sensi	Spec
0.3	123	86	61	30	**0.70**	0.80	0.59
0.5	144	82	63	11	**0.75**	0.93	0.57
0.7	151	61	86	02	**0.71**	0.99	0.43

Table 5 performance for 400 inputs

Vigi para	No of inputs = 400						
	TP	TN	FP	FN	Acc	Sens	Spec
0.3	146	132	90	32	**0.70**	0.82	0.60
0.5	167	119	100	14	**0.72**	0.92	0.54
0.7	176	85	134	05	**0.65**	0.97	0.39

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.114-117
ISSN: 2278-2419

V. CONCLUSION

The rate of prediction of breast cancer is analysed using ART1 network. It is found that approximately 70% of the data were classified correctly. The accuracy decreases as the vigilance parameter increases. So the vigilance parameter ρ can be choosen to be < 0.5. The future direction of this work is to improve the performance and the rate of prediction using ART1 network.

References

[1] Vikas Chaurasia, Saurabh Pal: A novel approach for Breast Cancer detection using data mining techniques, Intl Journal of Innovative Research in computer and communication Engineering,vol 2 , issue 1, jan 2014.page 2456 – 2465.

[2] Janghel.R.R., Anupam Shukla, Ritu Tiwari, and Rahul Kala.: Intelligent Decision Support System for Breast Cancer, Soft Computing and Expert System Laboratory, ABV- Indian Institute of Information Technology and Management, India.

[3] Marcano-Cedeno.A,Quintanilla- Dominguez.J,Andina. D.: WBCD breast cancer database classification applying artificial metaplasticity neural network, Expert Systems with Applications 38(2011)9573-9579.ELSEVIER.

[4] Seema Singh, Sunita Saini, Mandeep Singh.:Cancer Detection using Adaptive Network, Intl Journal of Advancements in Research & Technology,Vol 1,issue 4 , sep 2012 .

[5] Fawzi M.AL-Naima and Ali H. AL –Timeney: Resillient back propagation algorithm for breast biopsy classification based on ANN. www. intechopen.com

[6] Htet Thazin Tike Thein and Khin Mo Mo Tun.:An approach for breast cancer diagnosis classification using neural network: Advanced Computing:An International Journal (ACU),Vol.6, No.1, January 2015.

[7] Jiawei Han Micheline Kamber,Data Mining Concepts and Techniques, Edition 2

[8] Paulin.F, Dr. Santhakumaran.A.: Back propagation Neural Network by Comparing Hidden Neurons: Case study on Breast Cancer Diagnosis, Intl journal of Computer Applications,vol 2, no 4 , june 2010.

[9] Chandra Prasetyo Utomo, Asn Kardiana, Rika Yuliwulandari.: Breast cancer diagnosis using artificial neural networks with extreme learning techniques, Intl journal of advanced research in artificial intelligence, vol.3. No. 7, 2014. Page 10 -14.

[10] Breast Cancer Wisconsin Data

[11] http edu://archieve.ics.uci./ml/machinelanguage-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data.

[12] Sonia Narang,Harsh K Verma,Uday Sachdev: A review of Breast Cancer Detection using ART model of neural network, Intl Journal of Advanced research in Computer science and software Engineering, vol 2 , issue 10, oct 2012, page 311 – 318..

[13] http://www.myreaders.info/05Adaptive_Resonance_Theory.pdf

[14] K.Mumtaz et al,Evaluation of Neural network based classification systems for clinical cancer data classification, Journal of Computer Applications , vol 1, no 4 , 0ct – dec 2008.