ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

Early Diagnosis and Detection of Cancer using Artificial Neural Network based Prediction Algorithm

Alavi Kunhu[1], Nisi K, Sadeena Sabnam[2], Majida. A[3]

Associate Professor, Cochin College of Engineering and Technology, Valanchery, Kerala, India[1]

Assistant Professor, ECE Dept, Cochin College of Engineering and Technology, Valanchery, Kerala, India[2,3]

Abstract: Programmed cell death is called apoptosis and when this process breaks down, cancer begins to form. The chance of survival and successful treatment greatly increases with the early detection of cancer. So in this paper we are proposing an artificial neural network based prediction algorithm for the early detection of cancer. In our proposed work, we divide the available cancer data records into two group called train data records and test data records. Based on a prediction algorithm, we train three different artificial neural network models to predict the abnormalities in the patient data by employing sliding window method. Later on, we use this trained networks to test data records and the results are analyzed using various metrics parameters such as sensitivity, specificity, accuracy, precision and area under curve.

Keywords: Artificial neural networks; Sensitivity; Specificity; Accuracy; Precision.

I. INTRODUCTION

According to world health organization, 7 out of every 10 persons have a possibility of cancer occurrence. Early diagnosis is possible by recognizing possible warning signs and by taking prompt actions. This includes education to promote early diagnosis and employing screening tests. Screening refers to the use of simple tests across a healthy population, in order to identify individuals who have disease, but do not yet have the symptoms. For this, we are proposing an artificial neural network based prediction algorithm to properly diagnose as well as for detecting the possibility of cancer at a very early stage [1, 2, 3].

An artificial neural network mimics the basic operation of the brain. The brain is composed of neurons, which are the individual processing elements and neurons are connected by axons [4]. The human brain contains approximately 1011 neurons and each neuron connects to approximately 1000 other neurons. Similarly the artificial neural networks are a family of statistical learning algorithms generally presented as systems of interconnected neurons which can compute values from inputs, and are capable of machine learning as well as pattern recognition. Information is passed between the neurons and based on the structure and synapse weights, a network is provided.

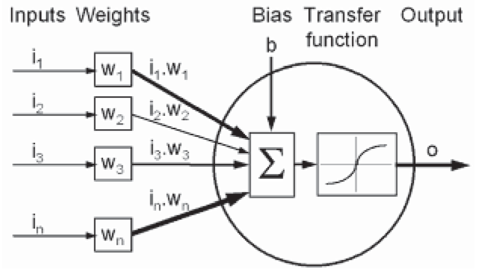

Figure 1. Generic model of Artificial Neural Network

Each input link i has an associated external input signal or stimulus xi and a corresponding weight wi, a sort of filter which is part of the linkage connecting the input to the neuron. The xi values can be real (+ or -), binary (0, 1), or bipolar (-1, 1). The weights, which model the synaptic neural connections, act to either increase or decrease the input signals to the neuron. The weights can also be binary or real-valued, but are usually assumed to be real. The outputs from the ANN can also be real-valued or binary or bipolar.

The neuron behaves as an activation or mapping function f (.) producing an output y = f(net), where net is the cumulative input stimuli to the neuron and f is typically a nonlinear function of net. Normally, net is often taken as the weighted sum of the inputs and f is typically a monotonic non-decreasing function of net and this can be summarized in the equation below.

$$Net=X_1W_1+X_2W_2+X_3W_3=\sum X_iW_i \quad .....(1)$$

The ANN can be a single layered or multi layered network [5]. Multi-layer networks have more complex, nonlinear relationships of input data to output results. Multiple-layer network consists of an input layer, an intermediate or hidden layer, and an output layer. The two input cells are named [u_1, u_2], two hidden cells [u_3, u_4], and one output cell [u_5]. Though input cells (u_1 and u_2) provide an input value to the network, hidden and output cells represent a function. The output cell u_5 is fed by two hidden cells u_3 and u_4 through weights $w_{5,3}$ and $w_{5,4}$ respectively. In a network with a hidden layer and the output layer, the hidden layer is computed first, and then the results of the hidden layers are used to compute the output layer.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

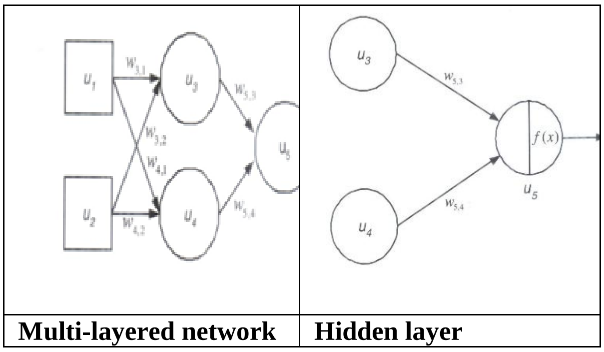

Figure 2. Multi layered network

Multilayered models generally use a sigmoid activation function, and therefore have continuous activations, mostly chosen from [0, 1] or [-1, 1].

A. Learning phase

Basically, there are two ways of learning in artificial neural network which are supervised and unsupervised learning. The overall concept of learning is to change the weights in the neuron in such a way that the error value is minimized for each pair of patterns of the fixed learning problem. As a stop criterion for the learning process, generally a total error E is used that is usually chosen to be the sum of the single error values. This total error is calculated anew after each epoch. An epoch is a complete run through the learning problem such that each pair of input /output patterns is processed once by the network using the learning algorithm. The learning process stops when the total error E is close enough to 0, or when the network is obviously not able to solve the learning problem. In supervised learning, the output is determined by comparing to the target pattern.

On the other hand, unsupervised learning algorithm works on a free learning problem which only contains input patterns. However, for this kind of learning procedure the network is also supposed to map similar input patterns to output that are similar to each other. In unsupervised learning algorithm, the units compete with each other and the principle of winner-takes-all is applied. It is trained without teaching signals or targets and it is only supplied with examples of the input patterns that it will solve eventually. This type of learning usually has an auxiliary cost function which needs to be minimized and the weights are modified where a cost function is minimized. At the end of the learning phase, the weights would have been adapted in such a manner such that similar patterns are clustered into a particular node.

So we are employing unsupervised learning procedure in the prediction algorithm used for the early detection of cancer. Here we are using the MATLAB high-performance language for technical computing integrated visualization, computation and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

The paper consists of V sections. The proposed method is explained in section II, training result analysis in section III, test result analysis in section IV and the section V contains the concluding remarks

II. PROPOSED METHOD

The proposed artificial neural network (ANN) based prediction algorithm is designed to accurately predict the patient's cancer output parameters like progression, based on patient's input records using sliding window method shown in Fig.3. In the case of sliding window method, the patient's record equally divided into 10 window group. Here we have used the average of 10 network group under various combinations of 10 windows to find out the predicted output result.

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
W2	W3	W4	W5	W6	W7	W8	W9	W10	W1
W3	W4	W5	W6	W7	W8	W9	W10	W1	W2
W4	W5	W6	W7	W8	W9	W10	W1	W2	W3
W5	W6	W7	W8	W9	W10	W1	W2	W3	W4
W6	W7	W8	W9	W10	W1	W2	W3	W4	W5
W7	W8	W9	W10	W1	W2	W3	W4	W5	W6
W8	W9	W10	W1	W2	W3	W4	W5	W6	W7
W9	W10	W1	W2	W3	W4	W5	W6	W7	W8
W10	W1	W2	W3	W4	W5	W6	W7	W8	W9

Figure 3. Sliding window method

The steps involved in the prediction algorithm are as shown below:

- The data collected from the patients are divided into two group called train data records and test data records
- The selected three artificial neural network functions such as newcf, newfftd and newfit are trained using train data records having 9 input parameters and one output parameter.
- The trained artificial neural network functions are then used to predict the output of the test data records having 9 input parameters.
- The performance of the trained artificial neural networks for both train data records and test data records are analysed by using various metrics parameters such as sensitivity, specificity, accuracy, precision, mean square error and area under curve.

The patient's data record distribution details are as shown below:

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

TABLE I. ANN NETWORK TRAIN AND TEST CASES

Case	Records distribution details
A	70% patient's data records used to train the ANN model 30% patient's data records used to test the ANN model
B	80% patient's data records used to train the ANN model 20% patient's data records used to test the ANN model
C	90% patient's data records used to train the ANN model 10% patient's data records used to test the ANN model

III.RESULTS AND ANALYSIS

Output performance of the proposed algorithm is analyzed using various metrics parameters like sensitivity, specificity, accuracy, precision, mean square error and area under curve [6, 7]. Sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity measures the proportion of actual positives which are correctly identified as such Specificity measures the proportion of negatives which are correctly. The analysis measures are calculated by the prior calculation of true positive, true negative, false positive and false negative.

- True Positive (TP): Sick people correctly diagnosed as sick
- False Positive (FP): Healthy people incorrectly identified as sick
- True Negative (TN): Healthy people correctly identified as healthy
- False Negative (FN): Sick people incorrectly identified as healthy

The Sensitivity (Sen) of a test is defined as the probability of a positive test given that the patient is ill.

$$Sen = TP/(TP+FN) \quad \ldots(2)$$

The specificity (Spe) of a test is defined as the probability of a negative test given that the patient is well.

$$Spe = TN/(TP+FN) \quad \ldots(3)$$

The accuracy (Acc) of a test is the proportion of true results (both true positive and true negative) in the population.

$$Acc = (TP+TN)/(TP+FP+FN+TN) \ldots(4)$$

[[

On the other hand, precision (Pre) defined as the proportion of the true positives against all the positive results (both true positives and false positive).

$$Pre = TP/(TP+FP) \ldots(5)$$

The proposed prediction algorithm is based three different artificial neural network functions such as feed-forward input time-delay back-propagation network (newfftd), cascade-forward back-propagation network (newcf) and fitting network (newfit) [8]. For both training and testing of our proposed ANN models, we have used progression data records having 9 input parameters (age, sex, other cancer, stage, grade, ever smoked, hMLH1, hMSH2 and Methyln) and 1 output parameter (progression). The schematic of the network training tool is as shown below. The network training tool gives an analysis of the performance, training state and regression. The performance plot gives an indication of the best performance attained in the training and also about the mean squared error involved. Regression is a variation between the expected output and target output. And finally the training state plot gives a status of the training done [9, 10].

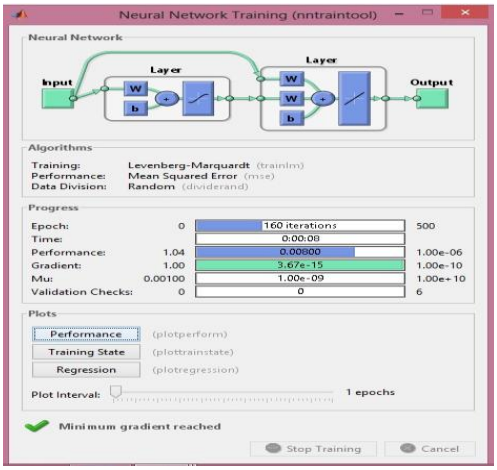

Figure 4. Schematic of the ANN training tool

The network training tool gives an analysis of the performance, training state and regression. The performance plot gives an indication of the best performance attained in the training and also about the mean squared error involved.Regression is a variation between the expected output and target output. And finally the training state plot gives a status of the training done.

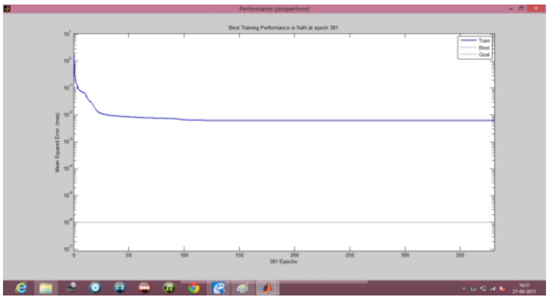

Figure 5. Performance plot of newcf 90-10 case

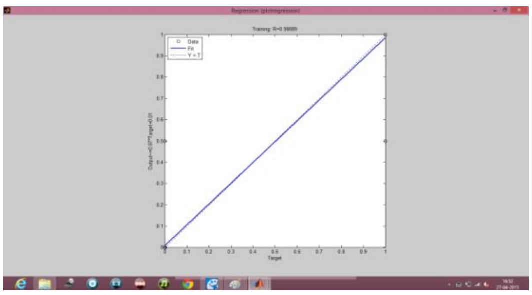

Figure 6. Regression plot of newcf 90-10 case

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

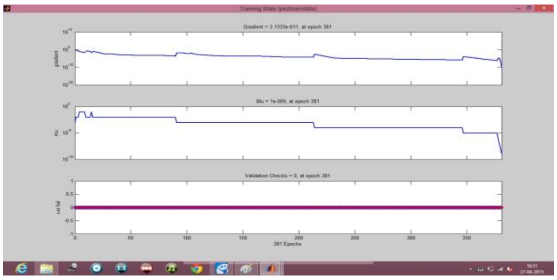

Figure 7. Training state status of newcf 90-10 case

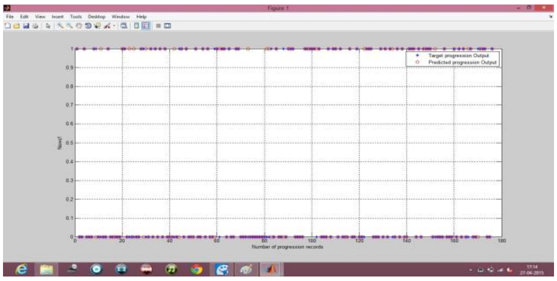

Figure 8. Train records predicted progression output plot

The progression output plot shown in Figure 8, gives an indication of the target progression output and predicted progression output. Based on the training process, we have 27 cases of the train results to analyze.

TABLE II. ANN TRAIN RESULTS

	Case	Sen	Spe	Acc	Pre	Mse	Auc
A	newcf	91.38	84.81	87.59	81.54	0.189	0.917
	newfftd	89.66	86.08	87.59	82.54	0.141	0.940
	newfit	89.66	89.87	89.78	86.67	0.118	0.951
B	newcf	96.72	96.84	96.79	95.16	0.040	0.988
	newfftd	96.72	98.95	98.08	98.33	0.035	0.989
	newfit	95.08	94.74	94.87	92.06	0.053	0.983
C	newcf	90.00	86.79	88.07	81.82	0.213	0.930
	newfftd	88.57	84.91	86.36	79.49	0.189	0.922
	newfit	85.71	89.62	88.07	84.51	0.159	0.940

Based on our training analysis we have come to the conclusion that newcf model is more efficient in terms of sensitivity. Also that as the train data increases, the performance also increases.

IV. TEST ANALYSIS

The trained network is finally used to test the data for the A, B and C cases for the three network models. We analyze the performance of the network on the basis of the area under curve, receiver operating characteristic plot, progression output data, sensitivity, specificity and so on.

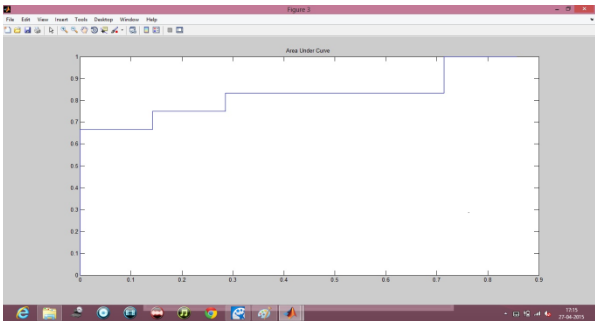

Figure 9. Test records area under curve for 90-10 newcf

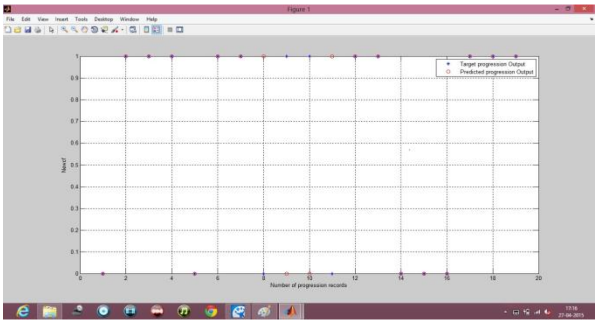

Figure 10. Test records predicted progression output plot

TABLE III. ANN TEST RESULTS

	Case	Sen	Spe	Acc	Pre	Mse	Auc
A	newcf	87.5	79.41	82.76	75.00	1.175	0.762
	newfftd	89.47	92.30	91.37	85.00	0.631	0.866
	newfit	88.46	90.62	89.65	88.46	0.654	0.863
B	newcf	81.25	95.65	89.74	92.85	0.217	0.816
	newfftd	89.47	95.00	92.30	94.44	0.089	0.921
	newfit	87.50	95.65	92.30	93.33	0.094	0.864
C	newcf	83.33	71.42	78.94	83.33	0.722	0.702
	newfftd	88.88	90.00	89.47	88.88	0.211	0.888
	newfit	91.66	57.14	78.94	78.57	0.430	0.660

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)
International Conference on Recent Trends in Computing and Communication (ICRTCC 2015)
Cochin College of Engineering & Technology
Vol. 4, Special Issue 1, June 2015

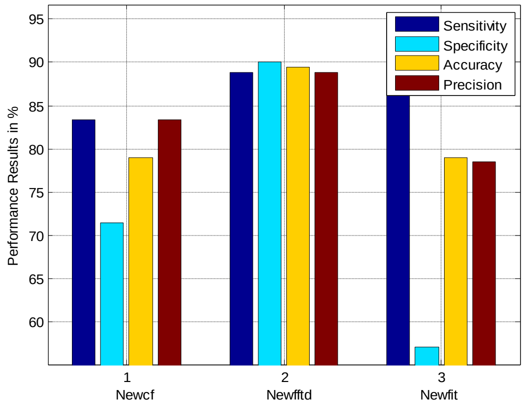

Figure 11. Performance analysis Case C test records predicted results

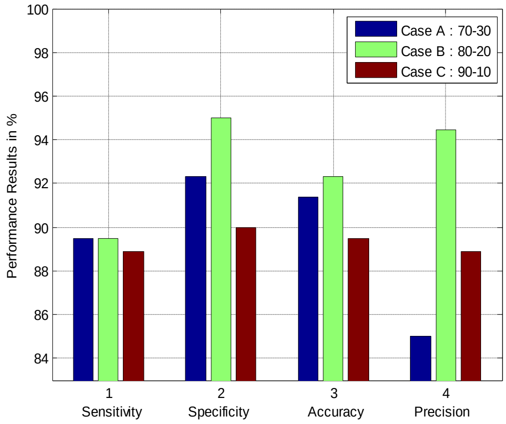

Figure 12. Comparison of predicted result performance of test records

From figure 11, we can see that newfftd method stands out in terms of highest sensitivity, specificity, accuracy, and precision. From figure 12, we can see that Case B gives better performance as compared to Case A and Case C.

V. CONCLUSION

This paper has presented a robust artificial neural network based prediction algorithm for the early diagnosis and detection of cancer. Three neural network models have been analyzed for 70-30, 80-20 and 90-10 train data and test data cases. The output performance is calculated on the basis of sensitivity, specificity, accuracy, precision and area under curve. On the basis of our analysis we have come to the conclusion that, as the train data increases the efficiency of the network goes on increasing. We expect to improve the efficiency of the network by considering more parameters and this will help to predict and to diagnose the possibility of cancer many years in advance.

ACKNOWLEDGMENT

The authors acknowledge all the reviewers and editor in charge for their valuable comments on improving this paper.

REFERENCES

[1] J Abdul Jaleel; Sibi Salim; R.B Aswin, "Artificial Neural Network Based detection of Skin cancer", International Journal of ISSN, Vol.1, Issue 3, September 2012.
[2] Ho Tak Lau; Adel Al-Jumaily, "Automatically Early Detection of Skin Cancer: Study Based on Neural Network Classification", IEEE International Conference of Soft Computing and Pattern Recognition, p.p: 375-380, December 2009.
[3] E. Gelenbe; Y. Feng; K.R.R. Krishnan, "Neural Network Methods for Volumetric Magnetic Resonance Imaging of the Human Skin", Proc. IEEE, 84, p.p:1488–1496, 1996.
[4] Jacek M. Zurada, "An Introduction to Artificial Neural Systems", West Publishing Company 1992.
[5] Golda. A, "Principles of Training multi-layer neural network using back propagation".
[6] Mahmoud Elgamal "Automatic Skin Cancer Image Classification", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 4, No. 3, 2013.
[7] Lisboa P.J; Etchells; T.A; Jarman I.H; Hane Aung M.S, "Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer", International Joint Conference on Neural Networks(IJCNN), p.p :2533-2538, 2007.
[8] Akram A. Moustafa, "Performance Evaluation of Artificial Neural Networks for Spatial Data Analysis", Contemporary Engineering Sciences, Vol. 4, No. 4, p.p.149 – 163, 2011.
[9] Zheru Chi; Zhongkang Lu; Fai-Hung Chan "Multi-channel handwritten digit recognition using neural networks", Circuit and Systems ISCAS'97 proceeding of IEEE International Symposium, Vol.1, p.p: 625-628, 1997.
[10] Cheng-Mei Chen ; Chien-Yeh Hsu ; Hung-Wen Chiu ; Hsiao-Hsien Rau, "Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees", Seventh International Conference on Natural Computation (ICNC), p.p :811-815, 2011.