ISBN 978-93-84468-11-8
Proceedings of 2014 International Conference on Artificial Intelligence & Manufacturing Engineering (ICAIME 2014)
Dubai, December 25-26, 2014, pp. 197-202

Performance of Cancer Prediction Based on Artificial Neural Network

Shayma Al-Ani[1], and Maysam Abbod[2]

[1] Department of Electronic and Computer engineering, Brunel University, London
[2] Department of Electronic and Computer engineering, Brunel University, London

Abstract: *Cancer has been known since of human history. The earliest written record regarding cancer can be dated back to circa 1600 BC by Egyptians. Cancer is a general condition which is subdivided into a group of conditions that are concerned with an abnormal growth in the cells within an organ or a tissue with the chance of spreading and invading other parts of the body. The number of cancer patients is increasing throughout the world,and thus emerges the necessity for new techniques to accurately predict the cancer behaviour to further improve the health status, by developing a new techniques based on intelligent systems.*

Keywords: *Accuracy, ANN, Cancer prediction, Ensemble.*

1. Introduction

Cancer has been known since of human history. Other names for cancer are malignant tumor or malignant neoplasm[1].

Genetic heritage, tobacco and alcohol intake, obesity, radiation exposure as well as having a poor and inactive lifestyle are some causes for abnormal cell growth and thus providing higher risks of getting cancer.

For such reasons, cancer is considered as one of the most dangerous and unpredictable diseases nowadays. A lot of research have been done on the prevention of cancer and cancer treatment[2]. Innovative and modern technology is being implemented in the goal of providing proper diagnosis, prediction and in some cases treatments for cancer.

Breast cancer is a type of cancer that affects the tissue of the breast. It may be apparent in different forms such as a lumps, change in breast shape and red patches in the skin among other symptoms affecting other parts such as pain in bones, shortness of breath and swollen of the lymph nodes. The risk factor of breast cancer includes inactive life and obesity, alcohol consumption, having children at a later age or not at all and hormone replacement therapy during menopause. Although the severity of the breast cancer depends upon the type of cancer and the person's age, the survival rates can reach up to 80 to 90%[3].

Artificial intelligence was one of the methods for approaching cancer and understanding its nature[3]. One of the most popular types of cancer being approached by artificial intelligence is breast cancer in females.

An artificial neural network combines artificial neurons in order to process information. A computational model inspired in the natural neurons.

An artificial neural network (ANN) is an imitation to the basic human brain operation and it is an interconnected neurons system that is capable of computing values using mathematical functions in which they determine the activation of the neuron. To adapt to the environmental changes, a learning system has to change itself. In addition, Multi-layers ANN are complex neural networks providing a nonlinear relationship of input-to-output results.

Multi-layer ANN comprises of an input layer, a hidden layer and an output layer as illustrated in Fig. 1. Basically, the input layer provides an input value to the network and each of the input cells has a weighting factor, which identifies the effect of the cell on the network. As for the hidden and output cells, they represent a function where the hidden layer is first computed, and then the results are used in computing the output layer.

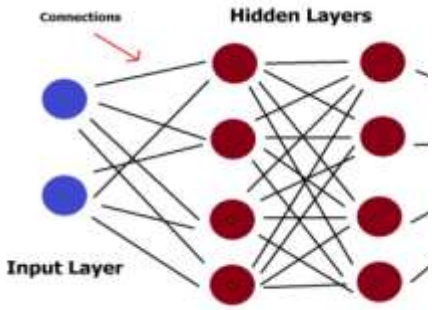

Fig. 1: Artificial neural network model.

The database for breast cancer was obtained from the University of Wisconsin, from Dr. William H. Wolberg. First of all, to represent instances, attributes are done on a scale of 2 to 10. Each instance has one of both possibilities; either benign or malignant. The size of the data set is only 369 instances and the collected results were in a single trial.

furthermore, the output data will be represented with the actual behavior of the tumor such as how long did and how many times the tumor went back to appear. in addition the time it took to more to other stages, the time the cancer spread and caused death and whether or not the cancer was the cause of death or there might be other causes of death (e.g. complications).

The samples reflect the grouping of the data through time order. The grouping information appears immediately though having been removed from the data itself.

As for the attributes, they include sample code number, clump thickness, uniformity of cell size and shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitosis and class (2 is benign and 4 is malignant).

2. Methodology

The proposed ANN based prediction algorithm accurately predicts the patient's cancer records output. ,by employing the ensemble method shown in Table I. In the case of ensemble method[4], the patient's record is equally divided into 10 window groups[5]. In this method, the average of 10 ANN network functions under different combinations of 10 window groups have been used in order to find out the predicted output[6], in addition to improving the prediction performance of a model with more accurate results[7].

The proposed model is trained by using three different ANN networks such as cascade-forward back propagation network (NEWCF), feed-forward input time-delay back propagation network (NEWFFTD), and fitting network (NEWFIT) ,each network is trained by using ensemble methods under different combination of groups by applying two methods which are the averaging and voting methods[8].

The averaging method is one of the major types of static committee machines. The network design for such method depends on the mean average of the networks. In addition, ensemble averaging depends on the mean average networks results. So in general, the whole idea of averaging method can be summarized by the following; generating N experts each having their initial values which are chosen from a random distribution. After that, each experts is trained separately and finally, they are combined and their values are averaged.

As for the voting method, it does not consider the level of significance by each network. This as a result, allows simple integration of all different sorts of network architectures. Majority voting is a simple voting method in which a group of unlabeled instance is performed depending on the class with the most frequent votes. This technique has been widely used to compare newly proposed method.

TABLE I: Sliding Window Method

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

In the case of Artificial Neural Network Model[9], the neuron behaves as an activation 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[10], where x_i indicate the inputs and w_i indicate the weighting parameters.

$$net = x_1w_1 + x_2w_2 + x_3w_3 = \sum_{i=1}^{3} x_iw_i \quad (1)$$

Output performances of the proposed algorithms are analysed using various parameters such as Sensitivity, Specificity, Accuracy, Roc, Area Under the Curve and MSE value[11].

The regression model is statistical model for estimating the relationships among variables. It includes many techniques for modelling and analyzing several variables and statistical models to compare it with the ANN model[12], output performances of them are analysed using various parameters[13], like Sensitivity, Specificity, Accuracy, Roc, AUC and MSE value .

3. Findings

The proposed ANN model has been trained by using three different ANN networks such as cascade-forward back propagation network (NEWCF), feed-forward input time-delay back propagation network (NEWFFTD), and fitting network (NEWFIT). First of all, the networks are randomly divided into two groups called training records and testing records.

Training records group contains about 70% of the total records, which are used to train the ANN by using 80% for training and 20% for validation of the ANN networks[4]. The trained ANN networks are used to predict the output parameter of testing records group which contain 30% of total records .

Moreover, three different methods have been used, average, voting, and regression model. Table II. shows the attributes variables used in the modeling analyses .

Table III. Table IV. Table V represent the performance of three methods for various ANN training networks in which it follows the principle of 70% Training and 30% Testing, while Tables VI, VII, VIII. represent the predicted patients records for three methods and three different ANN trained networks using ensemble method.

Sensitivity relates to the test's ability to identify positive results, which measures the proportion of actual positives which are correctly identified as such.

Specificity relates to the test's ability to identify negative results**,** which measures the proportion of negatives which are correctly identified .

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

Sensitivity=TP/(TP+FN)
Specificity=TN/(TN+FP)
Accuracy = (TP+TN)/ (TP+FP+TN+FN)

TABLE II: Attributes Variables Used In The Modeling Analyses

Input Variables
1. Sample code number: id number
2. Clump Thickness: 1 - 10
3. Uniformity of Cell Size: 1 - 10
4. Uniformity of Cell Shape: 1 - 10
5. Marginal Adhesion: 1 - 10
6. Single Epithelial Cell Size: 1 - 10
7. Bare Nuclei: 1 - 10
8. Bland Chromatin: 1 - 10
9.Normal Nucleoli: 1 - 10
10. Mitoses: 1 - 10
11. Class: (2 for benign, 4 for malignant)

TABLE III: Performance of ANN Networks Train Records Results Analysis

Methods	Sensitivity	Specificity	Accuracy
Average	97.101	97.778	97.549
Voting	98.551	97.778	98.039
Regression Model	91.304	92.519	95.0784

TABLE IV: Performance of ANN Networks Train Records Results Analysis MSE Value

Methods	MSE Value
Average	0.0625
Voting	0.0586
Regression Model	0. 1348

TABLE V: Performance of ANN Networks Train Records Results Analysis AUC

Methods	AUC
Average	0.9901
Voting	0.9919
Regression Model	0.9344

TABLE VI: Performance of ANN Networks Test Records Results Analysis

Methods	Sensitivity	Specificity	Accuracy
Average	97.101	97.778	97.549
Voting	98.551	97.778	98.039
Regression Model	91.304	95.519	93.0784

TABLE VII: Performance of ANN Networks Test Records Results Analysis MSE Value

Methods	MSE Value
Average	0.0772
Voting	0.0784
Regression Model	0.1492

TABLE VIII: Performance of ANN Networks Test Records Results Analysis AUC

Methods	AUC
Average	0.9911
Voting	0.9865
Regression Model	0.9367

"Fig. 2," and "Fig.3 ," indicate the ROC plot of breast cancer train records of average method and regression model for NEWCF,NEWFFTD and NEWCF networks, while "Fig. 4," and "Fig.5 ," indicate the ROC plot of breast cancer test records of average method and regression model.

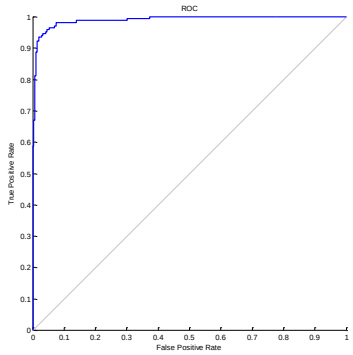

Fig. 2: Average method train case.

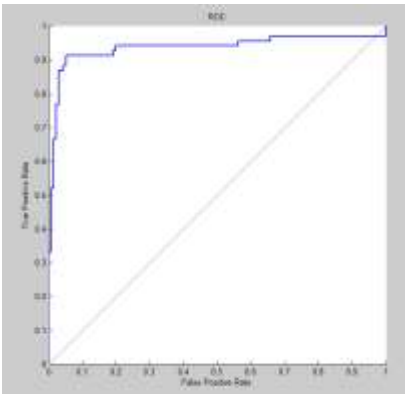

Fig. 3: Regression model train case.

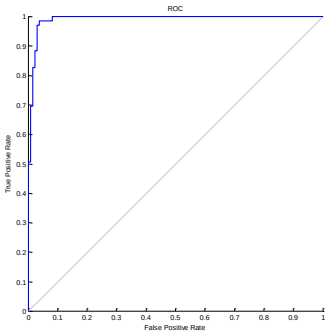

Fig. 4: Average method test case.

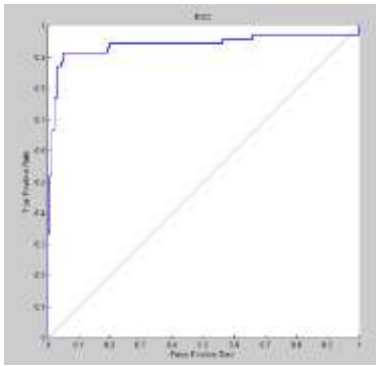

Fig. 5: Regression model test case.

4. Conclusion

The proposed ensemble model artificial neural network algorithm using two methods average, voting and for various artificial neural network functions such as feed-forward input time-delay back-propagation network, cascade-forward back-propagation network and radial basis network, accurately predicted patient's records. Output performances of records are analyzed using various parameters such as Sensitivity, Specificity, Accuracy ,Roc, Area Under the Curve and MSE value, and results show that artificial neural network methods obtain better predictive performance than could be obtained from regression models and that was all based on the different validations of the artificial neural networks.

5. References

[1] (2014). "What is Cancer". *National Cancer Institute*. [Online]. Available: http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer

[2] J.A. Cruz and D. S. Wishart. (Feb, 2007). Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*. [Online]. 2(2006), 59-77. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675494/.

[3] S.Christobel and J.Sunil (2009). Breast cancer (1. ed. ed.). Oxford: Oxford University Press. p. Chapter 13.

[4] P. Melville, "Creating Diverse Ensemble Classifiers," PhD proposal,

[5] Texas Univ., Texas, United States, 2003.

[6] Robi Polikar (2009) Ensemble learning. Scholarpedia, 4(1):2776.
http://dx.doi.org/10.4249/scholarpedia.2776

[7] Opitz, D.; Maclin, R. (1999). "Popular ensemble methods: An empirical study".Journal of Artificial Intelligence Research 11: 169–198.

[8] U, Naftaly, N. Intrator, and D. Horn. "Optimal ensemble averaging of neural networks." Network: Computation in Neural Systems 8, no. 3 (1997): 283–296.
http://dx.doi.org/10.1088/0954-898X/8/3/004

[9] L. Rokach. (Nov, 2009). Ensemble-based classifiers. Springer Science+Business Media. [Online]. 33(2010), 1-39. Available: http://www.ise.bgu.ac.il/faculty/liorr/AI.pdf

[10] Z. Chi, Z. Lu and F. Chan, "Multi-channel handwritten digit recognition using neural networks", Circuit and Systems ISCAS'97 proceeding of 1997 IEEE International Symposium, Vol.1, 1997, pp. 625-628. K. Tadashi; U. Junji; T. Shoichiro, "Hybrid GMDH-type neural network using artificial intelligence and its application to medical image diagnosis of liver cancer", System Integration (SII), 2011 IEEE/SICE International Symposium on 2011, pp. 1101-1106.

[11] P.J.G. Lisboa, T.A. Etchells, I.H. Jarman and M.S.H. Aung, "Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer", Neural Networks, 2007. IJCNN 2007. International Joint Conference on 2007, pp: 2533-2538.

[12] (2014). "Regression analysis". Wikipedia. [Online]. Available http://en.wikipedia.org/wiki/Regression_analysis#Regression_models.

[13] C.Chen, C. Hsu, H.Chiu and H.Rau, "Prediction of survival in patients with livercancer using artificial neural networks and classification and regression trees", Natural Computation (ICNC), 2011 Seventh International Conference on 2011, pp. 811-815.

[14] Janghel, R. R.; Shukla, Anupam; Tiwari, Ritu; Kala, Rahul, 2010. Breast cancer diagnosis using Artificial Neural Network models. Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on 2010, pp. 89-94