International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 12 | Dec-2017 www.irjet.net p-ISSN: 2395-0072

Applications of Artificial Neural Networks in Cancer Prediction

Sumit Khopkar[1], Neil Gupte[2], Anvay Sonpimple[3], Purva Kulkarni[4]

[1,2,3,4]*Student, Dept. of Computer Engineering, SKN College of Engineering, Maharashtra, India*

Abstract – *Cancer is one of the leading causes of death every year in the entire world. As of today there is limited research on the cure for cancer and hence it makes cancer as one of the primary concerns for deaths across the world. However machine learning techniques have provided us with a wide range of prevention and early detection provisions. Artificial Neural Networks (ANN) have been proved as one of the most efficient ways for successfully finding ways for prevention, early detection and prediction of prognosis period. This paper elaborates some of the methods for achieving the same through ANNs and also explains their results in this process.*

Key Words: **ANN, MEE, MaxEN, Genetic Algorithm, ensemble, Prediction, Prognosis**

1. INTRODUCTION

Cancer is known as a group of diseases involving abnormal cell growth and has a potential to spreading to other parts of the body[9]. In 2015 about 90.5 million people had cancer and it is observed that about 14.1 million people every year are diagnosed with cancer. The most common types of cancer are lung cancer, prostate cancer, colorectal cancer, stomach cancer, breast cancer, bladder cancer and cervical cancer [12]. In this paper we will be mainly covering the Artificial Neural Network (ANN) techniques applied on lung cancer, breast cancer and bladder cancer. Artificial Neural Networks work on the principle used in the neural networks used in brains which pass on information from one neuron to the next.

In this paper we have discussed a number of methods which are based on the ANN mechanism and are applied on breast, lung and bladder cancer to predict prognosis, recurrence and states of cancer. These methods include the Maximum Entropy Estimation (MEE), Multi-layer Perceptron (MLP), Genetic Algorithm (GA) and ensemble method. All of these methods were compared to other machine learning algorithms and it was concluded that ANN gave the best results amongst them. This has encouraged us to dive deeper into the concept of ANN and its effectiveness to perform prediction.

2. WORKING OF ARTIFICIAL NEURAL NETWORKS

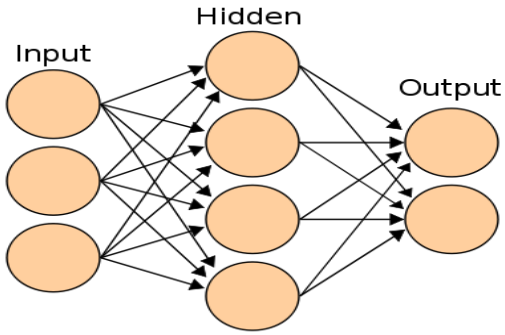

Fig -1: Artificial Neural Network layer structure

Artificial Neural Networks work on the principle used in the neural networks used in brains which pass on information from one neuron to the next [3]. Similarly in ANNs the information is passed on between layers of neural networks. Fig-1 represents the layer structure of ANN which consists of three different layers namely the input layer, the hidden layer and the output layer. The input layer takes the different inputs which are actually the independent variables in our dataset and pass it on to the hidden layer. The hidden layer is the backbone of the ANN model and is responsible for the computation of which parameters actually contribute to the prediction and pass it on to the output layer. The hidden layer does so by assigning weights to the input parameters. This is done with the help of a stochastic gradient descent (Fig-2) method which calculates weight to be assigned according to the error encountered in the calculation of predicted value with respect to the actual value [8].

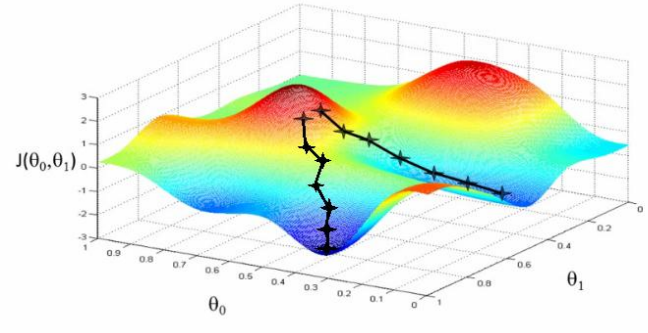

Fig -2: Stochastic Gradient Descent

And finally, the output layer receives the prediction value from the hidden layer and it produces the required output. This process is quite lengthy and requires large number of heavy computations. However, it has been proved to be

more efficient than most of the machine learning and deep learning techniques thus it proves its widespread popularity.

3. BREAST CANCER APPLICATIONS

3.1 Prognosis

Survival rates (prognosis) tell you what portion of people with the same type and stage of cancer are still alive a certain amount of time (usually 5 years) after they were diagnosed. They can't tell you how long you will live, but they may help give you a better understanding about how likely it is that your treatment will be successful. This is a pattern recognition problem which can be done with the help of probability distribution functions. One method known as the Maximum Entropy Estimation (MEE) which is used for multinomial distributions was proved to be very efficient [2]. MEE searches for constraints which are satisfied from a small set of data points and establishes correlations between the variables. This process requires huge computations. ANN helps in this by introducing parallelization into it with the help of Maximum Entropy Network model (MaxEN model). MaxEN consists of two modules of interconnected processing elements where each element computes simple operations. Fig-3 describes the working of the MaxEN testing module. The MaxEN testing module receives Lagrange Multipliers from the training module which it uses in order to calculate probability values.

There were a total of 4 constraints which were imposed which were mean, variance, correlation between outcome variable and risk factors and a second order correlation between the same.

Type 1:

$$< x_k >= \frac{1}{M} \sum_{p=1}^{M} x_{kp}$$

Type 2:

$$< x_k^2 >= \frac{1}{M} \sum_{p=1}^{M} x_{kp} x_{kp}$$

Type 3:

$$< x_1 x_k >= \frac{1}{M} \sum_{p=1}^{M} x_{1p} x_{kp}$$

Type 4:

$$< x_1 . x_k . x_u >= \frac{1}{M} \sum_{p=1}^{M} x_{1p} x_{kp} x_{up}$$

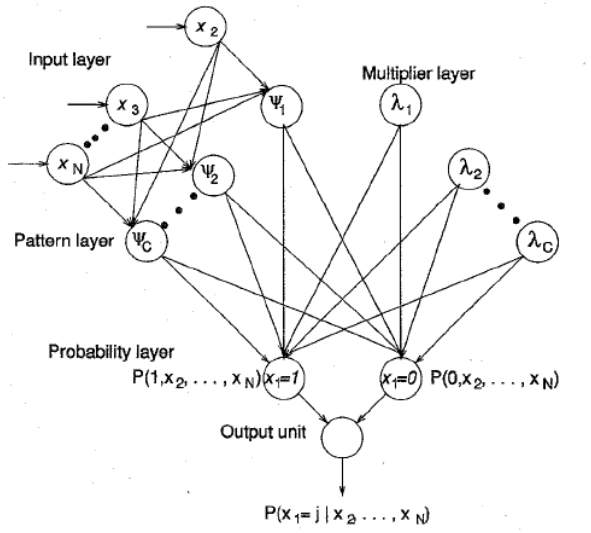

Fig -3: MaxEN Testing Module

The results of MEE were cross checked with that of Multi Layer Perceptrons (MLP) and Probabilistic Neural Networks (PNN) and it was found that the MEE model was the most effective. Fig-4 illustrates the effectiveness of using MEE with respect to MLPs and PNNs.

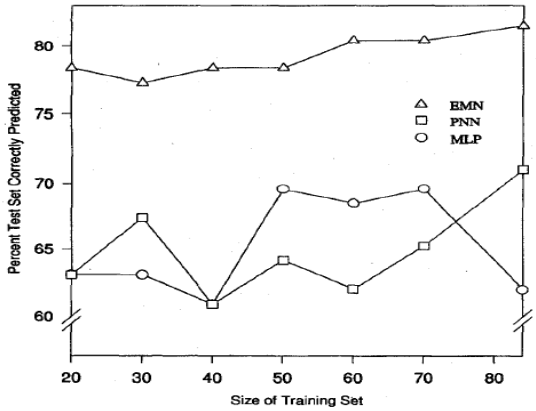

Fig -4: Comparison of MEE, MLP and PNN

3.2 Recurrence

If cancer is found after treatment, and after a period of time when the cancer couldn't be detected, it's called a cancer recurrence [11]. For prediction of cancer recurrence a data of 146 patients was obtained from the Clinical Centre of Kragujevac which was applied on classification techniques such as Naïve Bayes (NB), Decision Tree (DT) based on C4.5 algorithm, Support Vector Machine (SVM) with polynomial kernel, logistic regression (LR), K-Nearest Neighbors (K-NN) and ANNs [4]. ANN was constructed as an MLP with one hidden layer, unipolar sigmoid activation function was set in all neurons and back propagation with momentum was used as the learning algorithm.

Table -1: Results of testing of classifiers for the problem of cancer recurrence

Algorithms	AC	SENS	SPEC	AUC
ANN	0.9315	0.964	0.833	0.952
SVM	0.952	0.982	0.861	0.921
LR	0.911	0.964	0.75	0.96
DT	0.9657	0.991	0.889	0.833
NB	0.856	0.891	0.75	0.891

Accuracy (AC), Sensitivity (SENS), Specificity (SPEC) and Area under ROC curve (AUC) are used to measure the classifier performance based on the values obtained in the confusion matrix that contains information about the actual and predicted values. As we can see from Table-1, ANN gives the best results on the basis of AUC as parameter.

4. LUNG CANCER APPLICATIONS

Lung Cancer is one of the most fatal types of cancer, it is a serious problem faced by people of developed as well as developing countries. Nearly 160,000 lives are lost annually alone in USA. An ANN model was built which made use of genetic algorithm (GA) which successfully identified genes that classify lung cancer status with notable predictive performance [5]. Firstly the proper dataset was chosen which consisted of information of genes of normal lung patients, squamous cell lung carcinoma (SQ), small cell lung carcinoma (SM), pulmonary carcinoids (CO) and adenocarcinoma (AD) patients. Proper preprocessing of this data is done which helps us in eliminating non required data and helps in faster computations of result.

Then Artificial Neural network were used as prediction models to classify lung cancer. ANNs are highly adaptable learning machines which can detect non-linear relationships between the features and the sample classes.

Genetic algorithm is used for optimization of the developed Artificial Neural Network. The Genetic algorithm works in following way

At its core, a genetic algorithm

1. Creates a population of (randomly generated) members
2. Scores each member of the population based on some goal. This score is called a fitness function.
3. Selects and breeds the best members of the population to produce more like them
4. Mutates some members randomly to attempt to find even better candidates
5. Kills off the rest (survival of the fittest and all), and
6. Repeats from step 2. Each iteration through these steps is called a generation.

Through the optimized ANN network the following results were obtained-

160 sets of (chromosomes) with at most 200 generations each set for the classification problem were run. Majority (154 chromosomes) of the sets of analyses in this method satisfied the desired accuracy rate of at least 97% correct prediction. Only 6 chromosomes (no solution) obtained an accuracy rate of less than 97% (Figure 2). Forward selection method was used for this process.

Then confusion matrix was formed depicting the values of each type of cancer showing their specificity and sensitivity in the defined ANN model.

Table -2: Confusion matrix for model performance using ANN

Known Class	AD	CO	NL	SM	SQ
AD	0.979222	0.000257	0.009537	0.000000	0.010985
CO	0.106262	0.860034	0.002732	0.008225	0.022748
NL	0.071257	0.000000	0.926639	0.002104	0.000000
SM	0.078039	0.048174	0.078464	0.646439	0.148884
SQ	0.114092	0.010429	0.008352	0.042207	0.824920
	Overall: 93.66 %		Average: 84.75 %		

The system is able to classify the lung cancer types using gene expression with optimisation of artificial neural networks using genetic algorithm.the accuracy of the system using ANN is around 93.66% considering the various genetical input paramters. This is how ann can be used to predict the cancer type of a patient and ensuring proper prediction rate most of the times.

5. BLADDER CANCER APPLICATIONS

An ANN model was proposed to predict the patient's cancer records by employing an ensemble method. The ANN model made use of 10 window groups to find the predicted output. The proposed model was trained by three different ANN networks which were the cascade-forward back propagation network (NEWCF), feed-forward input time-delay back Propagation network (NEWFFTD), and fitting network (NEWFIT) and each of these networks were trained by ensemble method by applying averaging and voting methods [3]. Averaging method depends upon the mean average of networks. Majority voting is a voting method in which

unlabelled instances are performed depending upon the class with most frequent votes. The activation function which is the decider that which parameters are to be weighed more than the others is given as f(net) where net is the cumulative input stimuli to the activation function. x is the input and w is the weighing parameter.

$$net = x_1 w_1 + x_2 w_2 + x_3 w_3 = \sum_{i=1}^{3} x_i w_i$$

This model was tested against a regression model and the ANN model gave a much better performance.

Table -3: Performance of ANN networks test records results analysis.

Methods	Sensitivity	Specificity	Accuracy
Average	65.2174	74.2857	70.6897
Voting	60.8696	77.1429	70.6897
Regression Model	52.609	50	51.034

Table -4: Performance of ANN networks test records results analysis MSE and AUC values.

Methods	MSE Value	AUC
Average	0.1908	0.7280
Voting	0.1956	0.7193
Regression Model	0.1335	0.5745

5. CONCLUSIONS

Thus, from our survey we have discovered how ANN has proved to be one of the most efficient mechanisms for prediction of prognosis, recurrence and other issues related to cancer. It has given us a bigger motive to dig deeper into the concepts on ANN and its applications so as to utilize it on any further prediction problems we encounter in the future.

ACKNOWLEDGEMENT

We thank Prof. Saudagar S. Barde for assistance with machine learning and Artificial Neural Networks, and for comments that greatly improved the manuscript.

REFERENCES

[1] Topon Kumar Paul and Hitoshi Iba, "Prediction of Cancer Class With Majority Voting Genetic Programming Classifier Using Gene Expression Data.", IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND INFORMATICS, VOL 6, NO 2, APRIL JUNE 2009.

[2] Poh Lian Choong, Christopher J. S. Desilva, *Member, IEEE,* Hugh J. S. Dawluns, and Gregory F. Sterrett, "Entropy Maximization Networks: An Application to Breast Cancer Prognosis", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL 7, NO. 3, may 1996

[3] Shayma M. Al-Ani, Maysam Abbod, "Prediction of Cancer Behavior Based on Artificial Intelligence", New Developments in Circuits, Systems, Signal Processing, Communications and Computers, 2016

[4] Bojana R. Andjelkovic Cirkovic, Aleksander M. Cvetkovic, Srdjan M. Ninkovic and Nenad D. Filipovic, Member IEEE, "Prediction Models for Estimation of Survival Rate and Relapse for Breast Cancer Patients", Bioinformatics and Bioengineering (BIBE), 2015

[5] Joey Mark Diaz, Raymond Christopher Pinon, Geoffrey Solano, University of the Philippines, Manila, Philippines, "Lung Cancer Classification Using Genetic Algorithm to Optimize Prediction Models", Information, Intelligence, Systems and Applications, IISA, 2014

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

[7] Joseph, A.C., David, S.W., "Applications of machine learning in cancer prediction and prognosis", *Cancer Inf.*, 2006

[8] Kubat, M, ""Neural networks a comprehensive foundation by Simon Haykin, MacMillan", Cambridge University Press, 1999.

[9] "What is Cancer". National Cancer Institute. [Online]. http://www.cancer.gov/cancertopics/cancerlibrary/what-iscancer

[10] Article on Breast Cancer Survival Rates and Statistics. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html

[11] Article on Cancer Recurrence. https://www.cancer.org/treatment/survivorship-during-and-after-treatment/understanding-recurrence/what-is-cancer-recurrence.html

[12] Article on Cancer. https://en.wikipedia.org/wiki/Cancer