**Prediction of Breast Cancer using Machine Learning**

1. ***Abstract- The number of cases of Breast cancer seems to be constantly increasing worldwide from the year 1990 to 2013. The breast cancer incidence rate in Asia appears to be 37.5% and the mortality rate is 13.21% according to 2012 world cancer report. As per the recent statistics report of 2018, it has been estimated that the count of new cases of breast cancer is 14% in India. The percentage is higher when compared to lung cancer, oral cancer and other type of cancer. It has been identified that it occurs mostly in the age group between 41- 50 years and the distribution of this is found to be 42%, whereas it is 18%***

***,24% for the age groups of 31-40 and 51-60 respectively. It has got five stages of Survival which includes the survival rate of 100%, 98%, 88%, 52%, 16% for each stage respectively. Diagnosis in the early stage can reduce the death rate, even during the stage 3 if the cancer is predicted it can be treated. Hence with the help of machine learning using artificial neural network, the cancer can be predicted. In this paper Artificial Neural Network Algorithm is used to predict the breast cancer if it is benign or malignant. ANN has been a powerful tool for analyzing the data when there are non-linear interactions between the input and the output to be predicted. The results show that the accuracy of artificial neural network (ANN) for the prediction was better than other approaches. The result varies based on the number of iterations performed. The results are very competitive and can be used for diagnosis, prognosis, and treatment.***

1. ***Index Terms-Breast cancer, Wisconsin Dataset, Neural Network, Feed-Forward algorithm***
   1. **INTRODUCTION**

Breast cancer is known since ancient times. The visible symptoms at the later stages when the lumps turns to tumors are recorded in early period of time as the breast lumps tend to manifest themselves as visible tumors. The uncontrollable growth of cells in an organ is called a tumor which is termed as cancer. There are different forms of cancer, among which breast cancer is the most occurring disease. It is the leading cause for mortality between 40 and 55ages. Diagnosis in early stage can reduce the death rate. With the help of existing technologies such as data mining and machine learning, diagnosis can be made more accurate and medical errors can be minimized. The Wisconsin breast cancer database is used and analyzed.

# RESEARCH ELABORATIONS

*MACHINE LEARNING ALGORITHMSAND ITS TYPES*

Machine learning is a branch of artificial intelligence that deals with a variety of statistical, probabilistic and optimization techniques that allows computers to learn from past examples and to find patterns from large, noisy and complex data sets. It is concerned with designing and development of algorithms that allow computers to extend behaviors based on observation and documentation of patterns, like the sensor data or databases.

# Supervised learning

Data is labelled with a class or value. The goal is to predict the value or class of the labelled data. It is based on the datasets available. It includes classification and regression, Classification is the identification the categories to which the observation belongs based on the training set of data support vector machine, naïve Bayes, nearest neighbor falls under this. Regression is a statistical approach to find the relationship between the variables. It is used to predict the output from the given dataset decision tree, Neural networks comes underregression.

# Unsupervised learning

The data here is unlabeled or the value is unknown. The goal is to find the pattern or clustering groups and interpret data based on the input given. we should explore the data and find the structures in them. Data can be classified in groups according to their similarity.

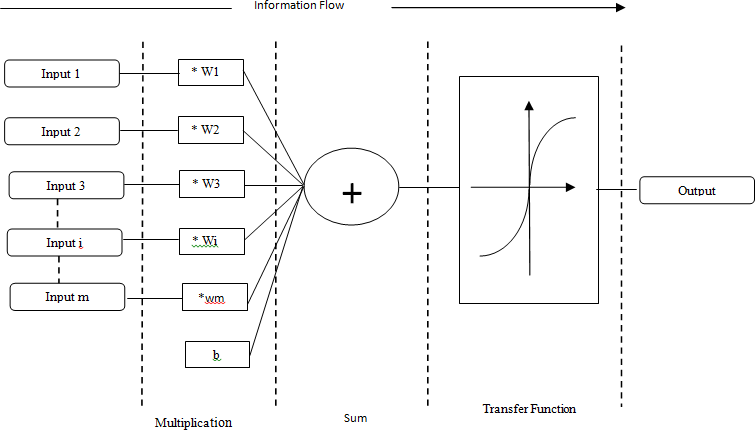
# Reinforcement learning

It is a technique which sets the framework of an artificial neural network, where data is not fed but are created by the actions in the environment. In the absence of a training dataset it bounds to learn from its experience itself. It deals with the actions that has to be taken by the artificial neural networks.

*Neural Networks and it’s learning rules*

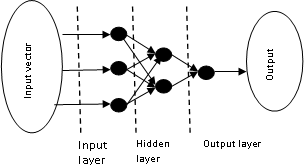
A neural network is a model which is similar to the brain that processes the information. It can be used to extract patterns from the complicated datasets. Many neural network models assume many simplifications over actual biological neural networks which are necessary to comprehend the properties and to attempt any mathematical analysis. Even if all the properties of the neurons are known, simplification is mandatory for analytical purpose. Neural networks can change the weights as a function of their performance. In ANNs, all the neurons operate together, which makes ANN to perform tasks at much faster rate. A neural network has to be trained prior to its usage. They are trained withtraining datasets and the solution is predictedwith testing dataset.

# Figure 1. Working flow of Artificial Neural Network Mathematical Model

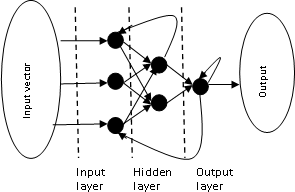
The basic aspect of a neural network is neuron. It is a mathematical model that has three sets of rules such as multiplication, summation and activation. The input will get multiplied with its weight and all the weighted input will get summed in the middle section. At the output section the activation part also called as transfer function will be there to decide what should be fired asoutput.

# Figure 2. Working principle of an artificial neuron Artificial Neural Network

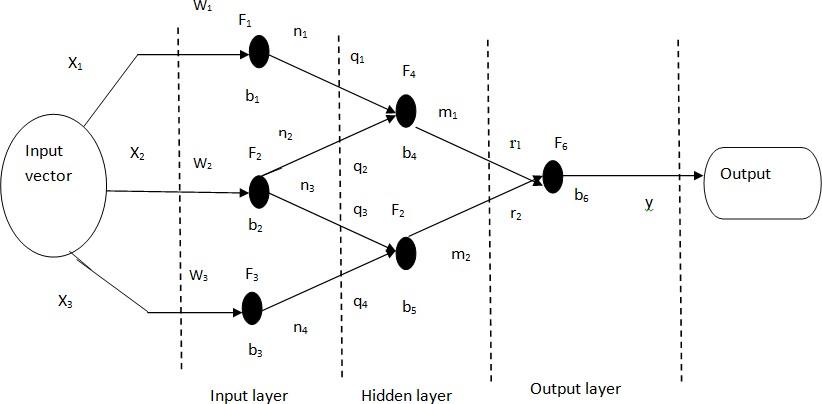
information can also flow in opposite direction as shown in figure 4



# Figure 3. Feed forward topology



**Figure 4. Recurrent topology Feed-forward Artificial Neural Networks**

**Figure 5. Feed forward Neural Network**

The Figure 5 shows the basic feed forward network topology. The Network contains feed-forward topology is called as feed-forward network. It will work in a restraint that the information flow has to be only in one direction with no back-loops.

𝑛1 = 𝐹1 𝑤1 𝑥1 + 𝑏1 (1)

𝑛2 = 𝐹2 𝑤2 𝑥2 + 𝑏2 (2)

𝑛3 = 𝐹2 𝑤2 𝑥2 + 𝑏2 (3)

𝑛4 = 𝐹3 𝑤3 𝑥3 + 𝑏3 (4)

𝑚1 = 𝐹4 𝑞1 𝑛1 + 𝑞2 𝑛2 + 𝑏4 (5)

𝑚2 = 𝐹5 𝑞3 𝑛3 + 𝑞4 𝑛4 + 𝑏5 (6)

𝑦 = 𝐹6 𝑟1 𝑚1 + 𝑟2𝑚2 + 𝑏6 (7)

𝑞1𝐹1 𝑤1 𝑥1 + 𝑏1 +

Based on the flow of information artificial neural

𝑟1 𝐹4 𝑞 𝐹 𝑤 𝑥

+ 𝑏 + 𝑏4 + ⋯

network is consist of two types. They are feed-forward topology and recurrent topology. In feed-forward topology

𝑦 = 𝐹6

… 𝑟

𝐹

2 2 2 2 2

𝑞3 𝐹2 𝑤2 𝑥2 + 𝑏2 +

(8)

+ 𝑏

the information is flow from input to output in only one direction as shown in Figure 3. Whereas in recurrent topology it takes more than one direction i.e., the .

2 5 𝑞4 𝐹3 𝑤3 𝑥3 + 𝑏3 + 𝑏5 6

# RESULTS AND DISCUSSION

The Dataset is obtained from UCI WISCONSIN Breast Cancer dataset. The figure 6 shows the Wisconsin Breast Cancer dataset. The Dataset consist of 569 entries. MATLAB software is used for simulating the output. The dataset has 30 features from that only 10 significant features are inserted into the MATLAB. The network is trained and tested in the ratio of 70-30%. After partitioning the dataset, the functions are given step by step using the nntool in the MATLAB is shown in figure 7.Then the network is created and it is shown in figure 8. For each and every training session the regression graph will also get varied. The Figure

9 shows the last regression graph obtained during final training session. Then performance measures are calculated by using the below formula. The performance measures like confusion matrix (shown in Table 2), accuracy, error, sensitivity, specificity, positive predictive value and negative predictive value can be calculated by using the equations (9), (10), (11), (12), (13), (14) respectively. Finally, table 1 shows the overall performance, measure of

# Figure 9. Training result

𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 = 𝑡𝑝 +𝑡𝑛 (9)

𝑇𝑜𝑡𝑎𝑙 𝑒𝑛𝑡𝑟𝑖𝑒𝑠

𝐸𝑟𝑟𝑜𝑟 = 1 − 𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 = 𝑓𝑝 +𝑓𝑛

(10)

feed-forward network model.

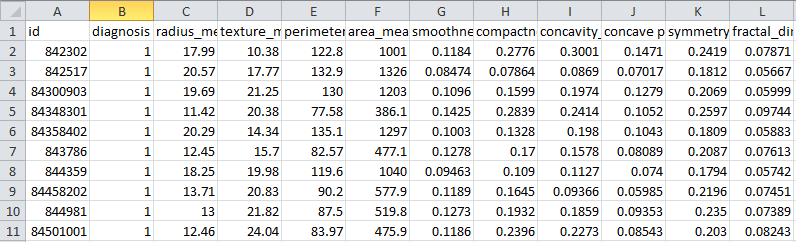
𝑆𝑒𝑛𝑠𝑖𝑡𝑖𝑣𝑖𝑡𝑦 =

𝑡𝑝

𝑡𝑝 +𝑓𝑛

(11)

𝑎𝑙𝑙 𝑒𝑛𝑡𝑟𝑖𝑒𝑠 𝑖𝑛 𝑡𝑕𝑒 𝑑𝑎𝑡𝑎𝑠𝑒𝑡

𝑆𝑝𝑒𝑐𝑖𝑓𝑖𝑐𝑖𝑡𝑦 = 𝑡𝑛

𝑓𝑝 +𝑡𝑛

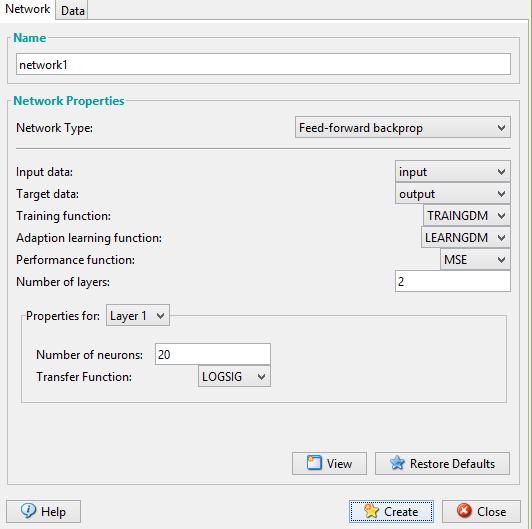
(12)

𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒 𝑃𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑣𝑒 𝑉𝑎𝑙𝑢𝑒 = 𝑡𝑝

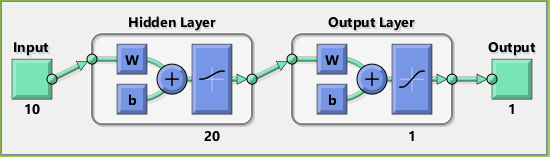
𝑡𝑝 +𝑓𝑝

(13)

# Figure 6. WISCONSIN Breast Cancer Dataset



**Figure 7. Creating Network or Data window**



**Figure 8. Custom Neural Network window**

𝑁𝑒𝑔𝑎𝑡𝑖𝑣𝑒 𝑃𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑒 𝑉𝑎𝑙𝑢𝑒 = 𝑡𝑛 (14)

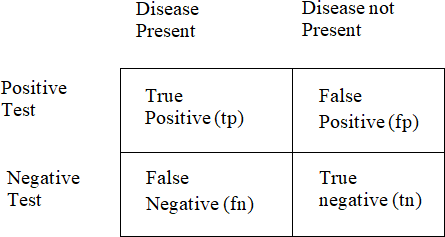
𝑡𝑛 +𝑓𝑛

Where tp, tn, fn, fp are found using the confusion matrix as shown in Table .2

# Table.1.Performance Measures of Feed-Forward model

|  |  |
| --- | --- |
|  | **Feed-Forward** |
| Accuracy | 0.8953 |
| Error | 0.1047 |
| Sensitivity | 0.7419 |
| Specificity | 0.7419 |
| Positive Predictive Value | 0.9583 |
| Negative Predictive Value | 0.8709 |

**Table.2. Confusion matrix**



* 1. **CONCLUSION**

Neural network has the ability toderive different kinds of samples. This characteristic seem to have an effect upon various aspects which includes the neurons count, connection type and the training extent of the network. Neural Network also has the advantage that it is able to find the lost values. Whereas logistic regression does not consider the lost values foradditional analysis. The ANN model performs better than logistic model. Given with enough running time, Neural networks are able to predict the output with high level accuracy. Being an easy and quick tool, It helps the health care labours to gauge the risk of breast cancer.

# REFERENCES

|  |  |
| --- | --- |
| 1. | Hiba Asria, Hajar Mousannif, Hassan Al Moatassime, Thomas Noel, “Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis”, *Procedia Computer Science 83, (2016), 1064-1069.* |
| 2. | Abien Fred M. Agarap, “On Breast Cancer Detection : An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset”, arXiv:1711.07831v1 [cs.LG] 20 Nov 2017. |
| 3. | Yasmeen Mourice George, et al.“Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Image”, IEEE SYSTEMS JOURNAL, VOL. 8, NO. 3, SEPTEMBER 2014 |
| 4. | Kalpana Kaushik, Anil Arora, “Breast Cancer Diagnosis using Artificial Neural Network”, International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol 7 issue 2 July 2016 |
| 5. | AutsuoHiga, “Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms”, International Journal of Computer Applications Technology and Research Volume 7–Issue 01, 23-27, 2018, ISSN:-2319–8656 |
| 6. | Abien Fred M. Agarap, “A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data”, arXiv:1709.03082v8 [cs.NE] 7 Feb 2019 |
| 7. | Esraa A. AL-Dreabi,et al. “Automated Detection of Breast Cancer Using Artificial Neural Networks and Fuzzy Logic”, International Journal of Sciences: Basic and Applied Research (IJSBAR) (2017) Volume 35, No 3, pp 109-120 |
| 8. | HtetThazin Tike Thein and Khin Mo Mo Tun, “AN APPROACH FOR BREAST CANCER DIAGNOSIS CLASSIFICATION USING  NEURAL NETWORK”, Advanced Computing: An International Journal (ACIJ), Vol.6, No.1, January 2015 |
| 9. | R.R.Janghel, Anupam Shukla,Ritu Tiwari, Rahul Kala, “Breast Cancer Diagnosis using Artificial Neural NetworkModels”, IEEE Xplore, 2010 |
| 10. | “UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set.” |