IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. V. (Jul.-Aug. 2017), PP 53-60
www.iosrjournals.org

Adaptive Neuro Fuzzy Inference System (ANFIS): MATLAB Simulation of Breast Cancer Experimental Data

Ejiofor C. I[1] & Okengwu U.A[2].
Department of Computer Science, University of Port-Harcourt, Port-Harcourt. Nigeria [1,2].
Corresponding Author: Ejiofor C. I

Abstract: *The accurate prediction of breast cancer has been an area of interest due to the complexities associated with experimental data for breast cancer. This paper has decided to explore the Adaptive Neuro Fuzzy Inference system (ANFIS) using Matrix Laboratory (MATLAB) for the simulation of Winconsin breast cancer experimental data. Twenty (20) Winconsin breast cancer dataset was used for the simulation, comprising of nine (09) attributes with diagnosing values identifying Malignant or Benign. MATLAB ANFIS simulation captured the fundamental editors inclusive of training, testing and ANFIS structure. The result of simulation showed that from all parameters of simulation; numbers of nodes had the highest values followed by total number of parameters, number of parameters, number of non-linear parameters and number of training data respectively.*
Keyword*: ANFIS, Breast Cancer, Simulation, Experimental data (dataset)*

Date of Submission: 12-07-2017 Date of acceptance: 14-08-2017

I. Introduction

Breast tumour is a generic name for malignancy and non-malignancy; with malignancy accommodating the propensity for propagation but lacking hyperplasia while non-malignancy lack the propensity for propagation but accommodating hyperplasia. Breast cancer; a form of tumour usually starts off within the inner lining of milk ducts or the lobules which supply milk for nursing mothers within the breast. Breast cancer originating from lobules is usually known as lobular carcinoma while one that develops from the ducts is called ductal carcinoma. Breast cancer undeniably is the most common cancer among female worldwide which account for 16% of all female cancer and 18.2 of all cancer death worldwide (MedicineNet, 2016).

Breast cancer is perceived as a malignant disease caused by uncontrolled growth of abnormal cells in the breast (WrongDiagnosis, 2011). These abnormal cells (old or damaged) are created for destruction or death after cell division with the aim of preventing cancerous growth. The statistical occurrence of cancerous cell in female breast is more common for caucasian women as opposed to African-American women. Usually, the risk factor for breast cancer includes being over 50 years of age, having a personal or family history of breast cancer. Other factor may include consuming hormones estrogen, starting menstruation before 12 year of age (early menstruation), starting menopause after age 55 and giving birth after 30 (late childbearing). Radiation may also pay an important role in obtaining cancerous growth, if expose repeated to radiation elements. These risk factors left unchecked will result in breast cancer, while breast cancer left untreated, multiply, spread and propagate (Oncolex, 2016).

A diagnosis of breast cancer can easily be delayed or missed due to minute breast lump affecting a mammogram, or lump may not be painful and might be ignored. In addition, some symptoms of breast cancer can resemble symptoms of other diseases and conditions. The treatment of breast cancer varies, depending on the individual case and the type and stage of the cancer. Treatment options may include surgery, radiation therapy, and chemotherapy (MedicineNet, 2016; Oncolex, 2016 and WrongDiagnosis, 2011).

However, early diagnosis is the fundamental key for prompt treatment of breast cancer which fosters huge success. Overtime, the complexities of breast cancer have been overwhelming. These complexities usually are sponsored by late detection of the disease, usually when several vital organs have been comprised due to continuous and active propagation. Another notable issue involves the accurate classification of breast cancer; differentiating between benign and malignancy cancerous growth. These two fundamental issues (promptness and accuracy) require novel methods in enhancing diagnosis (CancerQuest, 2016).

Till date, methods for the accurate and early detection of breast cancer are of utmost importance and are still an active area of current research (CancerQuest, 2016).
Standing the aforementioned, this paper intends to apply ANFIS MATLAB simulation for breast.

II. Applied Material

A Neuro-Fuzzy approach; Adaptive Neuro Fuzzy Inference System (ANFIS) is a combination of an Artificial Neural Network (ANN) and a Fuzzy Inference System (FIS) using the Takagi Sugeno Model and adaptive neural network learning approach capabilities. Figure 1 depicts the ANFIS model

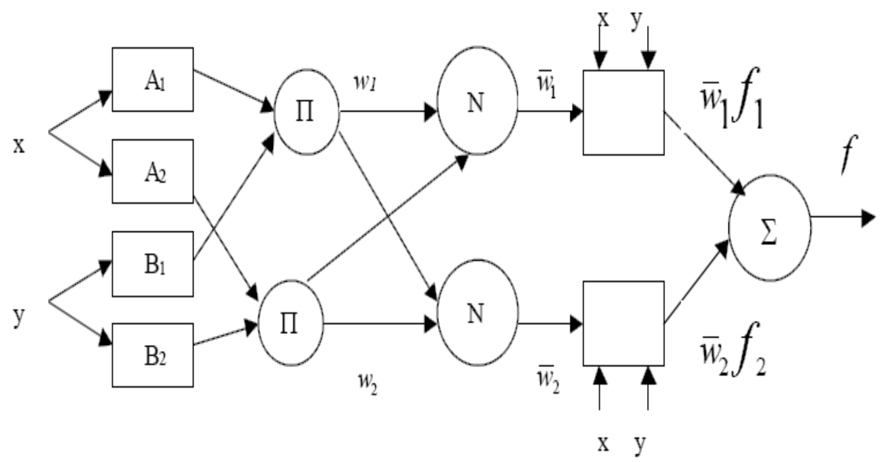

Figure 1: Structure of the ANFIS Network.

ANFIS is a six layers network (read as 1 – 6 or 0 -5), with distinct functional nodes performing specific function. The learning or training of ANFIS is structured using a hybrid approach, combining the Backward-Propagation Gradient Descent (BPGD) and a Least-Squares Estimator (LSE) method. This structure possesses computational capability of neural network with the explanatory power of fuzzy logic.

The model comprises of six integrate layers, which includes:

a. **Layer 0 (L0):** The Input layer is the first ANFIS architecture layer. This layer consists of input nodes which represents the input parameters.

b. **Layer 1 (L1):** This layer is called the membership function layer. The function of each node in this layer is maps the input of the linguistic variable from the input layer to variable linguistic values using an appropriate membership function.

c. **Layer 2 (L2):** This layer is called the rule layer. The node in this layer receives input from the membership function layer and calculates the firing strength of each rule using the min or prod operator.

d. **Layer 3 (L3):** This layer is called the normalization layer and the node in this layer receives input from the rule layer, computes the ratio of the incoming signal from the rule layer and it computes the output and the output of this layer.

e. **Layer 4 (L4):** This layer is known as the defuzzification. The nodes compute a parameter function on the normalization layer. Parameters in this layer are called consequent parameters.

f. **Layer 5 (L5):** This is known as the output layer. It comprises of a single node which sums all the incoming signals and produces the output of the ANFIS system

ANFIS: as a predictive model learns from a set of experimental data which are usually structured to a particular domain problem. This domain usually possesses parameter input and rules which are obtained from the experimental data. The antecedent parameters (ANFIS input) are obtained from the experimental data input fuzzified with an appropriate membership function. The consequent parameters are repeated obtained by adjusting the antecedent parameters utilizing the hybrid algorithm (Least Square Estimator: LSE and Back Propagation Gradient Descent: BPGD). During the training process, the experimental data is fed into the ANFIS model in cyclical form seen an epoch; comprises of both the forward pass (LSE) and the Backward Pass (BPGD). The adjustment of premise parameters and consequent parameters are captured as an epoch.

III. Methodology: Breast Cancer Prediction using Adapted Neuro-Fuzzy Inference System (BC-ANFIS).

The adopted approach implement breast Cancer prediction using ANFIS. The selected Wisconsin Breast Cancer Dataset obtained from UCI Machine Learning Repository was used for simulation (https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+ (original). The experimental data comprises of thirteen (13) columns. These columns includes cases, experimental code, nine (09) decision variables, diagnosis values (2, 4) and class of classification (Malignant or Benign). These columns cut-across 20 selected cases. Table 1 shows these experimental data.

Table 1: Wisconsin Breast Cancer Dataset

Cases	**Dataset Code**	**Clump Thickness**	**Uniformity of Cell Size**	**Uniformity of Cell Shape**	**Marginal Adhesion**	**Single Epithelial Cell Size**	**Bare Nuclei**	**Bland Chromatin**	**Normal Nucleoli**	**Mitoses**	**Classes of Diagnosis (2-Benign, 4-Malignant**	**Classification class**
1	1000025	5	1	1	1	2	1	3	1	1	2	Benign
2	1017122	8	10	10	8	7	10	9	7	1	4	Malignant
3	1002945	5	4	4	5	7	10	3	2	1	2	Benign
4	1044572	8	7	5	10	7	9	5	5	4	4	Malignant
5	1015425	3	1	1	1	2	2	3	1	1	2	Benign
6	1041801	5	3	3	3	2	3	4	4	1	4	Malignant
7	1016277	6	8	8	1	3	4	3	7	1	2	Benign
8	1047630	7	4	6	4	6	1	4	3	1	4	Malignant
9	1017023	4	1	1	3	2	1	3	1	1	2	Benign
10	1050670	10	7	7	6	4	10	4	1	2	4	Malignant
11	1018099	1	1	1	1	2	10	3	1	1	2	Benign
12	1054590	7	3	2	10	5	10	5	4	4	4	Malignant
13	1018561	2	1	2	1	2	1	3	1	1	2	Benign
14	1054593	10	5	5	3	6	7	7	10	1	4	Malignant
15	1033078	2	1	1	1	2	1	1	1	5	2	Benign
16	1057013	8	4	5	1	2	0	7	3	1	4	Malignant
17	1033078	4	2	1	1	2	1	1	1	1	2	Benign
18	1065726	5	2	3	4	2	7	3	6	1	4	Malignant
19	1035283	1	1	1	1	1	1	3	1	1	2	Benign
20	1072179	10	7	7	3	8	5	7	4	3	4	Malignant

The experimental data (Figure 3.1) served as the input base for BC-ANFIS. This model is an adopted ANFIS approach comprises of preprocessing, input, fuzzification, rule, normalization, defuzzification, (training; LSE/BPGD) output. Figure 2 depicts the BC-ANFIS.

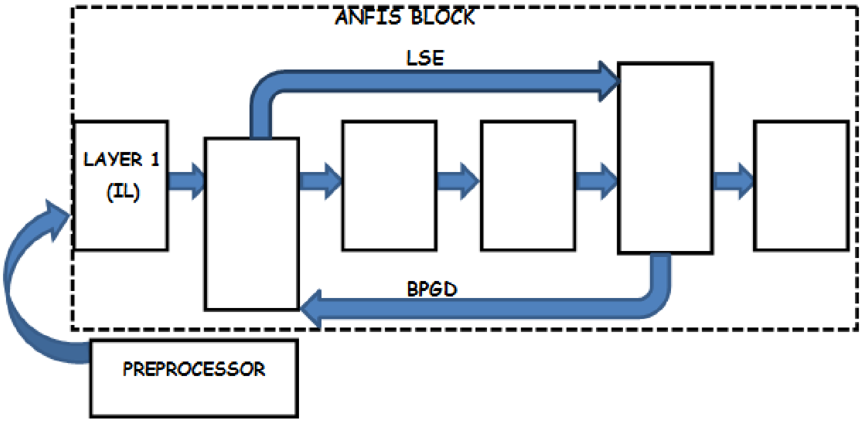

Figure 2: BC-ANFIS Model.

The BA-ANFIS model comprises of two main modules; Preprocessor block and the ANFIS block.

a. Preprocessor Block: The preprocessor block provides an avenue for preprocessing the crisp values ($x \geq 1$) into fuzzified noiseless experimental data ($x < 1$) prior to training. The preprocessing of the experimental data (Table 1) was handled using the equation 1 with diagnosis value the jogging up.

$$z' = \frac{z - min(u)}{max(u) - min(u)} \quad . \quad . \quad . \quad (Equation\ 1)$$

$Where: \ _{z}^{'}= Fuzzy\ \ Variable\ U\ in\ the\ ith\ case \qquad z = \ \ Variable\ U\ in\ the\ ith\ case.$

$min\ (u) = \ Minimum\ variable\ U\ in\ \ dataset \qquad \max(u) = \ Maximum\ variable\ U\ in\ dataset$

Table 2 shows the fuzzified noiseless experimental data.

Table 2: Preprocessed Fuzzified Wisconsin Breast Cancer Dataset

Cases	Dataset Code	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Classes of Diagnosis (2-Benign, 4-Malignant	Classification class
1	1000025	0.44	0	0	0	0.14	0	0.25	0	0	0.1	Benign
2	1017122	0.78	1	1	0.78	0.85	1	1	0.66	0	0.35	Malignant
3	1002945	0.44	0.33	0.33	0.44	0.85	1	0.25	0.11	0	0	Benign
4	1044572	0.78	0.66	0.44	1	0.85	0.89	0.5	0.44	0.75	0.25	Malignant
5	1015425	0.22	0	0	0	0.14	0.11	0.25	0	0	0.1	Benign
6	1041801	0.44	0.22	0.22	0.22	0.14	0.22	0.36	0.33	0	0.35	Malignant
7	1016277	0.55	0.78	0.77	0	0.29	0.33	0.25	0.66	0	0	Benign
8	1047630	0.66	0.33	0.55	0.33	0.71	0	0.36	0.22	0	0.25	Malignant
9	1017023	0.33	0	0	0.22	0.14	0	0.25	0	0	0.1	Benign
10	1050670	1	0.66	0.78	0.55	0.42	1	0.36	0	0.25	0.35	Malignant
11	1018099	0	0	0	0	0.14	1	0.25	0	0	0	Benign
12	1054590	0.78	0.22	0.11	1	0.57	1	0.36	0.33	0.75	0.25	Malignant
13	1018561	0.11	0	0.11	0	0.14	0	0.25	0	0	0.1	Benign
14	1054593	1	0.44	0.44	0.22	0.71	0.66	0.75	1	0	0.35	Malignant
15	1033078	0.11	0	0	0	0.14	0	0	0	0	0	Benign
16	1057013	0.77	0.33	0.44	0	0.14	0	0.75	0.22	0	0.25	Malignant
17	1033078	0.33	0.11	0	0	0.14	0	0	0	0	0.1	Benign
18	1065726	0.44	0.11	0.22	0.33	0.14	0.78	0.25	0.55	0	0.35	Malignant
19	1035283	0	0	0	0	0	0	0.25	0	0	0	Benign
20	1072179	1	0.78	0.78	0.22	1	0.44	0.75	0.33	0.5	0.25	Malignant

b. ANFIS Block: The ANFIS block comprises of six main layers. These layers include

i. Input Layer: Layer 1 (input layer) accept and transit the parameters of breast cancer tied to their associated values to next successive layer

ii. Membership Function Layer: Layer 2 (membership function) map various membership function to each input linguistic variables. The BC-ANFIS utilized Bell membership function due to the propensity for huge dataset. Equation 1 depicts the Bell Membership Function

$$\mu_{A_i}(x) = \frac{1}{1+[\frac{(x-c_i)^2}{a_i{}^2}]^{b_i}} \quad . \quad . \quad . \quad \text{Equation 2}$$

where {a, b, c} are called premise parameters.

iii. Rule Layer: Layer 3 (Rule layer) receives fuzzified range value from the input layer and calculates the firing strength of each rule. Equation 3 depicts the firing ANFIS strength.

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(x) \qquad i = 1,\ 2 \ . \quad . \quad . \quad \text{Equation 3}$$

iv. Normalization Layer: Layer 4 (Normalization layer) computes the normalized firing strength of each input signal coming from the rule layer. Equation 4 depicts the normalization firing strength.

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \qquad i = 1,\ 2 \ . \quad . \quad . \quad \text{Equation 4}$$

v. Defuzzification Layer: Layer 5 (Defuzzification layer) compute a parameter function on the normalized output forming the consequent parameters. Equation 5 highlight the defuzzification firing strength

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \ \{p_i, q_i, t_i\} \ . \quad . \quad . \quad \text{Equation 5}$$

vi. Output Layer: Layer 5 (Output layer) computes the overall values. This overall output is mapped to one of the two value "benign" or "malignant. The overall output is highlighted in Equation 6

$$O_{5,1} = overall\ output = \sum_i \overline{w}_i f_i \quad . \quad . \quad . \quad \text{Equation 6}$$

IV. Matrix Laboratory (MATLAB) Simulation

The BC-ANFIS was simulated using MATLAB taking cognizant of numerical data integration and open source capability of matlab. Seventy-five (75%: 15 cases) of experimental data was used for training while Twenty-Five (25%: 5 cases) was used for testing. Using generalized bell membership function the membership function was identified. Figure 3-8, depicts the ANFIS simulations.

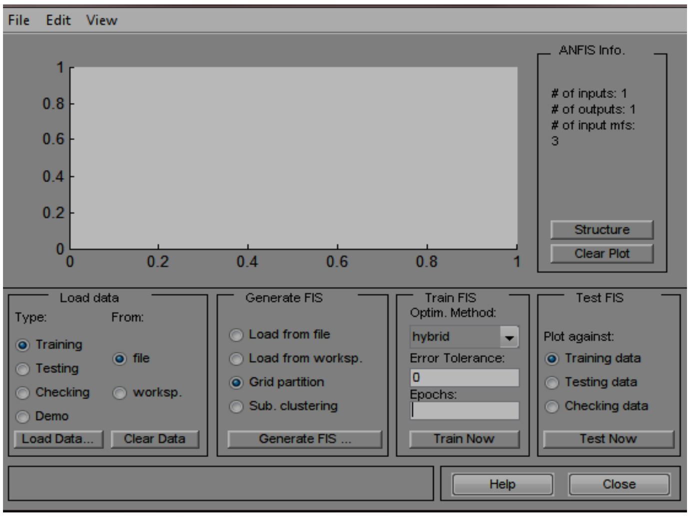

Figure 3: ANFIS training Editor

Figure 3, shows the ANFIS editor invoked with anfisedit command. It shows clearly the components of simulation.

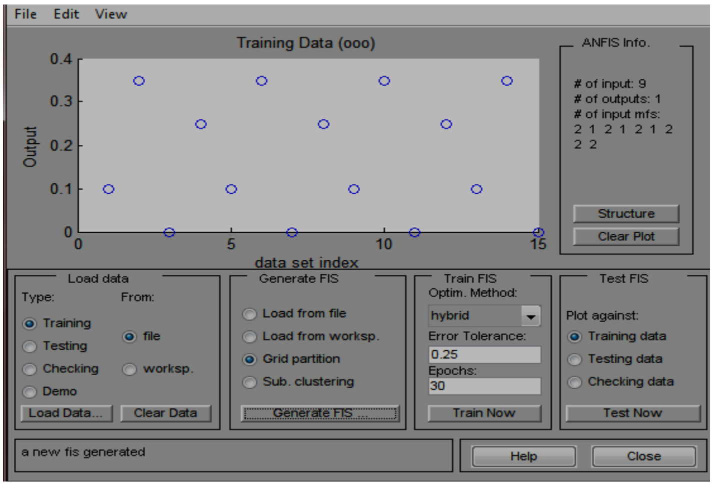

Figure 4: Training Data Editor

Figure 4, shows twenty experimental data portrayed across the editor. The training cases are dispersed across the graph plots showing training cases with varied patterns. Hybrid training was subsequently utilized in training the obtained data. At the far top right, the input is seen as nine (09) initial variables, permutated as 2 1 2 1 2 1 2 2 2 to producing 352 rules. This initial input produces a firing singleton output.

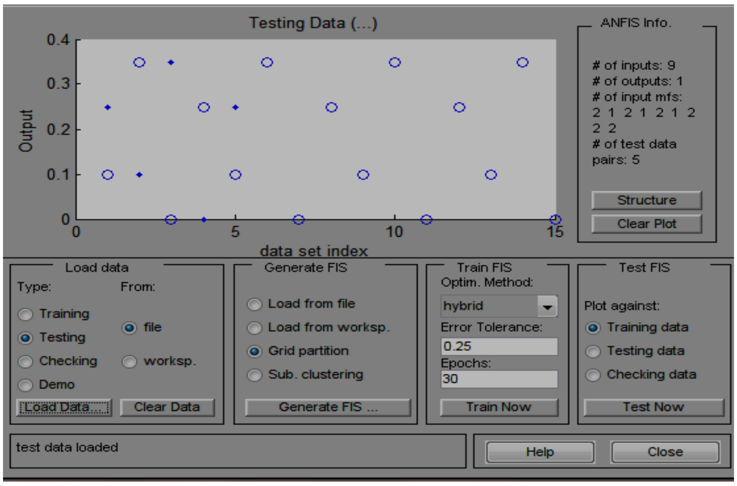

Figure 5: ANFIS Testing Data Editor

Figure 5, shows the testing error matched across the output plot with five (05) training data plot with an output of 0 to 0.4 with a dataset index of 0 to 15. The dotted plot matches the testing data referenced against the training data.

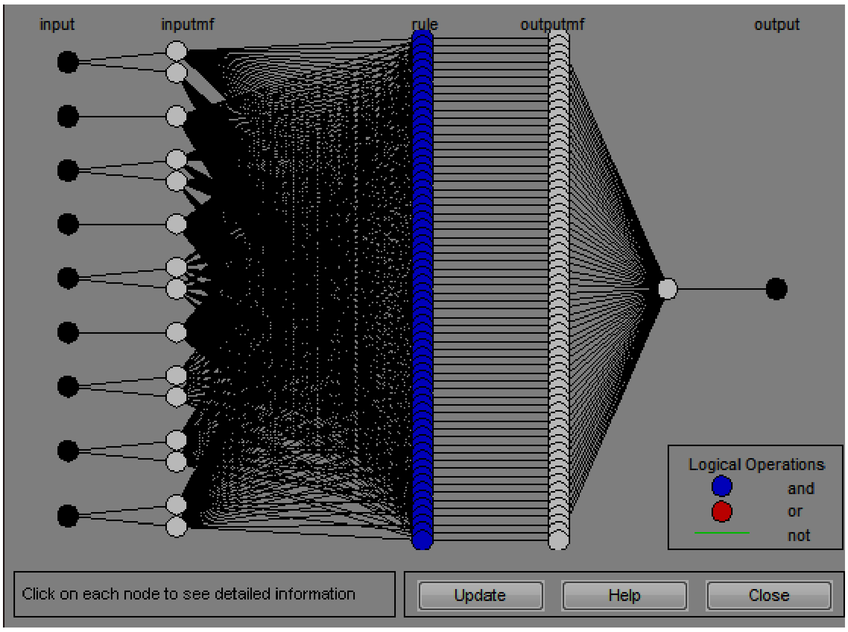

Figure 6: ANFIS Structure Editor

Figure 6, portray ANFIS structure encompasses all six layers, initiated with varied input parameters and terminates with an output. The nine integral inputs are captured on the far most left, succeed by membership layer, with rules layers succeeding the membership layer and rules precede the weighted normalization layer. The defuzzification initiated with middle of maximum, being a sugeno type inference system is captured within the fifth layer while the singleton output is seen as layer six been the far most right node (black).

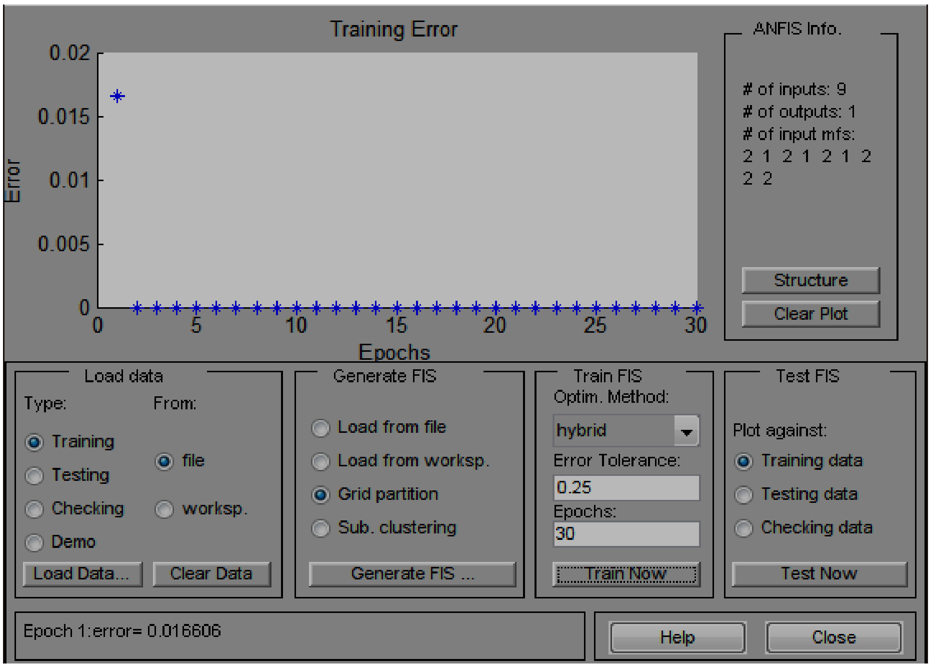

Figure 7: ANFIS Training Error Editor

Figure 7, portray the ANFIS training accommodated both across the horizontal and vertical bar of the editor which identify the training error. The training was initially set with epoch of 30 and tolerance error of 0.25. Overall training was attained at epoch 1 with 0.016606

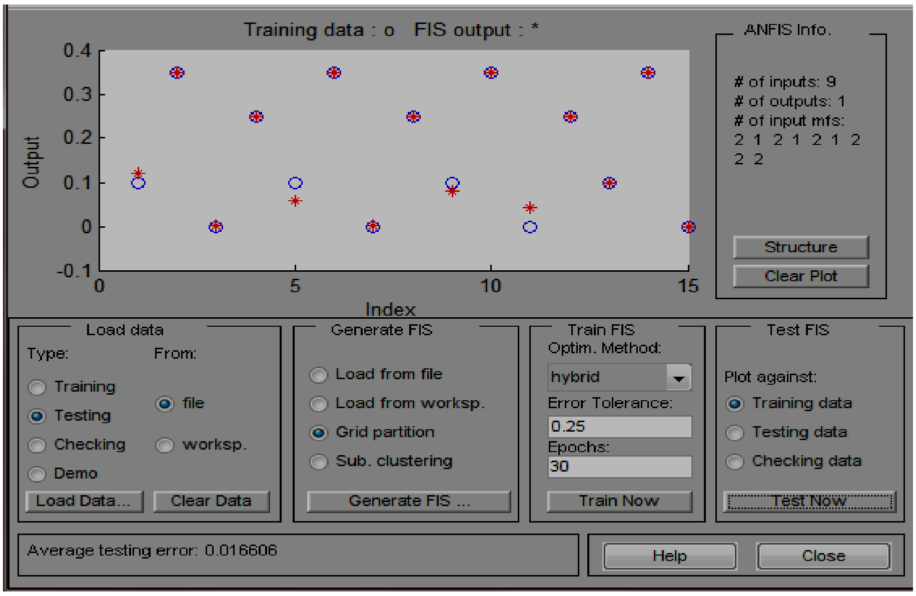

Figure 8: ANFIS Training and Testing Editor

Figure 8; portray the testing and training data matched across the output plot with 20 training data accounted for. The alignment of each testing data with training data highlights testing accuracy. The average accuracy testing error is 0.016606.

V. Matrix Laboratory (MATLAB) Result Analysis

The simulations captured with the aforementioned figures produce these results; Number of nodes: 170 (100%), Number of linear parameters: 64 (38%), Number of nonlinear parameters: 45 (26%), Total number of parameters: 109 (64%) and Number of training data pairs: 15 (9%). This result was set with a percentage range of 100% using the formula X/y * 100(where x, stands for expression value, y for the highest column expression value and 100, the benchmark percentage score. The graph of Figure 9 exemplify this values

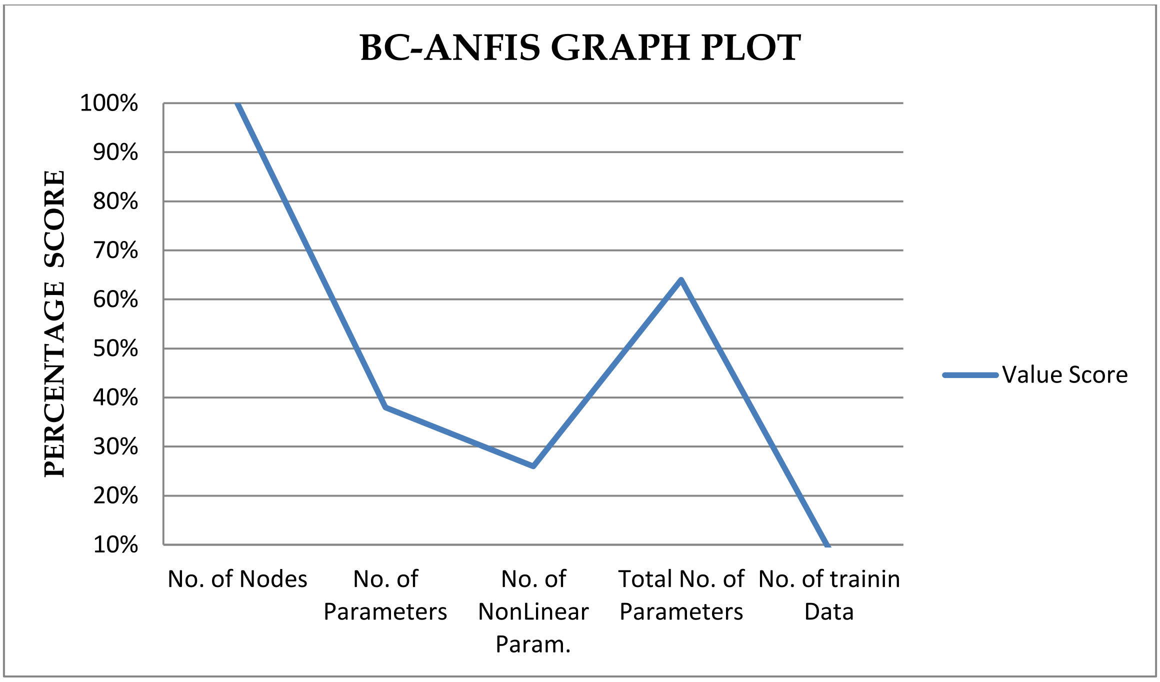

Figure 9: BC-ANFIS graph plot

Figure 9, captures the graph plot for MATLAB BC-ANFIS simulation. The graph plot with parameters: no. of nodes, no of parameters, no. of nonlinear parameter, total no of parameter and no. of training data. The graph show clearly in parentage that number of nodes from the first layer to the last layer had the highest percentage of 100%, followed by the total number of parameter equating to over 60%, succeeded by no. of parameters with over 35%, succeeded by no. of nonlinear parameters with over 25% and number of training data having the lowest score of about 9%. Therefore from the graph we can identify no of nodes as the most predominant value score.

VI. Conclusion

The simulation of breast cancer experimental data using Adaptive Neuro Fuzzy Inference System (BC-ANFIS) has been explored with graphical plot exemplified in this paper using Matrix Laboratory (MATLAB). The MATLAB ANFIS editors captured the training, testing, anfis structure and training error. This editor captures the integral neuro-fuzzy simulations while the graphical representation identified the paramount import of these parameters in terms of these simulations.

References

[1]. CancerQuest, (2016), Cancer Detection and Diagnosis, retrieved online from https://www.cancerquest.org/patients/detection-and diagnosis?gclid=
[2]. MedicineNet, (2016), Breast Cancer, retrieved online from http://www.medicinenet.com/breast_cancer_facts_stages/article.htm
[3]. Oncolex (2016), Breast Cancer symptoms, retrieved from http://oncolex.org/Breast-cancer/Background/Symptoms?gclid=CIC82KqtzNQCFZup7QodILEKlg
[4]. WrongDiagnosis (2011), "Breast cancer: Introduction/Causes and Symptoms", retrieved from http://www.wrongdiagnosis.com

Ejiofor C. I. "Adaptive Neuro Fuzzy Inference System (ANFIS): MATLAB Simulation of Breast Cancer Experimental Data." IOSR Journal of Computer Engineering (IOSR-JCE), vol. 19, no. 4, 2017, pp. 53–60.