A Novel Iterative Linear Regression Perceptron Classifier for Breast Cancer Prediction

Samuel Giftson Durai
Research Scholar, Dept. of CS
Bishop Heber College
Trichy-17, India

S. Hari Ganesh, PhD
Assistant Professor, Dept. of CS
H.H. The Rajah's College
Pudukottai -1, India

ABSTRACT

Breast cancer, the most common of types of cancer that threatens human life more specifically women can be diagnosed with classification techniques of data mining. This work is an extension of earlier implementation of breast cancer analysis of the author through iterative linear regressive classifier. The objective of this study is to make cent percent prediction accuracy in the diagnosis of breast cancer over the traditional Wisconsin dataset. The novelty of the paper includes the benefits of the previous ILRC and also takes the advantages of AI. The results of the proposed work are evaluated against the randmeasure and have proven that the results yield cent percent prediction accuracy in diagnosing breast cancer.

Keywords

Regression, perceptron, classification, data mining, linear functions

1. INTRODUCTION

Breast cancer, is a peculiar type of cancer that mostly affects women than any other human beings which is caused by two major factors called modifiable or non-modifiable. Modifiable factors are those that can be controlled like habitual and environmental issues. Non-modifiable factors are those that cannot be controlled like gender and family history [1]. According to a survey, 1 out of 28 women in India is prone to breast cancer as the early detection techniques on the presence of breast cancer are still lacking in the exact prediction of disease. Moreover, the lack of awareness, proactive measures and treatment facilities increases the risks of survival. Early detection of the syndromes may direct to overcome the breast cancer through appropriate treatment.

Many scientific researches have been proposed for the automatic prediction of breast cancer. Data mining is yet another scientific approach that encompasses numerous techniques and algorithms for analysing the hidden knowledge from large data sources [2]. Classification techniques of data mining acts a predictor of final results by processing the class labels of a dataset. The primary goal of classification algorithms is to build a knowledge prediction model from training data whose class labels are known and the model is then used to identify the class label of test data. Prediction of breast cancer is under the scope of classification algorithms as it categorizes the instances based on the factors that influence the disease. Different methods of classification use different types of algorithms in which the classification steps are varied from one another. Thus, not all algorithms provide the same classification accuracy and may have some advantages and disadvantages of its own. The prediction of breast cancer is either be 'benign' (non-cancerous) or 'malignant (cancerous)' instances based on the values of the attributes [3]. The objective of this research work is to implement two novel classification algorithms for predicting breast cancer with the following scopes:

- To propose classification algorithms that correctly classifies the benign and malignant instances of breast cancer
- To implement the proposed classification algorithms as software programs
- To validate the proposed classification algorithms theoretically and empirically
- To compare the proposed classification algorithms with state of the art breast cancer prediction algorithms in terms of accuracy

2. REVIEW OF LITERATURE

Venkatesan et al. [4] have analyzed the breast cancer data using four classification algorithms namely j48, Classification and Regression Trees (CART), Alternating Decision Tree (AD Tree), and Best First Tree (BF Tree). The authors have conducted the experiment through Weka tool. The classifier has applied for two test beds –cross validation which uses 10 folds with 9 folds used for training each classifier and 1 fold is used for testing and percentage split uses 2/3 of the dataset for training and 1/3 of the dataset for testing. The authors have claimed that the decision trees have a standard construct and easy to understand from which the rules can be extracted. The authors have also stated that j48 classifier has the highest accuracy with 99%.

Williams et al. [1] have focused at two data mining techniques namely naïve bayes and j48 decision trees to predict breast cancer risks in Nigerian patients. The analysis is made to determine the most efficient and effective model. The authors have collected the dataset from cancer registry of LASUTH, Ikeja in Lagos, Nigeria which contains 69 instances with 17 attributes along with the class label. The dataset holds 11 non-modifiable factors and five modifiable factors. The experiment is conducted through Weka and the authors have claimed j48 decision tree is better for the prediction of breast cancer risks with the values of accuracy (94.2%), precision, recall and error rates.

Majali et al. [5] have presented a diagnostic system using classification and association approach in data mining. The authors have used Frequent Pattern (FP) in association rule mining for classifying the patterns that are frequently found with benign and malignant instances. The authors have also used decision tree algorithm for predicting the possibility of cancer with respect to age. The authors have implemented Fp-growth algorithm for generating the frequent itemset without candidate generation which improves the performance of algorithm. The authors have claimed that their algorithm is able to achieve 94% of prediction accuracy.

Sivakami [6] has presented a disease status prediction model by employing a hybrid methodology of Decision Trees (DT) and Support Vector Machines (SVM). To alarm the severity of the disease the strategy of the system consists of two main parts namely information treatment and option extraction, and decision tree- support vector machines. The author has compared the results of the proposed model with Instance-based Learning (IBL), Sequential Minimal Optimization (SMO) and Naïve Bayes (NB) and has proven that proposed algorithms works better than the comparative algorithms with 91% of accuracy.

Sumbaly et al. [7] have discussed j48 decision tree classification algorithm for breast cancer diagnosis along with the summarization on the types of breast cancer, risk factors, disease symptoms and treatment. The authors have proven that the j48 algorithm is able to produce 94.5% of accuracy with correctly classified instances and have also suggested that neural network and digital mammography would be the alternative approaches for breast cancer prediction.

Thein et al. [8] have proposed an approach for distinguishing the classes of breast cancer through neural network. The authors have overcome the local optima issue of neural network differential evolution algorithm for determining the optimal value or near optimal value for ANN parameters. To overcome the issue longer training time and lower classification, the differential evolution algorithm is further collaborated with island based model. The island based model improves the accuracy and takes less training time by making an analysis between two different migration topologies.

3. METHODOLOGY

The proposed ILRPC (Iterative Linear Regressive Perceptron Classifier) combines the iterative linear regression algorithm with the concepts of perceptrons in neural networks so as to enhance the accuracy of breast cancer prediction. The contribution of linear regression models in feature selection is enormous for extracting the linear independent variable that influences the class label. Linear regressive classifier also found to be useful in predicting the class label of the given set of instances by computing the regression co-efficient of each attributes in the dataset. If the correlation between the dependent and independent variables are high, the linear regression models provide the best accuracy than the other classifiers else the transformation of attribute values will be meaningful for regression classification. [9] proposed a linear regressive classifier, where the data instances are transformed into linear form by multiplying the data values with regression – coefficients. The values are then summed, rounded up and compared with the class label. The results are then analyzed to find out the accuracy of the correctly classified instances. The result of linear regression model is able to achieve 99% of accuracy in predicting breast cancer. Thus, the linear regression model has been claimed as best model for breast cancer prediction. Still, the improvement of accuracy in linear regression model can be increased by iterating the model for n times to strengthen the linear relationship between the independent and dependent variables. As a consequence, this paper proposes an iterative linear regressive classifier for classifying the patients of breast cancer. The flow diagram of the proposed work is shown in Figure.

3.1. Multiple Linear Regression Model

Linear models act as a base for many machine learning algorithms. Linear regression technique, models one or more input variables that forms a linear relationship that fit into a line with respect to the output variable [10]. The linear regression model employs the least-squares for calculating the best-fitting line with the sample set that tries to minimize the sum of squares of the vertical distributions from each data object in the line. The model is then verified with the test data, for predicting class attribute by the line intercept values of each object. The advantage of using linear regressive classifier is its ability to classify multiple classes within the datasets. A linear regression model with two variables can be denoted as shown in Equation 1.

$$Y=a+bX, \qquad \ldots (1)$$

where ‘X’ is the input variable and ‘Y’ is the output variable. The slope of line is ‘b’, and ‘a’ is the line intercept. The slope can be calculated as shown in Equation 2.

$$b=r.Sy/Sx \qquad \ldots (2)$$

and the intercept can be calculated as shown in Equation 3.

$$a= My-bMx \qquad \ldots (3)$$

The proposed work has also encompasses the iterative multiple linear regression analysis which is an extension of simple linear regression analysis for assessing the association between two or more input variables and a single output variable. The multiple linear regression equation is as Equation 4.

$$Y=a+b_1X_1+b_2X_2+\ldots b_pX_p \qquad \ldots (4)$$

Where ‘Y’ is the output variable, X1to Xp are input variables ‘a’ is the value of ‘Y’ when all input variables are equal to zero, and b1 to bp are the estimated regression coefficients.

3.2. Multilayer Perceptron

A multilayer perceptron is a feed forward ANN model that maps the input data onto a set of appropriate outputs [11]. MLP can be represented as a directed graph where the layers involved in the models are fully connected to the next consecutive layers. The nodes except the input is a neuron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function [12]. The mathematical notation of the MLP is denoted in Equation 5.

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) \qquad \ldots (5)$$

Where
w denotes the weights for each attribute
x is the input
b is the bias and
φis the activation function.

Since the iterative transformation of breast cancer dataset has brought a strong linear relationship of independent variables to dependent variable, the dataset can further be applied onto MLP algorithm for enhancing the prediction accuracy of the disease by utilizing the weights that are retrieved by the neurons.

4. EXPERIMENTATION AND RESULT DISCUSSIONS

An experiment is conducted to evaluate the performance of the proposed iterative linear regressive perceptron classifier over breast cancer dataset. The dataset is taken from the UCI machine learning repository [13]. The dataset is created by Dr. William H. Wolberg, University of Wisconsin Hospitals, USA. Though the dataset contains 699 data points with 11 attributes, only 683 data points are taken into the consideration as the remaining data holds the missing values. The attribute that specifies the sample code number is also removed for the experiment. The description of the dataset is given in table.1.

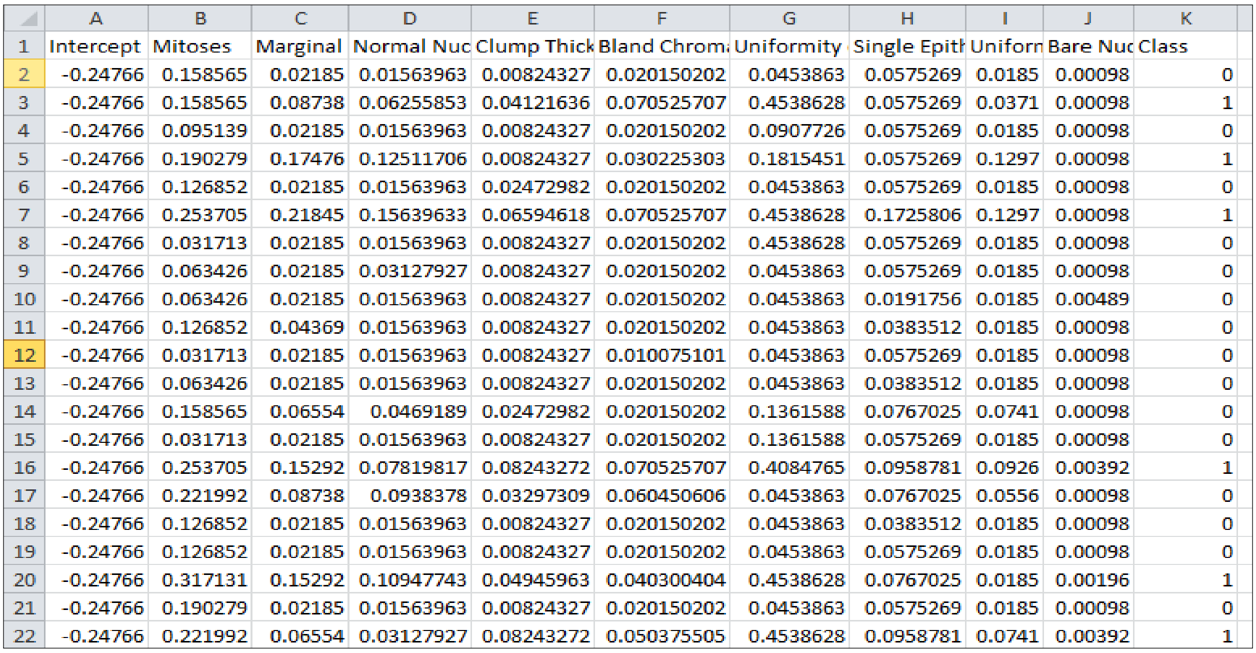
The highest correlation of iterative linear transformation of breast cancer is achieved in the second iteration with 0.94565977. Thus, the dataset that is transformed twice with the regression algorithm is pertained to perceptron algorithm.

Figure 1 depicts the iterative linear transformed breast cancer dataset.

Figure 2.shows the iterative linear breast cancer data into matlab 7.0. The dataset is loaded into open selection mode.

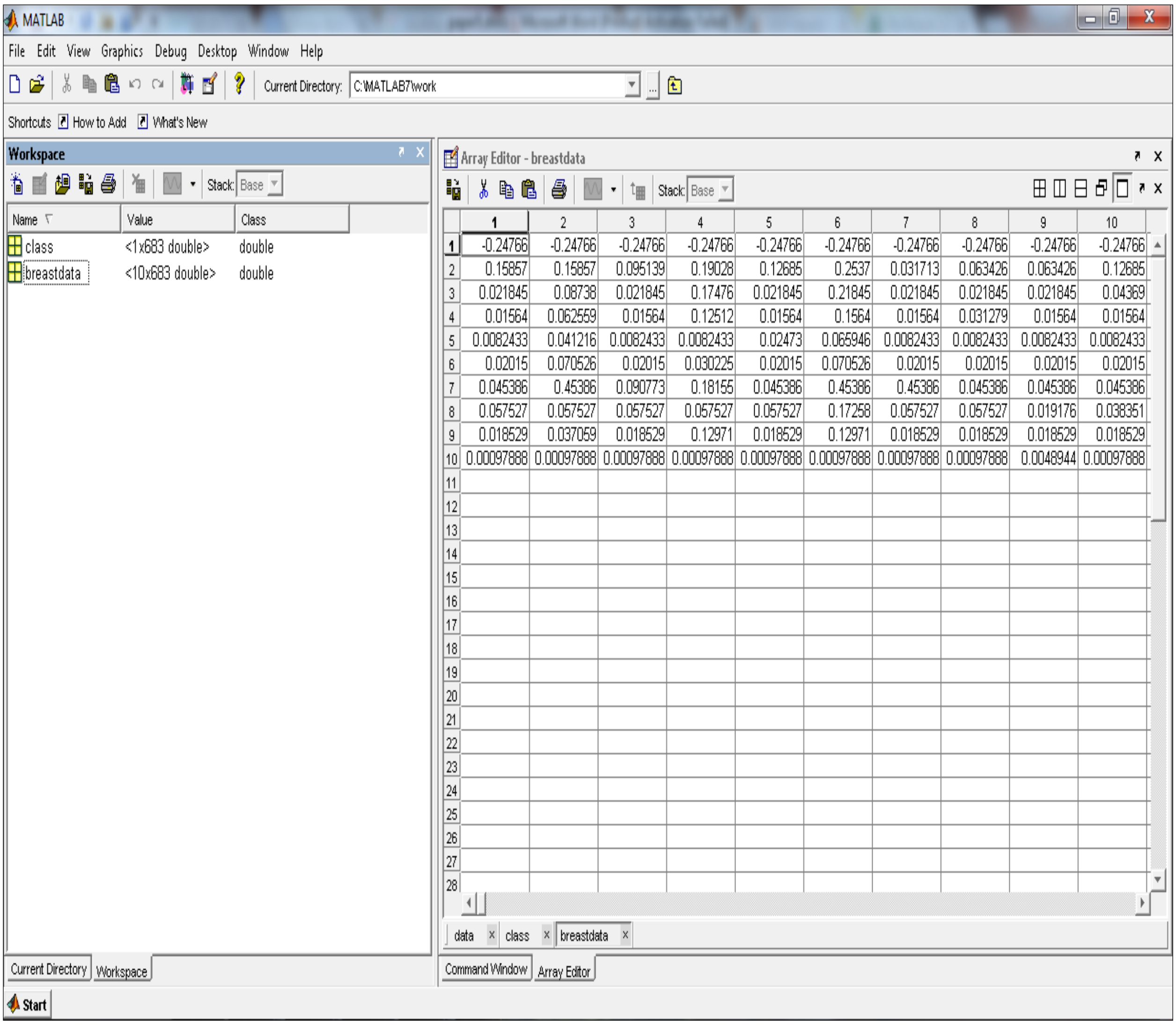
Table 1. Breast Cancer Dataset Description

S. No	**Attribute Name**	**Description**	**Range**
1	Mitoses	Describes the level of mitotic activity.	1-10
2	Marginal Adhesion	Quantifies how much cells are stick together on the outside of the epithelial.	1-10
3	Normal Nucleoli	Determines whether the nucleoli are small and hardly visible or larger, more visible, and more plentiful.	1-10
4	Clump Thickness	Assesses if cells are mono- or multi-layered.	1-10
5	Bland Chromatin	Rates the uniform "texture" of the nucleus in a range from fine to coarse.	1-10
6	Uniformity of Cell Shape	Estimates cell shapes equality and identifies marginal variances.	1-10
7	Single Epithelial Cell Size	Relates to cell uniformity, determines if epithelial cells are significantly enlarged.	1-10
8	Uniformity of Cell Size	Evaluates the consistency of sample cells size	1-10
9	Bare Nuclei	Calculates the ratio of the number of cells that are not surrounded by cytoplasm.	1-10



	A	B	C	D	E	F	G	H	I	J	K
1	Intercept	Mitoses	Marginal	Normal Nuc	Clump Thick	Bland Chroma	Uniformity	Single Epitl	Uniforn	Bare Nuc	Class
2	-0.24766	0.158565	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0575269	0.0185	0.00098	0
3	-0.24766	0.158565	0.08738	0.06255853	0.04121636	0.070525707	0.4538628	0.0575269	0.0371	0.00098	1
4	-0.24766	0.095139	0.02185	0.01563963	0.00824327	0.020150202	0.0907726	0.0575269	0.0185	0.00098	0
5	-0.24766	0.190279	0.17476	0.12511706	0.00824327	0.030225303	0.1815451	0.0575269	0.1297	0.00098	1
6	-0.24766	0.126852	0.02185	0.01563963	0.02472982	0.020150202	0.0453863	0.0575269	0.0185	0.00098	0
7	-0.24766	0.253705	0.21845	0.15639633	0.06594618	0.070525707	0.4538628	0.1725806	0.1297	0.00098	1
8	-0.24766	0.031713	0.02185	0.01563963	0.00824327	0.020150202	0.4538628	0.0575269	0.0185	0.00098	0
9	-0.24766	0.063426	0.02185	0.03127927	0.00824327	0.020150202	0.0453863	0.0575269	0.0185	0.00098	0
10	-0.24766	0.063426	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0191756	0.0185	0.00489	0
11	-0.24766	0.126852	0.04369	0.01563963	0.00824327	0.020150202	0.0453863	0.0383512	0.0185	0.00098	0
12	-0.24766	0.031713	0.02185	0.01563963	0.00824327	0.010075101	0.0453863	0.0575269	0.0185	0.00098	0
13	-0.24766	0.063426	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0383512	0.0185	0.00098	0
14	-0.24766	0.158565	0.06554	0.0469189	0.02472982	0.020150202	0.1361588	0.0767025	0.0741	0.00098	0
15	-0.24766	0.031713	0.02185	0.01563963	0.00824327	0.020150202	0.1361588	0.0575269	0.0185	0.00098	0
16	-0.24766	0.253705	0.15292	0.07819817	0.08243272	0.070525707	0.4084765	0.0958781	0.0926	0.00392	1
17	-0.24766	0.221992	0.08738	0.0938378	0.03297309	0.060450606	0.0453863	0.0767025	0.0556	0.00098	0
18	-0.24766	0.126852	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0383512	0.0185	0.00098	0
19	-0.24766	0.126852	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0575269	0.0185	0.00098	0
20	-0.24766	0.317131	0.15292	0.10947743	0.04945963	0.040300404	0.4538628	0.0767025	0.0185	0.00196	1
21	-0.24766	0.190279	0.02185	0.01563963	0.00824327	0.020150202	0.0453863	0.0575269	0.0185	0.00098	0
22	-0.24766	0.221992	0.06554	0.03127927	0.08243272	0.050375505	0.4538628	0.0958781	0.0741	0.00392	1

Fig 1: Iterative Linear Breast Cancer Dataset


MATLAB
File Edit View Graphics Debug Desktop Window Help
Current Directory: C:\MATLAB7\work
Shortcuts How to Add What's New
Workspace
Stack: Base
Array Editor - breastdata
Stack: Base
data class breastdata
Current Directory Workspace
Command Window Array Editor
Start

Name	Value	Class
class	<1x683 double>	double
breastdata	<10x683 double>	double

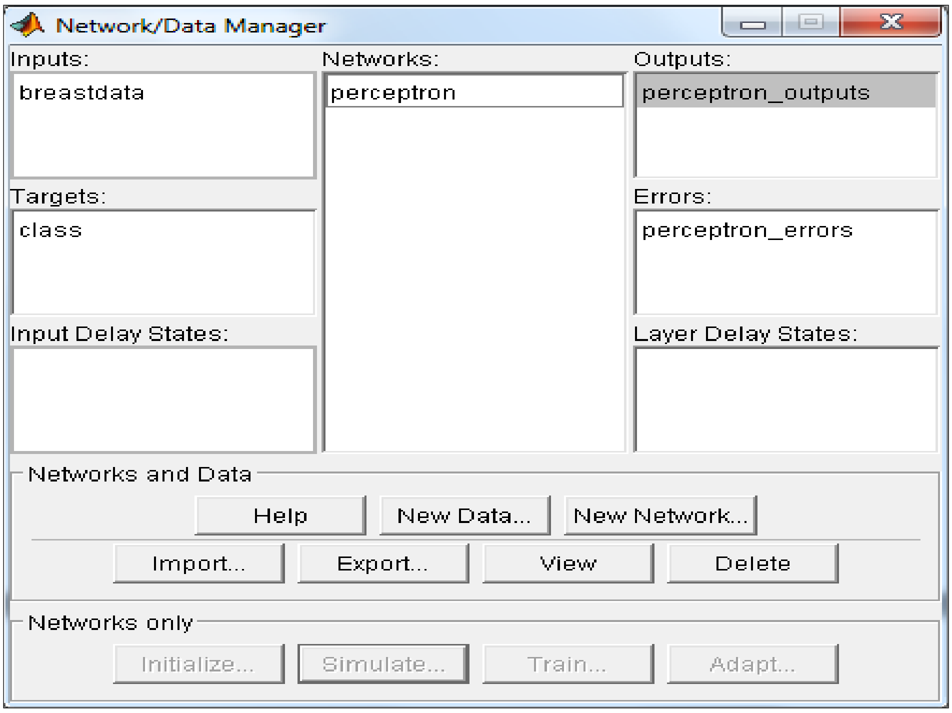
	1	2	3	4	5	6	7	8	9	10
1	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766	-0.24766
2	0.15857	0.15857	0.095139	0.19028	0.12685	0.2537	0.031713	0.063426	0.063426	0.12685
3	0.021845	0.08738	0.021845	0.17476	0.021845	0.21845	0.021845	0.021845	0.021845	0.04369
4	0.01564	0.062559	0.01564	0.12512	0.01564	0.1564	0.01564	0.031279	0.01564	0.01564
5	0.0082433	0.041216	0.0082433	0.0082433	0.02473	0.065946	0.0082433	0.0082433	0.0082433	0.0082433
6	0.02015	0.070526	0.02015	0.030225	0.02015	0.070526	0.02015	0.02015	0.02015	0.02015
7	0.045386	0.45386	0.090773	0.18155	0.045386	0.45386	0.45386	0.045386	0.045386	0.045386
8	0.057527	0.057527	0.057527	0.057527	0.057527	0.17258	0.057527	0.057527	0.019176	0.038351
9	0.018529	0.037059	0.018529	0.12971	0.018529	0.12971	0.018529	0.018529	0.018529	0.018529
10	0.00097888	0.00097888	0.00097888	0.00097888	0.00097888	0.00097888	0.00097888	0.00097888	0.0048944	0.00097888

Fig 2: Loading of Iterative Linear Breast Cancer Dataset

Once when the iterative linear breast cancer dataset is successfully loaded, the NN Tool of the matlab is selected for executing the perceptron algorithm. The NN tool consists of a network manager for managing the input, target, networks and outputs of the neural network. As a first step, the iterative linear breast cancer dataset is imported as inputs and class attribute is imported as targets. Figure 3 denotes the outlook of Network Manager of NNtool.

NN tool consists of numerous neural network algorithms to be used for machine learning. Since, the objective of the paper is to incorporate the iterative linear breast cancer dataset with perceptron algorithm, the network type is selected as perceptron in the creation of new network. The input ranges denote the ranges of inputs. The number of neurons denotes the number of hidden layers which is given as 1 for the illustration. The hardlim (hard limit) transfer function forces a neuron to output a 1 if its net input reaches a threshold, otherwise it outputs 0. This allows a neuron to make a decision or classification. It can say yes or no. This kind of neuron is often trained with the perceptron learning rule [12]. The perceptron weight and bias is obtained by learnp learning function. Figure 4 denotes the creation of new perceptron network.

The perceptron network is then initialized, simulated, trained and adapted with the iterative linear breast cancer dataset to get the desired output. As a result the network will come out with the weights that can be processed by the linear breast cancer dataset for predicting the future outputs. Moreover, the output of perceptron network will result perceptron outputs and perceptron errors which denotes the class outputs of iterative linear breast cancer data and possible error occurred with every single class variable in the dataset. Figures 5 to 12 denotes the manipulation of perceptron network with respect to iterative linear breast cancer data.



Inputs: breastdata

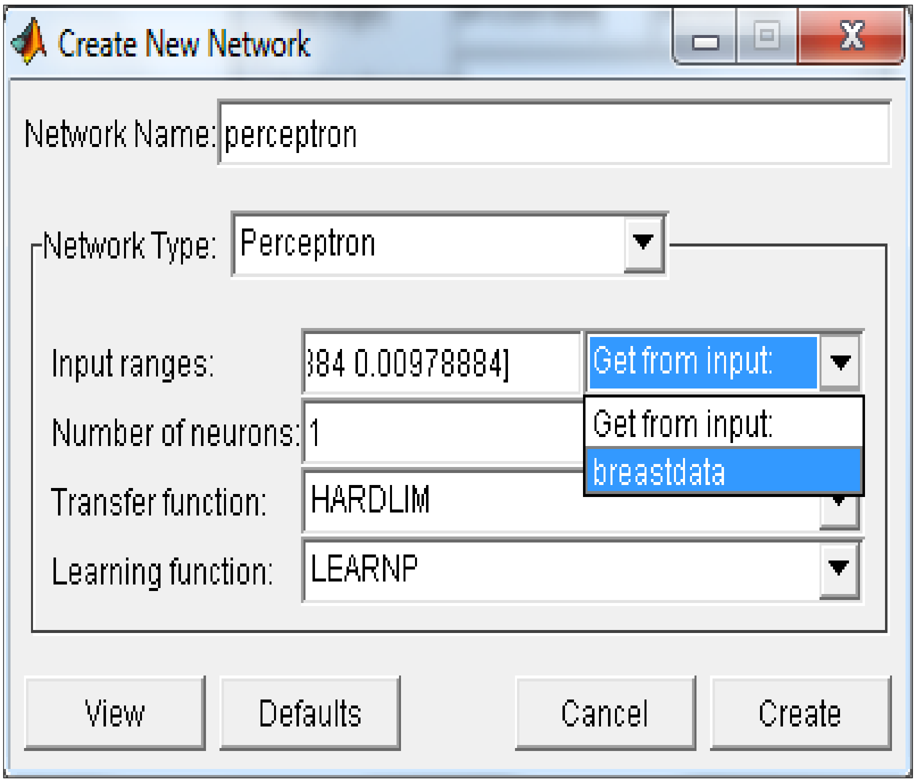
Targets: class

Networks: perceptron

Outputs: perceptron_outputs

Errors: perceptron_errors

Fig 3: NNTool Network Manager



Network Name: perceptron

Network Type: Perceptron

Input ranges: 384 0.00978884] Get from input:

Number of neurons: 1

Transfer function: HARDLIM

Learning function: LEARNP

Fig 4: Creation of New Perceptron Network

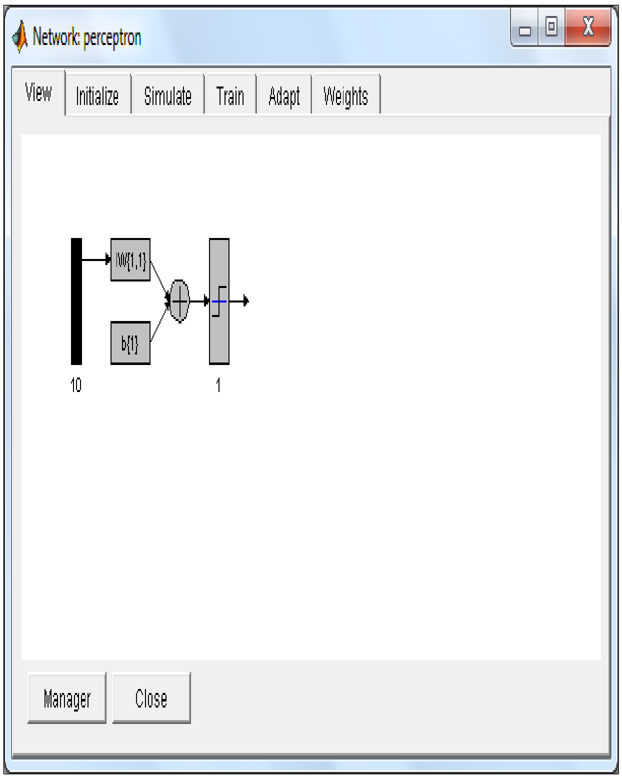
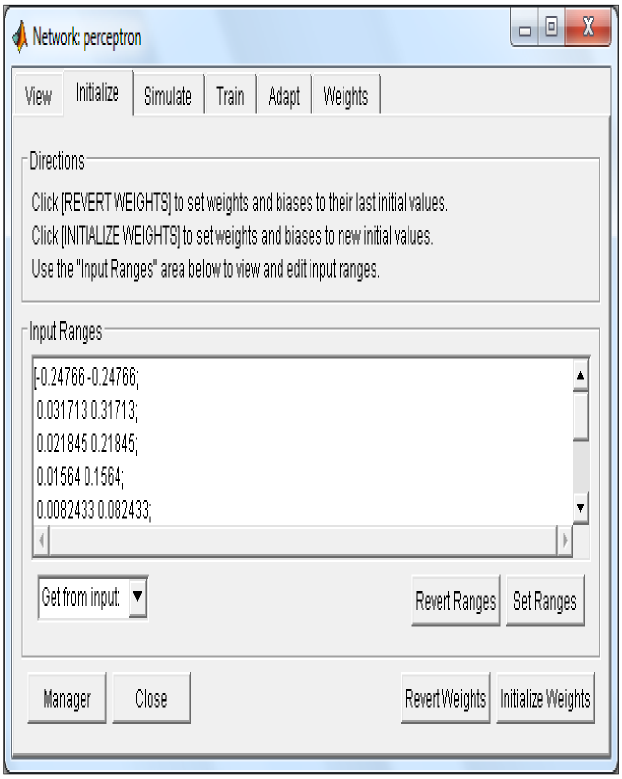


Fig 5:Perceptron Layers



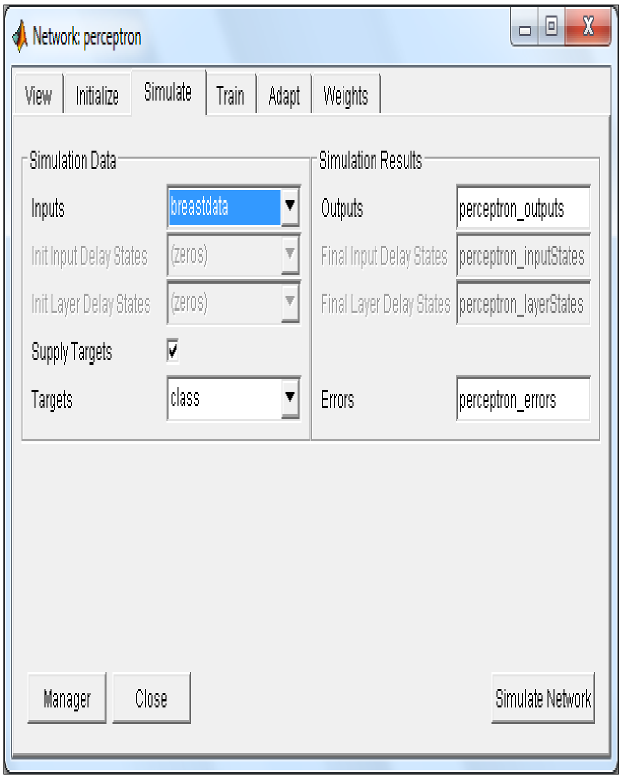
Directions

Click [REVERT WEIGHTS] to set weights and biases to their last initial values.
Click [INITIALIZE WEIGHTS] to set weights and biases to new initial values.
Use the "Input Ranges" area below to view and edit input ranges.

Input Ranges

```
[-0.24766 -0.24766;
0.031713 0.31713;
0.021845 0.21845;
0.01564 0.1564;
0.0082433 0.082433;
```

Fig 6:Perceptron Initialization



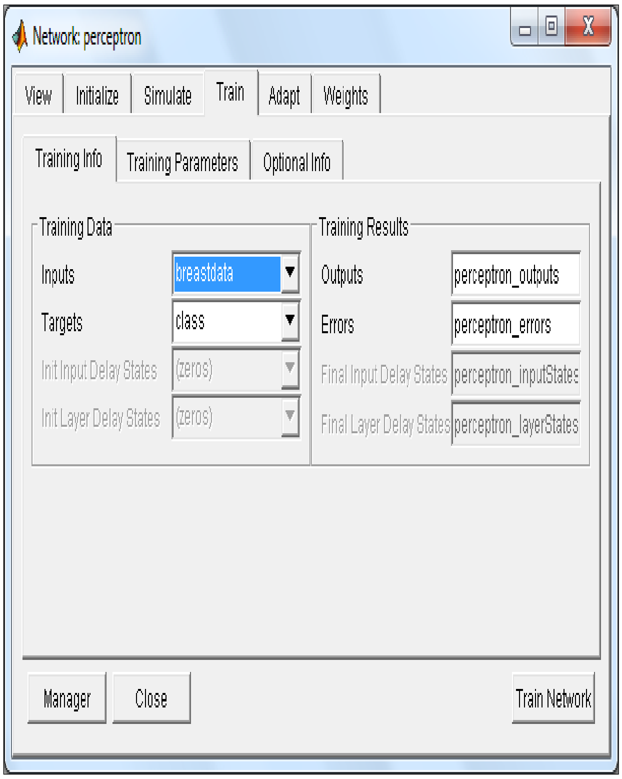
Simulation Data

Inputs: breastdata
Init Input Delay States: (zeros)
Init Layer Delay States: (zeros)
Supply Targets: ☑
Targets: class

Simulation Results

Outputs: perceptron_outputs
Final Input Delay States: perceptron_inputStates
Final Layer Delay States: perceptron_layerStates
Errors: perceptron_errors

Fig 7: Perceptron Simulation



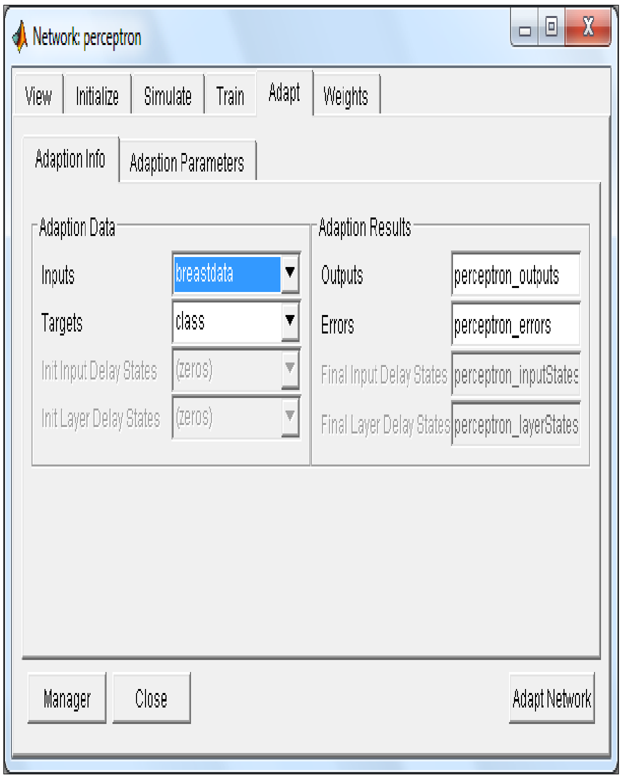
Training Data

Inputs: breastdata
Targets: class
Init Input Delay States: (zeros)
Init Layer Delay States: (zeros)

Training Results

Outputs: perceptron_outputs
Errors: perceptron_errors
Final Input Delay States: perceptron_inputStates
Final Layer Delay States: perceptron_layerStates

Fig 8: Perceptron Training



Adaption Data

Inputs: breastdata

Targets: class

Init Input Delay States: (zeros)

Init Layer Delay States: (zeros)

Adaption Results

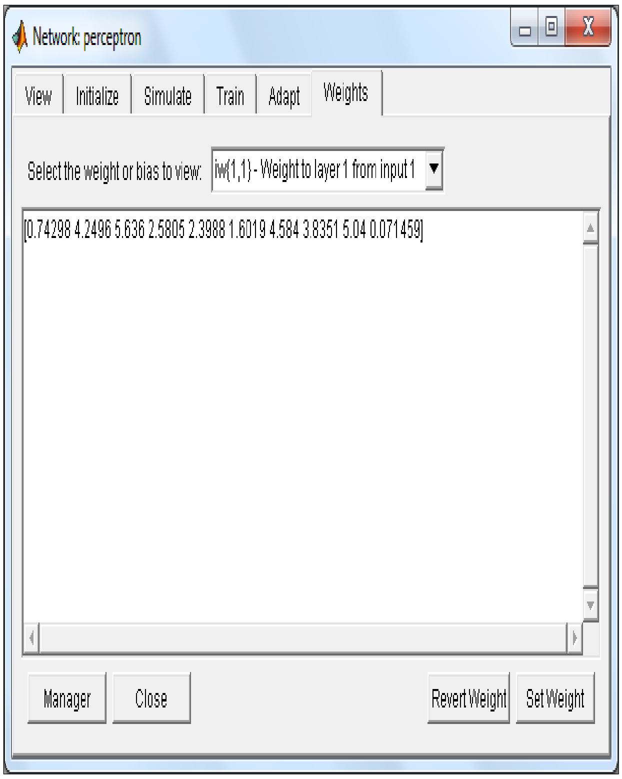
Outputs: perceptron_outputs

Errors: perceptron_errors

Final Input Delay States: perceptron_inputStates

Final Layer Delay States: perceptron_layerStates

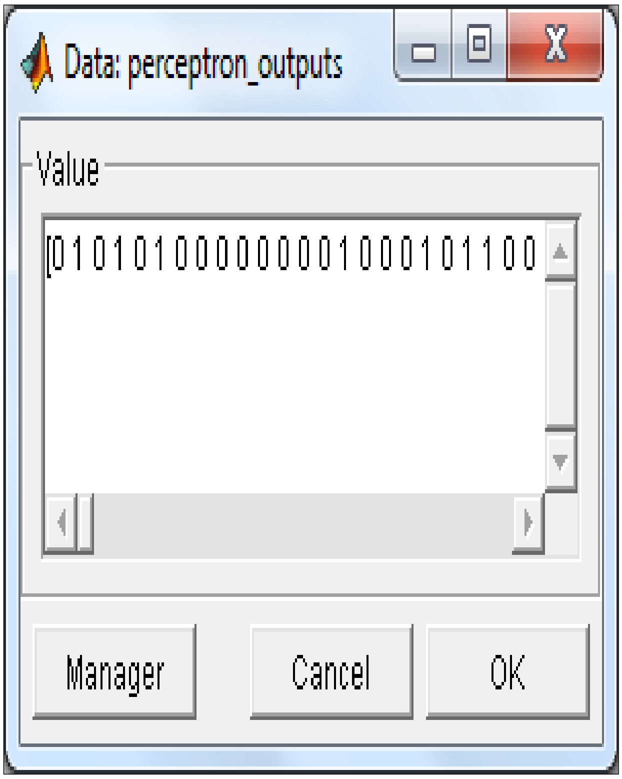
Fig 9: Perceptron Adaptation



Select the weight or bias to view: iw{1,1} - Weight to layer 1 from input 1

```
[0.74298 4.2496 5.636 2.5805 2.3988 1.6019 4.584 3.8351 5.04 0.071459]
```

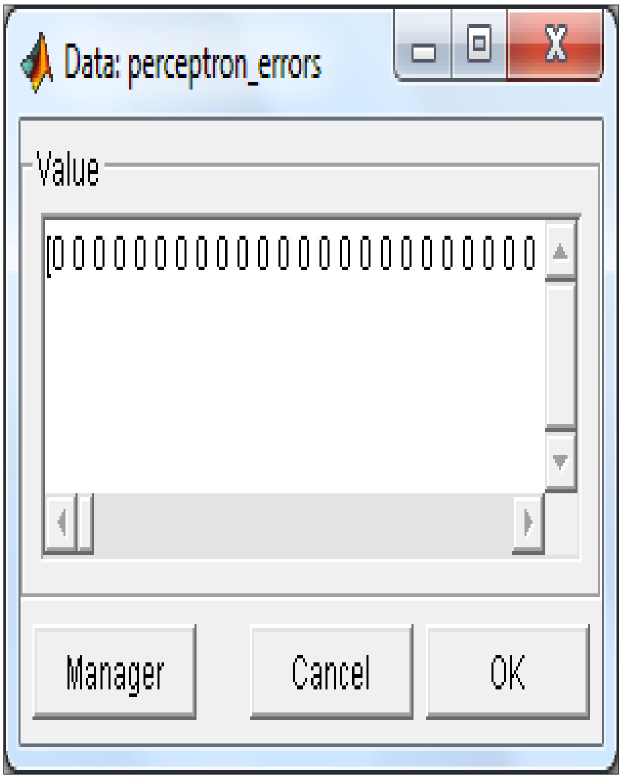
Fig 10: Perceptron Weights



Value

```
[0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 1 0 0
```

Fig 11: Perceptron Outputs



Value

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Fig 12: Perceptron errors

Since the error goal is reached to 0, the epochs of the perceptron network is stopped and has reached 0% error rate. The performance goal of the perceptron layer is shown in Figure 13.

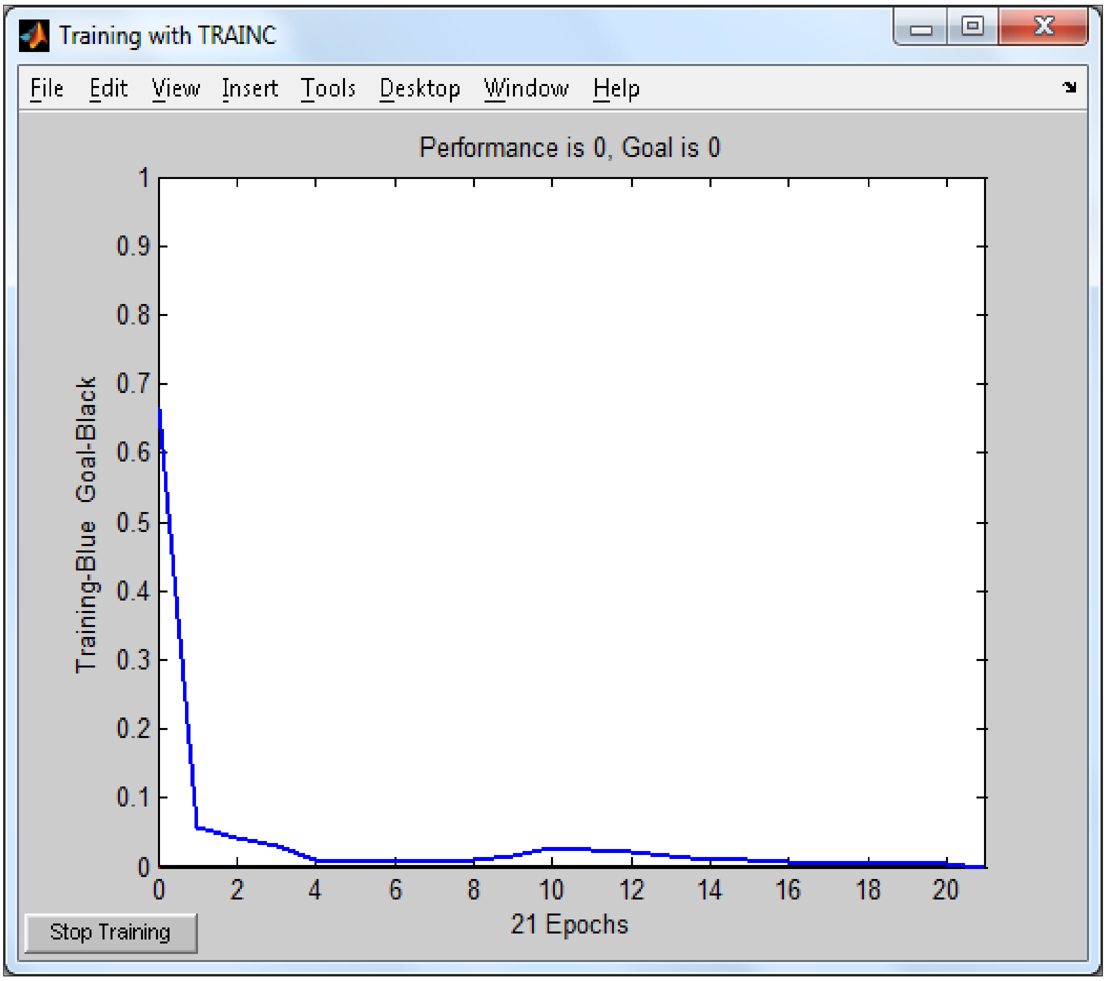


Fig 13: Performance goal of perceptron network

The prediction accuracy of the ILRPC is calculated using precision and recall metrics.

$$precision = \frac{true\ positive}{true\ positive\ + false\ positive} = \frac{228}{228+0} = 1$$

$$recall = \frac{true\ positive}{true\ positive + false\ negative} = \frac{228}{228+0} = 1$$

$$Accuracy = \frac{\sum true\ positive + \sum true\ negative}{\sum total\ population} = \frac{228+455}{683}$$
$$= 1$$

From the results, it has been observed that the prediction accuracy of the ILRPC is 1, which means the proposed work is able to diagnose the presence of breast cancer with 100% accuracy. Thus, the method has proven that it can be applied to breast cancer prediction. The prediction accuracy of the proposed classification algorithm is compared with the state of the art of classification algorithms and has shown in Table 2. The pictorial representation of accuracies obtained by conventional classification algorithms versus the proposed iterative linear regressive perceptron classifier is shown in Figure 14.

Table 2. Accuracy Analysis of the Proposed Classification Algorithm

S. No	Algorithm	Accuracy
1	BF Tree	95.46%
2	ID3	92.99%
3	J48	98.14%
4	SVM	96.4%
5	LRC	99.25%
6	ILRC	99.86%
7	ILRPC	100%

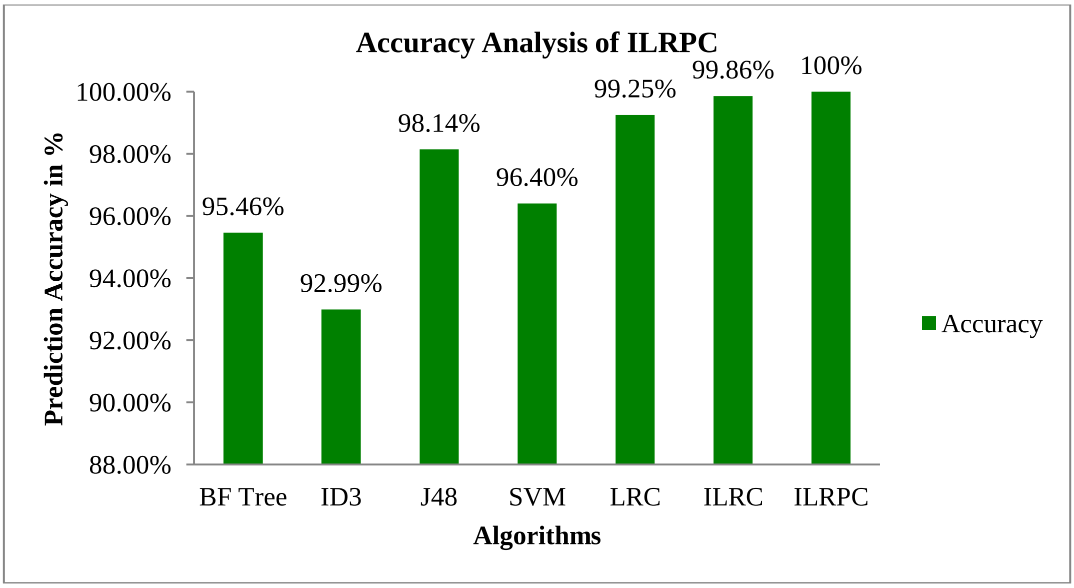


Fig 14: Accuracy Analysis of ILRPC

5. CONCLUSION

In this paper, a novel iterative linear regressive perceptron classifier is proposed to predict the breast cancerous patients. The prediction model proposed in this paper has proved to be the best model ever as it achieves a highest accuracy of 100% than the state –of-the-arts algorithms. The model employs an iterative approach to linear transformation of data using regression techniques and implements the processed dataset onto the perceptron algorithm which is the novelty of the paper. Thus, as a result the quality of the prediction is improved enormously by 100% prediction accuracy. Since, the iterations are incurred with the regression classifier, the time complexity may get increased. In future, the reduction of time complexity and the adaptability of the algorithm with various datasets have to be focused.

6. REFERENCES

[1] Williams, Kehinde, Peter Adebayo Idowu, Jeremiah AdemolaBalogun, and AdeniranIsholaOluwaranti. "Breast cancer risk prediction using data mining classification techniques." Transactions on Networks and Communications 3, no. 2 ,pp: 01-11, 2015.

[2] Joy Christy, S. Hari Ganesh, “Building Numerical Clusters Using Multidimensional Spherical Equation”, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 10, Issue No.82, pp:629-634, 2015.

[3] RadhanathPatra and ShankhaMitraSunan, “A Review on Different Computing Method for Breast Cancer Diagnosis Using Artificial Neural Network and Data mining Techniques.” International Journal of Advanced Research, ISSN: 2320-5407, pp: 598-610, 2016.

[4] Venkatesan, E., and T. Velmurugan. "Performance analysis of decision tree algorithms for breast cancer classification." Indian Journal of Science and Technology 8.29 , pp:1-8, 2015.

[5] Majali, Jaimini, et al. "Data Mining Techniques For Diagnosis And Prognosis Of Cancer." International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, pp:613-616, 2015

[6] Sivakami, K. "Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model." International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-5, pp:418-429, 2015

[7] Sumbaly, Ronak, N. Vishnusri, and S. Jeyalatha. "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique." International Journal of Computer Applications 98.10, pp:16-24, 2014.

[8] Thein, HtetThazinTike, and Khin Mo Mo Tun. "An Approach for Breast Cancer Diagnosis Classification Using Neural Network." Advanced Computing 6.1, pp: 1-10, 2015.

[9] Samuel Giftson, A. Joy Christy, S. Hari Ganesh. “Novel Linear Regressive Classifier for the Diagnosis of Breast Cancer”, 2nd World Congress on Computing and Communication Technologies (WCCCT 2016), St. Joseph’s College, 2nd and 3rd February 2017.

[10] Christy, A. Joy, and S. Hari Ganesh. "Linear Regressive Percentage Split Distribution Clustering." International Journal of Control Theory an Applications, ISSN: 0974-5572, Impact Factor: 1.891, Indexed in Scopus.

[11] https://www.hiit.fi/u/ahonkela/dippa/node41.html

[12] http://radio.feld.cvut.cz/matlab/toolbox/nnet/hardlim.html

[13] https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/