

# Prediction of Heart Diseases on the basis of the Cleveland Database

Dr. Kalpana Chansoria<sup>1</sup>

Sr. Lecturer, Government Polytechnic College, Mandla, Madhya Pradesh, India<sup>1</sup>

**ABSTRACT:** Angiography is the prime test to diagnose any heart disease. Neural Networks are statistical tool inspired from biological neural network which can be used to perform various computational tasks such as pattern recognition, predictions, function fitting etc. Proposed work tries to predict heart diseases i.e. Angiographic disease status using Artificial Neural Networks.

**KEYWORDS:** Angiography, Artificial Neural Networks, Heart Disease, Cleveland Database.

## I. INTRODUCTION

Cleveland data is a benchmark database for classification prepared by Cleveland Clinic Foundation. It provides various parameter related to heart anomalies and their Angiographic disease status. Artificial Neural Networks are statistical tools. Same as biological neural networks ANN has parameters as interconnecting weights denoting strength of neural connections by which it sums the weighted input and then passes through an activation function. This setup is inherently used for function fitting, pattern recognition, clustering, classification etc.

## II. RELATED WORK

In [3] Xiaoyong Chai and Li Deng and Qiang Yang and Charles X. Ling used dataset in the experiments and ran a 3-fold cross validation on these data sets. In [4] Gavin Brown shown An ensemble consisting of two networks, each with five hidden nodes, was trained using NC. [6] Jeroen Eggermont and Joost N. Kok and Walter A. Kusters finds decision trees that differ syntactically per fold and random seed. In [9] Shrikant Vyas modelled cognitive process for solving balance scale task using feed forward ANN.

## III. HEART DISEASES PREDICTION

In proposed work we have used feed forward neural network to classify different Angiographic diseases values. Angiography is a well known and broadly accepted methodology for detection of vulnerable conditions related to heart. For intended purpose dataset has been taken from the UCI machine learning repository. We have used batch processing ebpa algorithm to train our ANN. Matlab neural network tool box (nftool) is used to do the necessary classification task.

### 3.1 Cleveland Database

This database originally contains 76 attributes with 303 samples, but we have used a subset of 14 attributes. Used attributes are as follows

1. age: age in years
2. sex: sex (1 = male; 0 = female)
3. cp: chest pain type
  - Value 1: typical angina
  - Value 2: atypical angina
  - Value 3: non-anginal pain

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 9, September 2018

- Value 4: asymptomatic
- 4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
- 5. chol: serum cholesterol in mg/dl
- 6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- 7. restecg: resting electrocardiographic results
  - Value 0: normal
  - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- 8. thalach: maximum heart rate achieved
- 9. exang: exercise induced angina (1 = yes; 0 = no)
- 10. oldpeak = ST depression induced by exercise relative to rest
- 11. slope: the slope of the peak exercise ST segment
  - Value 1: upsloping
  - Value 2: flat
  - Value 3: downsloping
- 12. ca: number of major vessels (0-3) colored by flourosopy
- 13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
- 14. num: diagnosis of heart disease (angiographic disease status)
  - Value 0: < 50% diameter narrowing
  - Value 1: > 50% diameter narrowing

The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

### 3.2 CLASSIFICATION

We have used Matlab nftool tool box for classification of above data set. This is a two class classification problem 0 for no disease and 1 for presence of heart diseases. We have used 'ONE HOT' notation to represent our classes (viz. (0,1) for class 0, and (1,0) for class 1). After repeating experiment for various combination of number of neurons at hidden layer most efficient structure is found having 8 hidden layer neuron as shown in fig.1.

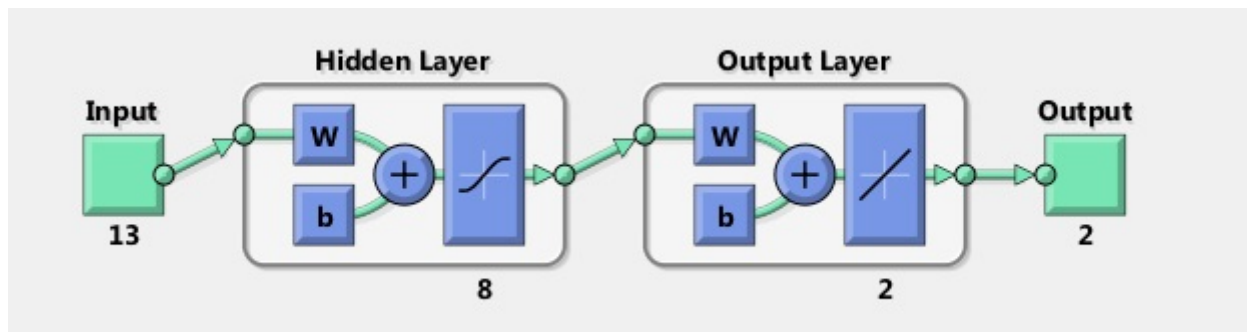


fig1: Proposed Structure of Artificial Neural Network

### IV. EXPERIMENTAL RESULTS

MATLAB version R2013a is used in our experiment. The Cleveland Database (downloaded from the UCI repository, [www.ics.uci.edu](http://www.ics.uci.edu)) has 303 X 75 matrix. After removing rows with missing values a 297 X 13 matrix is

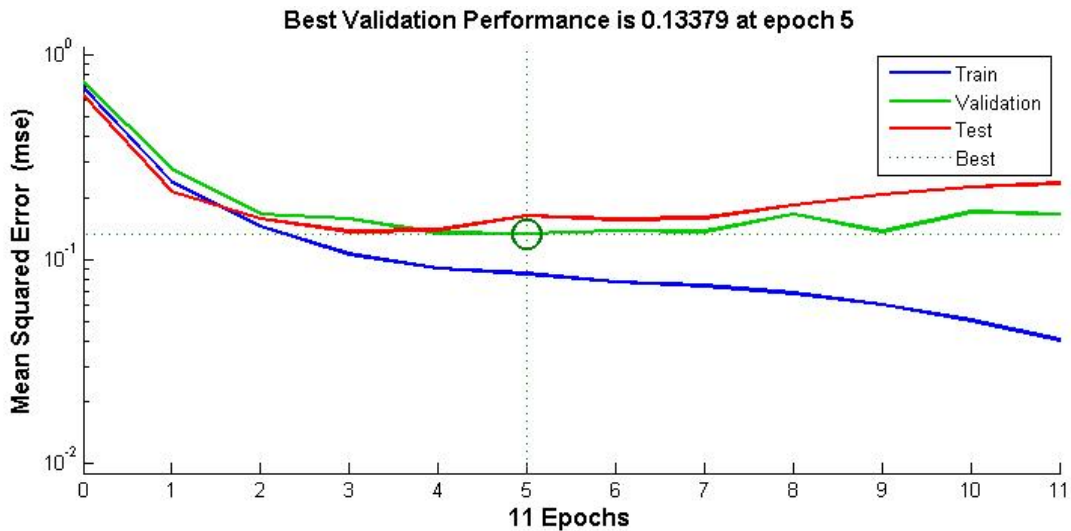
# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

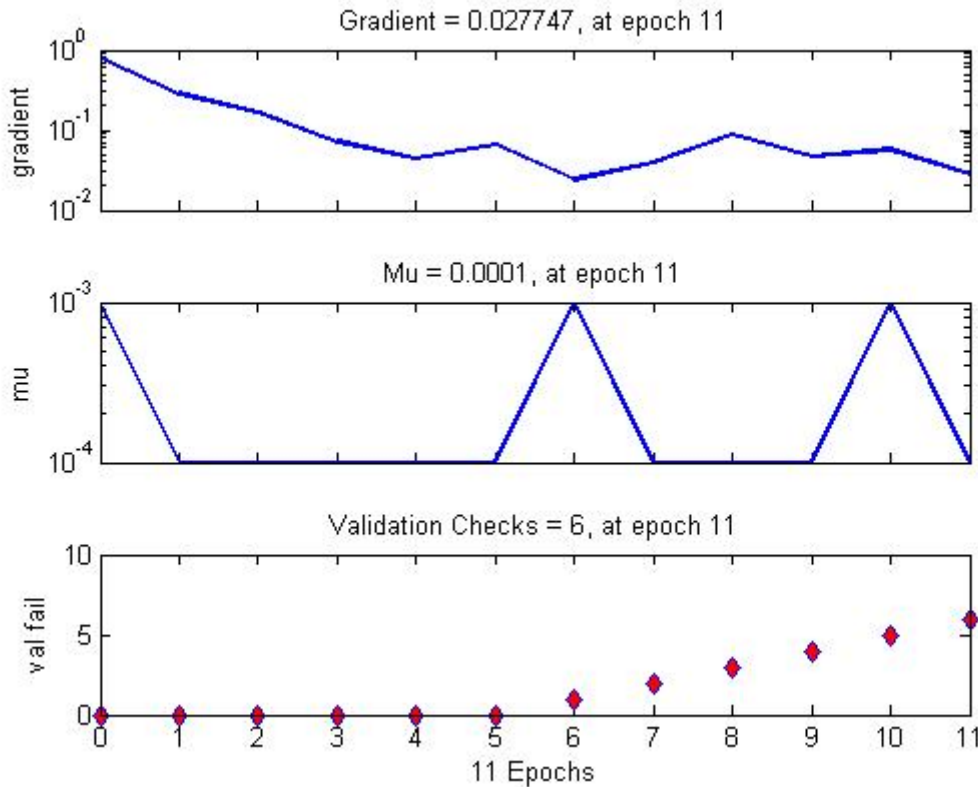
Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 9, September 2018

prepared, as input data and 297 X 2 matrix is prepared as output data. Out of these 297 samples, 70% sample were used for training purpose, 15% for validation and 15% for testing.



**Fig 2.1: Validation Performance**



**Fig 2.2 Gradient Analysis**

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 9, September 2018

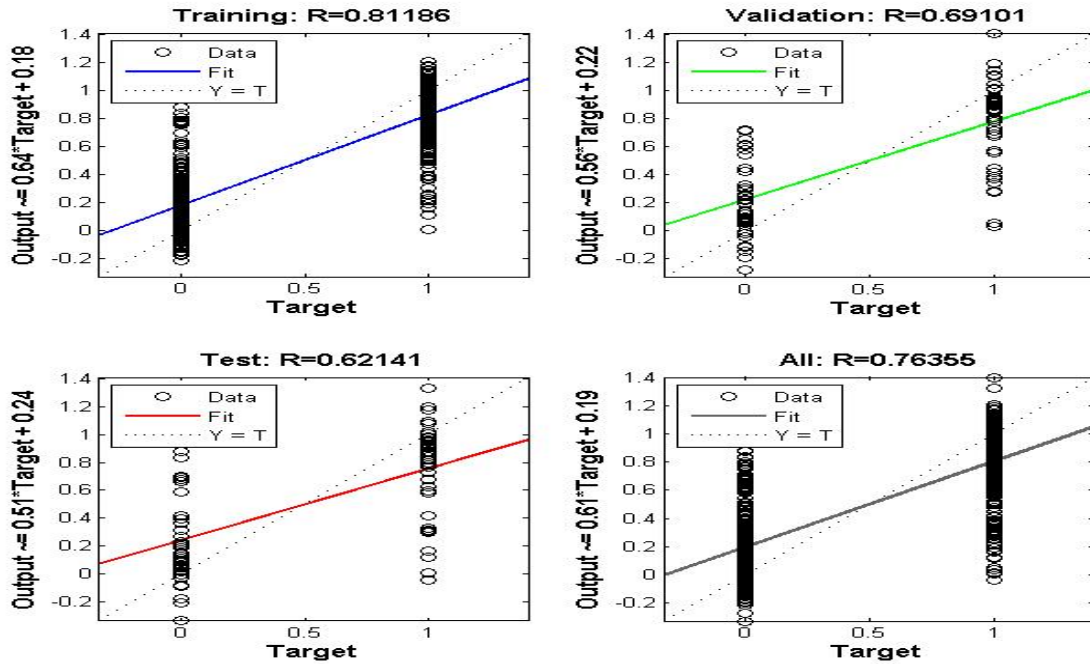


Fig 2.3: Regression Analysis

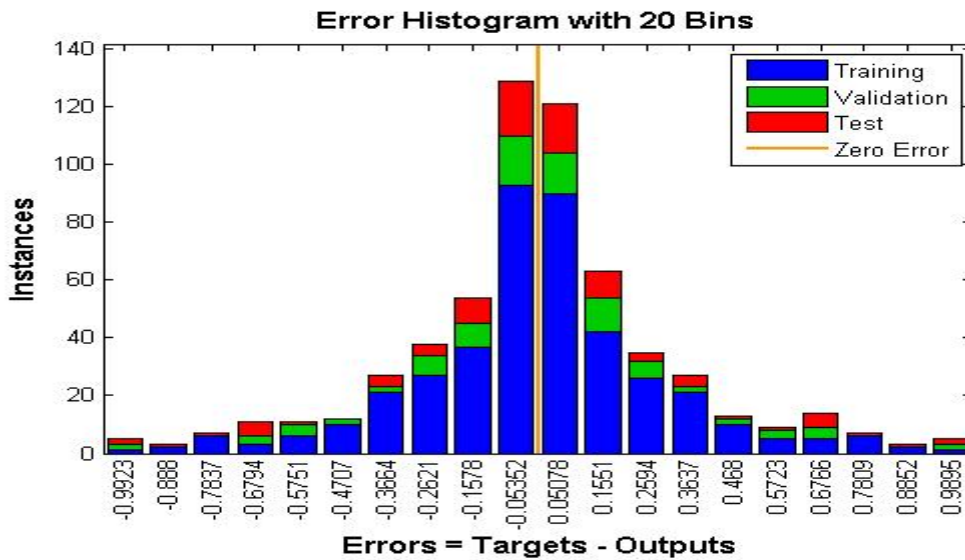


Fig 2.4: Error Histogram

# International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 9, September 2018

## V. CONCLUSION

As shown in result 2.1 most feasible solution could be found at 5<sup>th</sup> Epoch while training error keeps on decreasing and testing error starts to rise. Other results indicates positive correlation with initial results. So, from above results and discussion it can be concluded that an artificial neural network can be effectively used to classify multi class classification problems.

## REFERENCES

- [1] Zhi-Hua Zhou and Yuan Jiang. NeC4.5: Neural Ensemble Based C4.5. IEEE Trans. Knowl. Data Eng, 16. 2004.
- [2] Remco R. Bouckaert and Eibe Frank. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. PAKDD. 2004.
- [3] Xiaoyong Chai and Li Deng and Qiang Yang and Charles X. Ling. Test-Cost Sensitive Naive Bayes Classification. ICDM. 2004.
- [4] Gavin Brown. Diversity in Neural Network Ensembles. The University of Birmingham. 2004.
- [5] Kaizhu Huang and Haiqin Yang and Irwin King and Michael R. Lyu and Laiwan Chan. Biased Minimax Probability Machine for Medical Diagnosis. AMAI. 2004.
- [6] Jeroen Eggermont and Joost N. Kok and Walter A. Kusters. Genetic Programming for data classification: partitioning the search space. SAC. 2004.
- [7] David Page and Soumya Ray. Skewing: An Efficient Alternative to Lookahead for Decision Tree Induction. IJCAI. 2003.
- [8] Jinyan Li and Limsoon Wong. Using Rules to Analyse Bio-medical Data: A Comparison between C4.5 and PCL. WAIM. 2003.
- [9] Shrikant Vyas. Modeling Cognitive Process for solving Balance Scale Task using Feed Forward Neural Networks. Volume 5, Issue 6 (June 2016), PP.11-14.