A Comparison of Multilayer perceptron and Radial Basic Function for Disease Classification

Dhivya Devi. S [#1], Ravanan.R[*2]

[#]Research scholar, Department of Statistics, Presidency College, Chennai-05.
*The Principal, Presidency College, Chennai-05.
#thivyatvi89@gmail.com, *ravanastat@gmail.com

***Abstract*— Artificial Neural Networks (ANN) plays vital role in various fields like pattern recognition, identification, speech, vision, control system and classification. ANN is also trained to solve health problems and diagnosing various diseases in medical sciences. An appropriate method should be used to diagnose the disease that minimizes the errors in diagnosis. In this paper, Pima diabetic dataset has been used for the classification, multilayer perceptron and Radial basic function methods are proposed. Classification is one the most important method for classifying data set. Various real world problems in different fields like Management, Business, Industry, Science and Medicine can be solved using classification.

***Keywords*— ANN, Classification, Multilayer perceptron, Radial basic function.**

I. INTRODUCTION

Classification is one of the important techniques in data mining. Classification is a data mining function that assigns items in a collection to target data classes or concepts. The aim of classification is to accurately predict the target class for each case in the data. Classification is consist of predicting certain outcome based on input are given a data set, called a training set which contains set of attributes and respective outcome usually called prediction attributes. Artificial neural network is one of the known techniques for extraction of patterns. ANN allows nonlinear relations between independent and dependent variables and also ANN allow all possible interaction between the dependent variables. For these reasons mentioned above ANN technique is adopted for the classification of data.

II. DATA DESCRIPTION

This dataset is from the National Institute of Diabetics and Digestive and Kidney Disease. The Pima dataset is downloaded from UCI machine – learning repository which was collected by Peter Turney. The dataset is to diagnostically predict whether the patient has diabetes or not, based on the diagnostic measurement in the dataset. The same database has been used in many research works for the purpose of classification and testing algorithm. In particular, all the patients here are female at least 21 years. The total number of the sample is 768. The data set contains 9 attributes or features which are as follows

TABLE. 1. ATTRIBUTE INFORMATION

NO	ATTIRBUTE	DOMAIN
1.	PREGNANT	1-17.
2.	GULCOSE	56-197
3.	BLOOD PRESSURE	24-110
4.	SKIN FOLD THICKNESS	7-52
5.	INSULIN	15-846
6.	BMI	18.2-57.3
7.	DIABETICS PEDIGREE FUNCTION	0.0850-2.3290
8.	AGE	21-81
9.	CLASS VARIABLE	"+"VE FOR DIABETICS "-" VE FOR DIABETICS

These attributes measure the external appearance and the internal chromosome changes in 8 different scale. The two values in the class variable of Pima diabetics are

a. Positive for diabetics
b. Negative for diabetics, which is represented numerically by 1 and 0 respectively.

III. DESCRIPTIONS OF DATA SET

- Number of Instance :768
- Number of attributes : 8 and the Class variable
- Attribute 1 through 8 will represent the instance.
- Each instance has one of 2 possible classes: Positive or Negative.
- Class distributions: Positive 269 (35%) and Negative 499(65%).

TABLE II. A BRIEF DESCRIPTION OF DATASET

ATTRIBUTES	N	RANGE	MINIMUM	MAXIMUM	MEAN	SD
1	768	17	0	17	3.85	3.370
2	768	199	0	199	120.89	31.973
3	768	122	0	122	69.11	19.356
4	768	99	0	99	20.54	15.952
5	768	846	0	846	79.80	115.244
6	768	67.1000	.0000	67.1000	31.992578	7.8841603
7	768	2.3420	.0780	2.4200	.471876	.3313286
8	768	60	21	81	33.24	11.760

The data set was partitioned into two sets training set and testing set. Training consists in learning a relation between data and attributes from a part of training dataset, and testing consist in testing predictions of this relation on another part of the dataset. We use 80% of the example for training set and 20% for testing data.

IV. APLLIED NEURAL NETWORK STRUCTURE

The application of neural network plays a vital role in health care industry. The main objective of the paper is to represent how we can apply data mining techniques especially neural network in medical data The ANN is used to find the pattern recognition, classification. To obtain a good prediction especially with good classification, we should learn algorithms with good training set which rules are extracted the testing data classification.

4.1 ARCHITECTURE OF NEURAL NETWORK AND LEARNIG PHASE

Multi-Layer Perceptron and Radial Basic Function are the structures of network used in this paper. Each network consists of three layers such as Input layer, Hidden layer and Output layer. The features of the database consist of nine attributes parameters are the input of the network. The output layer consists of two classes corresponding to the type of diabetics either (1 for Positive or 0 for Negative).

The learning phase of network constitutes of back propagation algorithm. A supervised neural network based on back propagation algorithm is used here. It is necessary to conduct a learning phase algorithm to find the values of weights of the output of the neural network. The steps below represent the learning phase of neural network model.

1. Initialization →2. Learning **(ADD A NEURON ON THE HIDDEN LAYER)** → 3. Learning Error **(ADD A NEURON ON THE HIDDEN LAYER)** →4. End of Learning

4.2 ALGORITHIM USED IN THE STUDY

After the presentation of the data all the attributes is ready for the input of the neural network, different steps have been done to test the performance of the model.

- Selecting the architecture
- Initialization of neural network
- Neural Network training with back propagation algorithm
- Network testing
- Result

v. MULTI LAYER PRECEPTRON MODEL

The neural network is constitute one input layer which contains inputs corresponding to the nine attributes of the dataset, of one hidden layer with five neurons and of one out output containing two outputs corresponding to our two classes, Positive and Negative for diabetics.

VI. RADIAL BASIC FUNCTION

RBF is a different approach by viewing the design of neural network as a curve fitting problem in high dimensional spaces. Learning is equivalent to finding a surface in a multidimensional space that provides the best fit to the training data.

The construction RBF network in its most basic form involves three different layers. The input layer is consisting of source nodes. The second layer is hidden layer of high dimension. The last one is the output layer which supplies the response of the network to the activation patterns applied to input layer. The transformation from input layer to the hidden layer is non-linear whereas the transformation from hidden layer to the output layer is linear.

VII. SIMULATION RESULT

The simulation result was realized by SPSS Neural Network Tool box. Two different Neural Network structures Multi- Layer Perceptron and Radial Basic Function were applied to Pima Indian Diabetics database to show the performance of Neural Network on Pima Diabetic data. This study had using 615 training set and 153 testing set. The classification result for testing data set and training data set by MLP and RBF were given in tables,

TABLE 3.Experimental Results of PIMA Data using RBF

Training Samples	Testing Samples	Classification Efficiency in RBF Testing samples	Classification Efficiency in RBF Testing samples
462	306	77%	77%
526	242	78%	77%
609	159	78%	80%

TABLE .4 Experimental Results of PIMA Data using MLP

Training Samples	Testing Samples	Classification Efficiency in MLP Testing samples	Classification Efficiency in MLP Testing samples
457	311	75	76%
540	228	79	74%
610	158	75	78%

VIII. CONCLUSION

In this paper a new approach is developed to study the Pima diabetic data on Neural Network techniques.The objectives of the study is to create an effective tool for building neural models to help us making proper classification of two class variable of Pima diabetics data. A Detailed comparison between Multilayer networks MLP and RBF approaches showed in table 3 and table 4 respectively. RBF neural network is much more efficient than other model based on the MLP.

REFERENCES

[1]. A. Cichocki and R. Unbehauen, " Neural Networks for optimisation and signal processing," J. Wiley, Sons Ltd. And B.G. Teubner, Stuttgart, 1993.

[2]. Abdelaal Ahmed, Mohamed Medhat and FarouqWaelMuhamed, "Using data mining for assessing diagnosis of breast cancer," in Proc. International multi conference on computer science and information Technology, 2010, pp. 11-17.

[3]. Arun George Eapen, master thesis, "Application of Data Mining in Medical Applications", Waterloo, Ontario, Canada, 2004.

[4]. Bellaachia, Abdelghani and ErhanGuven, "Predicting Breast Cancer Survivability using Data Mining Techniques," Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining," 2006.

[5] .Burke H. B. Et al, "Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction", Cancer, 1997, vol.79, pp.857-862.

[6] .C.Meenakshi, M.Govindarajan and A.M.Sameeullah, "Detection of Breast Cancer using MLP and RBF Classifiers"‖, —IMS Manthan - Volume V, No. 1, June 2010.

[7] .Choi J.P., Han T.H. and Park R.W., "A Hybrid Bayesian Network Model for Predicting Breast Cancer Prognosis", J Korean Soc Med Inform, 2009,pp, 49-57.

[8]. Newton Cheung, "Machine Learning Techniques for Medical Analysis", Thesis for the Degree of Bachelor of Engineering, 2001.

*[9]. W.H. Wolberg and O.L. Mangasarian, "Multisurface method of Pattern Separation for Medical Diagnosis, Applied to Breast Cytology", in Proceedings of the National Academy of Sciences, **87**, pp. 9193-9196, U.S.A., December 1990.*