International Journal of Applied Research 2015; 1(10): 34-39

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2015; 1(10): 34-39
www.allresearchjournal.com
Received: 22-07-2015
Accepted: 24-08-2015

Prerana
Research Scholar, CSE Department, Prannath Parnami Institue of Management and Technology, Hisar, India.

Parveen Sehgal
Research Scholar, NIMS University, Jaipur, Rajasthan, India.

.

Correspondence
Prerana
Research Scholar, CSE Department, Prannath Parnami Institue of Management and Technology, Hisar, India.

Comparative Study of GD, LM and SCG Method of Neural Network for Thyroid Disease Diagnosis

Prerana, Parveen Sehgal

Abstract
This paper presents the comparative study of three main methods of neural network. Each of these methods are efficient in their working but on comparison there is lot of difference in their working in the field of early prediction of thyroid disease. The main purpose of this work is to propose the best technique for diagnosis of disease and to reduce the diagnosis time and improve the efficiency and accuracy.

Keywords: Artificial Neural Network, Thyroid disease, Gradient Descent, Levenberg Marquardt Method, Scaled Conjugate Gradient, Matlab

Introduction
Last century, the challenge was to develop new technologies that store large amount of data. Recently, the challenges are to effectively utilize the incredible amount of data and to obtain knowledge that benefits business, scientific, and government transactions by using subset of features rather than the whole features in the dataset. One of the major challenges in giving proper treatment is always fast and accurate diagnosis of the disease [1].
A lot of works have been done in medical diagnosis using different neural network techniques. But it had always been a tough task to identify the best technique for any diagnosis. Different techniques have their own limitations in terms of accuracy and time [2].
The thyroid gland is prone to several very distinct problems, some of which are extremely common. Production of too little thyroid hormone causes hypothyroidism or production of too much thyroid hormone causes hyperthyroidism.
Hypothyroidism, or an under active thyroid, has many causes.
Some of the causes are prior thyroid surgery, exposure to ionizing radiation, chronic inflammation of the thyroid (autoimmune thyroiditis), iodine deficiency, lack of enzymes to make thyroid hormone, and various kinds of medication [3].
Generally disease diagnosis is done according to and depends upon doctor's experience and knowledge. Diagnosis is based on signs, symptoms and physical examination of patient. Predictive data mining plays a vital role in disease diagnosis. Artificial neural network can be employed for the disease diagnosis, so as to improve the quality of diagnosis. Various data mining approaches exist for prediction modeling but here we employ trained artificial neural networks to develop a predictive system for disease diagnosis.
Thyroid gland secretes thyroid hormones to control the body's metabolous rate. The misfunction of thyroid hormone will leads to thyroid disorders. The thyroid or the thyroid gland is an endocrine gland. The thyroid gland releases thyroxine (T4) and triiodothyronine (T3) into the blood stream as the principal hormones. The two main functions of the thyroid hormone is to regulate the rate of metabolism and affect the growth. There are two most common problems of thyroid disorder or thyroid disease. They are Hyperthyroidism – releases too much thyroid hormone into the blood due to over active of thyroid and Hypothyroidism - when the thyroid is not active and releases too low thyroid hormone into the blood [4].

Literature Survey
A lot of work has already been done in this field that include methods like Decision Tree Splitting rules, Rough Set Approach with modified similarity relation. Empirical comparison

study has also been done using different data mining techniques. Besides, there are another diseases have been detected using Artificial Neural Network (ANN). The brief descriptions on the several disease diagnosis using several methods are given below:

M. R. Nazari Kousarrizi *et al.* proposed the methods of feature selection and classification for thyroid disease diagnosis, which is one of the most important classification problems [5]. Two common diseases of the thyroid gland, which releases thyroid hormones for regulating the rate of body's metabolism, are hyperthyroidism and hypothyroidism. Classification of these thyroid diseases is a considerable task. An important problem of pattern recognition is to extract or select feature set, which is included in the pre-processing stage. As a case study, Sequential forward selection and sequential backward selection, which are two well-known heuristic schemes, are employed for feature selection. The integrated feature selection and classification approach described has proved effective on these two thyroid datasets. Because the feature values are very much closed, thyroid disease is a hard problem to be solved for classification systems. In this study, two thyroid disease datasets are applied in the performance analysis of our proposed system. For the first dataset, we obtain a classification accuracy of 98.62% using SVM with feature subset which is the highest accuracy reached so far. moreover, the classification accuracy obtained from this structure shows an improvement of 1.39% in comparison with the best technique reported.

Jacqulin Margret *et al.* proposed the diagnosis of thyroid disease using decision tree splitting rules [6]. Various splitting rule for decision tree attribute selection had been analysed and compared. This helps to diagnosis the thyroid diseases through the extracted rules. From this work, it is clear, that normalized based splitting rules have high accuracy and sensitivity or true positive rate. This work can be extended for any medical datasets. Further enhancement can be made by using various optimization algorithms or rule extraction algorithms.

Anurag Upadhayay *et al.* performed the empirical comparison study by data mining classification algorithms (C 4.5, C5.0) for thyroid cancer set [7]. The main goal was to observe the impact of data mining technique on Thyroid Cancer. Two algorithms of Decision Tree technique, *C4.5* and C5.0 technique were implemented. And the following observation is noted down.

1. Size of tree for C4.5 was very large in compare to C5.0.
2. Both algorithms generated different rules.
3. After Pruning process in C5.0, Tree generated more accurate rule set.
4. Running Time of C5.0 was less as compared to C4.5.
5. C5.0 generated less train error as compared to the C4.5
6. Rule set Generated by the C5.0 algorithm is 6 and the confidence level of rules was more than 95%.

Farhad Soleimanian Gharehchopogh *et al.* performed a case study in diagnosis of thyroid disease using artificial neural network [8]. The importance of using ANNs to diagnose disease is to increase the accuracy of performance. The appropriate selection of artificial neural network architecture affects the network performance effectively to reach the high accuracy. By selecting a hidden layer and log-sigmoid activation function for hidden layer and 6 neurons in the hidden layer, classification accuracy of 98.6% is being achieved for Thyroid disease. The proposed method in this paper can be a solution to increase the performance of ANN.

S. B. Patel worked to predict the diagnosis of heart disease patients using classification mining techniques [9]. Three classification function techniques in data mining are compared for predicting heart disease with reduced number of attributes. These are Naïve Bayes, decision tree and classification by clustering. Genetic algorithm are also employed to determine the attributes which contribute more towards the diagnosis of heart ailments which reduces the number of tests in indirect way which are needed to be taken by a patient. 14 attributes are reduced to 6 attributes. Also, the observations conclude that the decision tree data mining technique outperforms other two data mining techniques after incorporating feature subset selection with relatively high model construction time. There is an algorithm named as Naïve Bayes, performs consistently before and after reduction of attributes with the same model construction time. Clustering via classification performs poor compared to other existing two methods. Inconsistencies and missing values were resolved before model construction.

Dilip Roy Chowdhury *et al.* [10] represent the use of artificial neural networks in predicting neonatal disease diagnosis. The proposed technique consist training of a MultiLayer Perceptron with a BP learning algorithm to recognize a pattern for the diagnosing and prediction of neonatal diseases. The most famous Back propagation algorithm was used to train the ANN architecture and the same has been tested for the various categories of neonatal disease. About 94 cases of different sign and symptoms parameter have been tested in this model. This study consist of ANN based prediction of neonatal disease and improves the diagnosis accuracy of 75% with higher stability.

E. Radwan *et al.* [11] used the concept of rough set theory in data discretization for continuous attribute values, data reduction and rule induction. Also, Rough sets try to cluster the Thyroid relation attributes in the presence of missing attribute values and build the Modified Similarity Relation that is dependent on the number of missing values with respect to the number of the whole defined attributes for each rule. Thus, the rule associated strength is measured. Moreover, The MSIM, modified similarity analysis relation, is used to classify rules contain missing attribute value, gaps, with respect to the number of the whole defined attributes for each rule. Also, constructing of discernibility matrix, deduction of the production rules, and reducts in the presence of the missing attribute value are used to extract the minimal set of productions rules that describe similarity relation among rules. Hence, feature selection reduces the dimensionality of the data, the size of the hypothesis space and allows classification algorithm to operate faster and more effectively.

Methodology Used

The software that is used in this research is MATLAB version 7.12.0. MATLAB is an interactive environment and high level language that enables to perform computationally intensive tasks faster than with traditional programming languages like C and C++. MATLAB can be used for a range of applications that include signal processing and communications, image processing, control systems, video processing, test and measurement, computational finance and biology.

Gradient Descent

Gradient descent method is also known as steepest descent. Gradient descent is a kind of iterative method that is provided with an initial point and it follows the negative of the gradient in order to move the point toward a critical point. This critical point is the required local minimum value. We are concerned only with local optimization of system because global optimization is challenging to compute.

It is an optimization algorithm that approaches a local minima of a function by taking the steps iteratively and these steps are directly proportional to the negative of the gradient of a function as the current point. Hence to calculate the derivative of cost function with respect to the weights and then change each weight by a small increment in the negative direction of the gradient is given as follows:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y}.\frac{\partial y}{\partial w} = -(y_{derived} - y)x = -\delta x$$

To reduce E by gradient descent, move or increment weights in the negative direction to the gradient i.e, $-(-\delta x) = +(\delta x)$. Weight value is calculated just by finding the first order derivative of cost function which is given by the formula:

$$w_{i+1} = w_i - \mu \nabla E(w)$$

Gradient descent is popular for very large-scale optimization problems because it is easy to implement and it can handle "black box" functions, and each iteration is cheaper than the other algorithms used for this purpose. Its major disadvantage is that it can take a long time to converge. A number of variants are present for this method. Two Basic variants are: 1. Levenberg Marquardt Method, 2.Scaled Conjugate Gradient**.**

Scaled Conjugate Gradient

Scaled Conjugate algorithm is a supervised learning algorithm. This works for feed forward neural networks. It is a member of class of Conjugate Gradient Methods (CGM). It works with second order information from neural network.

$$s_k = \frac{E^I(w_k + \sigma_k . p_k) - E^I(w_k)}{\sigma_k} + \lambda_k . p_k$$

SCG has two initial value parameters λ_k and σ_k. $E(w_{k+1}) \leq E(w_k)$ is a small number that is used to rectify the special case of converging to a function value of exactly zero. SCG is a batch learning method hence there will be no effect if parameters are being shuffled. Comparing the number of epochs is not relevant in the case of SCG and other algorithms like standard back propagation rather iteration can be checked out for comparing SCG with Standard Back propagation (BP). One iteration in SCG needs the calculation of two gradients, and in addition to this it requires one call to the error function, while one iteration in standard backpropagation needs the computation of one gradient and one call to the error function. Møller defines a *complexity unit* (cu) to be equivalent to the complexity of one forward passing of all patterns in the training set. Then computing the error costs 1 cu while computing the gradient can be estimated to cost 3 cu. According to Møller's metric, one iteration of SCG is as complex as around 10^{16} iterations of standard backpropagation (BP).

Levenberg Marquardt

The Levenberg-Marquardt (LM) algorithm is the most widely used optimization algorithm. Its performance is far more better than simple gradient descent and other conjugate gradient methods in a wide variety of problems. Nonlinear Least Squares is the problem for which the LM algorithm provides a solution. This method works on second order derivative. It works on curvature as well as on gradient of surface. Formulation Of Levenberg Marquardt Algorithm is given as follows:

$$w_{i+1} = w_i - (H + \lambda I)^{-1} d$$

Where λ is blending factor and I is identity matrix. As λ gets small the rule approaches the quadratic approximation update rule above. If λ is large, the rule approaches

$$w_{i+1} = w_i - \frac{1}{\lambda} d$$

Which is steepest descent. this algorithm adjusts the value of λ accordingly whether E is increasing or decreasing.

Experimental Results

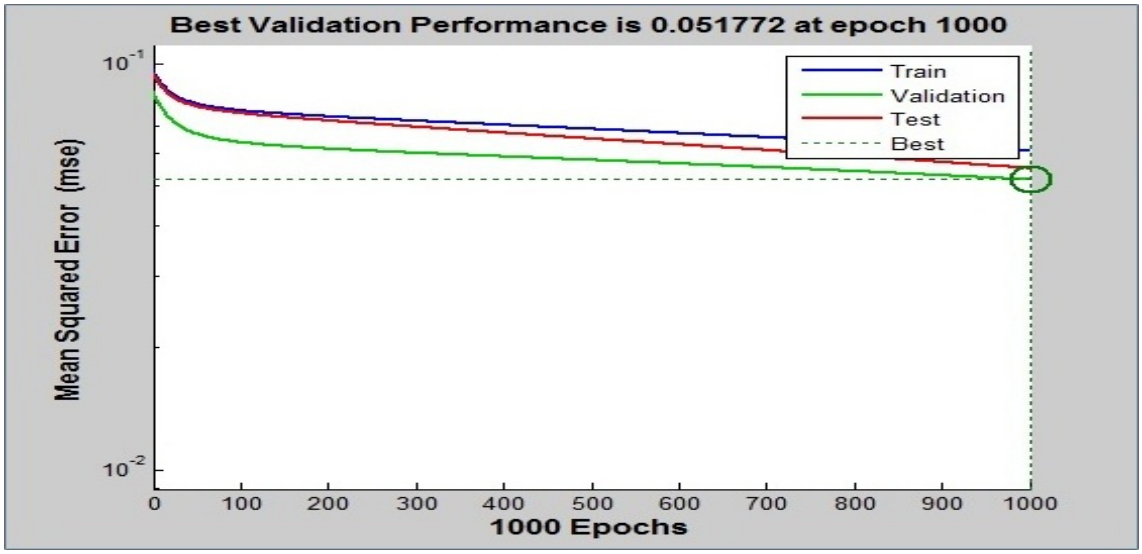

Fig 1: Training performance plot when employing gradient descent method

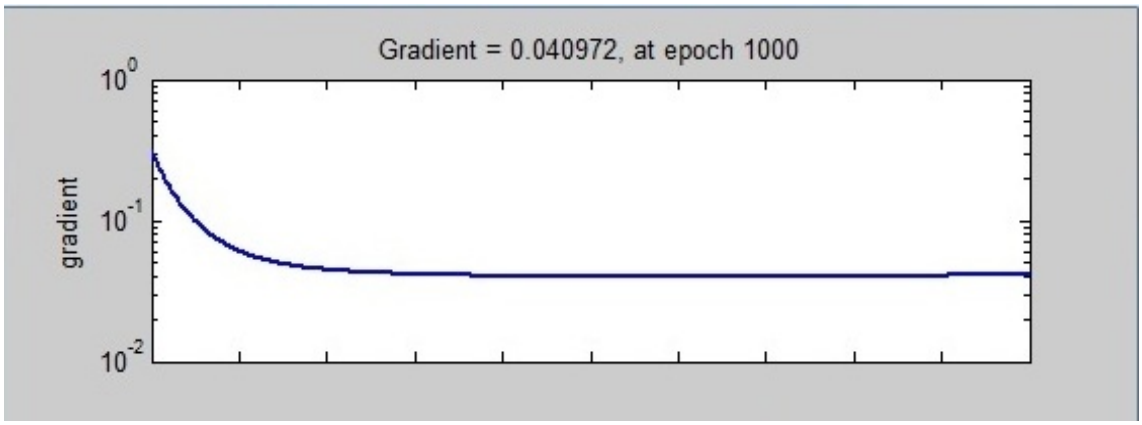

Fig 2: Gradient plot when employing gradient descent method

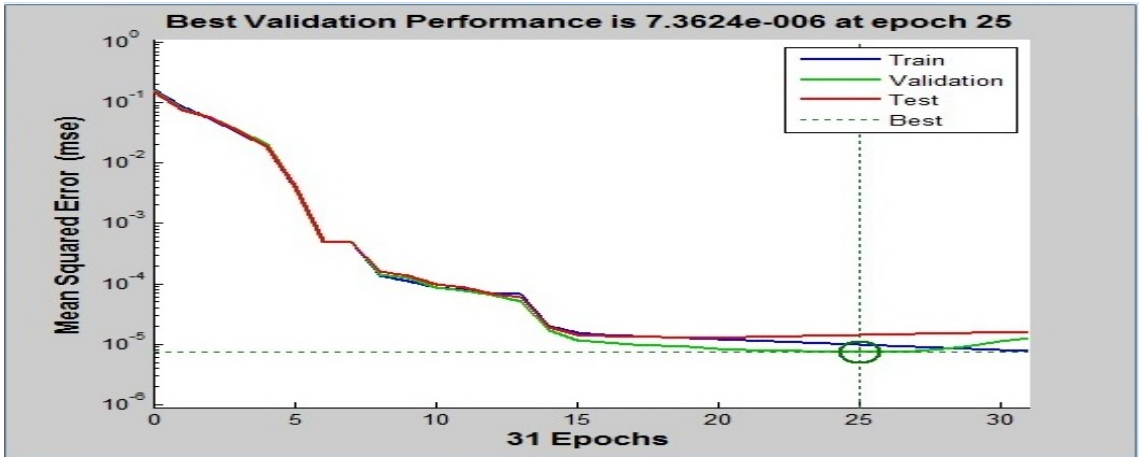

Fig 3: Training performance plot when employing Levenberg Marquard method

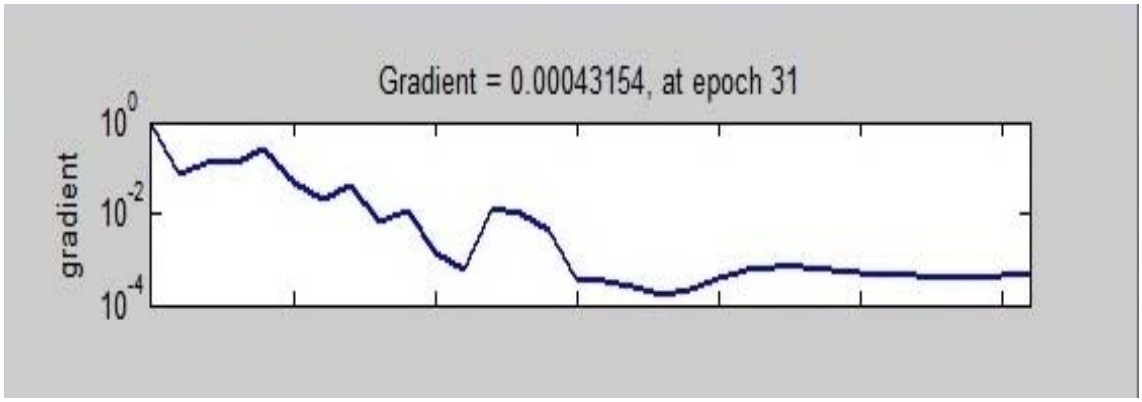

Fig 4: Gradient plot when employing Levenberg Marquardt method

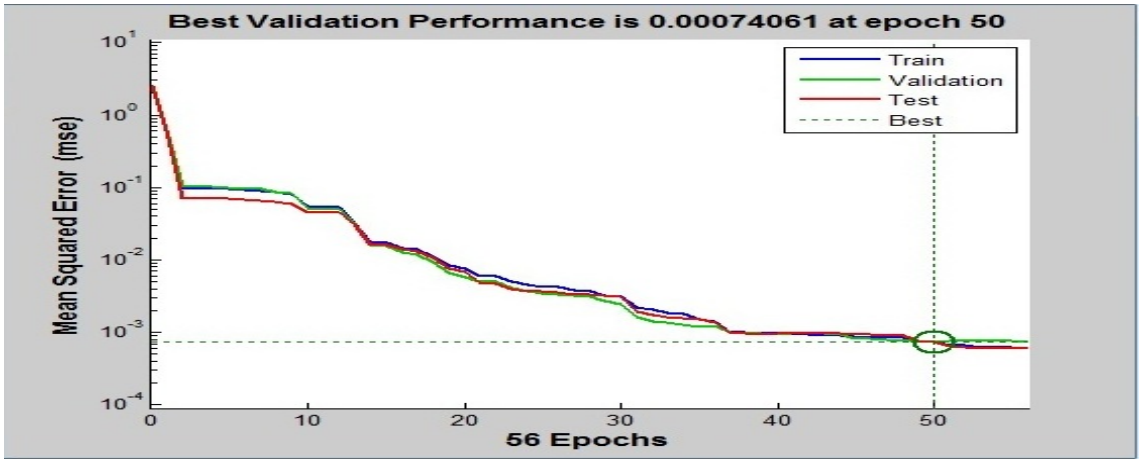

Fig 5: Training performance plot when employing Scaled Conjugate Gradient method

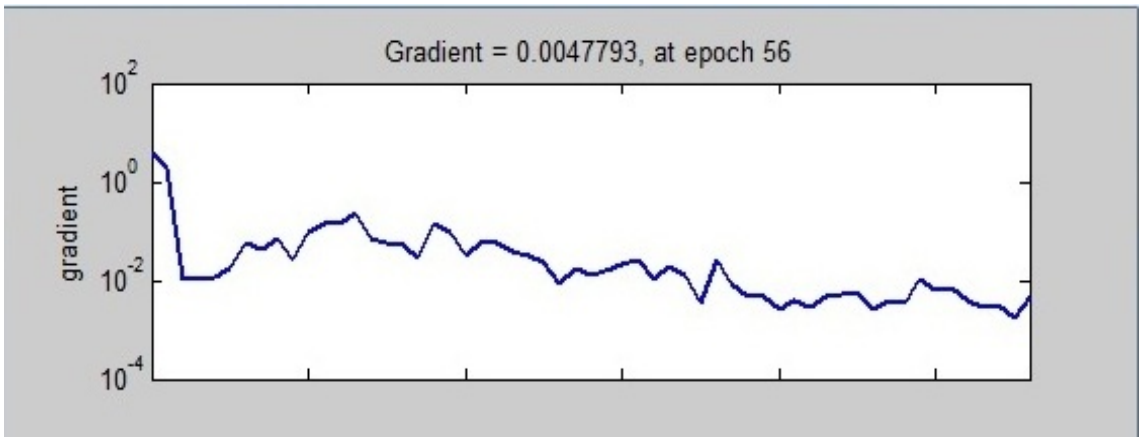

Fig 6: Gradient plot when employing Scaled Conjugate Gradient method

Table I: Comparative Study among GD, LM and SCG algorithm

Parameter used	G.D	LM	SCG
No. of iterations	1000	31	56
Time Taken	01:46	00:08	00:13
Gradient Value	0.040972	0.00043154	0.00047793
Validation Performance	0.051772	7.3624e-006	0.00074061
No. of epoch used for validation performance	1000	25	50

We have implemented thyroid disease diagnosis using predicting model in MATLAB Neural Network Toolbox software. 8 attributes were taken as input values for classify the patient of thyroid disease in classes from 1-3 values. Training has been done using Gradient Descent, Levenberg and Scaled Conjugate methods. Training Performance plot and gradient plots are given in fig. from 1 to 6.

Table 1 is showing the comparative study of these three algorithms.

The gradient plots depict the variation of error gradient verses number of epochs. These plots also present the initial and final values of the gradient values. It has been observed that Levenberg Marquardt method has shown a better training performance for achieving the set target in 31 epochs and Scaled conjugate gradient is showing a poor performance as it is achieving the set target value in 56 epochs.

Validation performance of Gradient Decent is coming out to be 0.051772 and for Levenberg Marquardt method it is 7.3624e-006 at epoch 25 while for scaled conjugate it is 0.00074061 at epoch 50.

There is a lot of difference in the gradient values of these three algorithms. For Gradient descent the gradient value is 0.040972, for Levenberg Marquardt the value is 0.00043154 and for Scaled Conjugate the value is 0.0047793.

Conclusion & Future Scope

It is clearly understood from the practical observations that Levenberg Marquardt Method works more efficiently than Scaled Conjugate and Gradient decent in all respect i.e., performance, time, gradient value etc.

While training the neural network with error back propagation in conjunction with gradient based training methods, from our experiments we conclude that Levenberg Marquardt method has shown a better performance in comparison with Scaled conjugate gradient method and Gradient Descent.

In addition, observations conclude that error accuracy limit achieved by Levenberg Marquardt method is superior and it trains the models to accuracy level of the order of 10^{-5} for the applied data set. Research can be extended at different angles like analysis and effects of varying network parameters like number of layers and number of neurons in hidden layers, learning rate, adaptive learning rate and other network parameters etc. Also, we can extend our research to find theoretical formulations for optimal values of these changed parameters.

References

1. Ashraf M, Chetty G, Tran D. Feature Selection Techniques on Thyroid, Hepatitis, and Breast Cancer Datasets, Feature Selection Techniques on Thyroid, Hepatitis, and Breast Cancer Datasets.
2. Abhishek GS. Mitra Thakur, Gupta D. Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis, IJCSIT. 2012; 3(3):3900-3904.
3. Zhang G, Berardi LV. An investigation of neural networks in thyroid function diagnosis, Health Care Management Science, 1998, 27-37. Available. http://www.endocrineweb.com/thyroid.html, (Accessed: 7 August 2007).
4. thyroid :www.wikipedia.org/thyroid
5. Nazari Kousarrizi MR, Seiti F, Teshnehlab M. An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification, International Journal of Electrical & Computer Sciences IJECS-IJENS 12(01).
6. Margret J, Lakshmipathi B, Kumar SA, Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules, International Journal of Computer Applications (0975 – 8887), April, 2012, 44(8).
7. Upadhayay A, Shukla S, Kumar S, Empirical Comparison by data mining Classification algorithms (C 4.5 & C 5.0) for thyroid cancer data set, International Journal of Computer Science & Communication Networks, ISSN:2249-5789, 3(1), 64-68.
8. Gharehchopogh FS, Molany M, Mokri FD. Using Artificial Neural Network In Diagnosis Of Thyroid Disease: A Case Study, International Journal on Computational Sciences & Applications (IJCSA).

August, 2013, 3(4).
9. Patel SB, Yadav PK, Dr. Shukla DP. Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques, (IOSR-JAVS), e-ISSN: 2319-2380, p-ISSN: 2319-2372 2013; 4(2):61-64. (www.iosrjournals.org).
10. Dilip Roy Chowdhury, Mridula Chatterjee, Samanta RK. An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE). 2011; 2(3).
11. Radwan E, Assiri AM, Thyroid Diagnosis Based Technique on Rough Sets With modified Similarity Relation, International Journal of Advanced Computer Science and Applications (IJACSA). 2013; 4(10).
12. Sen AK, Patel SB, Dr. Shukla DP. A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level, International Journal Of Engineering And Computer Science ISSN:2319-7242 2013; 2(9):2663-2671.
13. Bala S, Kumar K. A Literature Review on Kidney Disease Prediction using Data Mining Classification Technique, International Journal of Computer Science and Mobile Computing, 2014; 3(7):960-967.
14. Yadav C, Lade S, Suman MK. A Survey on Data Mining Techniques for the Diagnosis of Coronary Artery Disease, International Journal of Advanced Research in Computer Science and Software Engineering. ISSN: 2277 128X October, 2013, 3(10).
15. Bhatla N, Jyoti K. An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT). ISSN: 2278-0181 October, 2012, 1(8).
16. Dr. K. Usha Rani. Analysis of Heart Diseases Dataset Using Neural Network Approach, International Journal of Data Mining & Knowledge Management Process (IJDKP). 2011, 1(5).
17. Jesmin T, Ahmed K, Md. Zamilur Rahman, Md. Badrul Alam Miah. Brain Cancer Risk Prediction Tool Using Data Mining, International Journal of Computer Applications (0975 – 8887) January 2013, 61-12.
18. David Gil A, Johnson BM. Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, Global Journal of Computer Science and Technology, 63-71.
19. Sudhakar K, Dr. Manimekalai M. Study of Heart Disease Prediction using Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X January 2014, (4)1.
20. Nesma Ibrahim, Taher Hamza, Elsayd Radwan. An Evolutionary Machine Learning Algorithm for Classifying Thyroid Diseases Diagnoses, Egyptian Computer Science Journal. Jan 2011; 35(1):73-86.