An Artificial Neural Network Model for Neonatal Disease Diagnosis

Dilip Roy Chowdhury *diliproychowdhury@gmail.com*
Expert System Laboratory
Dept. of Computer Science & Application
University of North Bengal
Raja Rammuhunpur, 734013
West Bengal, India

Mridula Chatterjee *drmridulachatterjee@gmail.com*
Department of Pediatrics
NRS Medical College, Kolkata,
West Bengal, India.

R.K. Samanta *rksamantark@gmail.com*
Expert System Laboratory
Dept. of Computer Science & Application
University of North Bengal
Raja Rammuhunpur, 734013
West Bengal, India

Abstract

The significance of disease diagnosis by artificial intelligence is not obscure now a day. The increasing demand of Artificial Neural Network application for predicting the disease shows better performance in the field of medical decision making. This paper represents the use of artificial neural networks in predicting neonatal disease diagnosis. The proposed technique involves training a Multi Layer Perceptron with a BP learning algorithm to recognize a pattern for the diagnosing and prediction of neonatal diseases. A comparative study of using different training algorithm of MLP, Quick Propagation, Conjugate Gradient Descent, shows the higher prediction accuracy. The Backpropogation 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 exhibits ANN based prediction of neonatal disease and improves the diagnosis accuracy of 75% with higher stability.

Key words: Artificial Intelligence, Multi Layer Perceptron, Neural Network, Neonate

1. INTRODUCTION

Artificial Intelligence techniques consist of developing a computer based decision support system does somewhat that it were done by a human being. Several Neural Network Models are developed which helps doctors in diagnosing the patients more correctly and accurately. Neural networks provide a very general way of approaching problems. When the output of the network is categorical, it is performing prediction and when the output has discrete values, and then it is doing classification. Neural Network based Decision Support in medicine, particularly for the neonates, has at least the role of enhancing the consistency of care.

Among various phases of child development, Neonatal phase is considered to be one of the vital phases. In India, 30% to 40% babies are Low Birth Weight babies and about 10% to 12% of Indian babies are born less than 37 completed weeks (preterm). Thus, these babies are physically immature and cause the high neonatal mortality [1]. In a study, authors describe about prevalence diseases those are the major causes of deaths in the neonates in Terai region of West Bengal [2]. This mortality problem, especially in rural areas [3], can prevail over through fast

and accurate disease diagnosis and management of the newborn. In our earlier studies of data mining model development, several classification techniques have applied to get the maximum accuracy [4]. However, any ANN based model may be useful for classification of disease and even for taking necessary decision. This paper describes how artificial intelligence, for example artificial neural networks can improve this area of diagnosis.

The proposed model has the potential to cover rare conditions of all the exceptional symptoms of neonatal diseases to diagnose. The increasing range of neonatal patient information makes it feasible to more accurately quantify important experimental indicators, such as the relative likelihood for competing diagnoses or the clinical outcome. It is observed that, in few instances, computer-assisted diagnoses, particularly ANN based model have been claimed to be even more accurate than those decision taken by domain experts [5].

2. RELATED STUDIES OF ARTIFICIAL NEURAL NETWORK

There are several studies which have applied neural networks in the diagnosis of different disease. An artificial neural network trained on admission data can accurately predict the mortality risk for most preterm infants. However, the significant number of prediction failures renders it unsuitable or individual treatment decisions. In a study[6], the artificial neural network performed significantly better than a logistic regression model (area under the receiver operator curve 0.95 *vs* 0.92). Survival was associated with high morbidity if the predicted mortality risk was greater than .50. There were no preterm infants with a predicted mortality risk of greater than 0.80. The mortality risks of two non-survivors with birthweights >2000 g and severe congenital disease had largely been underestimated.

In another study [7], an effective arrhythmia classification algorithm used for the heart rate variability (HRV) signals. The proposed method is based on the Generalized Discriminant Analysis (GDA) feature reduction technique and the Multilayer Perceptron (MLP) neural network classifier. At first, nine linear and nonlinear features are extracted from the HRV signals and then these features are reduced to only three by GDA. Finally, the MLP neural network is used to classify the HRV signals. The proposed arrhythmia classification method is applied to input HRV signals, obtained from the MIT-BIH databases. Here, four types of the most life threatening cardiac arrhythmias including left bundle branch block, fist degree heart block, Supraventricular tachyarrhythmia and ventricular trigeminy can be discriminated by MLP and reduced features with the accuracy of 100%.

The study [8] of a functional model of ANN is proposed to aid existing diagnosis methods. This work investigated the use of Artificial Neural Networks (ANN) in predicting the Thrombo-embolic stroke disease. The Backpropogation algorithm was used to train the ANN architecture and the same has been tested for the various categories of stroke disease. This research work demonstrates that the ANN based prediction of stroke disease improves the diagnosis accuracy with higher consistency. This ANN exhibits good performance in the prediction of stroke disease in general and when the ANN was trained and tested after optimizing the input parameters, the overall predictive accuracy obtained was 89%.

As per the artificial neural networks in medicine world map[9], different universities, research centres, medical diagnostic centres are using ANN for medical diagnosis and management. Some studies are carried out using some combined architecture using ANN and different data mining techniques [10].

3. MLP NEURAL NETWORK MODEL

3.1 Structure of MLP

In medical decision making a variety of neural networks used for decision accuracy. MLPs are the simplest and commonly used neural network architectures programs due to their structural litheness, good representational capabilities and availability, with a large number of

programmable algorithms[11]. MLPs are feed forward neural networks and universal approximators, programmed with the standard back propagation algorithm. They are supervised networks so they require a desired response to be trained. They are able to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input-output map. Generally, an MLP consists of three layers: an input layer, an output layer and an intermediate or hidden layer. In this network, every neuron is connected to all neurons of the next layer, in other words, an MLP is a fully connected network[12]. Figure 1 shows the structure of a MLP network.

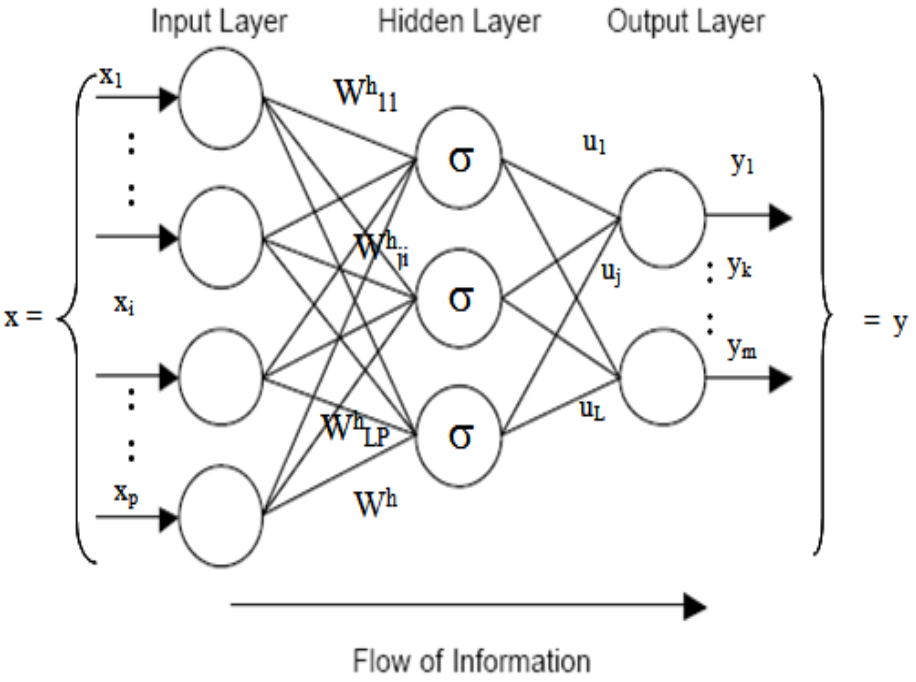

FIGURE 1: A structure of MLP Network

On the left this network has an input layer with three neurons, in the middle, one hidden layer with three neurons and an output layer on the right with two neurons. There is one neuron in the input layer for each predictor variable ($x_1 \ldots x_p$). In the case of categorical variables, *N*-1 neurons are used to represent the *N* categories of the variable.

3.2 MLP Input Layer

A vector of predictor variable values ($x_1 \ldots x_p$) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

The net calculation of input and output of the j hidden layer neurons are as follows:

$$net^h_j = \sum_{t=1}^{N+1} W_{ji} x_i$$

$$y_j = f\,(net^h_j)$$

3.3 MLP Hidden Layer

Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function σ. The outputs from the hidden layer are distributed to the output layer.

3.4 MLP Output Layer

The value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value *u, at time of a*rriving at a neuron in the output layer *j*. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network. If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single *y* value. For classification problems with categorical target variables, there are *N* neurons in the output layer producing *N* values, one for each of the *N* categories of the target variable.
Calculate the net inputs and outputs of the k output layer neurons are :

$$\text{net}^{o}_{k} = \sum_{j=1}^{j+1} V_{kj} y_j$$

$$Z_k = f(\text{net}^{0}_{k})$$

Update the weights in the output layer (for all k, j pairs)

$$v_{kj} \leftarrow v_{kj} + c\lambda\,(d_k - Z_k)\,Z_k\,(1 - Z_k)\,y_j$$

4. PROPOSED MODEL

4.1 Input Data

The data for this study have been collected from 94 patients who have symptoms of neonatal diseases. The data have been standardized so as to be error free in nature. All the cases are analyzed after careful scrutiny with the help of the pediatric expert. Table 1 below shows the various input parameters for the prediction of neonatal disease diagnosis.

Sl.No.	Parameters	Column Type
1	Birth_Term_Status	Categorical
2	Birth_Weight_Status	Categorical
3	Age_in_Hours>72	Categorical
4	Lathergy	Categorical
5	Refusual_to_Suck	Categorical
6	Poor_Cry	Categorical
7	Poor_Weight_gain	Categorical
8	Hypothalmia	Categorical
9	Sclerema	Categorical
10	Excessive_Jaundice	Categorical
11	Bleeding	Categorical
12	GI_Disorder	Categorical
13	Seizure	Categorical
14	Sluggish_Neonatal_Reflex	Categorical

TABLE 1: Input Parameters for Prediction Neonatal Disease

4.2 Feature Selection of Dataset

Data analysis information needed for correct data preprocessing. After data analysis, the values have been identified as missing, wrong type values or outliers and which columns were rejected as unconvertible for use with the neural network [13]. Feature selection methods are used to identify input columns that are not useful and do not contribute significantly to the performance of neural network. In this study, Genetic method is used for input feature selection. Genetic algorithms method [14] starts with a random population of input configurations. Input configuration determines what inputs are ignored during performance test. At each following step uses a process analogous to natural selection to select superior configurations and use them to generate a new population. Each step successively produces better input configuration. At the last step the best configuration is selected. The method is very time-consuming but good for determining mutually-required inputs and detecting interdependencies. This method use generalized regression neural networks (GRNN) or probabilistic neural networks (PNN) because they train quickly and proved to be sensitive to the irrelevant inputs. The removal of irrelevant inputs will improve the generalization performance of a neural network. Table 2 shows the finalized input parameters after applying feature selection method.

Code	Name of the Input Column	Input state	Importance %
C3	Age_in_Hours>72	Two-state	0.551381
C4	Lathergy	Two-state	12.344225
C6	Poor_Cry	Two-state	0.832139
C7	Poor_Weight_gain	Two-state	18.140229
C8	Hypothalmia	Two-state	15.23048
C9	Sclerema	Two-state	0.088902
C10	Excessive_Jaundice	Two-state	14.179179
C11	Bleeding	Two-state	4.159191
C12	GI_Disorder	Two-state	8.745518
C13	Seizure	Two-state	22.076618
C14	Sluggish_Neonatal_Reflex	Two-state	3.652138

TABLE 2: Percentage of Importance of Input Data after feature selection

4.3 Development of Neural Network Architecture

In this study, the multilayered feed-forward network architecture with 11 input nodes after feature selection of the input data, 5 hidden nodes, and 13 output nodes have been used for the neural network architecture. The numbers of input nodes are determined by the finalized data; the numbers of hidden nodes are determined through trial and error; and the numbers of output nodes are represented as a range showing the disease classification. The most widely used neural-network learning method is the Back Propagation algorithm [15]. Learning in a neural network involves modifying the weights and biases of the network in order to minimize a cost function. The cost function always includes an error term; a measure of how close the network's predictions are to the class labels for the examples in the training set. Additionally, it may include a complexity term that reacts a prior distribution over the values that the parameters can take. The activation function considered for each node in the network is the binary sigmoidal function defined (with $\sigma = 1$) as output = $1/(1+e^{-x})$, where x is the sum of the weighted inputs to that particular node. This is a common function used in many Back Propagation Network. This function limits the output of all nodes in the network to be between 0 and 1. Note all neural networks are basically trained until the error for each training iteration stopped decreasing. Figure 2 shows the architecture of the specialized network for the prediction of neonatal disease. The complete sets of final data (11 inputs) are presented to the generic network, in which the final diagnosis corresponds to output units.

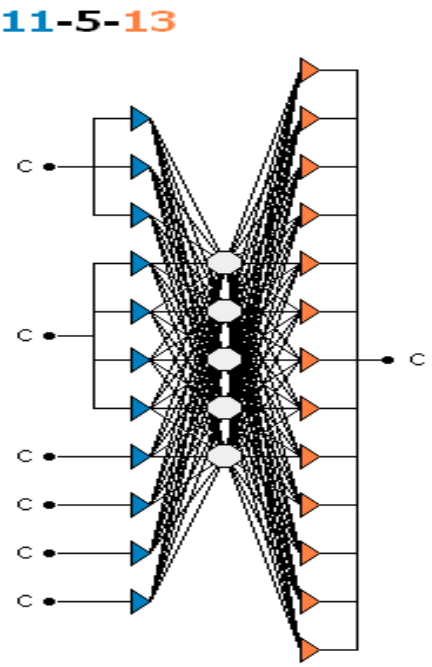

FIGURE 2: ANN Architecture for Neonatal Disease Diagnosis

The following are the results generated from the input given to the neural network after going through the process of careful training, validation and testing using NeuroIntelligence tool[16]. Table 3 shows the various categories of neonatal diseases and their classification and probability statistics.

Category	Probability
HIE_III	0.1702128
Hemorrhage	0.0106383
HIE_II	0.0425532
Hypo_Thalmia	0.0212766
Jaundice	0.0212766
Jaundice_BA	0.0319149
MD_Hypocalcemia	0.0957447
MD_Hypoglycemia	0.0319149
MD_Hypothermia	0.0319149
No_Disease	0.0851064
Others	0.0531915
Septicemia	0.3936170
Sizure_Disorder	0.0106383

TABLE 3: Category weights (prior probabilities)

4.4 Training Process of MLP Networks

In this context, our objectives of the training process was to find the set of weight values which will cause the output from the neural network to match the actual target values as closely as possible. We have faced several issues concerned in designing and training a multilayer perceptron network model. Some of the issues are:

i. To select the number of hidden layers to use in the network.
ii. To decide the number of neurons to be used in each hidden layer.
iii. Converging to an optimal solution in a reasonable period of time.

iv. Finding a globally optimal solution that avoids local minima.
v. Validating the neural network to test for overfitting.

4.5 Hidden Layers Selection

In my study one hidden layer is sufficient for the network. Two hidden layers are required for modeling data with discontinuities such as a saw tooth wave pattern. As we found that using two hidden layers rarely improves the model, and it may introduce a greater risk of converging to a local minima. So, Three layer models with one hidden layer are recommended for our study.

4.6 Deciding How Many Neurons to be Used in the Hidden Layers

The most significant characteristics of a multilayer perceptron network is to decide the number of neurons in the hidden layer. The network may be unable to model complex data, and the resulting fit will be poor, if an inadequate number of neurons are used in the network. Similarly, if too many neurons are used, the training time may become excessively long, and, worse, the network may overfit the data. When overfitting occurs, the network will begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data. Validation must be used to test for this. In view of the above our model consists of 5 neurons with one hidden layer.

5. RESULTS AND DISCUSSION

During data analysis, the column type is recognized. The last column is considered as the target or output one and other columns will be considered as input columns. The dataset is divided into training, validation and test sets. The Data have been analyzed using Neuro-intelligence tool [16].Table 4 shows the statistics of data partition sets.

Partition set using	Records	Percentage (%)
Total	94	100
Training Set	64	68
Validation Set	15	16
Test Set	15	16
Ignore Set	0	0

TABLE 4: Data Partition Set

To train a neural network is the process of setting the best weights on the inputs of each of the units. It has been proved that Genetic Algorithm and Back-Propagation neural network hybrids in selecting the input features for the neural network reveals the performance of ANN can be improved by selecting good combination of input variables [13]**.** Training set is considered to be the part of the input dataset used for neural network training and network weights adjustment**.** The validation set is parts of the data are used to tune network topology or network parameters other than weights. The validation set is used to choose the best network we have changed the number of units in the hidden layer. The test set is a part of the input data set used to test how well the neural network will perform on new data. The test set is used after the network is trained, to test what errors will occur during future network application.

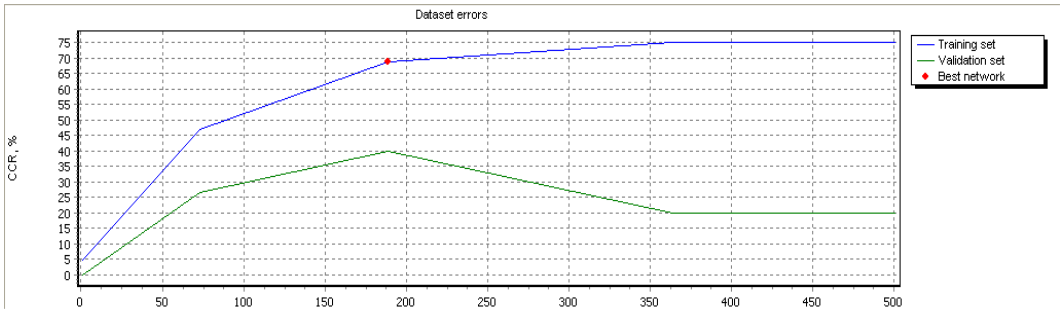

FIGURE 3: Error in Data Set

Figure 3 shows the various data set errors with respect to training set, validation set and the best network. It accomplishes the level of best network after training through repeated iterations. Correct Classification Rate for training and validation has been done to find the best network after a number of iterations. Table 5 shows the number of Iterations and CCR for training and validation as well.

Iteration	CCR (training)	CCR (validation)
73	46.875	26.666666
189	68.75	40
364	75	20

TABLE 5: Best Network on Iterations

The Network errors have been shown graphically in figure 4. We have tested the trained network with a test set, in which the outcomes are known but not provided to the network. We used diagnostic criteria and disease pattern status to train a neural network to classify individuals as diagnosed with disease name by several categories of neonatal disease.

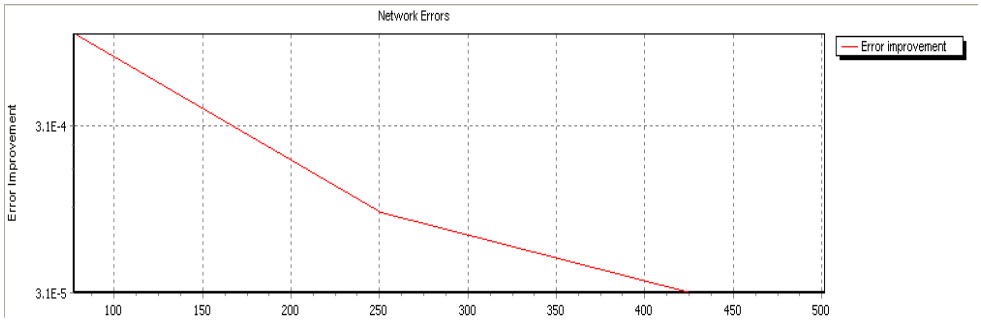

FIGURE 4: Network Error

The study shows that 39.36% of the respondents have the symptoms of Septicemia; 17.02% have the symptoms of HIE III; and 9.57% of the patients have the symptoms of Metabolic Disorder - Hypocalcemia. These are the most prevalent disease in the Terai region of North Bengal [2]. Table 6 shows the disease conformation percentage with category. Disease conformation is also presented in Fig. 5 representing disease vs. number of cases.

No. of cases	Name of Disease Conformation	Percentage (%)
1	Hemorrhage	1.06%
4	HIE_II	4.26%
16	HIE_III	17.02%
2	Hypo_Thalmia	2.13%
2	Jaundice	2.13%
3	Jaundice_BA	3.19%
9	MD_Hypocalcemia	9.57%
3	MD_Hypoglycemia	3.19%
3	MD_Hypothermia	3.19%
8	No_Disease	8.51%
5	Others	5.32%
37	Septicemia	39.36%
1	Sizure_Disorder	1.06%

TABLE 6: Disease Conformation Set

6. CONCLUSION

Neural network has been established of their potentials in many domains related with medical disease diagnosis and other application. Although, Neural networks never replace the human experts instead they can helpful for decision making, classifying, screening and also can be used by domain experts to cross-check their diagnosis. In our earlier studies on rough set based computing model [17] and soft computing model [18], we have established the accuracy of 71% for decision making of prevalence neonatal disease. This ANN MLP model proves the better results and helps the domain experts and even person related with the field to plan for a better diagnose and provide the patient with early diagnosis results as it performs realistically well even without retraining. As clinical decision making requires reasoning under uncertainty, expert systems and fuzzy logic will be suitable techniques for dealing with partial evidence and with uncertainty regarding the effects of proposed interventions. Neural Networks have been proven to produce better results compared to other techniques for the prediction tasks. Our study concludes with higher prediction result and when the Network has trained and tested after optimizing the input parameters, the overall predictive accuracy acquired was 75%.

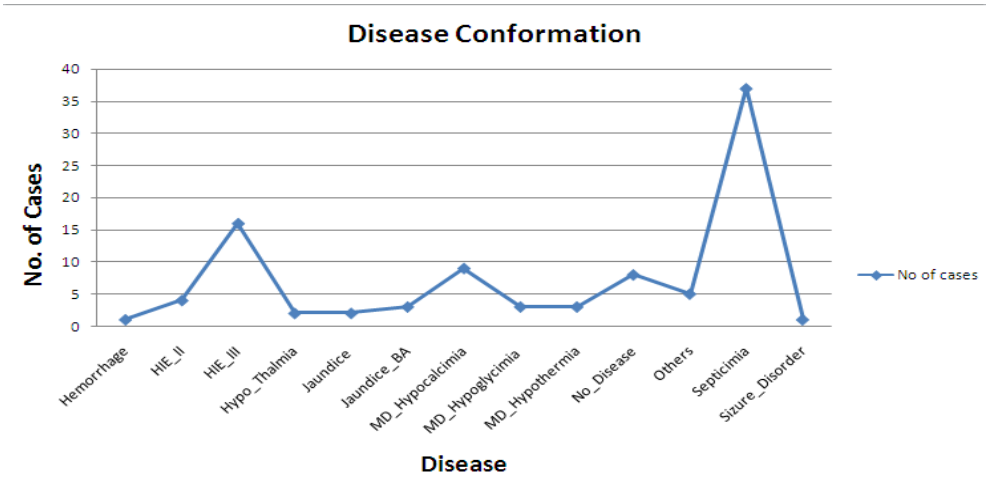

FIGURE 5: Various Neonatal Disease with no. of cases

A comparative study [19] is being presented in table. 7 to establish the relative suitability of ANN technique with other techniques such as RSES [20] and ROSSETTA [21]. The result of the table clearly demonstrates the superiority of ANN technique over other techniques explained earlier.

Tools	Methods/ Algorithms	Prediction Accuracy (%)
RSES[20]	Exhaustive without Reduct	70
	Genetic without Reduct	70
	Exhaustive with Reduct	70
	Genetic with Reduct	70
	Dynamic with Reduct	70
ROSETTA[21]	Genetic without Reduct	71.6
	Johnson(with approx. solutions) with Reduct	70.5
NEURO INTELLIGENCE [16]	ANN with MLP	75

TABLE 7: A Comparative Study of Different Techniques

7. REFERENCES

[1] D. Kumar, A.Verma, and V. K. Sehgal.(2007). "Neonatal mortality in India." Rural and Remote Health. [On-line]. 833(7). Available: www.rrh.org.au [October 8, 2009].

[2] D. R. Chowdhury, R.K Samanta and M. Chatterjee. "A Study of the status of new born in Terai region of West Bengal". Modelling, Measurement and Control C, France. vol. 68(1), pp. 44-52, 2007.

[3] T. A. Bang, A. R. Bang, S. Baitule, M. Deshmukh and H. M. Reddy. "Burden of Morbidities and the Unmet Need for Health Care in Rural Neonates - A Prospective Observational Study in Gadchiroli, India." Indian Journal of Pediatrics, vol. 38, pp. 952-965, 2001.

[4] D. R. Chowdhury, M. Chatterjee and R.K Samanta. "A Data Mining Model for Differential Diagnosis of Neonatal Disease." International Journal of Computing. Vol. 1(2), pp. 143-150, 2011.

[5] X. Qiua, N. Taob and Y. Tana, et al. "Constructing of the Risk Classification Model of Cervical Cancer by Artificial Neural Network. Expert Systems with Applications." An International Journal Archive. Vol. 32(4), pp. 1094-1099, 2007.

[6] B. Zernikow, K. Holtmannspoetter, E. Michel, W. Pielemeier, F. Hornschuh, A. Westermann, and K. Hennecke. "Artificial neural network for risk assessment in preterm neonates." Arch Dis Child Fetal Neonatal Ed. vol. 79(2), pp. F129–F134, 1998.

[7] F.Yaghouby, A. Ayatollahi and R. Soleimani, "Classification of Cardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal." World Applied Sciences Journal. vol 6.(11), pp. 1547-1554, 2009.

[8] D. Shanthi, G. Sahoo and N. Saravanan. "Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke." International Journals of Biometric and Bioinformatics (IJBB). vol 3(1), pp. 10-18, 2009.

[9] "Artificial Neural Networks in Medicine World Map," USENET: http://www.phil.gu.se/ann/annworld.html, [July 21, 2011].

[10] Z. H. Zhou and Y. Jiang. "Medical Diagnosis with C4.5 Rule Preceded by Artificial Neural Network Ensemble." IEEE Transaction on Information Technology in Biomedicine. vol 7(1), pp. 37-42, Mar. 2003.

[11] WG. Baxt. "Application of artificial neural networks to clinical medicine" Lancet. vol. 346 (8983), pp. 1135-1138, 1995.

[12] MR. Narasingarao , R. Manda, GR. Sridhar, K. Madhu and AA. Rao. "A Clinical Decision Support System Using Multilayer Perceptron Neural Network to Assess Well Being in Diabetes." Journal of the Association of Physicians of India. vol. 57, pp. 127-133, 2009.

[13] D. Shanthi, G. Sahoo and N. Saravanan. "Input Feature Selection using Hybrid Neuro-Genetic Approach in the diagnosis of Stroke." International Journal of Computer Science and Network Security. ISSN 1738-7906. vol. 8(12), pp. 99-107, 2008.

[14] F. Ahmad, A. M. Nor, Z. Hussain, R. Boudville and K. M. Osman. "Genetic Algorithm - Artificial Neural Network (GA-ANN) Hybrid Intelligence for Cancer Diagnosis." In Proc. Second International Conference on Computational Intelligence, Communication Systems and Networks, IEEE Computer Society, 2010, pp. 78-83.

[15] A. Blais and D. Mertz. An Introduction to Neural Networks – Pattern Learning with Back Propagation Algorithm. Gnosis Software Inc. 2001.

[16] "Neuro Intelligence using Alyuda", http://www.alyuda.com, 2008 [May 11, 2011].

[17] D. R. Chowdhury, M. Chatterjee and R.K. Samanta. "Rough Set Based Model for Neontal Disease diagnosis" International Conf. on Mathematics and Soft Computing, ICMSCAE, Panipath, India, 2010.

[18] D. R. Chowdhury, M. Chatterjee and R.K Samanta. "Neonatal Disease Diagnosis with Soft Computing", in Proc. International Conf. on Computing and System, ICCS, University of Burdwan, India, 2010, pp. 27-34.

[19] D. R. Chowdhury, R.K Samanta and M. Chatterjee. "Design and Development of an Expert System Model in Differential Diagnosis for Neonatal Disease". International Journal of Computing, vol 1(3), pp. 343-350, 2011.

[20] "RSES 2.2 User's Guide, Warsaw University," USENET: http://logic.mimuw.edu.pl/~rses [January 19, 2010].

[21] "The ROSETTA homepage". Internet: http://www.idi.ntnu.no/~aleks/rosetta/, [January 25, 2010].