CLASSIFYING THE INFERTILITY TREATMENT SUCCESS RATE BASED ON LEARNING FROM MULTI LAYER PERCEPTRON NETWORK

Dr. R. NANDHAKUMAR[1], Dr. M. DURAIRAJ[2]

[1]Assistant Professor, Pg and Research Department of Computer Science and Applications, Vivekanandha College of Arts And Sciences for Women (Autonomous), Tiruchengode, Namakkal, India.

[2]Assistant professor, Department of Computer Science and Engineering, Bharathidasan University, Tiruchirapalli, India.

ABSTRACT

Artificial Neural Network is permitted as an influential classifier for medical diagnosis for early recognition and success rate prediction of diseases. In this work, ANN is used for predicting the success rate in Infertility treatment. Multi-Layer Perceptron Network (MLPN) with Feed-forward back propagation learning algorithm is used as a classifier for classifying the patient record as success (1) or un success (0). In the proposed work, MLPN with 5 hidden neurons in the single hidden layer is used to predict the infertility treatment. This MLPN shows good performance in predicting the treatment success level with improved accuracy and less error. The MLPN is trained and tested with 18 and 10 features in the data set. The data set with 10 features is obtained by hybriding the Ant Colony Optimization Algorithm with Rough Set Theory.

Index Terms*—*Artificial Neural Network (ANN), Multilayer Perceptron, Back Propagation Learning, Hidden Neurons, Sensitivity, Specificity, Accuracy, False Positive Rate.

Introduction

Infertility is defined as the incapability to attain pregnancy after one year of unprotected intercourse. It is difficult a rising amount of married pairs about the planet. It is considered as the most important health problems in developing countries [ ]. WHO has also accredited infertility as a public health concern [1, 2]. 60-80% couples are infertile throughout the world among which 20-25% is from India [3]. The causes for the infertility are due to male factor, female factor, combination of both male and female factor and also due to unexplained factors. The effective ways like ART (Assisted Reproductive Technologies), In IVF (Vitro Fertilization) has evolved in medical field to deal with this cause. The achievement rate attained by this behaviour has been enlarged newly up to 10 percent. But it still fits only to the 40% of the people [4]. It is a challenging task for an embryologist to analyze and associate the large number of features. The technology is still lagging behind to improve the success rate in infertility treatment. To fill the gap of the technology lagging, Data Mining is introduced as an intelligent diagnostic and classification tool. The process of extracting value from the database is defined as Data Mining. To accomplish this aim, this paper expands a Multilayer Perceptron network and trains it to increase the success rate prediction. This work reduced the total number of features by combining Ant Colony Optimization Algorithm and Relative Reduct Algorithm.

The paper is organized as the portions: Section II examines several of previous studies carried out in calculating the achievement rate of IVF treatment. Section III shorts the data set used for the testing. Section IV illustrates how the Multilayer Perceptron Network (MLPN) is set for training. It also examines how the back propagation algorithm is employing for learning the network. Section V considers briefly about the results obtained. And finally the paper is concluded in Section VI.

Artificial Neural Network in Classification

Er, Yumusak and Temurtas [5] presented a comparative study based on multi-layer, probabilistic, knowledge vector normalization and general regression for chest disease diagnosis. Das, Turkoglu and Sengur [10] constructed a neural network ensemble methodology for diagnosing the heart disease. SAS enterprise miner 5.2 is used to construct the required network.

Moein, Monadjemi and Moallem [6] analyzed the procedure of medical diagnosis by converting it into machine implementable format. Symptoms from eight different diseases are taken and applied to MLP neural network. The results obtained support the role of data Fuzzification for the neural network based automatic medical diagnostic system. Lin [7] used Classification and Regression Tree (CART) and Case Based

Reasoning (CBR) techniques for intelligent diagnosis to diagnose the liver disease.

In this paper [8] the author's premeditated apply of neural network in data mining. It comprise the experiential that utilize of neural network is especially broad in data mining suitable to various qualities like Similar Performance, Self-organizing adaptive, Robustness and Fault Tolerance. It was utilized for business applications similar to pattern recognition, Classification, Prediction.

In this paper [9] the author discussed a data mining application of Neural Network (NN) is extremely broad. The NN can be exploited to model difficult associations among input and output or to discover patterns in data. Data mining procedure is composed by three key stages: Data preparation, Rule Extracting and Rules Assessments. The collection of neural network model and data mining technique can significantly raise the effectiveness of data mining techniques and it has been largely used [9].

This study [11] describes that suitable to an improvement in information technology, the field of Business Intelligence (BI) and Data Mining (DM) arose. The objective of BI/DM is to estimate knowledge from raw data. For Neural network (NN), flexible and Support Vector Machine's (SVMs) and nonlinear classification method, appropriate to analytical presentations they are capable. In general the outcomes achieved are aggressive. In mostly the NN for regression ones and SVM model for the categorization task [11].

In this paper [12] the author reported that the back propagation technique is used for supervised neural network and was developed for metal alloys. For the training of artificial networks it requests Unsupervised Learning, Supervised Learning and Reinforcement Learning [12]. The neural network can be separated into subsequent type of feed forward neural network and recurrent neural network.

The major aim of data mining [15] is to mine knowledge from enormous amount of data. The well-known approaches to data mining incorporate neural and symbolic form. Fuzzy neural network method is used as channel among numerical data and symbolic representation. By using fuzzy logic the authors expressed knowledge in such an approach which is accepted for the people to recognize.

Methodology

For classifying the infertility data set as success or un success, Multi Layer Perceptron Network (MLPN) is used. The network is trained and tested with the Back Propagation Learning Algorithm. The Training and Testing Algorithm used for learning the MLPN is as follows:

Step 0: *Initialize the weight and learning rate*

Step 1: *Perform Steps 2 to 9 when stopping condition is false*

Step 2: *Perform Steps 3-8 for each training pair*

Feed-Forward Phase (Phase I):

Step 3: *Every input element accepts input signal x_i and forwards it to the hidden unit (i=1 to n)*

Step 4: *Every hidden unit z_j (j=1 to n) summations its weighted contribution signals to compute net input:*

$$Z_{inj} = V_{oj} + \sum_{i=1}^{n} x_i v_{ij}$$

Compute outcome of the hidden unit by relating its establishment function over Z_{inj}

$Z_j = f(Z_{inj})$ where

$$f(Z_{inj}) = \frac{1}{1+e^{-\lambda Zinj}}$$

Step 5: *For every output unit y_k (k=1 to p), compute the net input:*

$$Y_{ink} = w_{ok} + \sum_{j=1}^{p} zjv_{jk}$$

And be relevant the activation function to analyse result signal

$$Y_k = f(Y_{ink})$$

$$f(Y_{ink}) = \frac{1}{1+e^{-\lambda Yiink}}$$

Back-propagation of error (Phase II):

Step 0: *Initialize the weights. The weights are occupied from the training algorithm*

Step 1: *Perform Steps 2-4 for each input vector.*

Step 2: *Set the activation of input unit for xi (i=1 to n).*

Step 3: *Compute the net input to hidden unit x and its output. For j=1 to p,*

$$z_{inj} = v_{0j} + \sum_{i=1}^{n} x_i v_{ij}$$

$$z_j = f(z_{inj})$$

Step 4: *Now add the output of the output layer unit. For k=1 to m,*

$$y_{ink} = w_{ok} + \sum_{j=1}^{p} z_j w_{jk}$$

$$y_k = f(y_{ink})$$

Sigmoidal activation function is used for calculating the output.

Step 6: *Every result element y_k (k=1 to m) accepts a objective pattern equivalent to the input training pattern and compute the error modification term:*

$$\delta_k = (t_k - y_k) f'(y_{ink})$$

where

$$f'(Y_{ink}) = \lambda\, f(Y_{ink})[-f(Y_{ink})]$$

On the beginning of the considered error modification term, the weight and unfairness is updated as follows:

$$\Delta W_{jk} = \alpha \delta_k z_j$$

$$\Delta w_{ok} = \alpha \delta_k$$

Also, transmit δ_k to the hidden layer backwards

Step 7: *Every hidden unit (z_j, j= 1 to p) calculations its delta inputs from the result units:*

$$\delta_{inj} = \sum_{k=1}^{m} \delta_k w_{jk}$$

The term δ_{inj} gets increased with the derivative of (Z_{inj}) to compute the error terms:

$$\delta_j = \delta_{inj} f'(Z_{inj})$$

where

$$f'(Z_{inj}) = \lambda\, f(Z_{inj})[1 - f(Z_{inj})]$$

On the starting point of planned δ_j, the weight and partiality are considered as :

$$\Delta V_{ij} = \alpha \delta_j x_i$$

$$\Delta V_{oj} = \alpha \delta_j$$

Weight and bias updation (Phase III):

Testing Algorithm:

Step 8: *Every outcome element (y_k, k= 1 to m) revises the unfairness and weights:*

$w_{jk}(new)=w_{jk}(old)+\Delta w_{jk}$

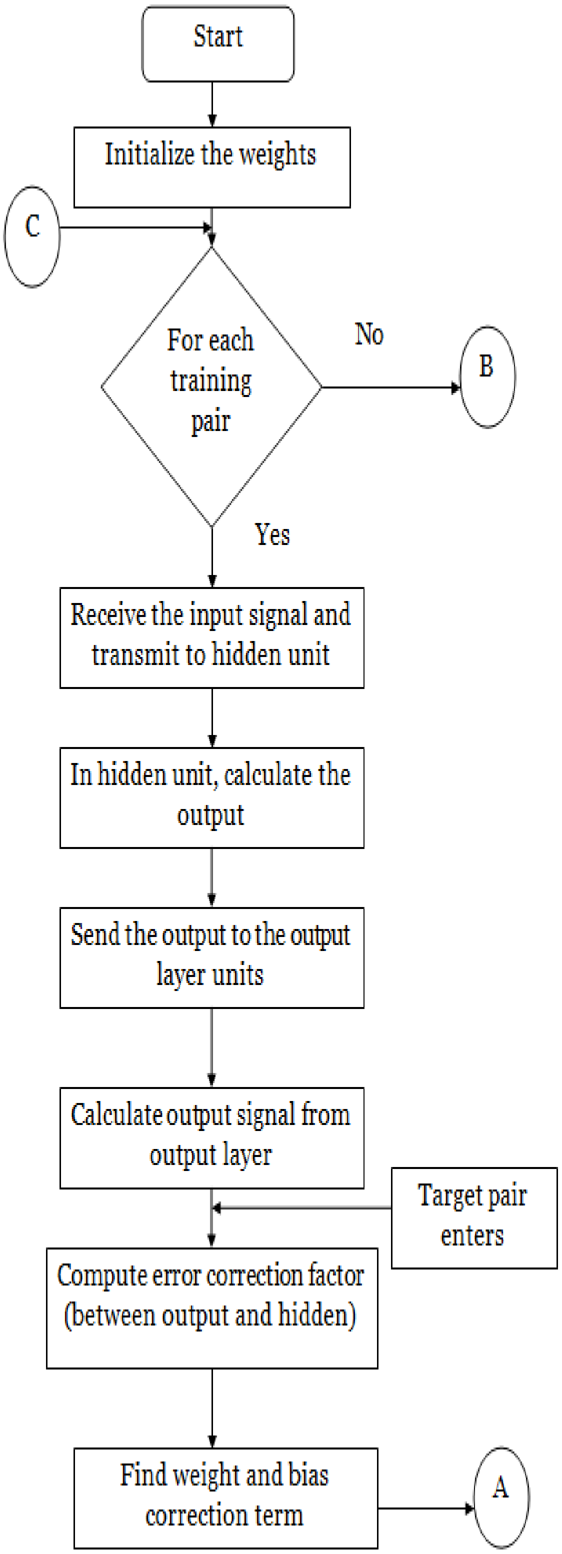
$w_{0k}(new)=w_{0k}(old)+\Delta w_{0k}$

Every hidden unit(z_j=j=1 to p) updates its unfairness and weights:

$v_{ij}(new)=v_{ij}(old)+\Delta v_{ij}$

$v_{0j}(new)=v_{0j}(old)+\Delta v_{0j}$

Step 9: *Ensure for the stopping condition. The stopping condition might be confident amount of epochs reached or when the definite result equals the objective output.*



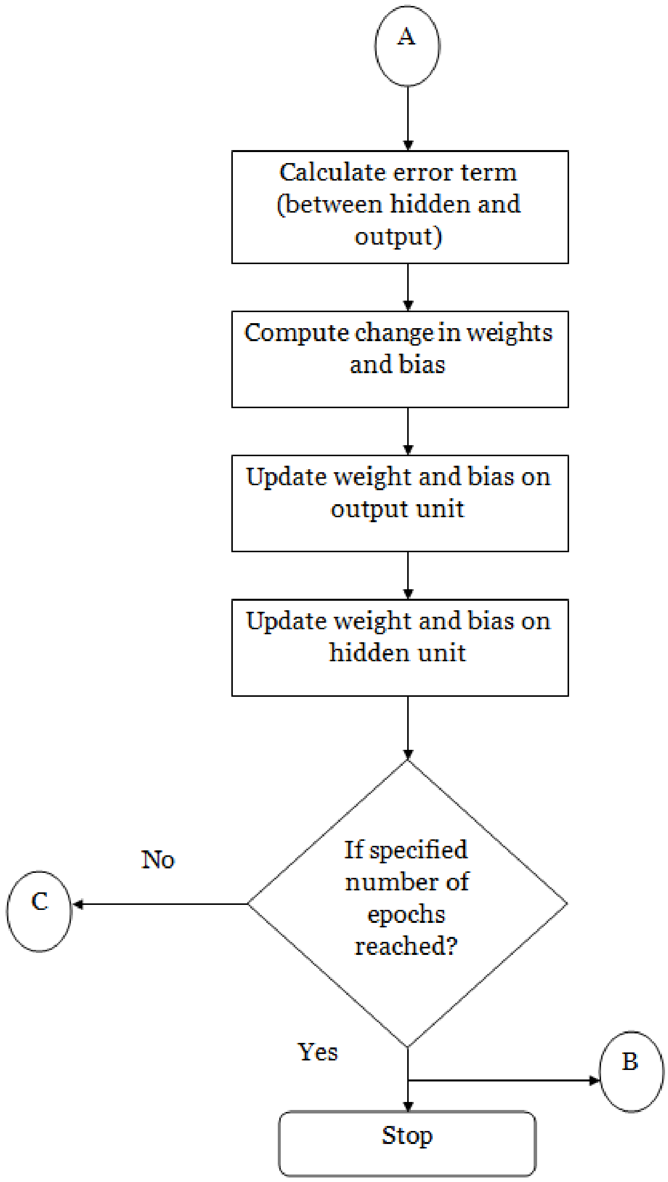


Figure 1: Training of the Back Propagation Learning Network

Data Set

The data set employed for the testing is gathered from different Hospitals, Fertility clinics, and Research centres in Tamil Nadu. This data set has 42 features. Between all the 42 features (attributes), 34 features is taken for the research based on the doctor's suggestion.

Attributes used for this work				
Name	Earlier Surgery	Endometriosis	Liquefaction Time	Male Factor Only
Unknown feature	Pre-Existing Symptoms of Depression	Tubal Infertility	Sperm Concentration	Severe Male Factor

Place	Fear And Negative Treatment Attitude	Ovulatory Factor	Sperm Motility	Female Factor Only
IVF Treatment	Psychological And Emotional Factors	Hormonal Factor	Sperm Vitality	Combined Factor
Miscarriage	Difficulty In Tolerating Negative Emotions For Extended Time	Cervical Factor	Sperm Morphology	Unknown Factor
Miscarriage Causes	Improbability	Unsolved feature	Amount of Oocytes recovered	Place
Medical Disorders	Strain Of Repeated Treatment	Semen Ejaculate Volume	Amount of Embryos Transferred	IVF Treatment

Table 1: Attributes used for this work

The list of features given in Table 2 is taken for reduction based on doctor's suggestion.

Semen Ejaculate Volume	
Liquefaction Time	
Sperm Concentration	Sperm Concentration
Sperm motility	Sperm motility
Sperm vitality	
Sperm morphology	
No. of oocytes retrieved	No. of oocytes retrieved
No. of embryos transferred	No. of embryos transferred
Male factor only	Male factor only
Severe male factor	
Female factor only	
Combined factor	
IVF Treatment	IVF Treatment

Table 2: List of attributes chosen for experimentation

Experimentation and Results

The proposed MLPN is trained by using the Scaled Conjugate Gradient (SCG) Algorithm. A number of algorithms like Levenberg-Marquardt (LM) Algorithm, Resilient back propagation RPROP (RP) Algorithm, BFGS quasi-Newton (BFG) Algorithm, Gradient descent with momentum and adaptive linear propagation (GDX) Algorithm are there to train the network. Table 3 discusses the features of the algorithm used for training the network. It is observed that the scg algorithm performs well even in increase in the number of weight and there is no decrease in the performance of the network. And it has only the modest memory requirement. This factors lead to use the scg Algorithm for training the MLPN.

Table 3: Back propagation learning algorithms

Algorithms	**Memory Requirements**	**Problem Applied**	**Convergence speed**	**Performance**
lm Algorithm	High	Function Approximation	Fastest	Performance degrade with the increase in weight
RP Algorithm	Small	Pattern Recognition	Fastest	Performance degrade with the increase in weight
SCG Algorithm	Modest	Function Approx	Fast	Performance doesn't

		imation and Pattern Recognition		grade. Perform well for network with large number of weight
BFG Algorithm	High	Function Approximation	Fastest	Performance degrade with the increase in weight
GDX Algorithm	Low	Function Approximation	Slower	Performance degrade

The Network has three layers namely, input layer, hidden layer and output layer. The input layer has 18 inputs and output layer has one input which depicts whether the results is success or unsuccessful. Fixing the amount of neurons in the hidden layer is an important task. If the network is tested with the low number of neurons in the hidden layer mean, there is chance for under fitting whereas enlarge in the amount of neurons which are not acceptable by the network will result in over fitting. The theory behind fixing number of hidden neurons in the hidden layer is:

[1] The finest dimension of the hidden layer must be among the dimension of the input and output layers.

[2] Amount of neurons in the hidden layer is equivalent to the mean of amount of neurons in the input layer and output layer

To discover the amount of hidden neurons, equation 1 is used.

$$N_h = \frac{N_s}{(alpha * (N_i + N_o))} -----> (1)$$

Where,

N_k = Amount of neurons in the hidden layer

N_i = Amount of neurons in the input layer

N_o = Amount of neurons in the output layer

Alpha = arbitrary scaling factor (value lies between 2-10)

The trial and error technique is used to fix the amount of neurons. The Table 4 depicts the value of Mean Squared Error (MSE) and Accuracy obtained in the case of training the network with the specified amount of neurons. The layer with number of neuron 5 shows more accuracy than the other ones. Figure 2 portrays the comparison in the accuracy value with the different number of neurons. The hidden layer with 5 neurons has the highest accuracy of 73.7%. Hence the MLPN network is trained using the SCG Algorithm with 5 numbers of neurons in the hidden layer. The hidden layer is chosen to be one in number. The number of hidden layers can be increased depending on the complexity of the problem. The MSE value should be lower as possible for producing promising results.

Table 4: Performance of the network with different number of hidden neurons

No. of neurons in the hidden Layer	**Time taken for execution (seconds)**	**MSE**			**Accuracy (%)**
		Training Value	Validation Value	Testing Value	
1	7	1.76779e-1	1.90260e-1	2.36664e-1	71.9
2	2	1.81819e-1	1.92901e-1	2.51144e-1	71.1

3	9	2.67856e-1	2.32561e-1	2.32337e-1	67.5
5	**4**	**1.82542e-1**	**1.66082e-1**	**2.42003e-1**	**73.7**
6	7	5.46731e-1	4.76213e-1	6.73211e-1	43.9

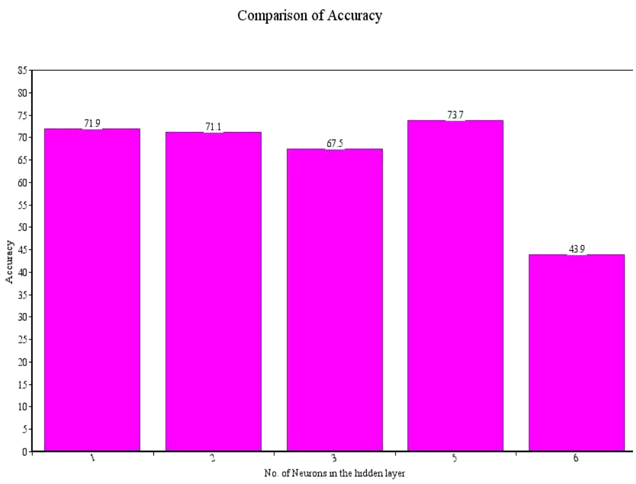


Figure 2 : Comparison of Accuracy

The metrics like Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Positive Rate (FPR), False Discovery Rate (FDR) and False Negative Rate (FNR) is used to estimate the performance of the network with altered amount of neurons.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad ----> (2)$$

Specificity is described as the determine of proportions of negatives that are accurately recognized. It is also symbolized as True Negative Rate.

$$Specificity = \frac{TN}{(TN + FP)} \quad ----> (3)$$

PPV is defined as

$$PPV = \frac{TP}{(TP + FP)} \quad ----> (4)$$

The Balance of PPV is the False Discovery Rate(FDR)

$$FDR = \frac{FP}{(TP + FP)}$$

$$or$$

$$FDR = 1 - PPV ----> (4)$$

The Negative Predictive Value (NPV) is defined as

$$NPV = \frac{TN}{(TN + FN)} \quad ----> (5)$$

False Positive Rate (FPR) (α) is defined as

$$FPR = \frac{FP}{(FP + TN)}$$

$$or$$

$$FPR = 1 - Specificity ----> (6)$$

FPR is the Type I error. It occurs when the null hypothesis is rejected even when it is true. FNR (β) is the Type II error which occurs when the null hypothesis is accepted, when it is false. It is defined as

$$FNR = \frac{FN}{(TP + FN)}$$

$$or$$

$$FNR = 1 - TPR \quad ----> (7)$$

Where,

TP- True Positive

FN- False Negative

FP- False Positive

TN- True Negative

Table 5 displays the results obtained when experimenting with different number of neurons. The results obtained with 5 hidden neurons shows promising results than the other neurons in the hidden layer. Figure 3 depicts the results obtained while training the network with different number of neurons in the hidden layer.

Table 5 : Results obtained using different number of neurons in the hidden layer

No. of neurons in the hidden Layer	Sensitivity	Specificity	PPV	NPV	FPR	FDR	FNR
1	0.62	0.84	0.83	0.63	0.16	0.17	0.37
2	0.61	0.83	0.83	0.62	0.16	0.17	0.38
3	0.58	0.82	0.83	0.56	0.17	0.17	0.42
5	**0.65**	**0.83**	**0.81**	**0.68**	**0.17**	**0.18**	**0.35**
6	0.43	1	1	0.03	0	0	0.57

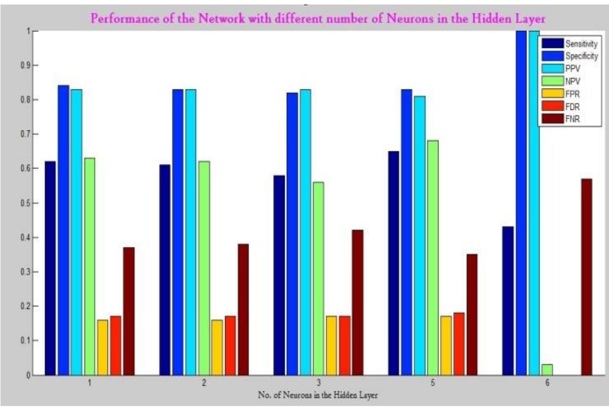


Figure 3: Results obtained using different number of neurons in the hidden layer

The Figure 4 shows the performance plot for the dissimilar number of neurons. In the Performance plot, the test and Validation curve should be the same. If the test curve increase over the validation curve, there is possible for the network to over fit. Training curve should decrease along with the increase in the number of iterations. It was observed that the performance plot for the network with 5 hidden neurons is better than the other ones.

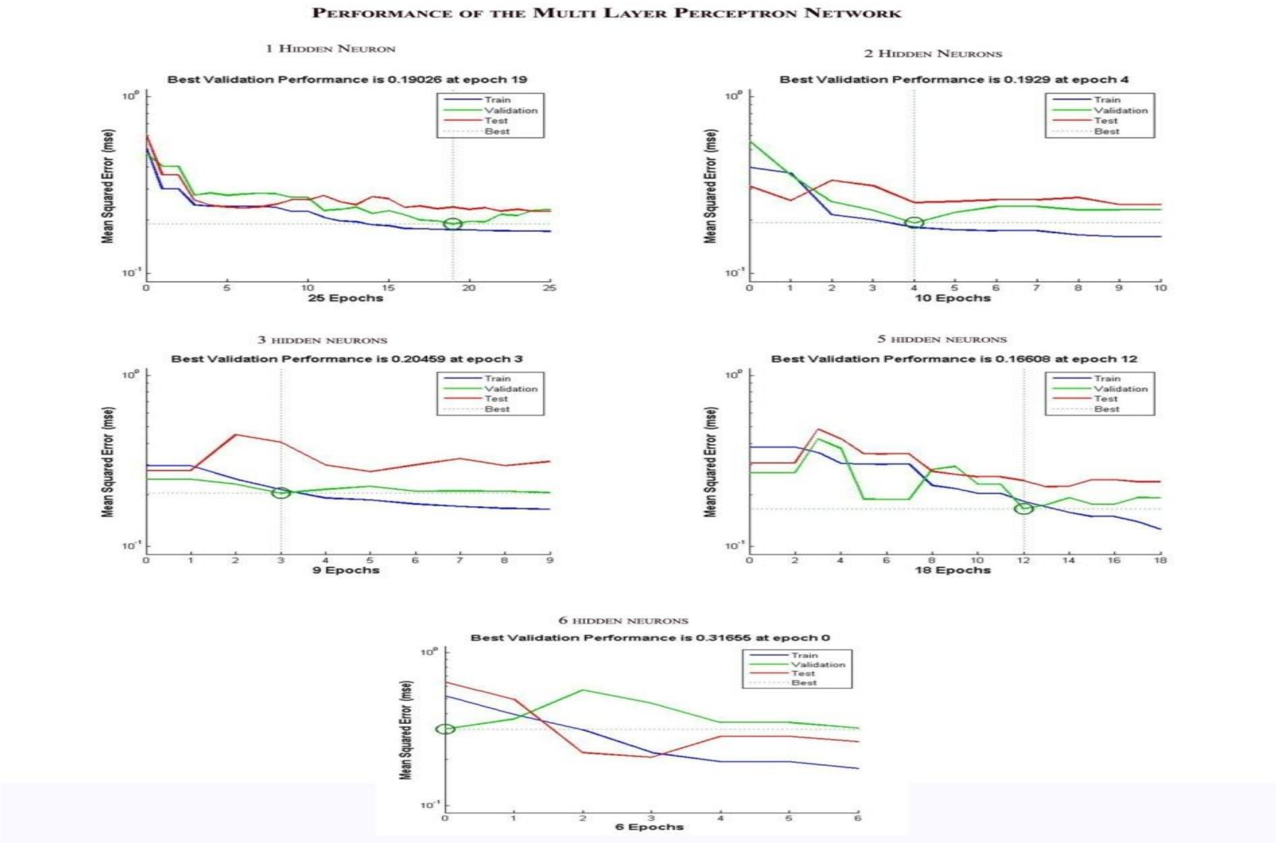


Figure 4 : Performance plot for various number of neurons in hidden layer

Figure 5 depicts the training state of the network. When the training MSE decreases, then the testing and validation MSE should decrease too. The validation fails when the validation MSE increase in Value.

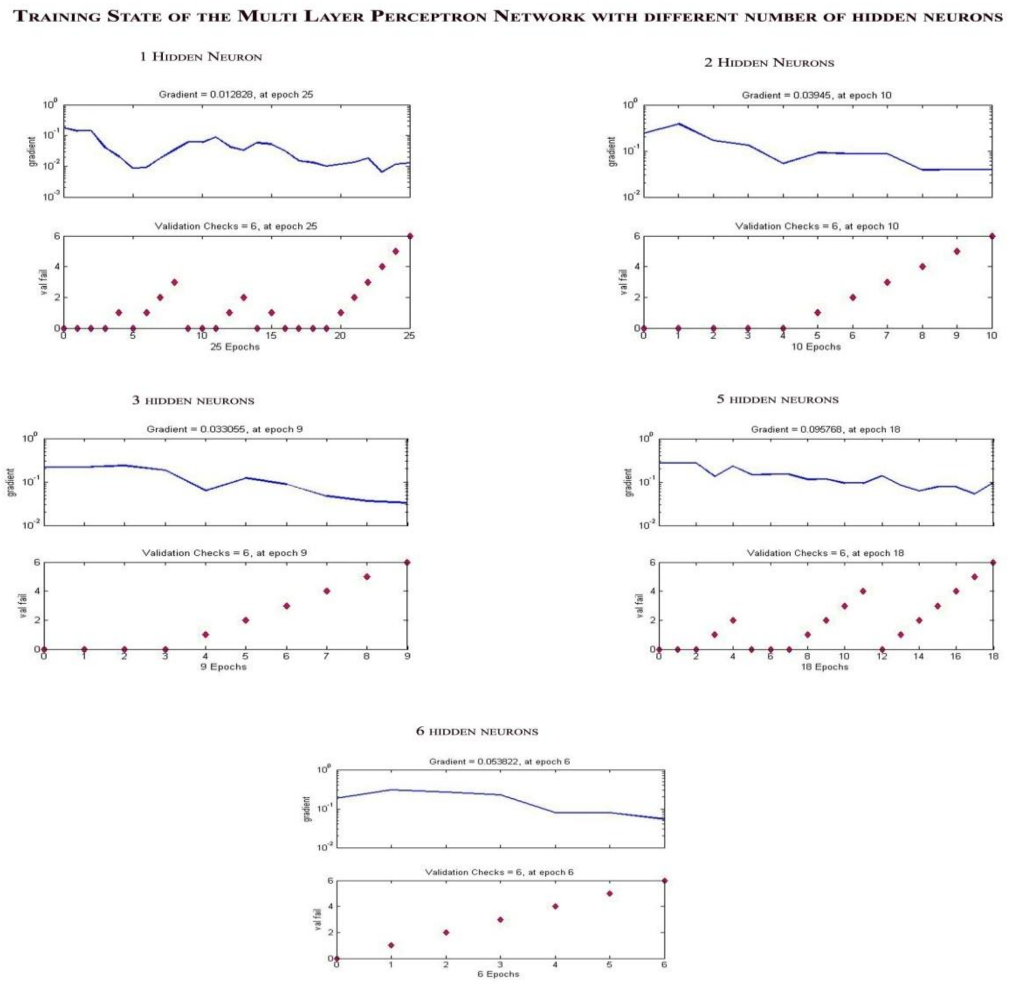


Figure 5: Training state for the different neurons

Figure 6 shows the confusion matrix plot obtained while training the network. The figure explictly proves that the accuracy obtained with 5 hidden neurons is higher than the other results. Figure 7 shows the Receiver Optimal Characteristic (ROC) Plot obtained while training the network. It is plotted with True Positive Rate with False Positive Rate. The Confusion Matrix and ROC plot obtained using the 5 hidden neurons network is promising while comparing the Confusion Matrix of other neurons.

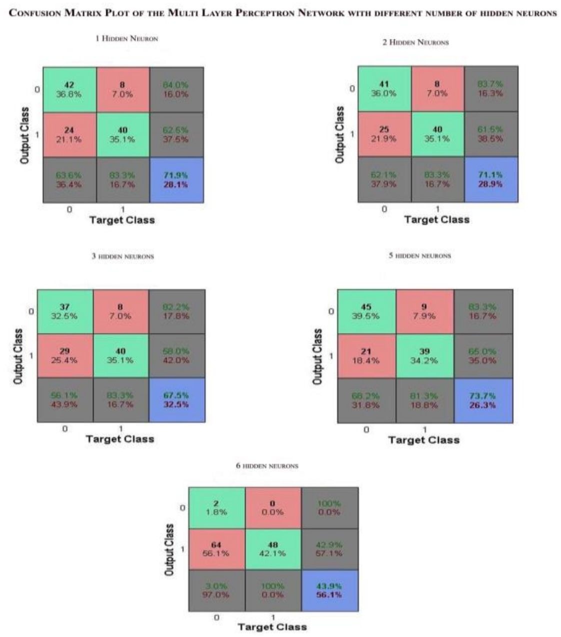


Figure 6: Confusion Matrix for the Network

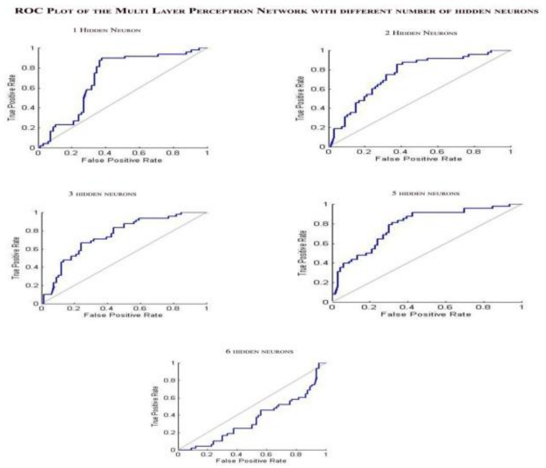


Figure 7: ROC Curve for the Network

The optimal data set obtained from the Algorithm is then trained by the MLPN with SCG as training algorithm. The network has 10 input neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. The results obtained are very promising. The results obtained from the network classified the outcome as Success or Unsuccess. The Figure 8 shows the results obtained from the optimal data set.

PERFORMANCE OF THE NETWORK WITH THE OPTIMAL DATA SET

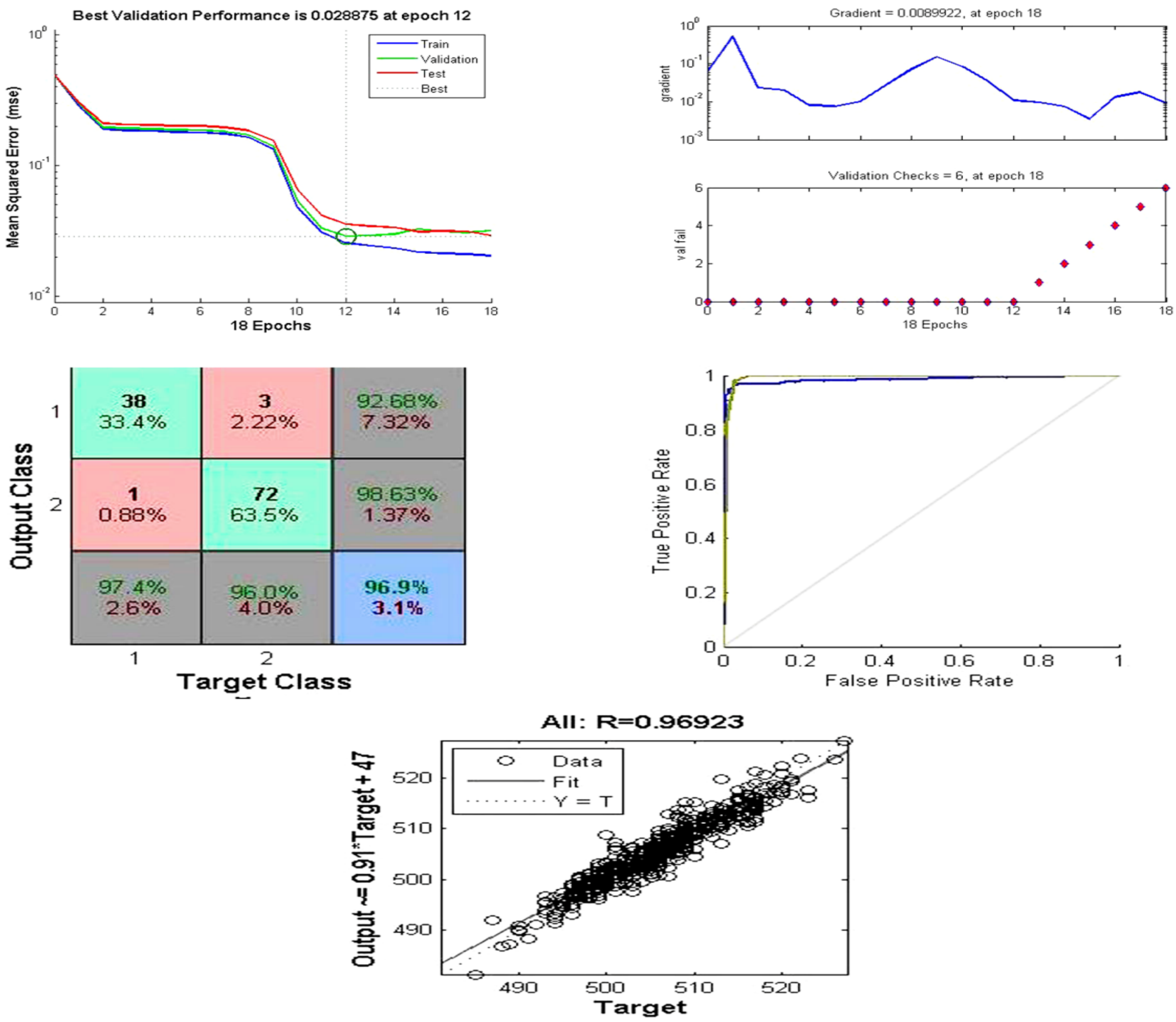


Figure 8: Performance of the Network with optimal data set obtained using hybrid Algorithm

Figure 9 depicts the comparison of the performance of the data set with 18 features and optimal features. The accuracy is increased to 96.9% from 73.7 %. This promising results obtained indicated the MLPN is best for classifying and recognizing the pattern of the network. The PPV, FNR, FDR, FPR is higher in the optimal data set and the NPV is reduced. The Sensitivity, Specificity and Accuracy in classification are also increased by training and testing the features of the optimal data set rather than the original data set with 18 features.

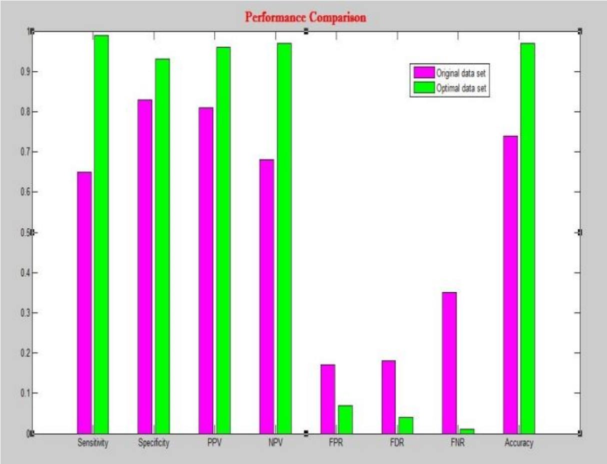


Figure 9: Comparison of the Performance

CONCLUSION

In this work, the MLPN is employed to predict the success rate of the Infertility treatment. It is observed that the success rate is improved 96.9%. While comparing the performance of the network with other metrics like Sensitivity, Specificity, TPR, FPR, PPV, NPV, FDR and FNR, the proposed MLPN with 5 hidden neurons in the hidden layer performs well. The number of hidden layers in the network depends upon the complexity of the problem. As there is possible for the problem size to increase, there is option to increase the number of hidden layers and test the network. This MLPN with Feed Forward SCG Back propagation Learning Algorithm perform well in classifying the patients for the infertility treatment with a promising accuracy of 96.9%.

References

1. Ombelet, Willem et al. "Infertility and the Provision of Infertility Medical Services in Developing Countries." *Human Reproduction Update* 14.6 (2008): 605–621. *PMC*. Web. 11 Aug. 2016.
2. Ombelet, W. "The Walking Egg Project: Universal Access to Infertility Care – from Dream to Reality." *Facts, Views & Vision in ObGyn* 5.2 (2013): 161–175.
3. Poongathai J, Gopenath TS, Manonayaki S, " Genetics of human male fertility", Singagapore Medical Journal, 50(4), 336-47, April 2009.
4. Milewski, R., Milewska, A. J., Czerniecki, J., Leśniewska, M., & Wołczyński, S. (2013). Analysis of the demographic profile of patients treated for infertility using assisted reproductive techniques in 2005–2010. Ginekologia Polska. 84(7), 609–614.
5. O. Er, N. Yumusak and F. Temurtas, "Chest disease diagnosis using artificial neural networks", Expert Systems with Applications, Vol.37, No.12, 2010, pp.7648-7655.
6. S. Moein, S. A. Monadjemi and P. Moallem, "A NovelFuzzy-Neural Based Medical Diagnosis System",International Journal of Biological & Medical Sciences,Vol.4, No.3, 2009, pp. 146-150.
7. R. Lin, "An intelligent model for liver disease diagnosis",Artificial Intelligence in Medicine, Vol.47, No.1, 2009, pp. 53-62.
8. P. Gaur, "Neural Networks in Data Mining", International Journal of Electronics and Computer Science Engineering (IJECSE, ISSN: 2277-1956), vol. 1, **(2012)**, pp. 1449-1453.
9. Z. Chen, "Research of Data Mining Based on Neural Network", E-Product E-Service and Entertainment (ICEEE), International Conference on:IEEE, **(2010)**,pp.1-3.
10. R. Das, I. Turkoglu and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles", Expert Systems with Applications, Vol.36, No.4, 2009, pp. 76757680.
11. P. Cortez, "Data mining with neural networks and support vector machines using the R/rminer tool", Advances in Data Mining. Applications and Theoretical Aspects: Springer, **(2010)**, pp. 572-583.
12. M. Kamrunnahar and M. Urquidi-Macdonald, "Prediction of corrosion behavior using neural network as a data mining tool", Corrosion Science, vol. 52, **(2010)**, pp. 669-677.
13. Y. Frayman and L. Wang, "Data mining using dynamically constructed recurrent fuzzy neural networks", Research and Development in Knowledge Discovery and Data Mining: Springer, **(1998)**, pp.122-131.