



An Improved Approach for Hidden Nodes Selection in Artificial Neural Network

H. N. Odikwa
Department of Computer Science
Wellspring University
Irhirhi Road, Off Airport
Road, Benin City Edo State

Nkechi Ifeanyi-Reuben
Department of Computer Science
Rhema University
Aba, Abia State

Osaki Miller Thom-Manuel
Information and Communication
Center
Ignatius Ajuri University, Rivers
State

ABSTRACT

Training of data in Artificial Neural Network (ANN) imposes a lot of difficulty when determining the number of hidden layers and nodes that will enhance the network convergence in a multi layer perceptron neural network. This study employed a different approach of choosing hidden layers and nodes by taken cognizance of the fact that it is the bedrock of easy network convergence in an artificial neural network by using a heuristic search function of means end analysis. The system adopted the means end analysis algorithm by using a forward and backward chaining to generate current operators and calculating their differences from the goal state which is the target value of the ANN. The system was trained using 500 prostate data and 100 diabetes patient diseases from Federal Medical Center Umuahia in Abia State, Nigeria to train the data in a neural network. The trained data was used for classification. The result revealed that means end analysis is promising to training data in an ANN and yielded accuracies 80%, 82%, 85% for hidden layers between 2 to 20 and hidden nodes between 2 to 6. The classification accuracies of 87%, 90%, 95%, 98% for prostate cancer disease were obtained for hidden layers of 30 to 60 and hidden nodes between 8 to 14. The classification accuracies for diabetes disease were 81%, 85%, 87% for hidden layers of 30 to 60 and 89%, 92.5%, 95%, 98.5% for hidden nodes between 8 to 14 for diabetes disease considering time and space trade-offs.

Keywords

Artificial Neural Network, Multi-layer Perceptron, Artificial Intelligence, Hidden Layers, Hidden Nodes, Prostate Cancer, Diabetes

1. INTRODUCTION

Over the years, the problem of choosing number of hidden nodes and layers has imposed serious problems in neural network architecture, training and network convergence. The ability to learn, memorize, generalize, prompted research in algorithmic modeling biological neural systems in [7]. The essence of having neural networks is to train data in order to mimic human beings and is attributed to a branch in Artificial Intelligence (AI) [12].

Artificial Neural Networks (ANNs) adopt the approach the human brain processes information [3]. An ANN has many processing units (neurons or nodes) functioning in unison. They are extremely interconnected by links (synapses) with weights. The Artificial Neural Network has an input layer, an output layer and many hidden layers as shown in Fig.1 [11].

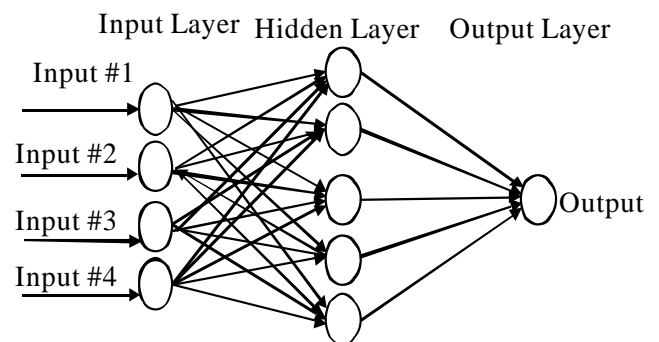


Fig. 1: An example of Artificial Neural Network [11][3]

Artificial Neural Network is machine learning paradigms that learns from past experiences and deploy what has been learnt to classify new data introduced into it [8]. Artificial neural networks can be referred as a prototype of the human brain. A neural network can be a replica of logic or reasoning base on the human brain. Generally, our brain comprises of an intensive set of nerve cells which are connected to each other. The brain cells in humans contain about 10 billion and 60 trillion connections (synapses). When the brain engages multiple nodes in analogous, it will do jobs very much faster than the greatest computers nowadays.

Normally, the neuron consist of soma which represents the cell body, a number of fibers called dendrites, an axon a single long fiber, and synapses that represents connections between cells. [4] illustrated the constituents of a neuron as shown in Fig.2 as follows;

Synapses weight: The inward bound signals from other nodes are trans-located on a neuron by special connections called the synapses. The synapses comprise of the chemical and electrical synapses.

Dendrite: It is a branch-like trees from the cell nucleus of the neuron (which is called soma) and thus receives signals electrically from many different sources, which in turn are transferred into the nucleus of the cell.

The axon: The axon transfers outgoing pulses. The axon is the component means where the pulses are transferred.

The soma: The exertion of the soma is to accumulate signals from synapses or dendrites through activation and inhibition. Likewise, an artificial Intelligence (ANN) is an interrelated group of artificial nodes which are dependent on a mathematical model to process information through nodes. ANN most at times is referred to adaptive system, in the sense

that it possesses the capability to change its structural based information processed in the neural network. ANN's structure has input layers, hidden layers, and output layers [3].

2. RELATED WORKS

Few researches have been made in the area of determining the best neural network hidden nodes. This deserves more attention as undesirable hidden nodes may result to either under-fitting or over-fitting of training data in a neural network and elongation of training time may be experienced [9]. In combating this menace resulting in neural network

training, [9] proposed a method of selecting hidden nodes using a named sales forecasting and calculating the mean square errors. This method was developed to predict the number of hidden layers by using only one sample means for training the artificial neural network. A research on varying the nodes and hidden layers in neural network architecture for a time frequency application yielded positive results that Entropy and Mean Square Error (MSE) is the decision criteria for the most optimum solution in choosing neural network hidden nodes [6].

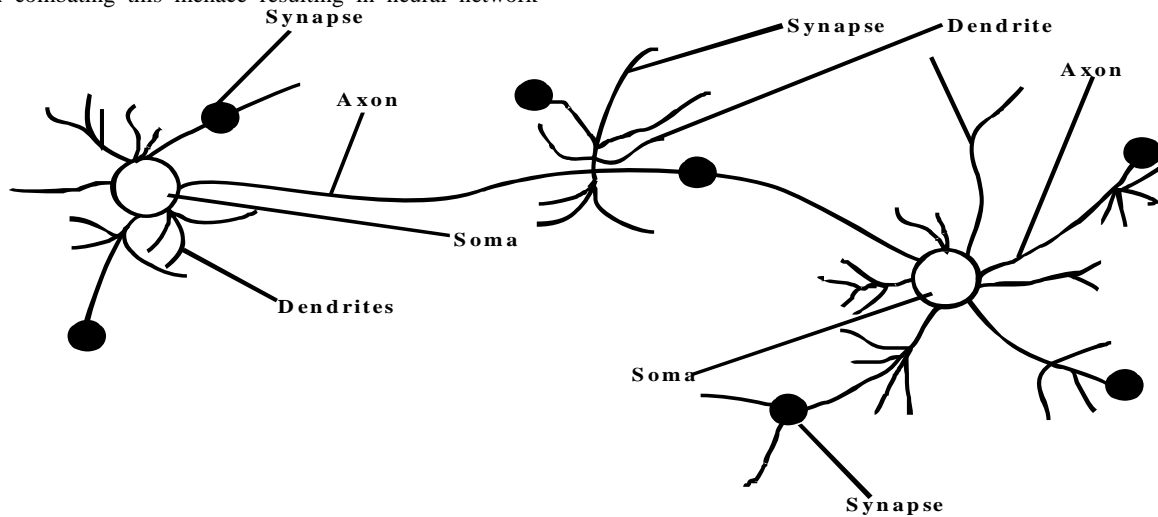


Fig. 2: Biological Neural Network [7]

A review was made that investigated the possibility of using two hidden layers in training neural network using feed forward and was found that the training generalized a better result than using one hidden layer [2]. This prompted other researches to intensify efforts on the optimum solution of finding the best hidden layers for neural networks which [10] deployed rational activation function in selecting the hidden nodes and layers. Yet this method yielded a promising result but not so reliable for larger datasets. Lasso regularization method was deployed which considered only outgoing weights for each hidden layer neuron. This development helped in identifying the number of redundant or unnecessary hidden layer nodes with the norm of outgoing weight connections with a predefined threshold value [6]. The redundant nodes identifications was modified by using accuracy-over-complexity fitness function with genetic algorithm to decide non-promising nodes in the neural network [13], while an extension of the model

[8] developed a decision support system with neural network perceptron to train and validate clinical data. The system adopted genetic algorithm for training data in neural network. The result shows that machine learning algorithm with Artificial Neural Network (ANN) and Genetic Algorithm (GA) outperforms other machine learning algorithms in clinical decision support system. It proved more reliable and robust in the classification of diseases.

3. MATERIALS AND METHOD

This work employed 500 datasets of prostate cancer patients and 100 datasets of diabetes patient diseases to train the artificial neural network. The data were normalized using the min-max method before the training. Different hidden layers and nodes were used in the training at different time from neuron 1 to N and hidden layers from 1 to n , as shown in Fig. 3 and Fig. 4 for the classification of both diseases.

The feed forward neural network is deployed to calculate the weights of the network while the mean end analysis of a heuristic function was to estimate the iteration of operators of each hidden layer and nodes to the target value of the neural network.

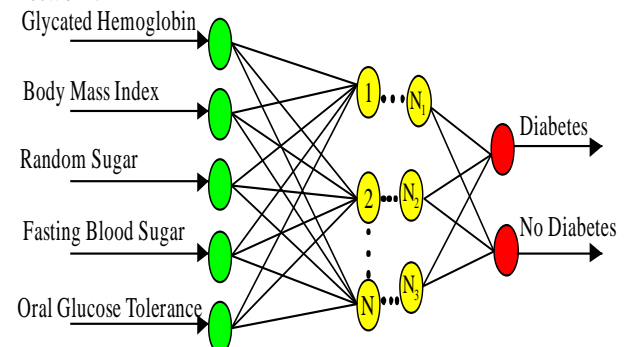


Fig. 3: Neural Network Architecture of Diabetes Disease

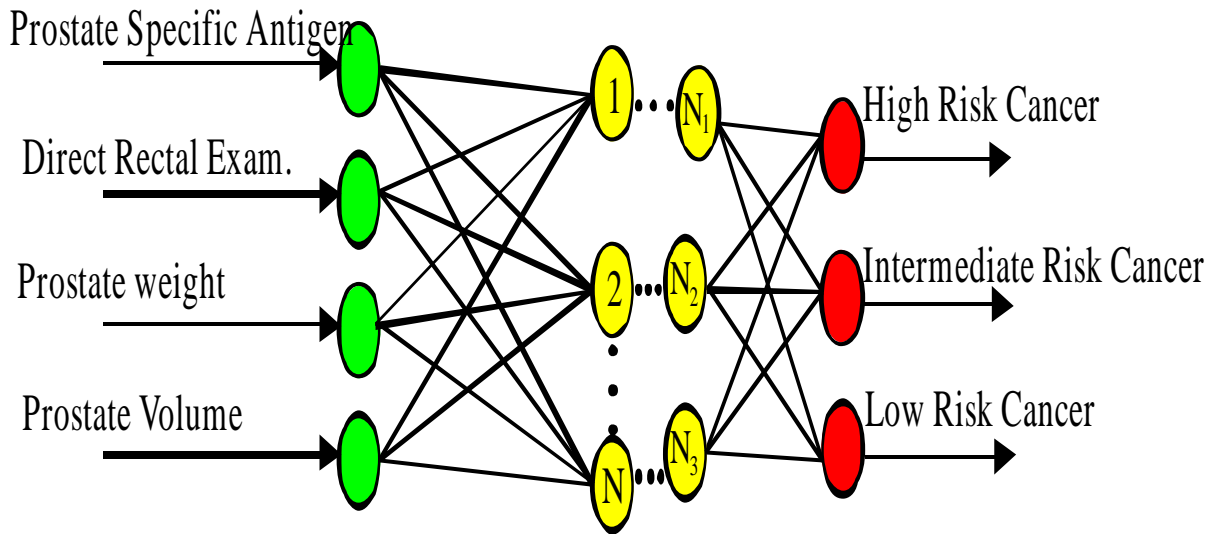


Fig. 4: Neural Network Architecture of Prostate Cancer Disease

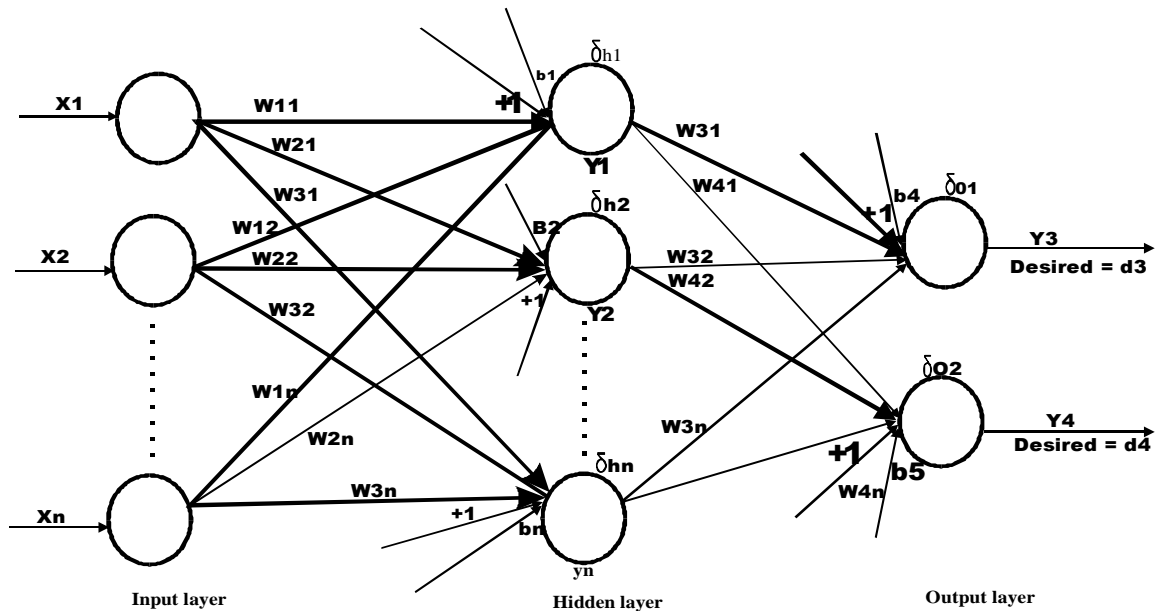


Fig. 5: General architectural structure of an artificial neural network

3.1 The Feed Forward Calculations

The Fig. 5 depicts the general architecture of an artificial neural network with the weights, bias and learning rates, input layers, hidden layers and output layers which serves as a base for calculating the feed forward pass using the algorithm in 3.2.1

3.2 Feed Forward Algorithm

The feed forward algorithm was deployed to calculate the weights in the neural network in Fig. 3 and Fig. 4 of the prostate and diabetes disease data. The mathematical model for the feed forward algorithm is given by [7] in equation (1)

$$Input\ j = x_j = \sum y_i w_{ij} \quad (1)$$

To determine the total of the input weights, hidden weights and output weights, equation (1) was expanded to equations (2), (3) and (4)

$$\phi_1 = (x_1 * w_{11}) + (x_2 * w_{12}) + \dots (x_n * w_{1n}) \dots (2)$$

$$\Omega_2 = (x_1 * w_{21}) + (x_2 * w_{22}) + \dots (x_n * w_{2n}) \quad (3)$$

$$\Omega_3 = (x_1 * w_{31}) + (x_2 * w_{32}) + \dots (x_n * w_{3n}) \quad (4)$$

Where x_1, x_2 and x_n are the inputs to neural network nodes and the w 's are the weights of the network.

The Output layer weights were calculated with equations (5) and (6)

$$\Phi_1 = (y_1 * w_{31}) + (y_2 * w_{32}) + \dots (y_n * w_{3n}) \quad (5)$$

$$\Phi_2 = (y_1 * w_{41}) + (y_2 * w_{42}) + \dots (y_n * w_{4n}) \quad (6)$$

In order to determine the error of the network, equation (7) was applied.



$$\delta_k = (1 - O_k) * O_k(T_k - O_k) \quad (7)$$

Where O_k is the calculated, T_k is the observed or actual output.

3.3 Back-Propagation Algorithm

The back-propagation algorithm is employed to adjusting the weights of the neural network that is, where genetic algorithm could not get to the actual output target, the back-propagation propagates the errors back to the neural network.

Calculate the local gradients ($\delta o_1, \delta o_2, \delta h_1, \delta h_2, \dots, \delta h_n$) using equation (8).

$$\delta h_n = \varphi'(1 * b_3 + x_1 * w_{3_1} + x_n * w_{3_n}) * (\delta o_1 * w_{3_n} + \delta o_2 * w_{4_n}) \dots (8)$$

The weights of the network were adjusted using the learning rule depicted in equation (9)

$$W_{new} = w_{old} + \eta * \delta * input \quad (9)$$

Where w_{jk} is the weight of the connection from unit j to unit k appropriate in the next layer and δ_k is the error of unit k and η is a constant that is called the learning rate, which takes the value from 0 and 1.

The biased errors were determined with equation (10); the introduction of bias in the network was to provide every node with a trainable constant. With equation (10), the bias errors can be determined.

$$b_n(n+1) = b_n(n) + \alpha * b_n(n-1) + \eta * \delta h_n(n) * 1 \quad (10)$$

3.4 Mean Square Error (MSE) of the Neural Network Training

To estimate the errors in the network, equation (11) is used in the calculation of the network errors using the mean square and results obtained.

$$MSE = \frac{\sqrt{(\sum_{i=1}^n (T_i - E_i)^2)}}{n} \quad (11)$$

Where T_i represents the predicted response, E_i denotes the target response and n is the number of observations (iterations).

3.5 Heuristic Function

Heuristic search is a type of generic search algorithm that uses hints or clues of information to get to the goal state. It is the study of the method and rules of discovery and invention [7].

The heuristic function is given in equation (12), employed in determining the values towards achieving the desired goals, in this case the target values of the neural network for every iteration,

$$h(n) = \sum^n (dist . to . goal) \quad (12)$$

Heuristic search refers to a search strategy that attempts to optimize a problem by iteratively improving the solution based on a given heuristic function or a cost. It is classified as an informed guess to find an acceptable solution.

3.6 Means End Analysis

Means end analysis is a problem solving technique. It is a form of heuristic function. It allows both backward and forward chaining by computing the current state to a goal state by computing their differences. To reduce the difference between current and goal state, an operator is used. In this paper the end means analysis is used to determine each of the neural network iteration differences from the target.

The algorithm of the Mean End Analysis used in the system is shown in Algorithm 1.

Algorithm 1: Means End Analysis Algorithm

1. Compare the current target value to the goal state value. If there exist no differences among them return;
2. Otherwise select the closest significant difference and reduce it until a successor is signaled.
3. Select an ‘as yet untested operator O that is relevant to the current difference’ if there exist no such operators, then signal failure.
4. Try to apply O to the new CURRENT. Produce descriptions of two states:
5. If (FIRST-PART \Leftarrow MEA (CURRENT, O-START)) And
 - O-START, a condition in which O’s preconditions are met and;
 - O-RESULT, the condition that would result if O were applied in O-Start

(LAST-PART \Leftarrow MEA (O-RESULT, GOAL))

Are successful, then signal success and return the result, of the concatenate FIRST-PART, O, and LAST-PART.

With equation (13), the difference to goal state is determined

$$G_{1_2} = S_1 - S_2 \quad (13)$$

Where s_1 is the current state of all the trainable hidden layer and nodes and s_2 is the goal state to the target value.

4. EXPERIMENTS AND RESULTS

Series of experiments were carried out to ascertain the best number of hidden layers and nodes to train data in artificial neural network; using means end analysis to determine the best solution approach. This approach is carried out by varying both the number of hidden layers and nodes by incrementing values in the hidden layers and nodes from 2 to 60 and 2 to 14 respectively for both prostate cancer and diabetes data as depicted in Table 1 and Table 2.



Table 1: Neural Network Training with different Hidden Layers and Nodes for Prostate Data

Sample data	No of hidden layers	No of hidden nodes	No of iterations	Training time(s)	Mean square error	Accuracy(%)
Prostate Cancer	2	2	100	30	0.0214	81
	10	4	150	50	0.0100	85
	20	6	250	70	0.0080	87
	30	8	500	80	0.0056	89
	40	10	800	180	0.0045	92.5
	50	12	1000	1800	0.0036	95
	60	14	1250	3000	0.0034	98.5

Table 2: Neural network training with different hidden layers and nodes for Diabetes data

Sample data	No of hidden layers	No of hidden nodes	No of iterations	Training time(s)	Mean square error	Accuracy (%)
Diabetes Disease	2	2	100	9	0.0254	65.0
	10	4	150	15	0.0200	70.2
	20	6	250	28	0.0060	78.4
	30	8	500	40	0.0032	82.0
	40	10	800	240	0.0021	89.1
	50	12	1000	2040	0.0019	92.0
	60	14	1250	3720	0.0013	97.9

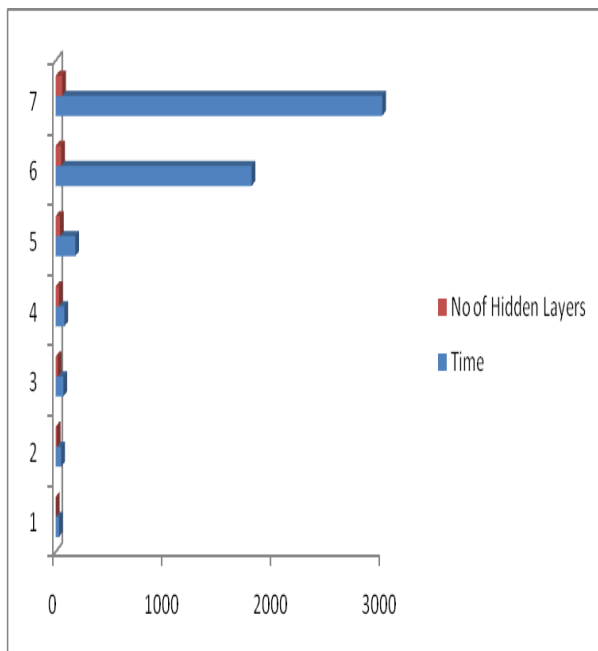


Fig. 6: Graphical representation of hidden layers and training time in NN with prostate cancer data

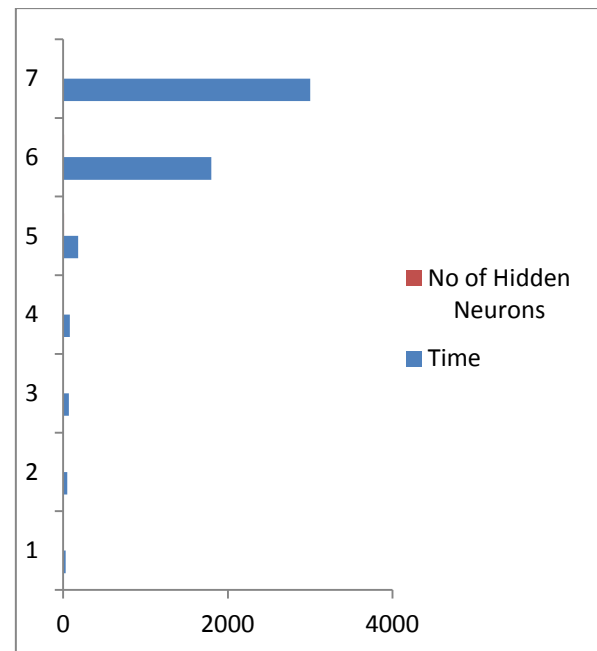


Fig. 7: Graphical representation of hidden nodes and training time in NN with prostate cancer data

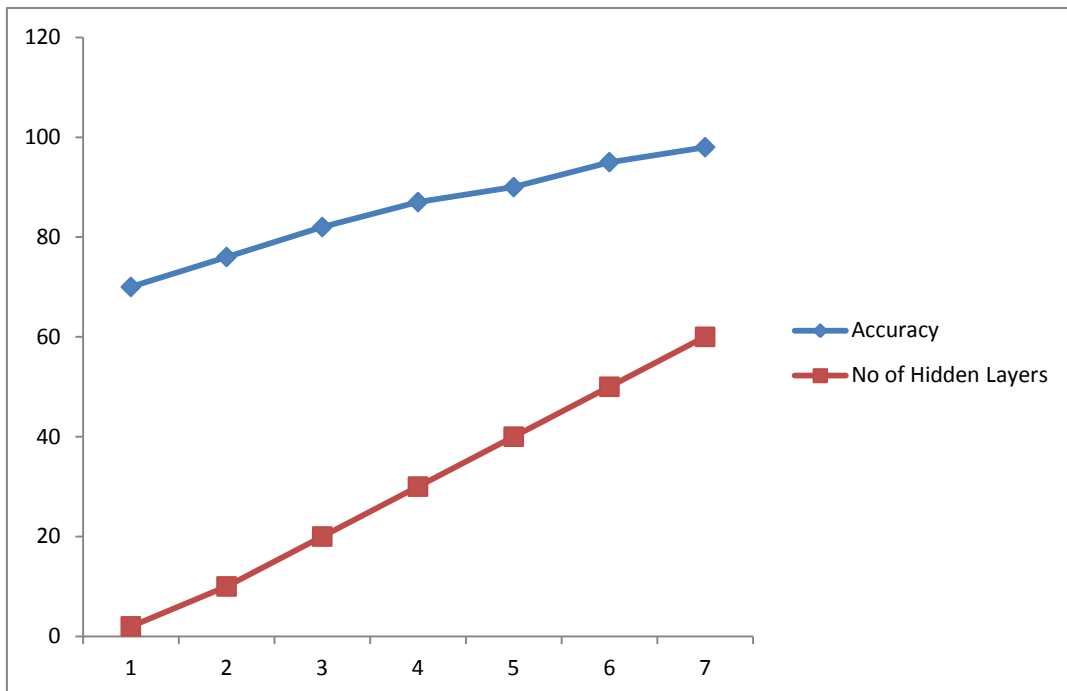


Fig. 8: Graphical representation of hidden layers and training accuracy in NN with prostate cancer data

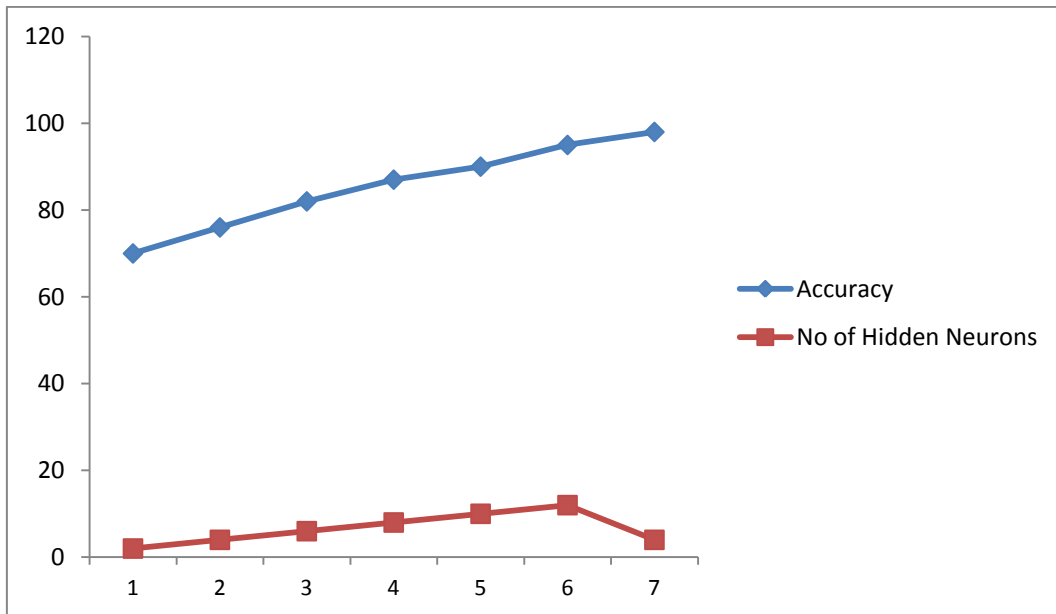


Fig. 9: Graphical representation of hidden nodes and training accuracy in NN with prostate cancer data

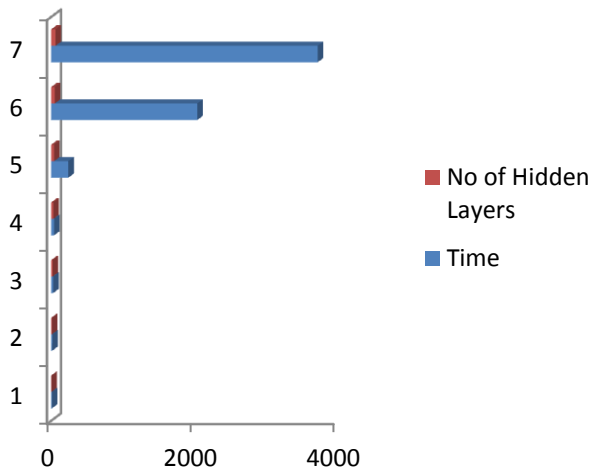


Fig. 10: Graphical representation of hidden layers and training time in NN with diabetes disease data

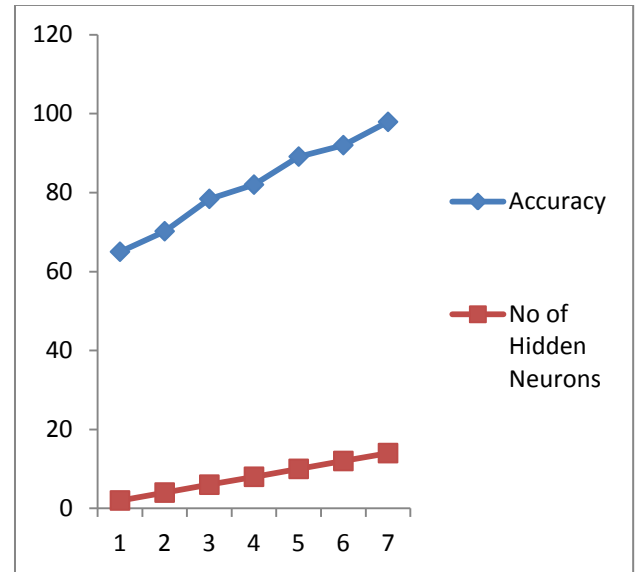


Fig. 13: Graphical representation of hidden nodes and training accuracy in NN with prostate cancer data

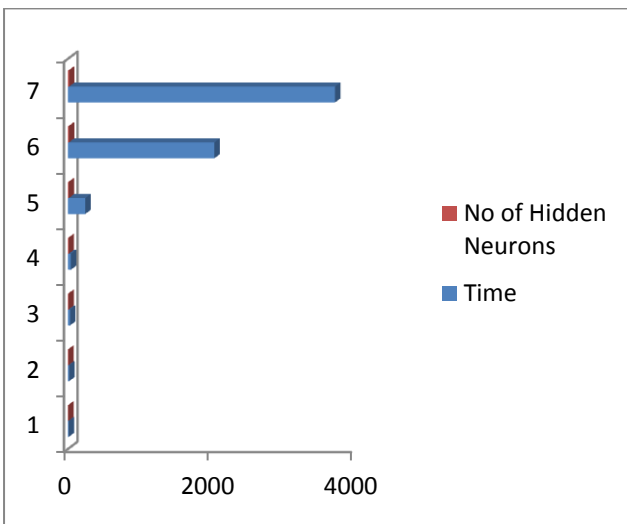


Fig. 11: Graphical representation of hidden nodes and training time in NN with prostate cancer data

5. TRADEOFFS IN SELECTING HIDDEN LAYER AND NODES

This work considered time trade-offs in choosing hidden layers and nodes in an Artificial Neural Network (ANN) multi-layer perceptron network. A good ANN architecture is one that limits the training time and converges at a limited time. But it is hard to have the both at the same time. The means end analysis employed the division of this problem into sub-goals. The first part of the goal is taken cognizance of smaller hidden layers and nodes that took less training time in expense of generating sufficient training accuracies. The second goal is using the second sub-goal of employing many hidden layers and nodes that gave sufficient accuracies in expense of training time and that is the trade-offs in selecting hidden layers and nodes in a neural network.

6. DISCUSSION OF RESULTS

Table 1 shows the results of the training of prostate cancer data with different hidden layers and nodes after applying means end analysis that satisfy the preconditions of selected operators that reduced the differences between the current state and goal state of the two divided sub-goals by applying equation (6) while the Fig. 6 and Fig. 7 show the graphical representation of hidden layers and training time in neural network with prostate cancer data which reviewed more number of hidden layers and nodes having higher accuracies in the range of 30 to 60 hidden layers and 8 to 14 hidden neurons. This fact proved that the more the hidden layers and nodes, the more the accuracy of classification generated but more time required for training the data to obtain network convergence. More so, with insufficient number of hidden nodes and layers converge faster as it requires less training time but sufficient accuracies are not ascertained. The accuracies of the data training are depicted in Fig. 8 and Fig. 9 for varying hidden layers and nodes computed as 81%, 85%, 87%, 89%, 92.5%, 95% and 98.5%. Consequently, Table 2, depicts the results of training diabetes data in a neural network with varying hidden layers between 2 to 60 and hidden nodes from 2 to 14 consecutively. Fig. 10 and Fig. 11 depict the graphical representation of hidden layers and training time, hidden nodes and training time with diabetes disease data in a

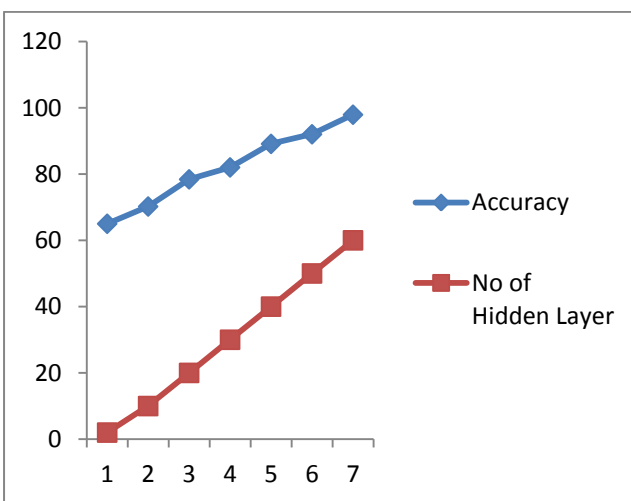


Fig. 12: Graphical representation of hidden layers and training accuracy in NN with prostate cancer data



neural network. The means end analysis was applied in choosing the best combination of hidden nodes and layers generated sufficient accuracies with more hidden nodes and layers ranging from 82% to 97.9% as shown in Fig. 12 and Fig. 13 respectively.

7. CONCLUSION

An improved model for selecting hidden nodes and layers in Artificial Neural Network (ANN) has successfully being designed and implemented. From the experiments and results carried out and analyzed, this work displayed reasonable improvement in choosing hidden layers and nodes by employing heuristic search function of means end analysis in selecting varying hidden nodes and layers that will generate sufficient classification accuracies. It is also deduced that combination of two or more hidden nodes and layers are highly recommended as a better architectural structure for any multi perceptron neural network considering the time and trade-off in data training. This fact has been proved with prostate cancer and diabetes disease data in this work.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge the efforts of Dr. O.O. Jegede of the Pathological Department, Federal Medical Center, Umuahia, and Dr. C.C. Chukwuegbo of PathConsult Medical Diagnostic Center, Umuahia, Abia State, Nigeria.

9. REFERENCES

- [1] Adel, E. (2009). Advanced applications for artificial neural network. Research Gate. Intech
- [2] Alan, J.T., Miltos, P., Simon, D.W., Saeed, M.G & Robert, E.M. (2018). Two hidden layers are usually better than one. *Institute of Electrical and Electronics Engineering*. 2(4): 244-252.
- [3] Benson-Emenike, M. E. & Ifeanyi-Reuben, N. J. (2018). Network Learning and Training of a Cascaded Link-Based Feed Forward Neural Network (CLBFFNN) in an Intelligent Trimodal Biometric System. *International Journal of Artificial Intelligence and Applications (IJAIA)*. 9(6): 27 – 47
- [4] David, R. (2005). Neural network architecture development process. *International journal of Adhoc, sensors and ubiquitous computing*. 10(7): 123-128.
- [5] Fabian, O., Kachi, G., & Okoli, H, B. (2010). Prospects of feed forward algorithm in Multilayer layer perceptron. *International journal of grid computing and applications*. 12(10): 34 – 40.
- [6] Hebtamu, Z.A., Wei, W. & Junhong, Z. (2018). Feed-forward Neural Networks with a Hidden layer regularization Method. *Journal of MDPI*. 4(12).
- [7] Imran, S., Ahmad, J., Syed, I.S & Faisal, M.K. (2016). Impact of Varying nodes and hidden layers in neural network architecture for a time frequency application. *Conference multi-topic of Institute of Electrical and Electronics engineering*.
- [8] Odikwa, H., Ugwu, C. & Inyiama, H. (2017). An Improved Model for Clinical Decision Support System. *International Journal of Artificial Intelligence and Applications (IJAIA)*. 8(6): 37-55.
- [9] Panchal, F.S., & Panchal, M. (2014). Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. *International journal of computer science and mobile computing*. 3(11): 455-464.
- [10] Saurabb, K.(2012). Approximating Number of Hidden layer Nodes in Multiple Hidden layer BPNN Architecture. *International Journal of Engineering trends and technology*. 3(6): 714-717
- [11] Serena, H. C., Anthony, J. J. and John, P. N. (2008). Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. Elsevier – Mathematics and Computers in Simulation. 78(2008): 379 - 400
- [12] Sreekanth, R. K. (2012). Artificial Intelligence Algorithms. *IOSR Journal of Computer Engineering (IOSRJCE)*. 6(3): 01-08.
- [13] Stathakis, D.(2009). How many hidden layers and nodes. *International journal of remote sensing*. 30(8): 2133-2147.