# PARTICLE SWARM OPTIMIZATION TRAINED NEURAL NETWORK FOR MEDICAL DIAGNOSIS

**Orven E. Llantos, Eugie L. Limpin, Macy P. Pellerin and Ellen L. Pates**

Department of Computer Science, School of Computer Studies
Mindanao State University-Iligan Institute of Technology, Iligan City, Philippines
orven.llantos@g.msuiit.edu.ph

## ABSTRACT

Multilayer Perceptron Neural Networks trained using Particle Swam Optimization (PSO) used for disease classification is utilized in this research. Experiments were conducted using dataset for acute diseases obtained from Mindanao State University-Iligan Institute of Technology (MSU-IIT) clinic and dataset for thyroid diseases from the UCI Machine Learning Data Repository as first and second dataset respectively. The data gathered from MSU-IIT clinic were normalized, disregarding repetitive data choosing only the most frequently occurring illness: Upper Respiratory Tract Infection (URTI), Systematic Viral Infection (SVI), and Acute Tonsillitis. The file from UCI Machine learning Data Repository consist of 5 attributes and 3 classes of diagnosis concerning thyroid diseases classification cases. There are 215 data samples in which 150 of it is used for training the network and the remaining 65 is for testing the network. The network trained with PSO can correctly classify diseases more than 90% of the time on the first dataset and more than 80% on the second dataset in greater than 600 and less than 1000 iterations. Results show that PSO can successfully optimize the weights of a Neural Network and produce good classification performance.

***Keywords:*** Artificial Neural Network, Multilayer Perceptron, Particle Swarm Optimization, Medical Diagnosis

## INTRODUCTION

Medical knowledge is constantly growing at a rapid pace. As a result, physicians find it more difficult to keep up with areas of medicine outside their specialization. Consultation with a specialist, who is not always readily available, is the traditional medical practice if the problem is beyond the attending physician's capabilities. This problem is compounded by the already present shortage of specialists (Arie, 2011; Boseley, 2011). This leads to longer waiting times for appointments and increased travel distances to get care which is not always acceptable, higher prices, and the weakening of health care in general (Walker, 2008).

One solution would be for higher learning institutions to train new doctors. However, one has to spend almost ten years in medical school to get a diploma. This simply suggests that the training for doctors and specialists in particular is lengthy and expensive. One has to undergo this procedure before one would even be qualified to diagnose a patient.

Computer programs were developed in an attempt to remedy these problems in as early as 1970's. These systems, also called Clinical Decision Support systems, according to Wyatt and Spiegelhalter (1991), are "active knowledge systems which use two or more items of patient data to generate case-specific advice". They are typically designed to integrate a medical knowledge base, patient data and an inference engine to generate case specific advice. Some of the earliest attempts are INTERNIST-I, DXplain, and QMR (Quick Medical Reference).

Recent studies used evolutionary algorithms in application to medical diagnosis (Hsieh, et. al., 2015; Behishti, et. al., 2014) for *myocardial infraction* or heart attack (Seenivasagam, et. al., 2016), prostate cancer (Sadoughi, et. al., 2015), cancer (Agrawal, et. al., 2015), diseases in urology (Moein, et.al., 2006), skin diseases (Bakpo, et. al, 2011) and tumor (Abdi, et. al., 2012) among others.

There are studies that have demonstrated the robustness of Particle Swarm Optimization as a training algorithm in diagnosis of specific diseases but the application of PSO in typical disease diagnosis is yet to be explored.

## ARTIFICIAL NEURAL NETWORKS AND MEDICAL DIAGNOSIS

According to Haykin (Haykin, 2005), an artificial neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity of storing experiential knowledge and making it available for use. It resembles the brain in two aspects:

1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, knows as synaptic weights, are used to store the acquired knowledge.

Artificial neural networks in medicine are mostly applied to classification problems or the task of assigning a patient to a small set of classes (Dybowski, *et. al.,* 2007) and they have been shown to be very effective (Lisboa, 2002). The works of Das, et.al (Das, et.al., 2009) and Zhang, et. al, (Zhang, et. al., 2008) are some examples of applications in medical diagnosis that utilized artificial neural networks.

In a study of medical diagnostic tool (Alexandridis, et. al., 2014), a neural network architecture adopts radial basis function (RBF) with evolutionary simulated annealing technique to optimize the result, and uses the non-symmetry fuzzy means (NSFM) for better prediction and shorter computational times. Given nine (9) datasets the neural networks predicted at a 95% confidence level using the Matthews Correlation Coefficient. On the other hand, a medical diagnostic tool (Hsieh, et. al., 2015) was developed using particle swarm optimization (PSO) to improve previous results of Hyper-Rectangular Composite Neural Networks (HRCNNs). The resulting neural network is described as PSO-based Fuzzy Hyper-Rectangular Composite Neural Network (PFHRCNN). It was tested using the database for Wisconsin breast cancer, liver disorders, and Parkinson's Disease from University of California at Irvine, Repository of Machine Learning Databases. Their study showed significance performance improvement in terms of prediction accuracy from their previous work.

Another study (Beheshti, et. al., 2014) discussed new learning algorithm that improved the diagnostic performance of neural network for medical diagnosis even with very noisy data. It introduced a meta-heuristic learning algorithm, centric (CAPSO), for more accurate predictions in a multi-layer perceptron neural network (MLP) tested on nine standard datasets of hepatitis, heart disease, Pima Indian diabetes, Wisconsin Prognostic breast cancer, Parkinson's disease, echocardiogram, liver disorders, laryngeal 1 and acute inflammations.

## DEVELOPING THE MEDICAL DIAGNOSIS APPLICATION

The constructed neural network is a multilayer perceptron neural network with PSO as its training algorithm. It consists of a single hidden layer with the number of nodes based on *Livingstone* and *Manallack's* suggestion in deciding how many hidden nodes a neural network will use. The researchers also used in training and testing of the network, different numbers of nodes for the hidden layer. Also, the activation functions utilized by the network are *Linear, Sigmoid*, and *Softmax activation function*.

In the discussion to follow, researchers applied the notion of the theoretical ideas supporting PSO and MLP in order to identify the network structure: the number of layers, nodes per layer, and activation function for each layer; other affecting factors like bias, inertial weight, acceleration, maximum and minimum velocity and position; swarm construction and how each particle contained in the swarm were topologically built, including how many iterations were allowed for the network and the minimum error it has to reach to stop training.

The network structure involves 3 layers: an input layer, one hidden layer and an output layer. For the network that is to deal with the MSU-IIT clinic dataset, the input layer contains 6 input nodes having one-to-one correspondence to the number of illness features dealt by the network. While for the thyroid dataset, the input layer contains 5 input nodes, corresponding to the 5 attributes relating to thyroid problem indications. The researchers varied the number of nodes for the hidden layer in different runs due to the idea that theoretically there was not any well established standard as to how many nodes are necessary in a hidden layer of a particular network that deals with a certain problem. The researchers performed runs with networks having 5, 10, and 15 hidden nodes. Lastly, number of nodes in the output layer is representative of the 3 classes of illnesses the network handles, which is the same for both structures of the networks dealing with MSU-IIT clinic dataset and the thyroid dataset.

The input layer activation function was chosen to be the *Identity Function*, this function was chosen due to the fact that each node in the input layer only receives a single value not necessarily to be modified as it will be propagated towards the succeeding layers.

The hidden layer uses the *Sigmoid* activation function, the most commonly used activation function for neural networks. When the h^th^ node of the hidden layer receives the total input that the previous layer propagated, this value will serve as the input to the function.

$$f(y_h) = \frac{1}{1 + e^{-y_h}}$$

The independent variable y in the equation is defined by

$$y_h = \sum_{i=1}^{5\ or\ 6} x_i(weigh\ t_{ih})$$

where $y_h$ is the h^th^ node in the hidden layer and the subscript *h* runs through values 1 to h, the total number of nodes in the hidden layer.

The *Softmax* function is employed in the output layer for it is suitable for classification problems. The variable q in the numerator is the net output to the node in the output layer which is a summation of the previously elaborated sigmoid function of the hidden layer. While the denominator indicates the total q of all the output nodes, thus the equation becomes a probability equation.

$$p_i = \frac{\exp(q_i)}{\sum_{j=1}^{n} \exp(q_j)},$$

With respect to PSO learning algorithm configuration, each Particle becomes a representation of a single neural network. Each particle has an attribute dimension, position and velocity. The particle dimension is equivalent to the number of synaptic weights of the network together with the bias, which equals to

$$d = 2(h * i) + h + i$$

For the network dealing with acute illnesses and thyroid diseases, the previous equation becomes

$$d = (6*h) + (h*3) + h + 3 = 10h + 3$$

and

$$d = (5*h) + (h*3) + h + 3 = 9h + 3$$

respectively, where *h* is the number of nodes in hidden layer. This parameter also determines the dimensionality of the particle's position which is a vector whose values are the weights of the network that the particle represents.

The velocity vectors has been modified and added a parameter inertial weight w, a value which has an effect of amplification or degradation effect of velocity change to the particle's position. Motivated by a research conducted by Shi and Eberhart, 0.9 was chosen to be the value of w, which was shown to yield a larger chance of convergence at presumably reasonable number of iterations. It allows only 90% of the computed velocity to be added to the previous position. The resulting equation is now:

$$v_n = v_n * w + c_1 * rand() * (gbest_n - x_n) + c_2 * ran() * (pbest_n - x_n)$$

The acceleration parameters, $c_1$ and $c_2$, were also varied from time to time to refine further results of the search. In addition, the network also has the attribute *pbest* that keeps track of the particles best position. Best position is determined by an attribute *pfitness*, which is calculated using Sum of Classification Errors (EC) for all pattern samples.

The Network Swarm has been built with 20 particles, which is equal to the swarm size. The Swarm's overall movement is, by a certain amount, dependent to an attribute *gbest*, whose structure is a vector of values with dimension equivalent to any particle's position's dimension. Since *gbest* changes through the training, one important factor triggers whether the *gbes*t should be updated or not is the *gfitness*. The swarm's *gfitness* was represented as the classification rate of the network. Classification rate is computed by comparing the total number of correct diagnosis versus the total diagnosis. The *gfitness* is then compared to a value *min_err*, a constraint which determines termination of the training process.

If (1 – *gfitness* ) <*min_err* : break

The *min_err* was chosen to be 0.03, which implies that the network's target is 97% accuracy, and then it stops iteration and halts. However, another constraint can also be set for the network to stop training: setting a maximum number of iteration *itermax*, which in the network training was set to 500. Swarm moves in the hyperspace exploring different values so as to minimize *min_err* in each epoch.

## EXPERIMENTAL RESULTS

Two datasets were used to evaluate the performance of Particle Swarm Optimization in training MLP neural networks. The first dataset taken from UCI Machine Learning repository contains five lab tests that determine whether a patient's thyroid belongs to the class euthyroidism, hypothyroidism or hyperthyroidism. Another dataset gathered from Mindanao State University clinic contains six input variables that separate the patient's sickness into three classes: Upper Respiratory Tract Infection, Systemic Viral Infection and Acute Tonsillitis.

Figure 1 shows the performance of the network for the thyroid dataset with varying hidden nodes in the first round. Shown in the graph are the three different structure of neural network having 10 hidden nodes (green), 15 hidden nodes (red), and 20 hidden nodes (blue), that progresses in its accuracy as the iteration continues. Figure 2, similarly, shows the performance of the network for the clinic dataset with varying hidden nodes in the first round. The graph also shows the progression of its accuracy as the networks with 5 hidden nodes (blue), 10 hidden nodes (green), and 15 hidden nodes (red) continues in iteration.

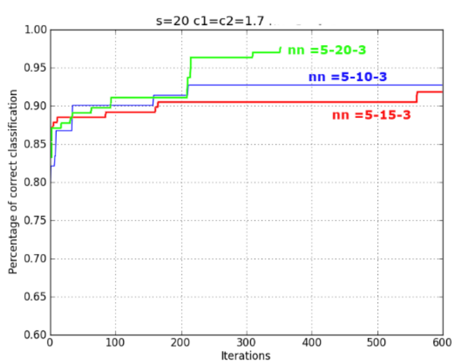


**Figure 1. Performance of the network for thyroid dataset with 10, 15 and 20 hidden neurons**

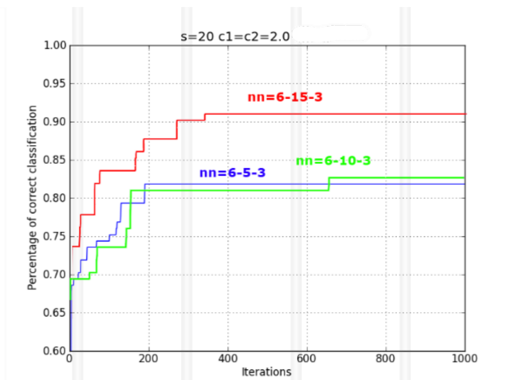


**Figure 2. Performance of the network for clinic dataset with 5, 10 and 15 hidden neurons.**

The conducted experiments showed that the neural networks were successfully trained on the two different datasets using Particle Swarm Optimization. On the first dataset (thyroid), it was found that a value of 1.7 for acceleration constant *c1* and *c2* and a 5-20-3 network produce good results. A value of 2.0 for *c1* and *c2* and 6-15-3 topology for the network that used the clinic dataset yields satisfactory results after 1000 iterations.

## CONCLUSION

The network trained with PSO can correctly classify diseases more than 90% of the time on the first dataset and more than 80% on the second dataset in less than 1000 iterations. Results showed that PSO can successfully optimize the weights of a Neural Network and produce good classification performance.

## REFERENCES

[1] Abdi, M.J., Hosseini, M.S., & Rezghi, M. A Novel Weighted Support Vector Machine Based on Particle Swarm Optimization for Gene Selection and Tumor Classification, *Computational and Mathematical Methods in Medicine*, 2012, 1-5.

[2] Agrawal, S & Agrawal J. Neural Network Techniques for Cancer Prediction: A Survey. *Procedia Computer Science*, 2015, 60, 769-774.

[3] Alexandridis A & Chondrodima E. A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing. *Journal of Biomedical Informatics*, 2014, 49, 61-72.

[4] Arie, S. (2011, September). World needs 3.5 million more health workers to deliver essential services. Retrieved 2011-12-12, from http://www.bmj.com/content/343/bmj.d6005.

[5] Bakpo, F.S. & Kabari, L.G. Diagnosing skin diseases Using an Artificial Neural Network, *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, 2011, 253-270.

[6] Beheshti, Z., Shamsuddin, S.M.H., Beheshti, E. Enhancement of artificial neural network learning using centripetal accelerated particle swarm optimization for medical diseases diagnosis. *Springer Soft Computing*, 2014, 18,11, 2253-2270.

[7] Boseley, S. (2011, September). The global health worker shortage needs to be addressed. Retrieved 2011-12-12, from http://www.guardian.co.uk/global-development/poverty-matters/2011/ feb/23/global-health-worker-shortage.

[8] Das, R., Turkoglu, I. & Sengur, A. Effective diagnosis of heart disease through neural networks ensembles, *Expert systems with applications*, 2009, Vol.36, No.4, pp. 7675-7680.

[9] Dybowski, R. & Gant, V. *Clinical applications of Artificial Neural Networks*, 2007, Cambridge University Press.

[10] Haykin, S.S. Neural Networks: A Comprehensive Foundation, 2005, Pearson Education (Singapore) Pre. Ltd.

[11] Hsieh, Y-Z, Su, M-C, & Wang, P-C. A PSO-based rule extractor for medical diagnosis.*Journal of Biomedical Informatics*, 2014, 49, 53-60.

[12] Lisboa, P. J. A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Networks, 2002, 15, 11–39.

[13] Moein, S., Monadjemi, S. A. & Moallem, P. A Novel Fuzzy-Neural Based Medical Diagnosis System, *International Journal of Biological & Medical Sciences*, 2009, 4(3), 146-150.

[14] Sadoughi F, Ghaderzadeh M, Solimany M & Fein R. An Intelligent System Based on Back Propagation Neural Network and Particle Swarm Optimization for Detection of Prostate Cancer from Benign Hyperplasia of Prostate. J Health Med Informat, 2014, 5(3), 1-5

[15] Seenivasagam, V. & Began, C. Myocardial infraction detection using intelligent algorithms. *Neural Network World,* 2016,26(1), 91-110.

[16] Walker, E. (2008, December 2). Nation faces daunting doctor shortage. Retrieved 2011-12-12, from http://www.medpagetoday.com/HospitalBasedMedicine/WorkForce/11980.
[17] Wyatt, J. & Spiegelhalter, D.J. Evaluating medical expert systems: what to test and how? *Medical Informatics*, 1991, 15, 3, 205-218.
[18] Zhang, G., Yan, P., Zhao, H. & Zhang, X. A computer aided diagnosis system in mammography using artificial neural networks, *In* International Conference on BioMedical Engineering and Informatics, 2008, Vol.2, pp. 823-826.