A NEURO-GENETIC ALGORITHM BASED MEDICAL DECISION SUPPORT SYSTEM FOR CONGENITAL HEART DISEASE DIAGNOSIS

ISSN: 2141 – 3290
www.wojast.com

[1]AMADIN, F. I., [2]EGWUATU, J. O.,
[3]OSAZUWA, W.A., AND [4]EKONG, V. E
[1,2]*Department of Computer Science, University of Benin, Nigeria.*
[3]*Department of Mathematics and Computer Science,*
Benson Idahosa University, Benin City
[4]*Department of Computer Science, University of Uyo, Nigeria*
[1]*frankamadin@uniben.edu,* [2]*joshegwuatu@gmail.com,* [3]
aosazuwa@biu,edu.ng, [4]*victoreekong@uniuyo.edu.ng*

ABSTRACT

One of the major causes of infant mortality is Congenital Heart Disease (CHD). In most cases, a proper diagnosis at an early stage will result in a significant chance of survival of the patient. Early diagnosis has always been lacking due to the fact that doctors are not equally skilled and may lack the required facilities needed for such diagnosis. In this paper, we seek to improve the performance of neural network based decision support systems on congenital heart disease by selecting the relevant features required for the diagnosis. The proposed system is designed and implemented using MATLAB with the implementation of a Back Propagation Neural Network (BPNN). Genetic Algorithm (GA) is employed in the selection of optimal features for the Neural Network (NN) training. The developed system based its prediction solely on the clinical symptoms of CHD. The proposed Neuro-genetic model identified and classified cases with a higher accuracy of 98% compared to a BPNN of 90%. This shows that the proposed model can improve the diagnosis comparable to that of the domain experts.

INTRODUCTION

Congenital Heart Defect is a problem that occurs at the developmental stage of a child's heart during pregnancy. These defects are the most common birth defects. Saxena (2005) observed that, nine out of every 1,000 babies born in India have CHD. It happens because of abnormal development of the fetus' heart during the early pregnancy. Esposito and Della (2008) identified some causes of CHD to be associated with genetic disorders such as Down syndrome.

The various types of CHD are classified into; holes in the heart, obstructed blood flow, abnormal blood vessels, heart valve abnormalities, a combination of defects (Congenital Heart Disease Index, 2016). Some infants are born with several heart defects, such as Tetralogy of Fallot; a hole in the ventricular septum, a narrowed passage between the right ventricle and pulmonary artery, a shift in the connection of the aorta to the heart, and thickened muscle in the right ventricle. This results in cyanosis (blueness), which may appear soon after birth, in infancy or later in childhood. These "blue babies" may have sudden episodes of severe cyanosis with rapid breathing. They may even become unconscious. During exercise, older children may become short of breath and faint. These symptoms occur because not enough blood flows to the lungs to supply the child's body with oxygen. If these defects are not recognized and treated properly at an early stage of the children, then they have to face many problems in their future.

In this study we propose a back propagation NN based model associated with a GA which aims at improving the prediction accuracy of CHD. The GANN method is applied to CHD dataset which has been obtained from a CHD data repository.

Artificial Neural Network (ANN) is an interconnected assembly of simple processing elements, unit or nodes, whose functionality is loosely based on the human brain neuron. The processing

ability of the network is stored in the interconnecting unit strengths, obtained by learning from a set of training patterns. ANN is computationally less intensive to suit complex applications though its structure is simple as stated in Liu and Liang (2005). Most ANN applications use Feed Forward (FF) architecture with gradient–based learning like Back Propagation (BP) algorithm (Yao and Liu, 1997) or modified BP algorithm (Hikawa, 2003). As the complexity of a network increases, the search space appears with more and more local optima and gradient–based learning may not always lead to global minima. Moreover BP needs complex operation, which restricts the search coverage. To improve the global convergence capability, an Evolutionary Algorithm (EA) can be used (Rumelhart and McClelland, 1986). Yao (1993) called this special class the "Evolutionary Artificial Neural Networks (EANN's)" in which evolution is another fundamental form of adaptation in addition to learning. EANN can be exploited to design the architecture, learn weight, adapt the learning rule and extract the rule from ANN (Martinetz *et al*., 1993). EA was broadly classified as Evolutionary Strategies (ES), Evolutionary Programming (EP) and Genetic Algorithms (GA), though many other types have emerged in the recent past (Palmes *et al.,* 2005). The capability of GA in the exploitation of information guides the direction of search towards feasible region and hence it converges at global optima.

GAs are search and optimization techniques based on the evolutionary ideas of natural selection and genetics (Goldberg, 1989). They follow the principle of survival of the fittest for better adaptation of species to their environment (Fogel, 1995). Both methods can be combined to get an optimal solution to a problem. Neuro-genetic approach is a hybrid of NN and GA. GA and NN may broadly be classified as non–invasive and invasive technique. Non–invasive method combines GA and gradient learning, while invasive method adapts the weight. The noninvasive method evolves the structure of the algorithm. Since it involves gradient method, proper initialization and network implementation is needed to overcome the local minima problem. On the other hand, invasive method uses GA for both weight and topology evolution of ANN.

MATERIALS AND METHOD

Dataset

The dataset comprises 200 cases of patients with different symptoms of CHD manifested by different CHD patients. Table 1 shows these symptoms and their categorization. Thirty three symptoms are categorized into two groups: General and Cardiologic signs.

In Table 2, the weighting scores allotted to different manifestation degrees ranging from 1 to 5 are indicated. A symptom with a weight of 1 may not manifest at all or may have low manifestation degree.

Optimization of Symptoms Using GA

All the thirty three symptoms may not manifest in all cases of the CHD. To determine related symptoms for CHD diagnosis, the following sets of mathematical equations are applied: If the set of symptoms (S) is given by Equation (1).

$$S=\{S_1, S_2, \ldots S_n\} \quad (1)$$

If *d* is the degree of each symptom of the disease, the following indices can be formulated:

i = 1 to n (where j is symptom index)

k = 1 to n (manifestation degree)

n = number of symptoms.

With GA our objective is to minimize the number of symptoms used for the prediction of the CHD. The objective function for this will be given as a quadratic error (cost) function which is derived from the mean weighted scores of the various symptoms given in Equation (2):

$$f(w_e) = 1/n \sum_{j=1}^{n} (w_t - w_j)^2 \quad (2)$$

Where w_t is a constant, the expected weighting score for a determinant symptom w_j and n the number of symptoms.

Equation (2) can be expanded to give Equation (3) while the error cost function can be written as Equation (4).

$$(w_t - w_j)^2 = {w_t}^2 - 2w_t w_j + {w_j}^2 \quad (3)$$

$$f(w_e) = \frac{1}{n}\sum_{j=1}^{n} |({w_t}^2 - 2w_t w_j + {w_j}^2)| \quad (4)$$

If n = 1, Equation (4) becomes Equation (5)

$$f(w_e) = {w_t}^2 - 2w_t w_j + {w_j}^2 \quad (5)$$

Equation (5) is subjected to the following constraints (C1 and C2)

C1: each of the symptoms must have high weighted score $3 < w_{jk} \leq 5$

C2: the highest manifestation degree is 100% and corresponds to $w_{jk} = 5$

Table 1: Symptom categorization of CHD

S/N	Category	Symptoms	Labels	Manifestation
1		Cough	S_1	V
2		Fatigue	S_2	A
3	General	Shortness of Breath	S_3	A
4		Vomiting	S_4	R
5		Headache	S_5	C
6		Diarrhea	S_6	R
7		Dyspnea	S_7	A
8		Cold	S_8	C
9		Problem with feeding	S_9	V
10		Easily tiring	S_{10}	A
11		Loss of weight	S_{11}	V
12		Chest pain	S_{12}	U
13		Dizziness	S_{13}	V
14		Clubbing	S_{14}	U
15		Fever	S_{15}	R
16		Syncope	S_{16}	V
17		Trauma	S_{17}	R
18		Sweating	S_{18}	C
19		Cyanosis	S_{19}	A
20		hypertension	S_{20}	A
21		Palpitations	S_{21}	U
22		Fluid build up	S_{22}	R
23		Anemia	S_{23}	R
24		Edema	S_{24}	U
25	Cardiologic	Chest Infection	S_{25}	U
26		Systolic BP	S_{26}	V
27		Diastolic BP	S_{27}	V
28		Increased Heart Beat	S_{28}	A
29		Cardiac Failure	S_{29}	A
30		Systolic Murmur	S_{30}	A
31		Diastolic Murmur	S_{31}	A
32		Regurgitation	S_{32}	V
33		Both Murmurs	S_{33}	A

[R=Rear, C= Common, U= Uncommon, V= Very common, A= Almost all]

Table 2: Manifestation weighting of various symptoms

Manifestation	Meaning	Label	Assigned
< 10%	Rear	R	1
10% to 34%	Uncommon	U	2
35% to 59%	Common	C	3
60% to 79%	Very common	V	4
80% to 99%	Almost all	A	5

Other components of the GA are specified in Table 3. Each chromosome is evaluated using the function shown in Equation (6).

$$F = 1/(1 + f(w_e)) \quad (6)$$

Table 3: GA component specification

Component	Value
Search method	GA
Population size	33
Encoding	Binary
Evaluation	Fitness function
Selection	Roulette wheel selection
Cross over function	Single point
Flip rate	0.01
Stoppage criteria	Till convergence to best solution is observed
Generation	50

To obtain the fitness value of each chromosome in the set of symptoms we calculate the average fitness of an individual chromosome using Equation (7) and Probability of the chromosome to be selected using Equation (8).

$$f_{Avg} = (\sum_{j=1}^{n} f_j)/n \quad (7)$$

$$p_j = f_j/f_{Avg} \quad (8)$$

The summary of the chromosome selection is shown in Table 4.

The expected count shows the number of a chromosome that can be selected for reproduction of offspring that will make up the population for the next generation. The integer part of the expected fitness determines the number of the chromosome that will be selected while the fractional part shows its chances of being repeated. Figure 1 shows the graphical representation of the expected counts of the various symptoms used in the GA.

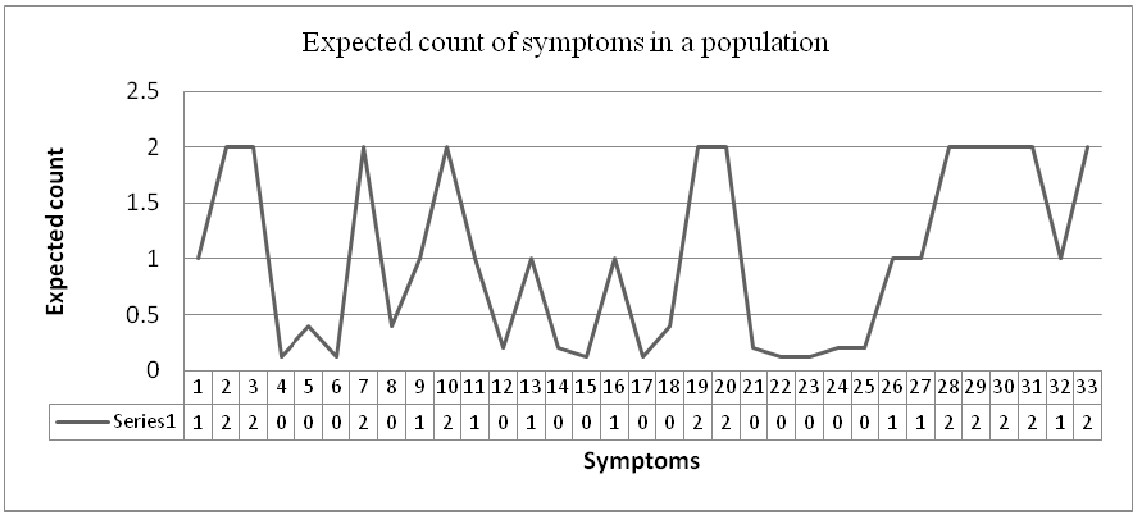

Figure 1: Graph of expected count of a symptom in the population

Table 4: Summary of Chromosome Selection of Initial Population

Chromosome ID	Weighting score (w_j)	Chromosome	Fitness (F) $F = 1/(1 + f(w_e))$	Probability (p) $p = f_i/f_{Avg}$	Expected count
1	4	00001	0.5000	1.0029	1
2	5	00000	1.0000	2.0057	2
3	5	00000	1.0000	2.0057	2
4	1	10000	0.0588	0.1179	0
5	3	00100	0.2000	0.4011	0
6	1	10000	0.0588	0.1179	0
7	5	00000	1.0000	2.0057	2
8	3	00100	0.2000	0.4011	0
9	4	00001	0.5000	1.0029	1
10	5	00000	1.0000	2.0057	2
11	4	00001	0.5000	1.0029	1
12	2	01001	0.1000	0.2006	0
13	4	00001	0.5000	1.0029	1
14	2	01001	0.1000	0.2006	0
15	1	10000	0.0588	0.1179	0
16	4	00001	0.5000	1.0029	1
17	1	10000	0.0588	0.1179	0
18	3	00100	0.2000	0.4011	0
19	5	00000	1.0000	2.0057	2
20	5	00000	1.0000	2.0057	2
21	2	01001	0.1000	0.2006	0
22	1	10000	0.0588	0.1179	0
23	1	10000	0.0588	0.1179	0
24	2	01001	0.1000	0.2006	0
25	2	01001	0.1000	0.2006	0
26	4	00001	0.5000	1.0029	1
27	4	00001	0.5000	1.0029	1
28	5	00000	1.0000	2.0057	2
29	5	00000	1.0000	2.0057	2
30	5	00000	1.0000	2.0057	2
31	5	00000	1.0000	2.0057	2
32	4	00001	0.5000	1.0029	1
33	5	00000	1.0000	2.0057	2

[f_t =16.4528, f_{avg} = 0.49857 and N=33.0]

The GA produces the best combination of input features to provide the solution with less computational complexity but more accuracy. The selected attributes using GA are shown in Table 5.

Table 5: Optimal Input Parameters

S/N	Symptom	Symptom code
1	Fatigue	S_2
2	Shortness of breath	S_3
3	Dyspnea	S_7
4	Easy Tiring	S_{10}
5	Cyanosis	S_{19}
6	Hypertension	S_{20}
7	High Systolic BP	S_{26}
8	High Diastolic BP	S_{27}
9	Heat beat	S_{28}
10	Cardiac Failure	S_{29}
11	Systolic Murmur	S_{30}
12	Diastolic Murmur	S_{31}
13	Both Murmur	S_{32}

GANN Architecture

A hybrid GANN developed in Amadin *et al.* (2015) was utilized for evolving the optimal symptoms for the diagnosis of CHD (Figure 2).

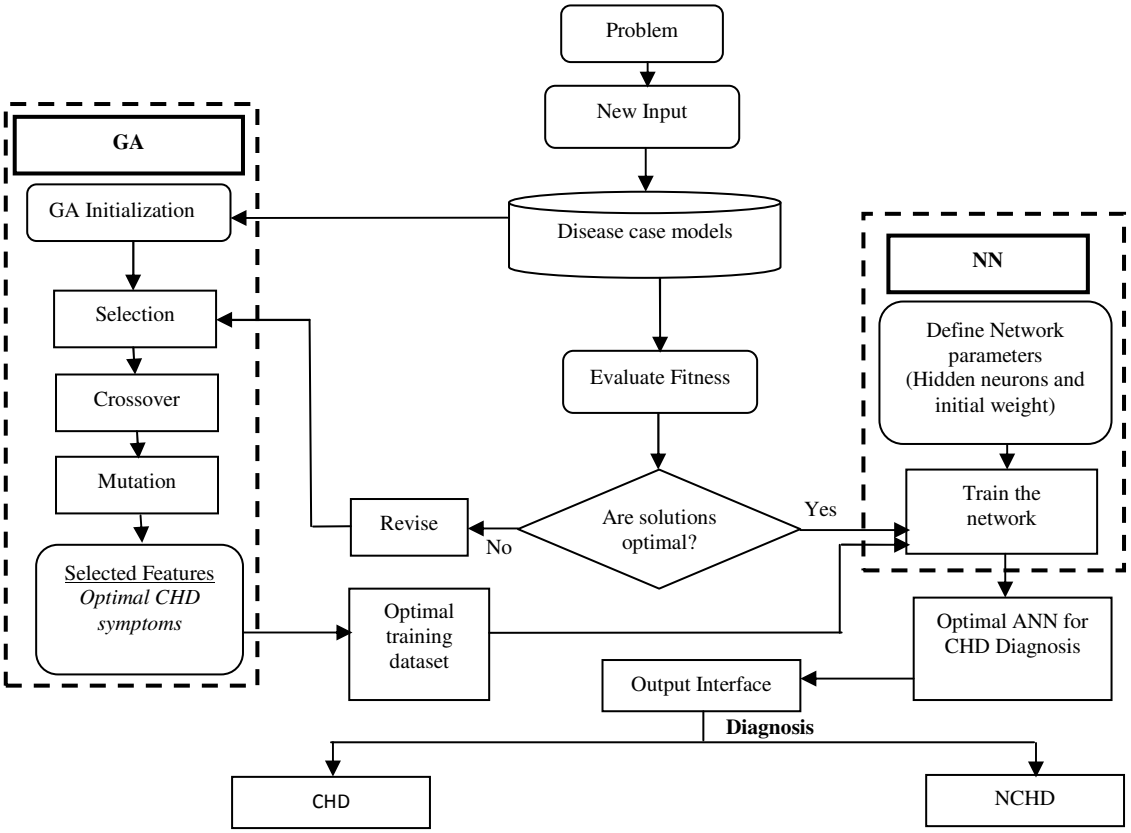

Figure 2: GANN process for the diagnosis of CHD

EXPERIMENT AND RESULTS

The system was implemented with MATLAB (R2014b). The NN and global optimization tool boxes were deployed in the system in simulating the GANN model. The system used seventy percent of the data (140 samples) for training. Testing and evaluation were carried out with 30 records (15%) each. In every training session, GA selects training samples randomly from the entire dataset thereby generating different values of Mean Square Error (MSE) as indicated in Tables 6 and 7.

The input data are preprocessed and are encoded within the range -1 to 1 where most of the inputs are of two-states. The data is partitioned and Table 6 shows the number of records in the training set, validation set and test set.

Table 6: Data partition for training, validation and test

S/N	Data partition set	Record	Percentage (%)
1.	Training	140	70.0
2.	Validation set	30	15.0
3.	Test set	30	15.0
4.	Ignored set	0	0.0
	Total	200	100.0

Table 7: Error rates on training of the network

Method	Training	Testing
Correlation coefficient	0.9967	0.9989
Mean absolute error	0.0029	0.0017
Root mean square error	0.0204	0.0195
Relative absolute error	1.0659	0.399
Root relative absolute error	5.3576	2.0009

Table 8 shows the average prediction accuracy of Neuro-Genetic approach and the BPNN approach. The average predictive accuracy is the corresponding percentage values of the regression values of the training, validation and test values. The result shows clearly that the hybrid GANN system predicts better than the BPNN with an accuracy of 98% and 89.8% respectively. This accuracy is as a result of the reduction in the input parameters obtained by the GA and thus eliminated the chances of removing vital parameters.

Table 8: Average Predictive Accuracy

Approach	Training	Validation	Testing
BPNN	89.2%	88%	90%
Neuro-Genetic	90.1%	91.1%	98%

CONCLUSION AND RECOMMENDATION

In this paper, we have presented a medical decision support system for congenital heart disease diagnosis based on Neuro-Genetic approach. Our goal was to improve the performance of CHD NN based MDSS by optimizing the input parameters to reduce training difficulty arising from redundant input features, reduce computational effort, enhance comprehensibility and increase the accuracy of predictions of the NN based system. Experimental results from a dataset of 200 sets of CHD symptoms showed that the system can give better prediction accuracy (98%) that will support medical expert in their decision making.

REFERENCES

Amadin, F. I, Egwuatu, J. O. and Ekong, V. E. (2015). A hybrid neuro-genetic approach for efficient Marburg virus disease diagnosis, *World Journal of Applied Science & Tech.*, 7 (2), 182-183.

Congenital Heart Disease Index, Available at: http://www.medicinenet. com/congenital_ heart_disease/ (Retrieved March 15, 2016)

Esposito, M. and Della, G. V (2008). *Congenital Heart Disease*: *An Ontology – Based Approach for the Examination of the Cardiovascular System*, Springer Link Publications, ISSN 0302-9743, 5177, pp. 509-516.

Fogel, B. D. (1995). Using Evolutionary Programming to Create Neural Networks that are Capable of Playing Tic-Tac-Toe, In *Proc. of IEEE International Conference of Neural Networks*, San Francisco, 2, pp. 875-880.

Goldberg, E. D., (1989). Zen and the Art of Genetic Algorithms, *In Proceedings of the Third International Conference on Genetic Algorithms*, pp. 80-85, Morgan Kaufmann.

Hikawa, H. (2003). A Digital hardware pulse–mode neuron with piecewise–linear activation function, *IEEE Transactions on Neural Networks*, 14 (15): 1028–1037.

Liu, J., and Liang, D. (2005). A Survey of FPGA–Based Hardware Implementation of ANNs. *Proc. IEEE Conf*., pp 915–918.

Martinetz, T. M., Berkovich, S. G., and Schulten, K. J. (1993). Neural Network for Vector Quantization and its Application to Time–series Prediction. *IEEE Trans. on Neural Networks*, 4: 558–569.

Palmes, P. P., Hayasaka, T., and Usui, S. (2005). Mutation–Based Genetic Neural Network. *IEEE Trans. on Neural Networks*, 16 (3): 587–600.

Rumelhart, D., and McClelland, J. (1986). *Parallel distributed processing: Explorations in microstructure of cognition*. Cambridge, MA: MIT Press.

Saxena, A., (2005). Congenital Heart Disease in India: a Status Report, *Indian Journal of Pediatrics*, 72: 595-598.

Vanisree, K. and Jyothi, S. (2011). Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks, *International Journal of Computer Applications* (0975 – 8887), 19 (6).

Yao, X., and Liu, Y. (1997). A New evolutionary system for evolving artificial neural networks, *IEEE Transactions on Neural Networks*, 8 (3), pp. 694–713.