# NEURAL NETWORK AND GENETIC ALGORITM FOR THE PREDICTION OF HEART DISEASES

Gigih Prasetyo Cahyono, Hani Ramadhan, Kartika Candra Kirana
5113201005, 5113201907, 5113201012
Pasca Sarjana Teknik Informatika
Institut Teknologi Sepuluh November, Surabaya

**Abstract**

Many survey showed that the mortality of heart disease increased, so some researches tried to search the alternative for predict the heart disease. The purpose of this research is integrated the genetic algorithm and back propagation for prediction the heart disease. Genetic Algorithm (GA) is used to select a good subset of features, so it could to optimize the connection weights of ANN that applied in prediction of heart disease.

**Keywords** : Back Propagation, Genetic Algorithm, Heart Disease

## 1. Introduction

The National Health Survey showed that the mortality that caused by heart disease around 26.3% in the year of 2001. Then, 17.5 million (30%) people died because of heart disease at 2005 (Andarmoyo & Nasriati, 2012). The mortality of heart disease increased up to 44% at 2007 (Dewayani, 2007).

Many researches tried to search the alternative for predict the heart disease. Adeli & Neshat (2010) design a fuzzy expert system for heart disease diagnosis. This system has 13 input field, 44 rule system, and one output. Input fields are chest pain type, blood pressure, cholesterol, resting blood sugar, maximum heart rate, resting electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan, sex and age. The output field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4 (distinguish presence (values 1, 2, 3, 4)). This experimental results showed that the accuracy of this system around 84%. Soni et al (2011) used weighted associative classifier (WAC) for hearth diagnosis. This system is very close to Adeli & Neshat's system, where it has 13 input field, some role and one output field. The role of this method based on the classification association rule. The accuracy is found to be 81.51% for WAC.

Kamruzzaman et al (2004) applied a modified feed forward neural network constructive algorithm to diagnose some disease, such as heart disease. This system also has 13 input field and one output field. The efficiency of proposed algorithm on this research is 86,47% for heart disease diagnose.

Dhanwani et al (2012) explored the hybrid model for stroke disease that integrated genetic algorithm and back propagation algorithm. The main idea is selecting a good subset of features, without sacrificing accuracy. In this research, a new hybrid model of Neural Networks and Genetic Algorithm to initialize and optimize the connection weights of ANN so as to improve the performance of the ANN that applied in prediction of stroke disease. The accuracy is found to be 98.67%. That is the best result from the others. We assumed that the stroke diagnosis and heart diagnosis are similar where both of them has some inputs and one outputs, so, this study tried to integrated genetic algorithm and back propagation for prediction the heart disease.

## 2. Hybrid Model of Back Propagation and Genetic Algorithm

Artificial neural network (ANN) is a paradigm of information processing that inspired by biological neural networks. The element of this system composed by some neurons that connected with each other' s. ANN trained a dataset like human for identify a specific purpose. The advantage of ANN are : (1) adaptive, (2) self-organization, and (3) real time operation. However, ANN also has some weakness: (1) that's not effective for numerical operation with high precision, (2) that's not efficient for logic, symbolic or algorithm operation, and (3) that's needed the long time to process the big data. There are many supervised learning algorithm of ANN. One of them is back propagation. (Sutoyo et al, 2011)

### 2.1 Back Propagation

Back propagation is a method of training artificial neural networks that included as a supervised learning method. The network learns from many inputs for a desired output. This method required a dataset that is used to making up the training set. It is most useful for feed-forward networks. Back propagation also required the activation function (Kusumadewi, 2004).

The architecture of back propagation involved the following: (1) input layer, (2) hidden layer, and (3) output layer. Figure 1 showed the architecture of back propagation.

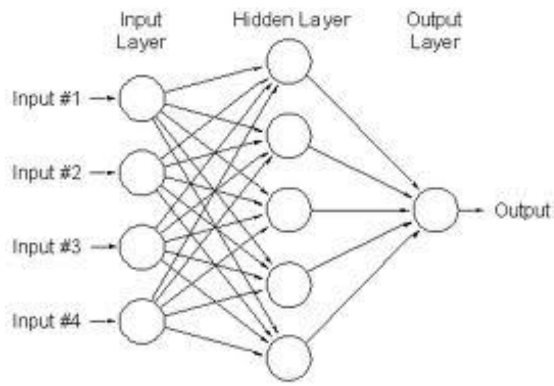



**Figure 1. The Architecture of Back Propagation**

The back propagation learning algorithm can be divided into three phases: propagation and weight update.

1. Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.
    a. Every input unit ($x_i$, i=1, 2, 3, …, n) got $x_i$ signal and send the signal to hidden layer
    b. Every hidden unit ($z_j$, j=1, 2, 3, …, p) sum the weight of input and implemented activation function to calculate output signal.

    $$z\_in_j = v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \quad [1]$$

    $$z_i = f(z\_in_j) \quad [2]$$

    c. Every output unit ($y_k$, k=1, 2, 3, …, m) calculate the weight of input and implemented activation function to calculate output signal

    $$y\_in_k = w_{0k} + \sum_{i=1}^{n} z_i w_{ij} \quad [3]$$

    $$y_i = f(y\_in_k) \quad [4]$$

2. Backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons.

a. every output unit ($y_k$, k=1, 2, 3, …, m) got pattern that appropriate with the training input patter, then calculate error with Equation 5. After that, count the weigh and refraction based on Equation 6 and 7.

$$\delta_k = (t_k - y_k) f'(y\_in_k) \quad [5]$$

$$\Delta w_{jk} = \alpha \, \delta_k \, z_j \quad [6]$$

$$\Delta w_{0k} = \alpha \, \delta_k \quad [7]$$

b. Every hidden unit ($z_j$, j=1, 2, 3, …, p) count input delta

$$\delta\_in_j = \sum_{k=1}^{m} \delta_k w_{jk} \quad [8]$$

$$\delta_j = \delta\_in_j \, f'(z\_in_j) \quad [9]$$

3. Weight update,for each weight-synapse follow the following steps:
   a. Multiply its output delta and input activation to get the gradient of the weight.
   b. Subtract a ratio (percentage) of the gradient from the weight. (Sutoyo et al, 2011)

This ratio (percentage) influences the speed and quality of learning; it is called the learning rate. The greater the ratio, the faster the neuron trains; the lower the ratio, the more accurate the training is. The sign of the gradient of a weight indicates where the error is increasing, this is why the weight must be updated in the opposite direction. Repeat phase 1 and 2 until the performance of the network is satisfactory (Wikipedia, 2013).

## 2.2 Genetic Algorithm

A genetic algorithm (GA) is a heuristic search that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover (Sutoyo, 2011).

In a genetic algorithm, a population of candidate solutions (called individuals) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties (its chromosomes or genotype) which can be mutated and altered. The solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible.

The evolution usually starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated. the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's genome is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm (Wikipedia, 2013). Based on the above explanation, Figure 2 showed the cycle of genetic algorithm

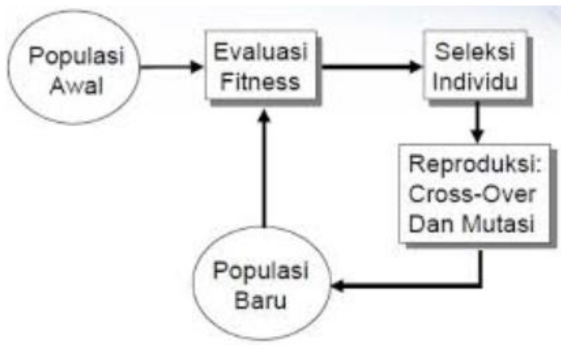



**Figure 2. The Cycle Of Genetic Algorithm (Sutoyo, 2011)**

Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. Once the genetic representation and the fitness function are defined, a GA proceeds to initialize a population of solutions and then to improve it through repetitive application of the mutation, crossover, inversion and selection operators.

### 2.3 Hybrid Model

The main idea of neural networks and genetic algorithm integration is selecting a good subset of features, without sacrificing accuracy. In this research, a new hybrid model of neural networks and genetic algorithm used to initialize and optimize the connection weights of ANN so as to improve the performance of the ANN.

## 3 Experiment

This system has 13 input field and one output. Table 1 described input fields.

**Table 1. Input Description**

| No | Feature | Description | Classification |
|---|---|---|---|
| 1 | Age | Age is defined in years. Based on the data sheet, vulnerable data used in this study were between 25-70 years old. | 25 $\leq$ age < 35 age = 0.2<br>35 $\leq$ age < 45 age = 0.5<br>45 $\leq$ age < 66 age = 0.7<br>$\geq$ 66 age = 0.7 |
| 2 | Sex | Men has a higher heart disease risk than women. The comparasion of abnormal heart risk between woman and men are 1:17 and 1:5 (Andarmoyo & Nasriati, 2012). | Men = 0.5<br>Woman = 0.2 |
| 3 | Chest pain type | Typical angina<br>(1) Substernal chest discomfort with a characteristic quality and duration that is<br>(2) provoked by exertion or emotional stress and<br>(3) relieved by rest or nitroglycerin<br>Atypical angina<br>• Meets 2 of the above characteristics<br>Noncardiac chest pain<br>• Meets ≤1 of the typical angina characteristics<br>asymptomatic<br>• without typical angina characteristics<br>(American Heart Association,1999) | typical angina = 0.9<br>atypical angina = 0.7<br>non-anginal pain = 0.5<br>asymptomatic = 0.3 |
| 4 | blood pressure | for the age <45. the increasing blood pressure have no significant influence on heart disease (Andarmoyo & Nasriati, 2012). | Normal (<130/<85) = 0.3<br>Higher Normal (130-139/85-89) = 0.5<br>Hypertension (>140/>90) = 0.7 |
| 5 | Cholestrol | Cholestrol have no significant influence on heart disease (Andarmoyo & Nasriati, 2012) | Normal (<200) = 0.2<br>Abnormal ($\geq$200) = 0.5 |
| 6 | Resting blood sugar | Total cardiac diabetes 3.2 times the amount of cardiac non-diabetic (Andarmoyo & Nasriati, 2012). | Diabetes = 0.8<br>Normal = 0.4 |
| 7 | Resting | ECG showed the risk of heart disease | Normal = 0.0 |

| | | | |
|---|---|---|---|
| | electrocardiography (ECG) | (American Heart Association, 1999) | ST-T wave abnormality = 0.9<br>Probable or definite left ventricular hypertrophy by Estes' criteria = 0.9 |
| 8 | Thalac | maximum heart rate achieved. Calculate thalac with : 100 < thalach < 220 – Umur | Normal = 0.0<br>Abnormal = 0.9 |
| 9 | Exang | exercise induced angina | Yes/1.0 = 0.5<br>No/0.0 = 0.5 |
| 10 | Oldpeak | ST depression induced by exercise relative to rest | Yes/1.0 = 0.5<br>No/0.0 = 0.5 |
| 11 | Slope | The slope of the peak exercise ST segment | Upsloping Value = 0.5<br>Flat Value = 0.5<br>Downsloping = 0.5 |
| 12 | CA | The number of major vessels colored by floursopy | [0,1,2,3] = 0.5 |
| 13 | Thalium Scan | - | Not found cold spot = 0.0<br>Found cold spot = 0.9 |

The output field refers to the presence of heart disease in the patient. It is integer valued from 0 (the patient have no the risk of heart disease) to 1 (the patient have the risk of heart disease).

The step of hybrid model involved the following: (1) First, define the network, (2) generate initialization randomly, (3) train the network and compute the fitness, (4) do selection, crossover, or mutation to choose the best feature, (5) train the network using back propagation, (6) study the performance by presenting testing pattern to get the optimal ANN for heart diseases prediction. The system has implemented in java platform and used UCI dataset. Figure 1. showed the hybrid algorithm flowchart of back propagation and genetic algorithm.

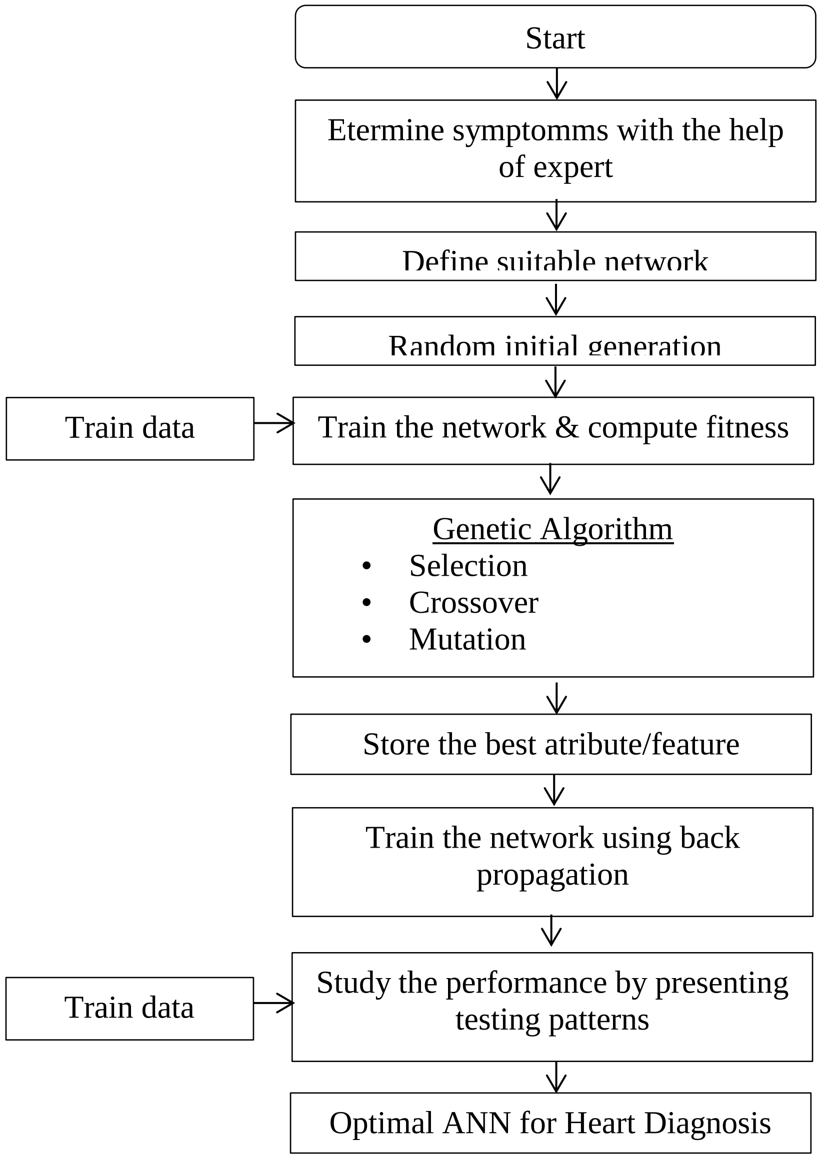



**Figure 3. Scheme of The Hybrid Algorithm**

## 4 Result

Figure 4 showed the GUI of heart disease prediction. The system used 202 training data and 74 testing data. Table 2 showed genetic algorithm parameter. Table 3 showed the data partition set.

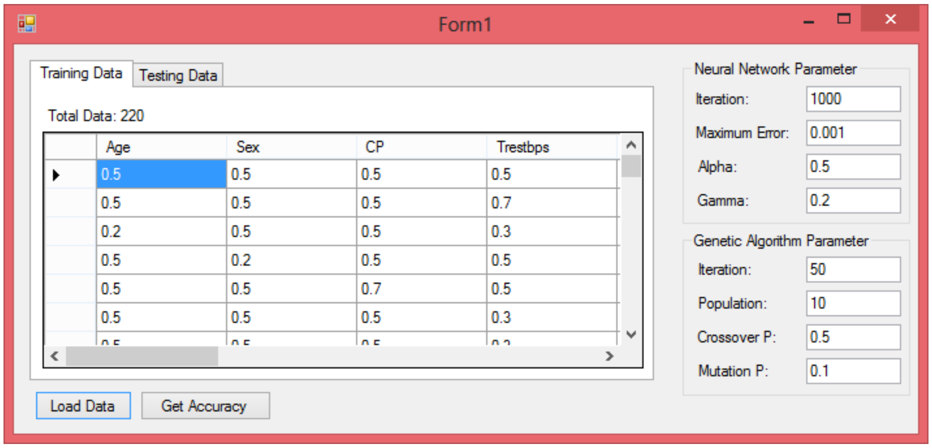
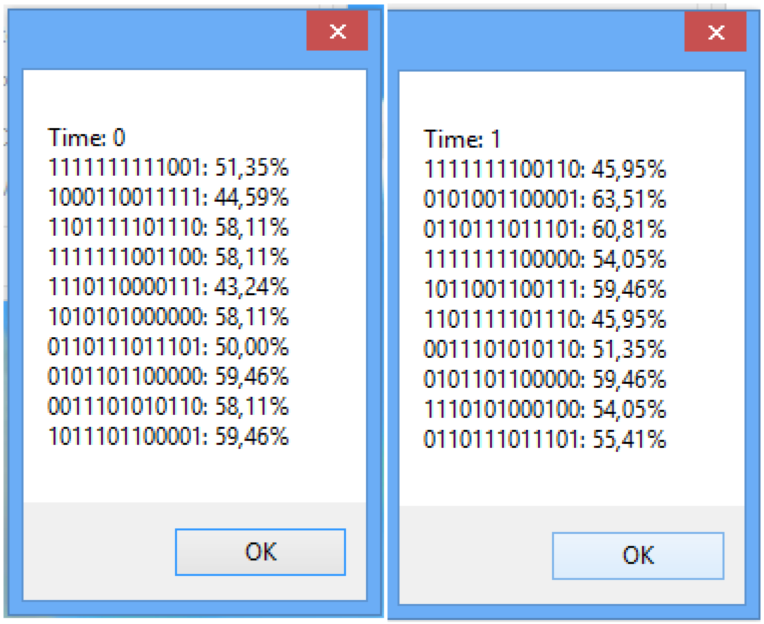



**Figure 4.** GUI of Heart Disease Prediction




(a) (b)

**Figure 5.** Comparasion of Before and After Genetic Algorithm

**Table 2.** Genetic Algorithm Parameter

| Search Method | Genetic Algorithm |
|---|---|
| Population size | 10 |
| Iteration | 50 |
| Crossover probability | 0.5 |
| Mutation probability | 0.1 |

**Table 3.** Data Partition

| Search Method | Percentage (%) |
|---|---|
| Before integrated with GA | 59.46 |
| After integrated with GA | 68.92 |

**Table 4.** Data Partition

| Search Method | Record | Percentage (%) |
|---|---|---|
| Training set | 220 | - |
| Testing set | 74 | 68.92 |

## 5 Conclusion

From the result, we get various results of accuracy. Those results are the cause of integration of Genetic Algorithm and Backpropagation Neural Network. It may not give a significant improvement because of the similarity of features by the range of data and the number of training data. But, we may conclude that the integration of Genetic Algorithm improves the efficiency of the Neural Network Algorithm by selecting significant features.

## References

Adeli, Ali & Neshat, Mehdi (2010) A Fuzzy Expert System for Heart Disease Diagnosis. IMECS. Vol.1.

Andarmoyo, Sulistyo & Nasriati, Ririn (2012) Faktor Resiko Kejadian PJK (Penyakit Jantung Koroner) pada Usia Muda. Hibah Penelitian Kementerian Pendidikan dan Kebudayaan.

American Heart Association (1999) A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients With Chronic Stable Angina) available from : http://circ.ahajournals.org/content/99/21/2829.full (Accessed 8/12/2013)

Dhanwani, D. W. A., 2013. "Study of Hybrid Genetic Algorithm using Artificial Neural Network in Data Mining for he Diagnosis of Stroke Disease". International Journal of Computational Engineering Research, 3(4), pp. 95-100.

Kamruzzaman, S. M et al (2004). “Medical Diagnosis Using Neural Network ”. ICECE. Vol.3,Hal 538-540

Kusumadewi, Sri. 2004. Membangun Jaringan Saraf Tiruan Menggunakan Matlab dan Excel Link.Graha Ilmu: Yogjakarta

Sony Jyoti, e. a., International Journal on Computer Science and Engineering . 2011. Intelligent and Effective Heart Disease Predictin System using Weighted Associative Classifiers, 3(6), pp. 2385-2392.

Sutoyo et al. 2011. Kecerdasan Buatan. Andi: Yogjakarta

Wikipedia. 2013. Back Propagation, (Online) http://en.wikipedia.org/wiki/Backpropagation 2013 (diakses 14 Desember 2013)

Wikipedia. 2013. Genetic Algorithm, (Online) http://en.wikipedia.org/wiki/Genetic_algorithm 2013 (Accessed 14 Desember 2013)