IJISCS (International Journal of Information System and Computer Science)

Full Paper
eISSN 2598-246X
pISSN 2598-0793

NEURAL NETWORK OPTIMIZATION WITH GENETIC ALGORITHM FOR HEART DISEASE PREDICTION

M Agus Badruzaman Al Khoir[1], Sriyanto[2]
[1,2] Department of Informatics Engineering, Institute of Informatics and Business Darmajaya Lampung, Indonesia
[1.2] Jl. ZA. Natural Fence No. 93, Gedong Meneng, Kec. Rajabasa, Bandar Lampung City Lampung, Indonesia

*Corresponding author
khoir65@gmail.com

Article history:
Received July 12, 2022
Revised August 26, 2022
Accepted August 29, 2022

Keywords:
Prediction;
Coronary Heart Disease;
Neural Network;
Genetic Algorithm.

Abstract
Coronary Heart Disease (CHD) is a contributor to the number 1 cause of death in the world besides cardiovascular disease. The tendency of Indonesian people who do not know and ignore coronary heart disease is a factor that causes Indonesia to be high as a contributor to deaths caused by coronary heart disease. With this research, it is expected to produce new predictions of heart disease using genetic optimization of neural networks with better prediction results and can obtain algorithms with new percentage values in predicting coronary heart disease. Genetic optimization of neural network is used because the algorithm follows the human nervous system which has the characteristics of parallel processing, processing elements in large quantities and fault tolerance. The results of the research carried out are the accuracy obtained by 82.18% and increased to 83.50% after using genetic algorithm optimization, from these results it can be concluded that the neural network algorithm can be better if it is supported by genetic algorithm optimization.

1.0 INTRODUCTION

Coronary heart or coronary artery disease is a disease caused by the buildup of plaque in the blood vessels which results in blockage of the main blood vessels of the heart or damage to the main blood vessels of the heart. Based on the statistics agency, coronary heart disease contributes 45% of deaths every year from a total of 9.4 million deaths in the world. The number of deaths caused by coronary heart disease is estimated to increase by 23.3 million in 2030. Based on WHO data, Indonesia is one of the countries with a high mortality rate due to coronary heart disease, which is 12.9%. The high percentage of coronary heart disease in Indonesia is largely due to the unhealthy lifestyle of the Indonesian people. A study conducted by the Ministry of Health stated that diet and lack of physical activity as well as the tendency of Indonesian people who do not know and ignore coronary heart disease are factors that cause Indonesia's high rate of death due to coronary heart disease.

Based on the research of Duwi Cahya Putri Buani (2021), the results of research using genetic algorithms for heart failure prediction are condensed that the nave Bayes algorithm has maximum performance in making predictions if before making predictions feature selection is carried out first before applying genetic algorithms. The results of trials using the same data obtained a comparison of the level of accuracy of 69.60% in the first trial without using genetic algorithms. The second trial was carried out after feature selection with genetic algorithms, in the second trial the accuracy increased to 96.67% [1] . Research conducted by Debi Setiawan, et al (2019), the results of the study show that genetic algorithms are able to

predict autoimmune diseases in the first iteration, this has been seen with a value of 18.88 taken from the genetic algorithm process. [2] Research conducted by Ramalia Noratama and Debi Setiawan (2021) showed that genetic algorithms can predict SLE, from 30 patients, 14 patients with SLE were not detected, 9 patients were detected with SLE that attacks the brain (Cerebral), 6 patients were detected with SLE that attacks the kidneys (renal), 1 patient was detected with SLE that attacks the skin (dermatological). If SLE is indicated, the patient is advised to go to the hospital for further examination. With the application of genetic algorithms in the prediction of SLE disease, it can reduce the mortality rate due to PTM, especially due to SLE disease. The results of this genetic algorithm calculation will be developed into an application. It is hoped that the applications created can be used by the community. So that people can do their own Lupus check using the application [3] . Bakthiar Rifai (2013) tested a heart disease prediction model using a neural network algorithm using data from patients who were adequately treated for the disease. This model produces accuracy, precision, *recall* and AUC values resulting in 91.45% accuracy, 92.79% precision and 0.937 AUC value using the *Neural Network algorithm*.

Several studies that have been conducted have focused on genetic algorithms combined with other algorithms for prediction of heart failure, autoimmune, SLE and heart disease, based on several previous studies genetic algorithms have a high level of accuracy and are able to support other algorithms in the prediction process. This study follows up on previous research conducted by bahtiar rifai (2013) to improve accuracy in predicting heart disease by using *Genetic Algorithm Optimization* on *Neural Networks*.

This research focuses on Neural Network Optimization using Genetic Algorithm to Predict Heart Disease using public data obtained from Kaagle. With this research, it is expected to produce new predictions of heart disease using genetic optimization of neural networks with better prediction results and can obtain algorithms with new percentage values in predicting coronary heart disease.

2.0 THEORETICAL

2.1. Optimization

According to the Big Indonesian Dictionary (Depdikbud: 1995: 628) optimization comes from the word optimal which means the best, the highest. Optimization is also interpreted as a measure where all needs can be met from the activities carried out. According to Winardi (1996:363) optimization is a measure that causes the achievement of goals. In general, optimization is the search for the best value from the available functions in a given context [4].

John Holland who came from the University of Michigan in 1975 introduced genetic algorithms or soft computing models, where genetic algorithms are heuristic search techniques based on biological evolutionary mechanisms that are modified from Darwin's theory and genetic operations on chromosomes and are often used to solve optimization problems. The search for genetic algorithms uses cycles. The cycle will continue to look for the best fitness based on the predetermined role. The cycle of the genetic algorithm was first introduced by David Goldberg. When the best fitness value has been found, the process will stop. The genetic algorithm process starts from the initial population generation to the formation of a new population.

2.2. Neural Network

Neural Network is a category of Soft Computing science, adopting the ability of the human brain that is able to provide stimulation, process, and provide output. The output is obtained from a variety of stimulation and processes that occur in the human brain. Neural Network is an AI (Atrifacial Intelligence) method whose concept imitates the nervous network system that exists in the human body, where nodes are formed that are interconnected with each other. The nodes are connected through a link called the term weight.

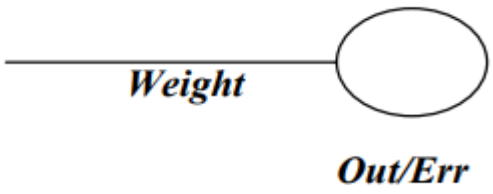

Figure 1 Node and Weight

The basic idea of the algorithm follows the human nervous system which has the characteristics of parallel processing, processing elements in large quantities and fault tolerance [5] . Neural network is built from many nodes/units which are connected by direct link. Links from one unit to another are used to propagate activation from the first unit to the next unit. Each link has a numerical weight. This weight determines the strength and marker of a connectivity.

2.3. Genetic Algorithm

Genetic Algorithms have been known since 1975 and have been widely used in business, engineering and scientific applications. This algorithm can be used to get the right solution for the optimal problem from one variable or multiple variables. Before this algorithm is run, what problem you want to optimize must be stated in the objective function, which is known as the fitness function. If the fitness value is greater, then the resulting system is getting better. Although initially all fitness values are likely very small (because this algorithm generates them randomly), some will be higher than others. Chromosomes with a high fitness value will provide a high probability of reproducing in the next generation. So that for each generation in the evolutionary process, the fitness function, which simulates natural selection, will push the population towards increasing fitness [6] .

Genetic Algorithm is a technique to find the optimal solution of a problem that has many solutions. This technique will search from several solutions obtained to get the best solution according to predetermined criteria or what is known as the fitness function. This algorithm is included in the group of evolutionary algorithms using Darwin's evolutionary approach in the field of Biology such as inheritance, natural selection, gene mutation and combination (crossover). Because it is an optimal search technique in the field of computer science, this algorithm is also included in the metaheuristic algorithm group.

3.0 METHODOLOGY

3.1. Research Flow

At the research stage, it begins with a literature review to study Neural Network and Genetics in predicting heart disease, followed by data collection on heart disease. Initial processing is carried out as processing or compiling the data obtained so that it can be followed up at the experimental stage and model testing is carried out using the Genetic Algorithm and Neural Network. At the end of the study , evaluation and validation were carried out to ensure the results of the study. The flow of the research carried out can be seen in Figure 3.1.

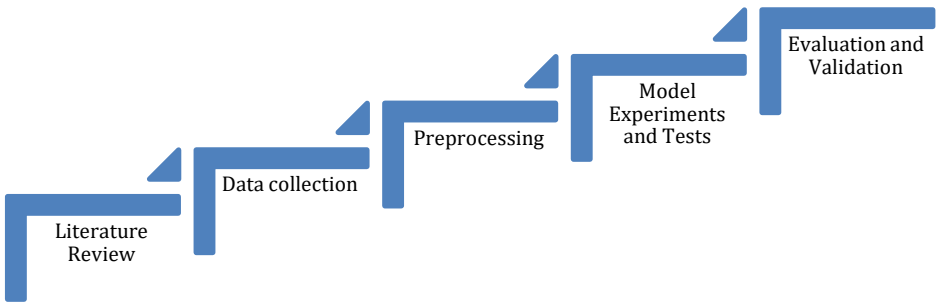

Figure 2 Research Flow

3.2. Method of collecting data

The data collection method is an important thing in research and is a strategy or method used by researchers in collecting the data needed in their research. Data collection methods used in this study are:

a. Literature review
   The literature review is carried out by reading, citing and making notes sourced from library materials that support and are related to research in this case regarding Naïve Bayes data mining , Random Forest and logistic regretion.
b. Documentation

The source of the data that the research uses is taken from Kaagle which is free heart disease data and can be downloaded publicly and will be processed using Rapid Miner.

3.3. Neural Network Algorithm

Determination The learning rate value is by testing the Learning Rate input. Performance evaluation which serves to test the performance of the classifier which consists of recall, precision, and accuracy. A recall is a positive data set that is correctly classified as positive data. Precision is a data set classified as positive that is really positive. Accuracy is the accuracy of data classification determined from training data [7] . The formula used is:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FP} \quad (2)$$

$$Akurasi = \frac{(TN+TP)}{(TN_{FN}+TP+FP)} \quad (3)$$

4.0 RESULANTS AND DISCUSSION

4.1. Testing Data Using Neural Network

In this study, the data used amounted to 304 data divided into two for training data and testing data. And the rest of the data is used for testing. For testing, the system will take a number of data randomly from the dataset according to the number of input testing data entered in the form by the user. The training data is modeled using Neural Network parameters.

Row No.	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope
1	63	1	3	145	233	1	0	150	0	2.300	0
2	37	1	2	130	250	0	1	187	0	3.500	0
3	41	0	1	130	204	0	0	172	0	1.400	2
4	56	1	1	120	236	0	1	178	0	0.800	2
5	57	0	0	120	354	0	1	163	1	0.600	2
6	57	1	0	140	192	0	1	148	0	0.400	1
7	56	0	1	140	294	0	0	153	0	1.300	1
8	44	1	1	120	263	0	1	173	0	0	2
9	52	1	2	172	199	1	1	162	0	0.500	2
10	57	1	2	150	168	0	1	174	0	1.600	2
11	54	1	0	140	239	0	1	160	0	1.200	2
12	48	0	2	130	275	0	1	139	0	0.200	2
13	49	1	1	130	266	0	1	171	0	0.600	2
14	64	1	3	110	211	0	0	144	1	1.800	1
15	58	0	3	150	283	1	0	162	0	1	2
16	50	0	2	120	219	0	1	158	0	1.600	1
17	58	0	2	120	340	0	1	172	0	0	2
18	66	0	3	150	226	0	1	114	0	2.600	0
19	43	1	0	150	247	0	1	171	0	1.500	2
20	69	0	3	140	239	0	1	151	0	1.800	2
21	59	1	0	135	234	0	1	161	0	0.500	1
22	44	1	2	130	233	0	1	179	1	0.400	2
23	42	1	0	140	226	0	1	178	0	0	2

Figure 3. Test data input

In the data training process, the dataset used is 70 percent of the total modeled dataset as shown in Figure 4.1 below.

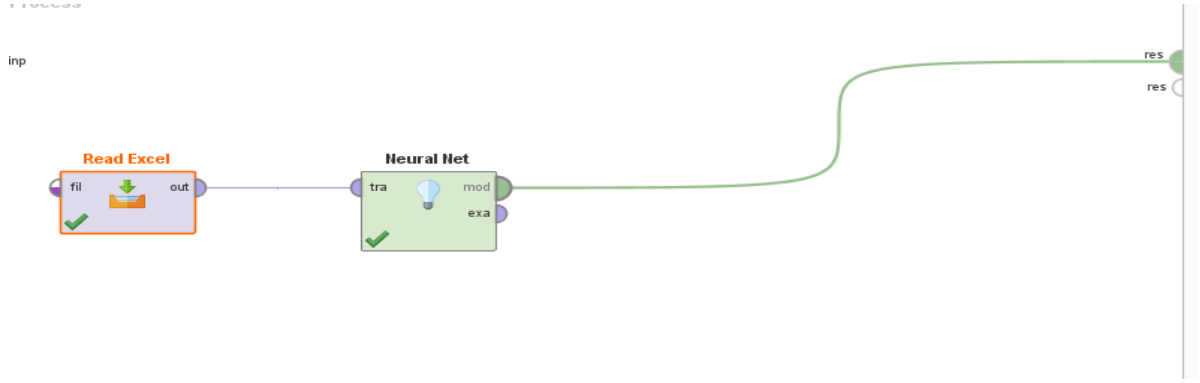

Figure 4. Training Set Process

In the use of *Neural Network* , the network is given an *input* pattern as a *training pattern* and the results of the pattern are formed in the process to process in the *hidden layer* to the output section. The highest accuracy results found were 82.53% using 9 inputs, 1 hidden layer with 7 attributes and 2 outputs, as shown in Figure 5.

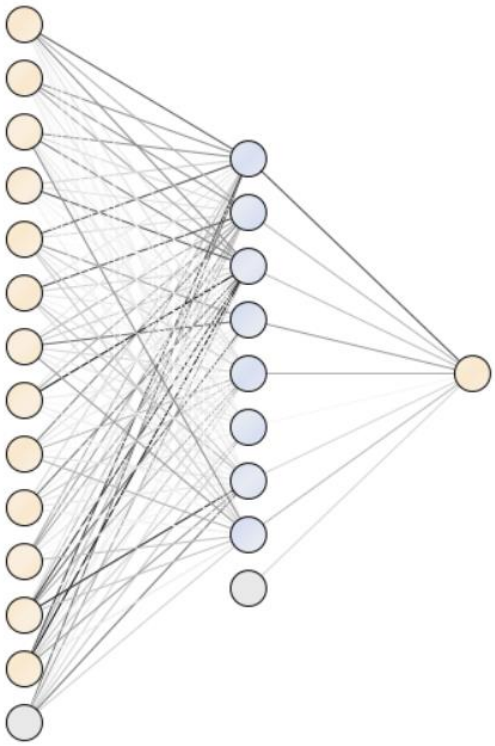

Figure 5. Neural Network

To calculate better accuracy, validation is carried out first using cross validation

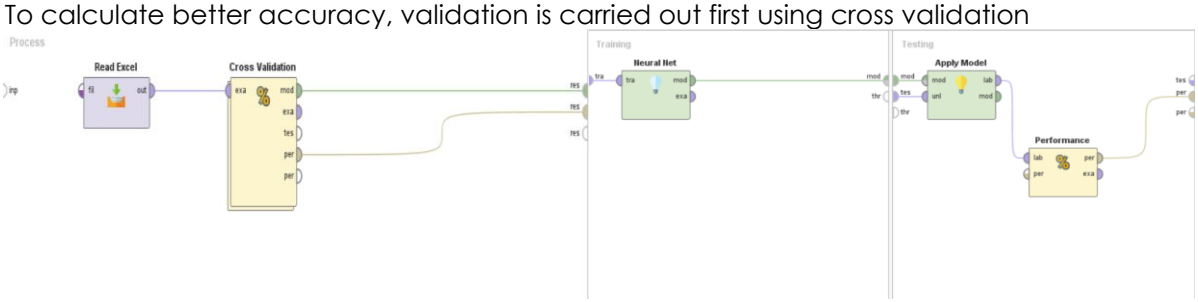

Figure 6. Cross Validation Process

After the cross validation process is carried out, the accuracy results are obtained as shown in Figure 7.
accuracy: 82.19% +/- 8.47% (micro average: 82.18%)

	true Sakit	true Tidak	class precision
pred. Sakit	141	30	82.46%
pred. Tidak	24	108	81.82%
class recall	85.45%	78.26%	

Figure 7. Accuracy Results

Figure 7 shows a table with an accuracy value of 82.18%, with a prediction accuracy of 82.46% for heart disease and a prediction for no heart disease at 81.82%.

4.2. Genetic Algorithm Optimization

After testing using the Eural Network, the next step will be to optimize the genetic algorithm for comparison of accuracy results. The genetic algorithm optimization process can be seen in Figure 8 below.

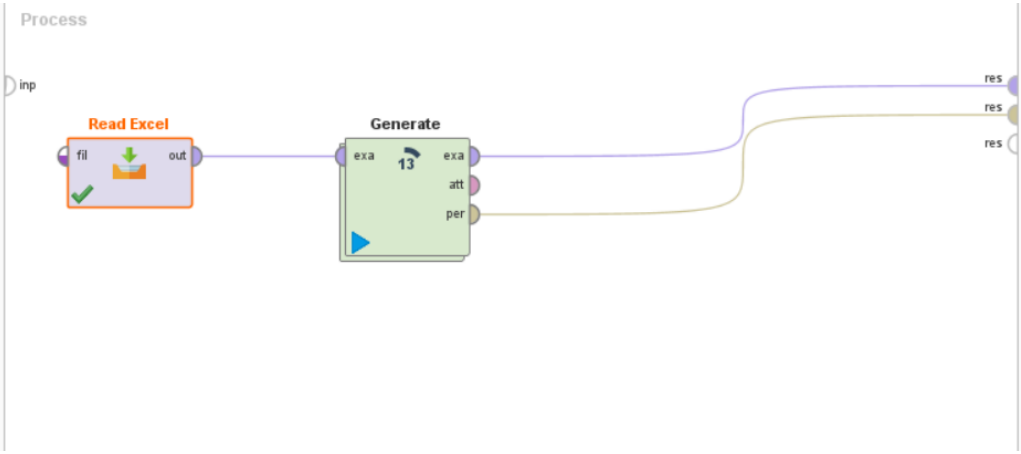

Figure 8. Optimization Process

After the optimization process is carried out, the accuracy results are obtained as shown in Figure 9 below

accuracy: 83.45% +/- 7.40% (micro average: 83.50%)

	true Sakit	true Tidak	class precision
pred. Sakit	148	33	81.77%
pred. Tidak	17	105	86.07%
class recall	89.70%	76.09%	

Figure 9. Genetic Algorithm Accuracy Results

Figure 10 shows a table with an accuracy value of 83.50%, with an accuracy of prediction of heart disease of 81.77% and prediction of not having heart disease of 86.07%.

4.3. Results Comparison

After testing the data using the classification model, it can be seen in table 2 the level of accuracy of the neural network method, and the optimization of the genetic algorithm.

Table 2. Analysis of Results

	Neural Network	Genetic Algorithm Optimization
Precision Class		
- **Pred. Sick**	- 82.46%	- 81.77%
- **Pred. Not**	- 81.82%	- 86.07%
Class Recall		
- **True. Sick**	- 85.45%	- 89.70%
- **True. Not**	- 78.26%	- 76.09%
Accuracy	82.18%	83.50%

Based on the tests that have been carried out using a neural network with optimization of the genetic algorithm, the level of accuracy obtained increases. Before using the optimization of the genetic algorithm, the accuracy obtained was 82.18% and increased to 83.50% after using the optimization of the genetic algorithm, from these results it can be concluded that the neural network algorithm can be better if it is supported by optimization of the genetic algorithm.

5.0 CONCLUSION

The conclusion obtained from this research is that the Tree, Random Forest, and Logistic Regression methods can be used to predict heart disease. The use of genetic algorithm optimization with neural network gives quite good results where the accuracy value is 82.18% and becomes 83.50% after using genetic algorithm optimization. From the test data that has been carried out by training and testing, it can be seen that the level of accuracy of testing data for predicting heart disease is getting better after using genetic algorithm optimization. Based on the results of this study, it can be a contribution to the relevant institutions, but there are several things that the authors can suggest for further research, namely further exploration of heart disease prediction data using other models that have many features that need to be tried, considering the problem is more complex than just identify patients with heart disease.

REFERENCES

[1] D. Cahya and P. Buani, "Application of Naïve Bayes Algorithm with Feature Selection of Genetic Algorithm for Predicting Heart Failure," vol. 9, no. 2, pp. 43–48, 2021.

[2] D. Setiawan, RN Putri, and R. Suryanita, "Debi Setiawan, 2) Ramalia Noratama Putri, 3) Reni Suryanita," vol. 4, no. 1, pp. 8–19, 2019.

[3] RN Putri *et al.* , "Prediction of Systemic Lupus Erythematosus Disease Using Genetic Algorithms," *Digits. Zo. J. Teknol. inf. commune.* , vol. 12, no. 1, pp. 19–31, 2021.

[4] S. 2009, "Optimizing Economic Value," pp. 1–70, 2009.

[5] A. Budiarti, "Chapter 2 theoretical foundations," *App. and anal. lit. Fasilkom UI* , pp. 4–25, 2006.

[6] LW Kusuma, "Design of an expert system for preparing food menus for pulmonary tuberculosis patients with genetic algorithms," Multimedia Nusantara University, 2014.

[7] N. Purwati, R. Nurlistiani, N. Purwati, R. Nurlistiani, and O. Devinsen, "DATA MINING WITH NEURAL NETWORK ALGORITHM AND DATA VISUALIZATION," vol. 20, no. 2, 2020.