

Effect of Genetic Algorithm on Prediction of Heart Disease Stadium using Fuzzy Hierarchical Model

Dian Palupi Rini^{1*}, Defrian Afandi², Desty Rodiah²

¹Master of Computer, Faculty of Computer Science, Sriwijaya University

²Department of Informatic Engineering, Faculty of Computer Science, Sriwijaya University

*dprini@unsri.ac.id

ABSTRACT

The Fuzzy Hierarchical Model method can be used to predict the stage of heart disease. The use of the Fuzzy Hierarchical Model on complex problems is still not optimal because it is difficult to find a fuzzy set that provides a more optimal solution. This method can be improved by changing the membership function constraints using Genetic Algorithm to get better predictions. Tests carried out using 282 heart disease patient data resulted in a Root Mean Squared Error (RMSE) value of 0.55 using the best Genetic Algorithm parameters, including population size of 140, number of generations of 125, and a combination of cross-over rate and mutation rate of 0.4 and 0.6 whereas the RMSE value generated by the Fuzzy Hierarchical Model before being optimized by the Genetic Algorithm was 0.89. These results indicate an increase in the predictive value of the Fuzzy Hierarchical Model after being optimized using the Genetic Algorithm.

Keywords: Prediction, Heart Disease, Genetic Algorithm, Fuzzy Hierarchical Model.

1. INTRODUCTION

Diagnosis of the stage of heart disease depends on risks and symptoms such as age, blood pressure, blood sugar levels, smoking, and others. Risk factors for heart disease are numerous and uncertain and there is no clear line between healthy and diagnosed disease, making it difficult for medical personnel to diagnose heart disease accurately [1].

Fuzzy logic is more flexible to vagueness and problem change. Therefore, the process of diagnosis of the disease is very suitable using fuzzy logic because of the vagueness of symptoms and diseases [2]. Fuzzy logic methods such as the Fuzzy Hierarchical Model method can be used for the diagnosis of heart disease contained in the Pramasari study [3]. However, the use of fuzzy logic to complex problems can result in performance that sometimes lacks maximum results.

A common problem found is the difficulty in obtaining fuzzy sets and fuzzy rules that are suitable for solutions to problems [4]. Changes in the membership function can affect the performance results of the fuzzy method used so that the fuzzy method can be optimized by changing its membership function [5]. One optimization method was suitable for the optimization of fuzzy method membership functions, namely Genetic Algorithms. The use of genetic algorithms provides an approach to adjust membership functions automatically and comparatively.

Dian Palupi Rini, Defrian Afandi, Desty Rodiah
Effect of Genetic Algorithm on Prediction of
Heart Disease Stadium using Fuzzy Hierarchical Model

Based on the explanation above, this study will optimize the limitations of membership function in the Fuzzy Hierarchical Model by using Genetic Algorithms in case of studies for the results of predictions of the stage of heart disease.

2. MATERIAL AND METHODS

This study describes the process of generation of chromosomes with genetic algorithms based on input data as parameters in the form of population size, the number of generations, crossover rate value, and mutation rate value so as to produce the best membership function limits for heart disease risk factors and get optimal results for prediction of heart disease diagnosis in the form of categorical values whether healthy (stage 0), stage I, stage II, stage III, or stage IV with Fuzzy Hierarchical Model as the basis for calculation.

2.1 HEART DISEASE

Heart disease has stage levels as follows [3].

1. Stage I, experiencing a narrowing of the heart only, has not suffered damage to the heart muscle. Symptoms are still not visible so many people have not realized that they have been exposed to stage I. Examples of patients suffering from hypertension (high blood pressure), a family history of cardiomyopathy.
2. Stage II, patients have heart disease related to kidney failure but have not shown symptoms clearly.
3. Stage III, Symptoms of heart disease experienced by patients are visible.
4. Stage IV, a condition in which a heart failure patient needs further treatment, such as heart transplant surgery.

2.2 FUZZY HIERARCHICAL MODEL

Fuzzy Hierarchical Model is a fuzzy logic method that can simplify the rule that will be used in the system aimed at speeding up the calculation process because the number of input variables can affect the number of rules obtained and will later be used in determining the output results [6]. In the working stages of the Fuzzy Hierarchical method This model, input variables are processed gradually according to the specified hierarchical part and stored in temporary variables. These temporary variables will then be reprocessed with input variables or other temporary variables to produce outputs that in this study are the results of a diagnosis of stage of heart disease.

2.3 FUZZIFICATION

Fuzzification is the process of converting crisp values into fuzzy values in the form of fuzzy set linguistic variables using the membership function [2].

2.3.1 MEMBERSHIP FUNCTION

The membership function is a graph that represents the magnitude of the membership degrees of each input variable located at intervals 0 and 1 [7]. This

study used the S curve membership function (sigmoid). The S curve representation (sigmoid) is a growth and depreciation curve. The S-curve representation of growth and depreciation can be seen in Figure 1.

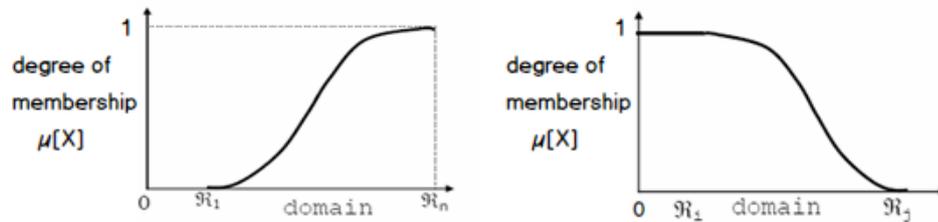


FIGURE 1. S-Curve Growth and S-Curve Depreciation

The membership function of the growth and depreciation curve can be seen in equations 1 and 2.

$$S(x, \alpha, \beta, \gamma) = \begin{cases} 0 & ; x \leq \alpha \\ 2 \left(\frac{x-\alpha}{\gamma-\alpha} \right)^2 & ; \alpha \leq x \leq \beta \\ 1 - 2 \left(\frac{\gamma-x}{\gamma-\alpha} \right)^2 & ; \beta \leq x \leq \gamma \\ 1 & ; x \geq \gamma \end{cases} \quad (1)$$

$$S(x, \alpha, \beta, \gamma) = \begin{cases} 1 & ; x \leq \alpha \\ 1 - 2 \left(\frac{x-\alpha}{\gamma-\alpha} \right)^2 & ; \alpha \leq x \leq \beta \\ 2 \left(\frac{\gamma-x}{\gamma-\alpha} \right)^2 & ; \beta \leq x \leq \gamma \\ 0 & ; x \geq \gamma \end{cases} \quad (2)$$

2.4 FUZZY INFERENCE SYSTEM

A fuzzy Inference System (FIS) is a system that makes decisions based on fuzzy set theory, fuzzy rule base, and logic concepts in fuzzy [8]. This study used the Tsukamoto method as its inference system. FIS Tsukamoto is based on the concept of monotonous reasoning. For each consequent to the rule in the form of IF-THEN must be represented into a fuzzy set with monotonous membership functions [9]. The inference output of each rule is given expressly (crisp) based on α -predicate. Calculates α -predicate based on equation 3.

$$\alpha\text{-predicate} = \text{MIN} (\mu_A[X], \mu_B[Y]) \quad (3)$$

2.5 DEFUZZIFICATION

Defuzzification is the process of converting values from fuzzy sets into a single value only [10]. Defuzzification uses the center average defuzzifier method to determine the output value (Z), which can be seen in equation 4.

$$Z = \frac{\sum_{i=1}^n \alpha_i z_i}{\sum_{i=1}^n \alpha_i} \quad (4)$$

2.6 ROOT MEAN SQUARE (RMSE)

Root Mean Square Error (RMSE) is a measurement method by measuring the difference in the value of a model's prediction against the actual value of a variable. If the RMSE value is getting smaller then the prediction of a model is more valid. The RMSE calculation can be seen in equation 5.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (A_i - P_i)^2}{n}} \quad (5)$$

Where A_i is the actual output, P_i is the predicted output, and n is the amount of data.

2.7 GENETIC ALGORITHM

Genetic algorithms are heuristic search techniques based on the evolutionary ideas of natural selection and genetics. The process of solving problems using genetic algorithms can be seen with the following stages [8]:

a. Chromosomal representation and early population generation

The representation of chromosomes is formed based on the number of membership function limits of each variable.

b. Reproduction

The stages of reproductive process are crossover and mutation.

- Crossover

The study used a one-cut-point crossover technique. This is done by randomly taking two parents from the population in exchange for the value of their chromosomes.

- Mutation

The study used a random mutation technique. This method works by increasing or subtracting the value of randomly selected genes with the smallest random number.

c. Evaluation

At this stage to calculate the fitness value of each chromosome. Fitness calculations are shown in equation 6.

$$fitness = \frac{1}{RMSE} \quad (6)$$

d. Selection

The selection used is elitism selection. Elitism selection is the combination of strings and offspring in one population, then sorting the largest to smallest fitness values. The individual with the greatest fitness value will be selected as the best individual.

2.8 METHODOLOGY

2.8.1 DATA COLLECTION

The data collected in this study was obtained from the dataset site UCI Machine Learning Repository: Heart Disease Data Set. Data in the form of risk factors for heart disease. This data set has 76 attributes.

2.8.2 PREPROCESS

Processing the data obtained before being used as test data in the next activity. At this stage, a selection of attributes will be used as input data. Some attributes in the dataset are null and for criteria used in fuzzy in the form of values that have a range, so based on expert conclusions and after the selection of attributes, 11 attributes are obtained as input data and 1 attribute of heart disease diagnosis results that will be used to measure the error rate of the system to be built. The 11 attributes of these risk factors are age, maximum heart rate, oldpeak, systolic, diastolic, cholesterol, blood sugar, number of cigarettes, length of smoking, physical activity, and duration of exercise.

2.8.3 BEST INDIVIDUAL SEARCH WITH FUZZY HIERARCHICAL MODEL AND GENETIC ALGORITHM

The test data generated at the pre-processing stage will be used for calculations with the Fuzzy Hierarchical Model to be described in Figure 2.

Dian Palupi Rini, Defrian Afandi, Desty Rodiah
Effect of Genetic Algorithm on Prediction of
Heart Disease Stadium using Fuzzy Hierarchical Model

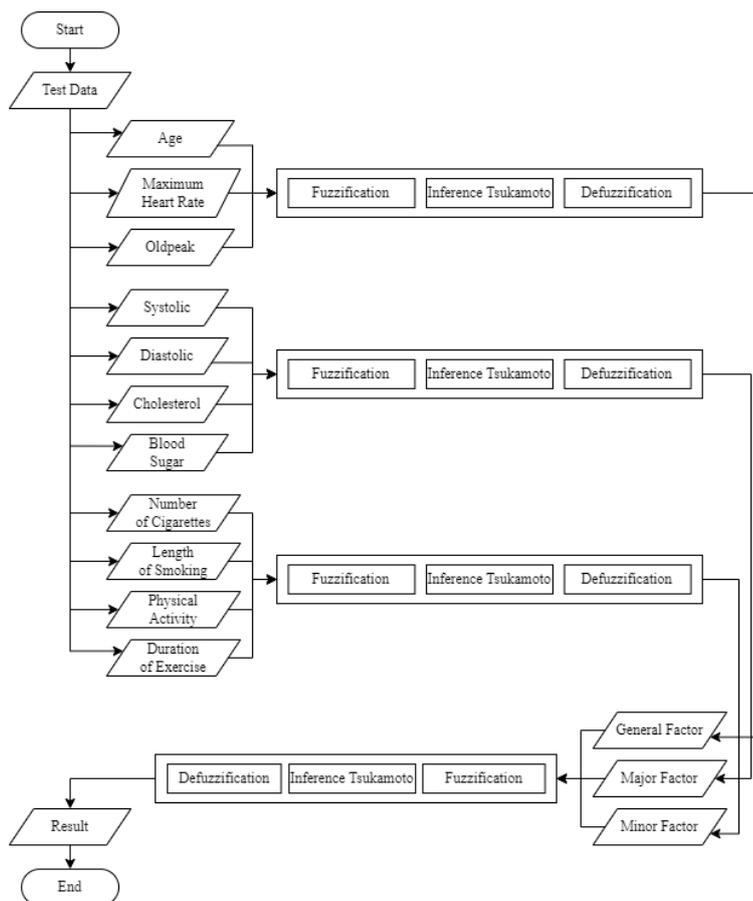


FIGURE 2. Calculation Process of Fuzzy Hierarchical Model

Based on the calculation process fuzzy hierarchical model in Figure 2 is described as follows:

- a. Defines the interval value limit of each set of variables to create an equation of the input and output variable membership function, thus generating the membership curve of each input and output variable. This limitation is also obtained based on chromosomes resulting from genetic algorithms.
- b. After that the fuzzification process, where the input value of each category of risk factors in the form of numerical data from the dataset is converted into fuzzy input using the membership function equation that has been created.
- c. The next stage of the inference process is to calculate values based on rules that have been set by experts.
- d. The final stage is the process of defuzzification using the center average defuzzyfier method that converts fuzzy values into numeric values again which then the numerical data will be categorized into stage 0 (healthy), 1, 2, 3, and 4 category data.

The fuzzy Hierarchical Method is used as an evaluation calculation on genetic algorithms. The best individuals are obtained in the process of Genetic Algorithms based on the highest fitness scores. Once acquired, this best individual will be used as a limitation of the new membership function in the Fuzzy Hierarchical Model method. The best individual search process will be described in Figure 3.

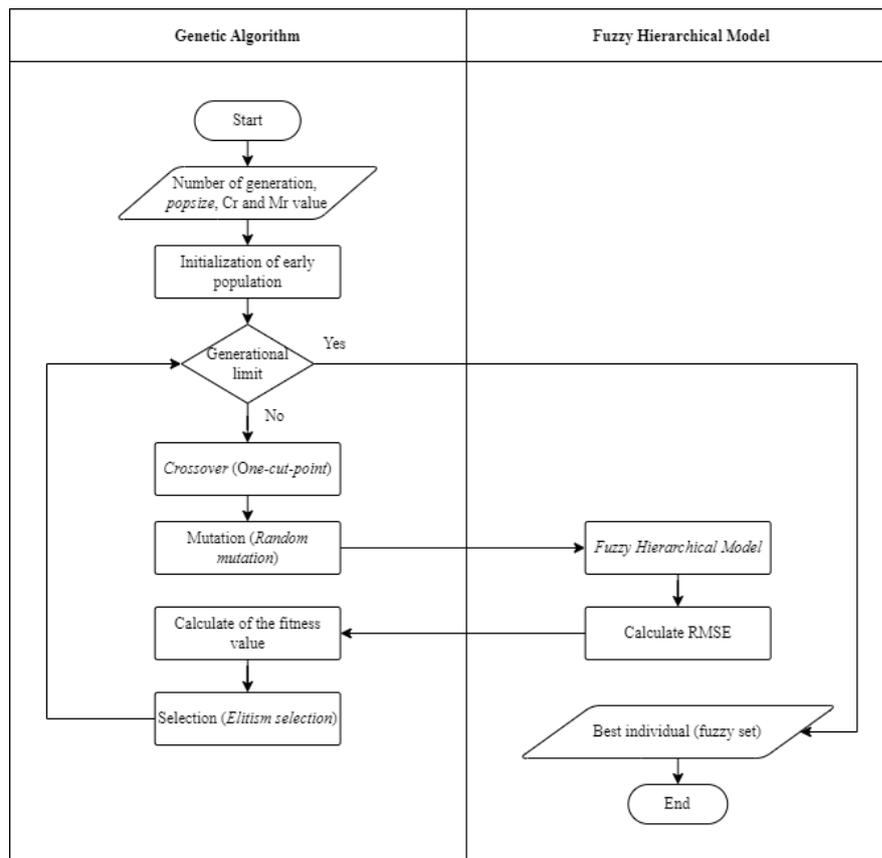


FIGURE 3. Best Individual Search with Fuzzy Hierarchical Model and Genetic Algorithm

Based on the best individual search with Genetic Algorithms and Fuzzy Hierarchical Models in Fig. 3 is described as follows:

- a. Determining the parameters of genetic algorithms are the number of generations, population size, crossover rate value, and mutation rate value.
- b. Generates the initial population randomly according to the specified population.
- c. Forming a new population with the following steps:
 1. Perform a crossover process on the selected parent to get offspring based on the random selection of existing chromosomes.
 2. Perform the mutation process in the randomly selected parent based on the existing chromosomes.
- d. Perform the Fuzzy hierarchical model process (described in Figure 2) until defuzzification results are found to calculate RMSE.
- e. Perform a calculation of the fitness value of each chromosome.
- f. Make selection after combining the parent and child individuals in determining the individual to continue the next process using the method of elitism selection.
- g. If it has not reached the generational limit, then using the new population for the next generation is obtained from the selection of individuals as much as the

initial population that has been determined and then performs the process of points c to f.

- h. If it has reached the generation limit then the best chromosome is obtained based on fitness values.

3. RESULT AND DISCUSSION

3.1 POPULATION SIZE TESTING

The test used a population size with multiples of 20 from 20 to 200. Other parameters specified are the value of $cr = 0.5$, $mr = 0.5$, and the number of generations = 100. Each condition of population size was carried out 5 times, then taken the average fitness value. The highest average fitness score will be taken as the best measure of population. Here are the results of population size tests shown in Figure 4.

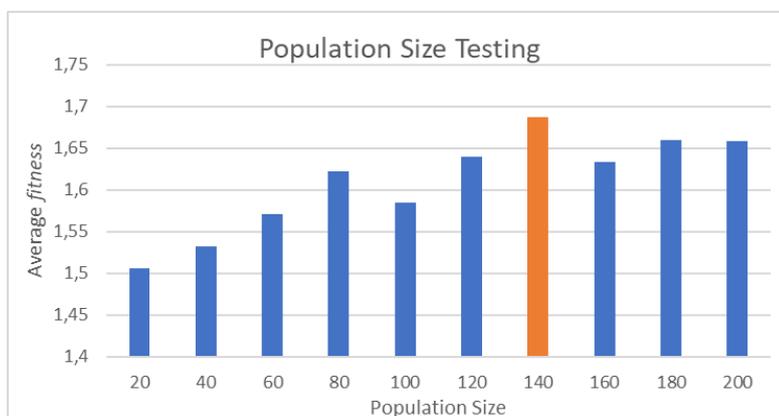


FIGURE 4. Population Size Testing Result

The best population size was obtained in the input of 140 population that produced the highest average fitness scores among other population sizes.

3.2 CR AND MR VALUE COMBINATION TESTING

The combination of cr and mr values used in this test is a multiple of 0.1 between 0 and 1 with cr (crossover rate) value from 0.1 to 0.9 and mr (mutation rate) value from 0.9 to 0.1. Other parameters set are the number of generations = 100 and population size based on the best population size obtained in the population size test. Each condition combination of cr and mr values is done 5 times the experiment, then take the average value of fitness. The highest average value will be taken as a measure of the best cr and mr values. Here are the test results of a combination of cr and mr values shown in Figure 5.

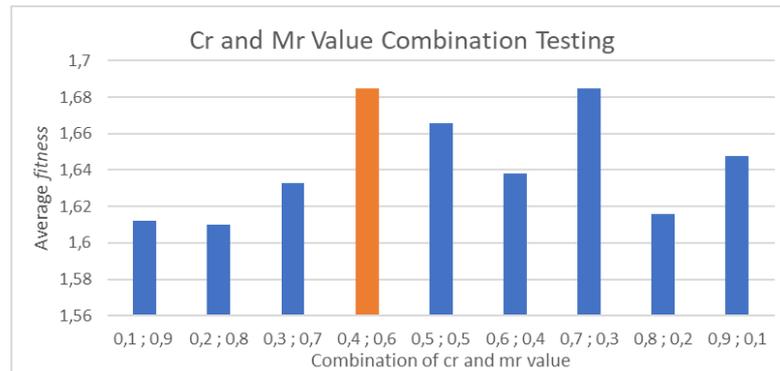


FIGURE 5. Cr and Mr Value Combination Testing Result

Obtained a combination of the best cr and mr values in input values cr = 0.4 and mr = 0.6 which produces the average value of the highest fitness value in the combination of cr and mr values.

3.3 GENERATION NUMBER TESTING

The number of generations used in this test is a multiple of 25 from 25 to 250. Other parameters set are population size = 140, value cr = 0.4, and mr = 0.6 based on best population size and a combination of the best cr and mr values in previous tests. Each condition inputs the number of generations carried out 5 times the experiment, then taken the average value of fitness. The highest average value will be taken as the best number of generations. Here are the results of the generational test shown in Figure 6.

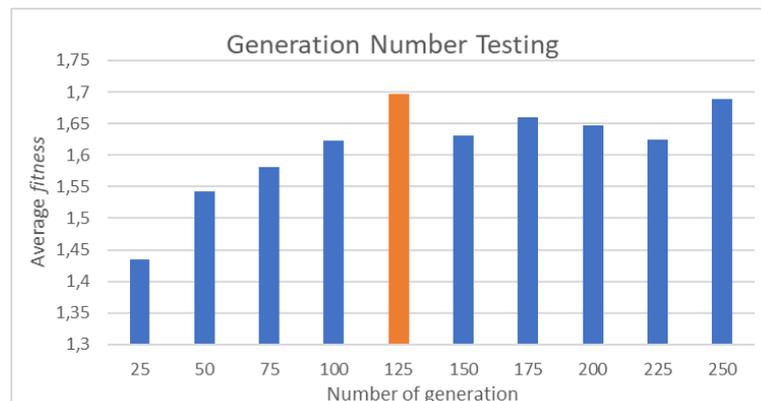


FIGURE 6. Generation Number Testing Result

The best number of generations at 125 entries resulted in the highest average fitness score among the number of other generations.

3.4 ANALYSIS OF RESEARCH RESULT

The tests that have been done in the previous explanation get the best Genetic Algorithm parameters with population size = 140, a combination of values cr = 0.4

Dian Palupi Rini, Defrian Afandi, Desty Rodiah
Effect of Genetic Algorithm on Prediction of
Heart Disease Stadium using Fuzzy Hierarchical Model

and $mr = 0.6$, and the number of generations = 125 that provide the best solution results. The best individual is taken from 5 attempts based on the parameters of the best genetic algorithm. The best individuals to use as limitations to the new membership function can be seen in Table 1.

TABLE 1.
Best Individual

Age		Maximum Heart Rate				Oldpeak					
35	42	57	60	126	171	172	187	0,7	2,1	2,4	3,2

Systolic				Diastolic				Cholesterol				Blood Sugar			
132	139	141	85	142	142	150	147	87	93	95	99	157	251	297	298

Number of Cigarettes				Length of Smoking				Physical Activity				Duration of Exercise			
9	22	22	35	4	6	8	8	1,3	4	6	6,8	16	32	42	43

This new membership function limitation will be used for heart disease prediction with the Fuzzy Hierarchical Model method and calculate the error value of the prediction results with RMSE measurements. The combination of the Fuzzy Hierarchical Model and Genetic Algorithm results in a lower RMSE than the Fuzzy Hierarchical Model which is not combined with the Genetic Algorithm. A comparison of the results of the tests conducted can be seen in Table 2.

TABLE 2.
Comparison of Test Result

Number	Id Patient	Age	Maximum Heart Rate	Oldpeak	Systolic	Diastolic	Cholesterol	Blood Sugar	Number of Cigarettes	Length of Smoking	Physical Activity	Duration of Exercise	Expert Diagnostic Result (Actual)	System Diagnostic Result (Predict)	Error (Actual-Predict)	Error Squared (Actual-Predict) ²
--------	------------	-----	--------------------	---------	----------	-----------	-------------	-------------	----------------------	-------------------	-------------------	----------------------	-----------------------------------	------------------------------------	------------------------	---

													FHM	FHM + GA	FHM	FHM + GA	FHM	FHM + GA	
1	1	63	150	2,3	145	85	233	100	50	20	13	10,5	1	1	1	0	0	0	0
2	2	67	108	1,5	160	90	286	115	40	40	13	9,5	2	1	1	1	1	1	1
3	3	67	129	2,6	120	80	229	100	20	35	10	8,5	1	1	1	0	0	0	0
4	4	37	187	2,5	130	78	250	98	0	0	17	13	0	0	0	0	0	0	0
5	6	41	172	1,4	130	86	204	106	0	0	9	7	0	0	0	0	0	0	0
6	7	56	178	0,8	120	75	236	99	20	20	16	11,3	0	1	1	1	1	1	1
7	8	62	160	3,6	140	100	268	220	0	0	7	6	2	1	1	1	1	1	1
...
282	298	47	179	0	130	80	253	132	0	0	11	9,5	0	0	0	0	0	0	0
Average																		0,79	0,30
RMSE																		0,89	0,55

4. CONCLUSION

Genetic algorithms can improve predictions of the Fuzzy Hierarchical Model method by updating the membership function limitations obtained based on the best individuals on the parameters of the Genetic Algorithm with a population size of 140, the number of generations 125, and the combination of crossover rate and mutation rate of 0.4 and 0.6. This increase in prediction is evidenced by the resulting error value (RMSE) lower than before combined with the Genetic Algorithm.

Test result on this study showed that the Fuzzy Hierarchical Model method optimized by the Genetic Algorithm resulted in an error value (RMSE) of 0.55 lower than the Fuzzy Hierarchical Model method without genetic algorithm optimization which resulted in an error value (RMSE) of 0.89.

REFERENCES

- [1] A. Adeli and M. Neshat, "A Fuzzy Expert System for Heart Disease Diagnosis," in *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS)*, 2010, I. 134–139.
- [2] D. Kurnianingtyas, W. F. Mahmudy, and A. W. Widodo, "Optimization Degree Membership from Fuzzy Tsukamoto Uses Genetic Algorithm for The Diagnosis of Beef Cow Disease," *Journal of Information Technology and Computer Science*, Vol. 4, No. 1, 8–18, 2017.

Dian Palupi Rini, Defrian Afandi, Desty Rodiah
Effect of Genetic Algorithm on Prediction of
Heart Disease Stadium using Fuzzy Hierarchical Model

- [3] D. Pramasari, "Classification of Stages of Heart Disease Using the Fuzzy Hierarchical Model Method," S.Kom. thesis, Sriwijaya University, Palembang, 2019.
- [4] A. Alharbi and F. Tchier, "A Fuzzy-Genetic Algorithm Method for the Breast Cancer Diagnosis Problem," in *The Ninth International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP)*, 2015, 122–127.
- [5] A. A. A. Esmin and G. Lambert-Torres, "Evolutionary Computation Based Fuzzy Membership Functions Optimization," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2007, 823–828.
- [6] J. Al Hafiz, "Prediction of Stage Of Heart Disease With Fuzzy Method Sugeno Model Hierarchy And Certainty Factor," S.Kom. thesis, Sriwijaya University, Palembang, 2020.
- [7] S. Fadilah, "Fuzzy Tsukamoto's Membership Degree Optimization Uses Genetic Algorithms For Weather Prediction," S.T. thesis, Sultan Syarif Kasim Islamic University, Riau, 2019.
- [8] N. Z. Ilahiyah, D. E. Ratnawati, and C. Dewi, "Optimization of Limitations of Fuzzy Tsukamoto Membership Function Uses Genetic Algorithms For CreditWorthiness (Case Study: PD. BPR. Lamongan Regional Bank)," *Journal of Information Technology Development and Computer Science*, Vol. 2, No. 9, 2957–2964, 2018.
- [9] M. K. Ardi, B. D. Setiawan, and M. A. Fauzi, "Tsukamoto's Fuzzy Inference System Optimization Uses Genetic Algorithms To Find Out The Length of Flush Time In Strawberry Plants," *Journal of Information Technology Development and Computer Science*, Vol. 2, No. 10, 3802–3810, 2018.
- [10] F. A. Fathurrachman, F. A. Bachtiar, and I. Cholissodin, "Optimization of Fuzzy Tsukamoto Membership Function with Genetic Algorithms on Gold Price Forecasting for Stock Trading," *Journal of Information Technology Development and Computer Science*, Vol. 3, No. 4, 3939–3948, 2019.