BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0571–0580
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0571-0580

FUZZY LOGISTIC REGRESSION APPLICATION ON PREDICTIONS CORONARY HEART DISEASE

Vera Febriani [1], Dian Lestari[2*], Sri Mardiyati[3], Oktavia Lilyasari[4]

[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia*
Pondok Cina, Beji, Depok, 16424, Indonesia
[4]*Department Cardiology and Vascular Medicine, Faculty of Medicine, Universitas Indonesia*
National Cardiovascular Center Harapan Kita
Jl. Jendral S. Parman Kav. 8, Slipi, Jakarta, 11420, Indonesia

*Corresponding author's e-mail: ** **dian.lestari@sci.ui.ac.id**

Article History:

Received: 10th November 2022
Revised: 11th February 2023
Accepted: 19th February 2023

Keywords:

Least Square;
Fuzzy Logic;
Mean Degree of Membership;
Logistic Regression

ABSTRACT

According to the World Health Organization (WHO) in 2015, 70% of cardiac deaths were caused by coronary heart disease (CHD). Based on WHO data in 2017, 17.5 million deaths were recorded, equivalent to 30% of the total deaths in the world caused by coronary heart disease. Coronary heart disease is a disorder of heart function caused by plaque that accumulates in arterial blood vessels so that it interferes with the supply of oxygen to the heart tissue. This causes reduced blood flow to the heart muscle and oxygen deficiency occurs. In more serious circumstances, it can result in a heart attack. Risk factors for coronary heart disease include age, gender, hypertension, cholesterol, heredity, diabetes mellitus, obesity, dyslipidemia, smoking and lack of physical activity. If a person's chances of suffering from coronary heart disease can be predicted early based on the existing risk factors, then the mortality rate of coronary heart disease can be suppressed. The objective of this study is to build a model that can predict the possibility of a patient suffering from coronary heart disease. The study used the Fuzzy Logistic Regression model. This model was used to maximize prediction results in which data size was limited. The least square method was used to estimate the value of the parameter. We obtained data from National Cardiovascular Center Harapan Kita, Jakarta. Evaluation with the Mean Degree of Membership method showed that the model built was feasible and good enough to predict coronary heart disease. By using the confusion matrix, the accuracy of the prediction model is 80.00%, with a specificity of 42.85% and a sensitivity of 100%.

How to cite this article:

V. Febriani, D. Lestari, S Mardiyati and O. Lilyasari., "FUZZY LOGISTIC REGRESSION APPLICATION ON PREDICTIONS CORONARY HEART DISEASE," *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0571-0580, March 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Acces

1. INTRODUCTION

One's probability to develop coronary heart disease are influenced by several risk factors. There are two types of risk factors, namely uncontrollable risk factors such as genetics, age, gender, and controllable risk factors such as smoking, diet, exercise, cholesterol, and blood pressure[1][2]. Many studies have been conducted before to predict the incidence of coronary heart disease through the existing risk factors in patients. Sellappan Palaniappan and Rafiah Awang in 2008 utilized data mining techniques such as Decision Tree, Naïve Bayes, and Neural Network to detect heart disease. Of the 909-patient data and fifteen risk factors observed, an Intelligent Heart Disease Prediction System (IHDPS) model was built to detect heart disease. This model can assist doctors in making decisions to perform actions on patients [3]. In 2014, Banu and Gomathy grouped the data using the K-means Clustering algorithm and implemented the Maximal Frequent Itemset Algorithm (MAFIA). This study succeeded in building a model that can predict the risk of heart attack in patients [4]. Febriani et.al in 2021 predicted readmissions after CABG operations using logistic regression by paying attention to eighteen risk factors. Of the eighteen risk factors obtained, only four were the most significant [5]. Haleh Ayatollahi in 2019 observed 1324 patient data coming from AJA University of medical science. A total of twenty-five risk factors were observed using two methods, namely the Support Vector Machine (SVM) and the Neural Network Algorithm [6]. Other studies related to coronary heart disease have also been conducted [7] - [14].

With time, science has experienced a very rapid development. Predicting a person's chances of suffering from a particular disease is not only done by the experts in the medical field, but nowadays researchers in the field of science can also conduct empirical data analysis. In this study, observational data on the risk factors of heart disease were divided into two types: numerical and categorical data. Numerical data includes weight, age, blood pressure, cholesterol levels, the number of heart rates per unit time. Meanwhile, categorical data includes gender, history of diabetes mellitus, smoking or not. The risk factor was a predictor variable (input) that would be used to predict the response variable (output), which was the chance of a person suffering from coronary heart disease. The data used in the research was retrieved from National Cardiovascular Center Harapan Kita, Jakarta. Fuzzy Logistic Regression Model was utitlized to predict the possibility of a patient suffering from coronary heart disease. We used least square method, Mean Degree of Membership method, and Confusion Matrix to estimate the value of the parameter, to test model feasibility, to predict the accuracy, respectively.

Logistic regression is a statistical modeling method used to describe the functional relationship between predictor variables (inputs) and response variables (outputs). In logistic regression, predictor variables (inputs) can be categorical, numerical or both. Each input value has a major influence in determining the predicted result. Therefore, the Logistic Regression model is very suitable to analyze and predict existing problems. Borowiak and Saphiro in 2014 wrote in their book titled "Financial and Actuarial Statistics: An Introduction" that one of the limitations of regression logistics models is when faced with small data, this causes the accuracy of predictions to decrease [15]. In Indonesia, the database storage system in hospitals is not yet matched the developed countries, thus data collection for this research is conducted manually that leads to limitation in obtaining the data. To overcome these limitations, this study will combine statistical methods and reasoning techniques or soft computing. Soft computing is a model approach using artificial intelligence. One example of a soft computing model is Fuzzy Logic. Fuzzy's theory of logic was first introduced in 1965 by a professor from the University of California, Prof. Lotfi A. Zadeh. Fuzzy logic is the logic that states estimate not certainties, much like human thinking. With Fuzzy logic, it allows tolerance in the grouping of values to produce a more rational output. Therefore, the output obtained by the software measurement becomes more accurate [16]. The previous researchers who used the fuzzy model includes Rosma Moh Dom et.al in 2012. In their study, with data from research center of the University of Malaya, oral cancer was predicted based on eight risk factors. The results showed that this model can work well in predicting oral cancer [17]. Furthermore, in 2011, Saadeh Pourahmad et.al predicted the possibility of suffering from diabetes by applying a fuzzy logistic regression model. The results of the evaluation of the model using the mean degree of membership and mean square error showed satisfactory results [18]. In 2018, Nimet Yapici Pehlivan and Aynur Sahin introduced an integrated fuzzy logistic regression approach to describe the relationship between crisp inputs and fuzzy binary output [19]. In addition, there are other studies that combine regression analysis and fuzzy logic [20],[21]. With this background, this research has combined the logistic regression model and the fuzzy logic model, hereinafter referred to as the fuzzy logistic regression model to predict the possibility of a patient suffering from coronary heart disease.

2. RESEARCH METHODS

2.1 Logistic Regression Model

Logistic regression is a regression model used to analyze the relationship between predictor variables (annotated with x) and response variables (annotated with y), where the response variable has only two possibilities, namely success and failure expressed by 0 (failed) and 1 (success). The following is a general equation of logistic regression [22].

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \dots, m \tag{1}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{2}$$

With:

$\boldsymbol{Y} = (Y_1, \quad Y_2, \quad \dots \quad Y_m)^T$ be vector $m \times 1$ of response variable.

$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mk} \end{pmatrix}$ be $m \times k$ matrix of m subjects of k predictor variables.

$\boldsymbol{\beta} = (\beta_1, \quad \beta_2, \quad \dots \quad \beta_k)^T$ be $k \times 1$ vector of the regression coefficient.

$\boldsymbol{\varepsilon} = (\varepsilon_1, \quad \varepsilon_2, \quad \cdots \quad \varepsilon_m)^T$ be $m \times 1$ vector showing the error value.

Assuming $E(\varepsilon_i)$=0. So: $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ (3)

Suppose a binary response variable is defined as follows:

$$Y_i = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases}$$

Y_i assumed to be distributed Bernoulli. Y_i has the following probability distribution:

Tabel 1. Probability of response variables

Y_i	Probability
1	$P(y_i = 1) = \pi_i$
0	$P(y_i = 0) = 1 - \pi_i$

From Table 1 the chance of success is stated with one and the chance of failure is expressed with zero. The probability value is π_i to succeed, and $1 - \pi_i$ to failed. Assuming $E(\varepsilon_i) = 0$ the expectation value for the response variable is obtained, namely: $E(Y_i) = (1 - \pi_i)0 + \pi_i.1 = \pi_i$

So that:

$$E(Y_i) = E(\boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i) = E(\boldsymbol{x}_i'\boldsymbol{\beta}) + E(\varepsilon_i) = \boldsymbol{x}_i' \tag{4}$$

Because Y is a binary response variable that is valued at 0 and 1 hence the value $E(Y_i)$ has limitations in the form of:

$$0 \le E(Y_i) = \pi_i \le 1 \tag{5}$$

In logistic regression, the relationship between predictor variables and response variables is non-linear. To make it linear can be conducted transformations on variables, so that the relationship between the predictor variable and the response variable becomes linear. The logit transformation for the need for linearity is defined as follows:

$$\boldsymbol{x}_i'\boldsymbol{\beta} = \ln\frac{\pi}{1-\pi} \tag{6}$$

So that relationship between π_i and $\boldsymbol{x}_i'$ defined as:

$$\pi_i = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} \tag{7}$$

2.2 Fuzzy Logistic Regression Model

Suppose a dataset $X_i = (X_{i1}, X_{i2}, \dots, X_{ij}, \dots, X_{ik}), i = 1, 2, \dots, m$ where X_{ij} is a predictor variable j for subject i, and m represents the number of subjects. Fuzzy Logistic Regression is defined as [23]:

$$\widetilde{W}_i = \tilde{A}_0 + \tilde{A}_1 X_{i1} + \cdots + \tilde{A}_n X_{ik}, i = 1, \dots, m. \tag{8}$$

With $\widetilde{W}_i$ is a response variable and $\tilde{A}_0, \tilde{A}_1, \dots \tilde{A}_k$ is a research parameter that in this case is referred to as the fuzzy coefficient.

In fuzzy logistic regression models, the function used as an approach is a triangular function. So, the fuzzy coefficient will also be expressed in triangular form:

$$A_j = (a_j^c, s_j^L, s_j^R)_T \ , \ j = 0, 1, 2, \dots, k \tag{9}$$

With its membership function:

$$\mu_{A_j}(\alpha_j) = 1 - \frac{|a_j - \alpha_j|}{s_j}, a_j - s_j \le \alpha_j \le a_j + s_j, \forall j = 0, 1, 2, \dots, k. \tag{10}$$

And with:

a = central point can be written as a_j^c
s = scatter point which consists of a left spread s_j^L and right spread s_j^R
j = expressing the predictor variable j
α = input variables

So that the fuzzy coefficient can also be written into:

$$A_j = (a_j, s_j)_T, \ j = 0,1,2, \dots, k \tag{11}$$

Since the fuzzy coefficient is triangular, the fuzzy output $\widetilde{W}_i, i = 1, \dots, m$ later it will also be triangular. So that $\widetilde{W}_i = (f_{ia}^c(x), f_{is}^L(x), f_{is}^R(x))_T$, where:

$$f_{ia}^c(x) = a_0^c + a_1^c x_{i1} + \cdots + a_n^c x_{ik}, \tag{12}$$
$$f_{is}^L(x) = s_0^L + s_1^L x_{i1} + \cdots + s_n^L x_{ik}, \tag{13}$$
$$f_{is}^R(x) = s_0^R + s_1^R x_{i1} + \cdots + s_n^R x_{ik}, \tag{14}$$

As in the Logistic Regression model written in Equation (6), in the fuzzy logistic regression model $\widetilde{W}_i$ also defined logit transformations namely:

$$\widetilde{W}_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right) \tag{15}$$

One of the fundamental differences of logistic regression models and fuzzy logistic regression is in the use of the term probability. The logistic regression model expresses the chance of success with probability. Whereas in the fuzzy logistic regression model, the term used is possibilistic or the probability of success which is annotated as μ_i . Value $\frac{\mu_i}{1-\mu_i}$ referred to as possibilistic odds.

Based on **Equation (9)**, **Equation (10)** and **Equation (15)** then the membership function for $\widetilde{W}_i = (f_{ia}^c(x), f_{is}^L(x), f_{is}^R(x))_T$ be:

$$\exp\left(\widetilde{W}_i(x)\right) = \left(\frac{\mu_i}{1-\mu_i}\right)(x) = \begin{cases} 1 - \frac{f_{ia}^c(x) - \ln(x)}{f_{is}^L(x)}, f_{ia}^c(x) - f_{is}^L(x) \le \ln(x) \le f_{ia}^c(x), \\ 1 - \frac{\ln(x) - f_{ia}^c(x)}{f_{is}^R(x)}, f_{ia}^c(x) < \ln(x) \le f_{ia}^c(x) + f_{is}^R(x). \end{cases} \quad (16)$$

$$\mu_i(x) = \widetilde{W}_i\left(\ln\frac{x}{1-x}\right) = \begin{cases} 1 - \frac{f_{ia}^c(x) - \ln\frac{x}{1-x}}{f_{is}^L(x)}, f_{ia}^c(x) - f_{is}^L(x) \le \ln\frac{x}{1-x} \le f_{ia}^c(x), \\ 1 - \frac{\ln\frac{x}{1-x} - f_{ia}^c(x)}{f_{is}^R(x)}, f_{ia}^c(x) < \ln\frac{x}{1-x} \le f_{ia}^c(x) + f_{is}^R(x). \end{cases} \quad (17)$$

The model to be built is obtained by substituting **Equation (11)** to **Equation (8)** so that:

$$\widetilde{W}_i = (\alpha_0, s_0)_T + (\alpha_1, s_1)_T x_{i1} + (\alpha_2, s_2)_T x_{i2} + \cdots + (\alpha_k, s_k)_T x_{ik} \quad (18)$$

Where:

$\widetilde{W}_i = (W_{1i}, W_{2i})^T$ is a vector 2×1 of subject i .

$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mk} \end{pmatrix}$ be $m \times k$ matrix of k predictor variables of m subjects.

$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_k)^T$ be $k \times 1$ vector of the regression coefficient with $\beta_j = (\alpha_j, s_j)$.

2.3 Least Square Method

One method that can be used to estimate the value of a parameter is the Least Square method. The Least Square method is a parameter estimation method that works by minimizing the sum of the residue (the difference between the actual value and its estimated value) squared [24], [25]. The Least Square method was first introduced by Legendre in 1805 [26]. The functions to be minimized in the least square method are:

$$SSE = \sum_{i=1}^{m} \left(d\left(\widetilde{w}_i, \widetilde{W}_i\right)\right)^2 \quad (19)$$

With $\widetilde{w}_i$ states the estimation model and $\widetilde{W}_i$ states the observation model.

2.4 Mean Degree of Membership

Suppose a fuzzy regression logistic model comes from m observation. MDM (Mean Degree of Membership) is the average value of its membership degree [25]. The range of Mean Degree of Membership values is at intervals of 0 and 1. A Mean Degree of Membership greater than 0.5 indicates that the model built is feasible or appropriate [20]. The hypotheses used for the conformity test of this model are:

H_0: fuzzy logistic regression models good enough to predict coronary heart disease.

H_1: fuzzy logistic regression models are not good enough to predict coronary heart disease.

Test statistics:

$$MDM = \frac{1}{m}\sum_{i=1}^{m} \widetilde{W}_i(\omega_i) = \frac{1}{m}\sum_{i=1}^{m} \exp\left(\widetilde{W}_i\left(\left(\frac{\mu_i}{1-\mu_i}\right)\right)\right) \quad (20)$$

Based on decision making:

H_0 accepted if the Mean Degree of Membership > 0.5

H_0 rejected if the Mean Degree of Membership < 0.5

2.5 Confusion Matrix

Confusion matrix is one of the methods that can be used to calculate accuracy. Confusion matrix states the comparison of the number of correctly predicted test data and the number of incorrectly predicted test data [27]. From the confusion matrix, the accuracy, specificity, and sensitivity of predictions can be calculated.

3. RESULTS AND DISCUSSION

In this study, the data used was obtained from National Cardiovascular Center Harapan Kita Jakarta in 2019. A total of twenty data samples consisting of ten research variables, namely age, gender, BMI (Body Mass Index), sodium, TDS (Systolic Blood Pressure), TDD (Diastolic Blood Pressure), BPM (Beats Per Minutes), hemoglobin, potassium, and blood glucose were retrieved from the medical records.

Coding research variables are presented in Table 2. The patient observation data at NCCHK is shown in Table 3.

Table 2. Coding Research Variables

No	Variable	Coding	Category
1	Gender	0	Woman
		1	Man
2	CHD	0	Non-CHD
		1	CHD

Table 3. Patient Observation Data at The National Cardiovascular Center Harapan Kita Jakarta

No	Gender	Age	BMI	Na	TDS	TDD	Cp	Hb	K	GDS	CHD
1	1	60	27,39	142	154	82	75	12,6	4,55	135	**1**
2	0	66	20,2	136	179	94	67	10,9	3,5	213	0
3	1	50	27,55	142	146	81	59	11,1	4	96	**1**
4	1	58	23,59	128	187	103	70	12,2	3,5	119	**1**
5	1	63	24,98	110	124	68	84	10,2	4,6	234	**1**
6	1	63	23,81	141	175	97	62	11,9	3,7	193	0
7	1	72	22,77	135	184	88	72	14,8	2,7	426	**1**
8	1	60	28,91	137	148	83	130	15	4,8	403	**1**
9	1	68	24,61	140	148	81	60	13,7	3,9	164	0
10	0	67	23,05	141	99	58	56	8,4	3,4	144	0
11	1	62	24,84	138	123	64	65	12,3	3,5	197	**1**
12	1	55	28,67	141	137	77	83	11,2	4,1	135	**1**
13	0	62	29,78	145	166	85	83	8,4	4,4	202	0
14	0	70	20	141	101	75	94	16,2	2,7	226	0
15	1	53	29,76	121	104	55	96	6,8	4,6	126	**1**
16	1	66	23,88	140	109	48	84	7,4	4,05	277	**1**
17	1	56	30,8	134	141	81	67	14,3	4,5	191	**1**
18	1	69	24,44	132	128	68	68	12,7	5,28	153	**1**
19	1	63	23,26	141	141	78	65	13,8	4,5	102	**1**
20	1	64	32,83	137	231	104	84	10,2	2,5	251	0

3.1 Building the Model

Based on Equation (18) X_1 = Gender, X_2 = Age, X_3 = BMI, X_4 = Na, X_5 = TDS, X_6 = TDD, X_7 = BPM, X_8 = Hb, X_9 = K, X_{10} = GDS then the model built is:

$$\widetilde{W}_i = \tilde{A}_0 + \tilde{A}_1 Gender_i + \tilde{A}_2 Age_i + \tilde{A}_3 BMI_i + \tilde{A}_4 Na_i + \tilde{A}_5 TDS_i + \tilde{A}_6 TDD_i + \tilde{A}_7 BPM_i + \tilde{A}_8 Hb_i + \tilde{A}_9 K_i + \tilde{A}_{10} GDS_i, \text{dengan } i = 1, \dots, 20. \quad (21)$$

3.2 Parameter Estimation a and s using the Least Square method

Using the help of R software, the parameter is obtained $\boldsymbol{a}$ and $\boldsymbol{s}$:

$$\boldsymbol{a} = (8.99 \quad 0{,}1476 \quad -0.087744 \quad -0.08315 \quad -0.016 \quad 0.019165 \quad -0.05205 \quad 0.00198 \quad 0.1356584 \quad 0.178836 \quad 0.00105\ )^T$$

$$\boldsymbol{s} = (8.99 \quad 0{,}1476 \quad -0.09292 \quad -0.08315 \quad -0.016 \quad 0.019165 \quad -0.05205 \quad 0.00198 \quad 0.1356584 \quad 0.178836 \quad 0.00105\ )^T$$

3.3 Predicting Coronary Heart Disease

From Equation (21) and by substituting the value a and s obtained from the parameter estimation process, the general model for the prediction of coronary heart disease in patients at National Cardiovascular Center Harapan Kita Jakarta is:

$$\begin{aligned}\widetilde{W}_i = {} & (8.99\,,8.99) + (0.1476{,}0.1476)_T \times Gender_i + (-0.087744, -0.09292)_T \times Age_i \\ & + (-0.08315, -0.08315)_T \times BMI_i + (-0.016, -0.016)_T \times Na_i \\ & + (0.019165, 0.019165)_T \times TDS_i + (-0.05205, -0.05205)_T \times TDD_i \\ & + (0.00198{,}0.00198)_T \times BPM_i + (0.1356584, 0.1356584)_T \times Hb_i \\ & + (0.178836{,}0.178836)_T \times K_i + (0.00105{,}0.00105)_T \times GDS_i\ , \text{for } i = 1, \dots, 20.\end{aligned} \quad (22)$$

Equation (17) can determine possibilistic values for each patient. As an example, the following data described the calculation process for patient number 8.

Table 4. Patient Observation Data number 8

No	x_{i1} Gender	x_{i2} Age	x_{i3} BMI	x_{i4} Na	x_{i5} TDS	x_{i6} TDD	x_{i7} BPM	x_{i8} HB	x_{i10} K	x_{i11} GDS	Y_i PJK
8	1	60	28.91	137	148	83	130	15	4,8	403	**1**

$$\begin{aligned}\widetilde{W}_8 = {} & (8.99\,,8.99) + (0.1476{,}0.1476)_T \times 1 + (-0.087744, -0.09292)_T \times 60 \\ & + (-0.08315, -0.08315)_T \times 28.91 + (-0.016, -0.016)_T \times 137 \\ & + (0.019165, 0.019165)_T \times 148 + (-0.05205, -0.05205)_T \times 83 \\ & + (0.00198{,}0.00198)_T \times 130 + (0.1356584, 0.1356584)_T \times 15 \\ & + (0.178836{,}0.178836)_T \times 4.8 + (0.00105{,}0.00105)_T \times 403 \\ & = (1.366912{,}1.0563635)_T\end{aligned}$$

Substitution of patient prediction model eight to Equation (16) and Equation (17). So that:

$$\mu_8(x) = \widetilde{W}_8\left(\ln\frac{x}{1-x}\right) = \begin{cases} 1 - \dfrac{1.366912 - \ln\frac{x}{1-x}}{1.0563635}, 0.577019 \le x \le \mathbf{0.796880} \\ 1 - \dfrac{\ln\frac{x}{1-x} - 1.366912}{1.0563635}, \mathbf{0.796880} < x \le 0.918585 \end{cases} \quad (23)$$

Hence, the possibilistic value that represents the magnitude of the possibility of patient number 8 with 60 years of age, pre-obsesity according to the BMI category, has a history of hypertension and diabetes mellitus, and above normal pulses to suffer from coronary heart disease is 0.796880. The probability value obtained from the calculation is at intervals of 0 and 1. By using the midpoint at 0.5, for the possibility value less than of 0.5, it will be categorized as a non-CHD patient and for the possibility value more than 0.5, it will be categorized as a CHD patient.

From the calculation results, the MDM value for the built model is 0.65776 > 0.50 then H_0 Accepted. Therefore, fuzzy logistic regression models are good enough to predict coronary heart disease. By identifying the prediction results using the confusion matrix, we obtained 80.00% accuracy, 42.85% specificity and 100% sensitivity. These numbers indicate that the model built is good enough.

4. CONCLUSIONS

In this study, a fuzzy logistic regression model has been applied to predict the possibility of a person suffering from coronary heart disease. This combination of the regression logistic model and the fuzzy logic model maximize prediction results that have limitations in data size. The evaluation using the Mean Degree of Membership method show that the model built is feasible and good enough to predict coronary heart disease. By using the confusion matrix, we found that the model built is good enough with the accuracy of the prediction model is 80.00%, a specificity of 42.85% and a sensitivity of 100%.

AKNOWLEDGEMENT

This work is fully funded by research grant from Universitas Indonesia.

REFERENCES

[1] Northside Hospital, "3 Risk factors for heart disease you can't control-and 5 you can". Available from https://northsidecvi.com/article-3-risk-factors-for-heart-disease-you-cant-control-and-5-you-can, [Accessed: June 10, 2022].

[2] World Heart Federation, *Cardiovascular Risk Factors*. Available from https://world-heart-federation.org/world-heart-day/cvd-causes-conditions/risk-factors/, January 2022.

[3] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *AICCSA 08 - 6th IEEE/ACS Int. Conf. Comput. Syst. Appl.*, pp. 108–115, 22 April 2008, doi: 10.1109/AICCSA.2008.4493524.

[4] M. A. Nishara Banu and B. Gomathy, "Disease forecasting system using data mining methods," *Proceedings - 2014 International Conference on Intelligent Computing Applications*, ICICA 2014, pp. 130–133. https://doi.org/10.1109/ICICA.2014.36, 2014.

[5] V. Febriani, D. Lestari, S. Mardiyati, and S. Devila, "Predicting readmission risk after coronary artery bypass graft surgery using logistic regression model," *in J. Phys. Conf. Ser.*, vol. 1725, no. 1, 2021, doi: 10.1088/1742-6596/1725/1/012083.

[6] H. Ayatollahi, L. Gholamhosseini, and M. Salehi, "Predicting coronary artery disease: A comparison between two data mining algorithms", *BMC Public Health*, 19:448, April 2019, doi: 10.1186/s12889-019-6721-5.

[7] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad, and G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," *Proceedings of ICDLT 2019: 2019 3rd International Conference on Deep Learning Technologies*, 5 July 2019, doi: 10.1145/3342999.3343015.

[8] P. I. Aaronson, and J. P. T. Ward, *At a Glance: Sistem Kardiovaskular*, 3th ed, Jakarta: Penerbit Erlangga, 2010.

[9] T. R. Barbeau, *Cardiovascular physiology*. Florence: Francis Marion University, 2004.

[10] S. Agarwal, C. Schechter, A. Zaman, "Assessment of functional status and quality of life after percutaneous coronary revascularization in octogenarians," *Age and Ageing*, Vol. 38, Issue 6, pp. 748-751, November 2009. https://doi.org/10.1093/ageing/afp174.

[11] H. H. Gray, K. D. Dawkins, J. M. Morgan, and I. A. Simpson, *Lecture Notes: Kardiologi*, 4th ed., Jakarta: Erlangga, 2005.

[12] R. Shoufiah. "Hubungan Faktor Resiko dan Karakteristik Penderita dengan Kejadian Penyakit Jantung Koroner," *Mahakam Nursing Journal*, Vol 1, No. 1, pp. 17-26, Mei 2016.

[13] I. Soeharto, *Pencegahan Dan Penyembuhan Penyakit Jantung Koroner*, Jakarta: Gramedia Pustaka Utama, 2001.

[14] A. F. Yahya, *Menaklukkan Pembunuh no.1: Mencegah dan Mengatasi Penyakit Jantung Koroner Secara Tepat*, Bandung: PT Mizan Pustaka, 2010.

[15] D. S. Borowiak, and A. F. Saphiro, *Financial and Actuarial Statistics: An Introduction*, 2nd ed. New York: Chapman and Hall/CRC, 2020.

[16] L. A. Zadeh, Fuzzy sets. *Information and control*. Vol. 8, pp. 338-353, 1965.

[17] R. M. Dom, B. Abidin, S. A. Kareem, S. M. Ismail, and N. M. Daud, "Determining the critical success factors of oral cancer susceptibility prediction in Malaysia using fuzzy models," *Sains Malaysiana*, vol. 41, no. 5, pp. 633–640, 2012.

[18] S. Pourahmad, S. M. T. Ayatollahi, and S. M. Taheri, "Fuzzy logistic regression: A new possibilistic model and its application in clinical vague status," Iran. *J. Fuzzy Syst.*, vol. 8, no. 1, pp. 1–17, 2011.

[19] N. Y. Pehlivan and A. Sahin, "An Integrated approach for fuzzy logistic regression," *İstatistikçiler Dergisiİstatistik ve Aktüerya*, vol. 11, no. 1, pp. 42–54, 2018.

[20] S. Pourahmad, S. M. T. Ayatollahi, S. M. Taheri, and Z. H. Agahi, “Fuzzy logistic regression based on the least squares approach with application in clinical studies,” *Comput. Math. with Appl*., vol. 62, no. 9, pp. 3353–3365, 2011, doi: 10.1016/j.camwa.2011.08.050.

[21] J. Mohammadi and S. M. Taheri, “Pedomodels fitting with fuzzy least squares regression,” *Iran. J. Fuzzy Syst*., vol. 1, no. 2, pp. 45–61, 2004.

[22] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, second ed, New York: John Wiley and Sons, 2000.

[23] H. Tanaka, S. Uejima, and K. Asai, “Linear Regression Analysis with Fuzzy Model”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol: 12, Issue: 6, November 1982, page: 903-907. http://dx.doi.org/10.1109/TSMC.1982.4308925.

[24] D. C. Montgomery, *Introduction to Linear Regression Analysis*, 5th ed., USA: John Wiley & sons, Inc, 2012.

[25] W. Mendenhall, and T. Sincich, *A Second Course in Statistics: Regression Analysis*, 7th ed., Boston: Pearson Education Limited, 2012.

[26] B. S. Gildeh, and D. Gien, A goodness of fit index to reliability analysis in fuzzy model. Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation: 78-83, 2002.

[27] B. Ratner, *Statistical and machine-learning data mining*, 3rd ed., Boca Raton: CRC Press, 2017.