ISSN 1 746-7233, England, UK
World Journal of Modelling and Simulation
Vol. 13 (2017) No. 2, pp. 83-93

A fuzzy linear regression model for identifying risk factors of coronary heart disease*

Lazim Abdullah[1 †]

School of Informatics and Applied Mathematics Universiti Malaysia Terengganu

(*Received July 30 2015, Accepted February 11 2017*)

Abstract. Coronary heart disease is a major cause of morbidity and mortality in the modern society. Many risk factors for coronary heart disease have been discussed and identified by medical fraternity. However, the magnitudes of risk factors, particularly in predicting the disease are remained unknown and inconclusive. The purpose of this study was to develop fuzzy regression prediction model and to investigate the contribution of risk factors that affect coronary heart disease based on the knowledge of flexibility of individual's risk habits and risk factors. Four risk factors, namely body mass index, cholesterol reading, systolic blood pressure and serum fasting glucose level were identified and investigated using the matrix-driven multivariate fuzzy linear regression model. One hundred and thirty patients' data supplied by a government funded university hospital in Peninsular Malaysia were tested using the matrix-based fuzzy regression model. Apart from coefficients of the regression equation, variations in risk of coronary heart disease caused by the risk factors were investigated using coefficient of determination with flexibility of α- cuts of the model. Finally, the performance of the regression model was measured using the area under receiver operating characteristic curve analysis and a comparative analysis. The analyses indicate that body mass index was the dominant and the strongest predictor to the likelihood of developing coronary heart disease. The multiple risk factors did contribute to the coronary heart disease, but the model was not sensitive to the flexibility of α-cuts. The areas under receiver operating characteristic curves for all predictors were greater than 0.7 indicate that the fuzzy regression model was reliable. This investigation is not only affirms the numerous adverse effects of the risk factors to coronary heart disease but also provides evidence on the feasibility of using the flexibility of fuzzy numbers for obtaining fuzzy linear regression model.

Keywords: chronic heart disease, risk factors, linear regression, fuzzy numbers, alpha-cuts

1 Introduction

Coronary heart disease (CHD) is one of the leading causes of death in many countries of the world including Malaysia. The disease has become a major cause of morbidity and mortality in the modern society. According to the National Heart Association of Malaysia[22], CHD is on the rise in Malaysia despite improvement in health services and facilities. CHD was the second leading cause of death in 2006, accounting for 15.5 per cent of those who died in government hospitals. By 2010, the disease is projected to be the leading cause of death in Malaysia and other developing countries. CHD is caused by a narrowing of the coronary arteries, which results in a decreased supply of blood and oxygen to the heart. There are many risk factors associated with CHD. Some of the risk factors associated with CHD are heredity, being male, advancing age, cigarette smoking, high blood pressure, diabetes, obesity, lack of physical activity, and abnormal blood cholesterol levels. Recently, Menotti et al.[21] conclude that behavioral factors, including cigarette smoking, physical activity and diet are strong predictors of lifetime incidence of common heart diseases. Atwa et al.[3] evaluate the effect

* The author wishes to acknowledge the research student, Ho Boo Yeng for assisting in data collection.
† Corresponding author. *E-mail address*: lazim_m@umt.edu.my.

of sex on the outcome of congenital heart diseases in children and suggest that females are more vulnerable to hospitalization due to chest infection than males in children. There are two types of risk factors of CHD which are risk factors cannot control such as age, gender and family history, while the risk factors can control such as diabetes, high blood pressure and overweight[34]. Despite these full set of risk factors of CHD, the most influential factor is still inconclusive. Some research suggests that obesity is the main factors due to its numerous adverse effects on general, and especially, cardiovascular health[16]. Obesity has been implicated as one of the major risk factors for hypertension, heart failure, and coronary CHD[17, 30]. It is observed that all these factors are inter-correlated thereby the most influential risk factors are difficult to . Therefore the actual risk factors remain to be unspoken and inconclusive. As a result of these unanswerable situations, many researches with various approaches have been embarked to search clue for dominant risk factors of CHD.

One of the popular approaches in analyzing relationships between a response variable and explanatory variables is multiple linear regressions. The multiple linear regression approaches have been tested successfully in many medical sciences. The relationship between cardiovascular disease-predisposing genes and blood pressure, for example, was analyzed by Tang, et al.[32] with variance analysis and multiple linear regression analysis. Malave et al.[18] examine heart rate variability in relation to circulating levels of TNF, TNF receptors, and norepinephrine in patients with heart failure and in control subjects using multiple linear regression analysis. Hanley, et al.[13] assess the associations between proinsulin and cardiovascular risk factors using correlation and multiple linear regression analyses. Koji et al.[15] investigate the influence of metabolic syndrome on the relationship between arterial stiffness and the risk of coronary artery disease using multiple linear regression analysis. Carels[8] examine the association between disease severity, functional status, depression and daily quality of life in congestive heart failure patients using multiple linear regression. Multiple linear regression analysis was used to evaluate the associations between symptoms (i.e., anxiety, depression, pain) and disease-specific and generic QOL scores controlling for demo- graphic (i.e., age, comorbidity) and clinical (i.e., physical function) characteristics[7]. Recently, McNabney et al.[20] suggest that in best fit multiple regression models, congestive heart failure and diabetes mellitus were not significant predictors of the evaluated care utilization measures. Despite Bellazzi and Zupan[6] reported that the relationships between the predictor and the response variables are normally governed by complex mathematical equations, multiple linear regression has been continuously used by researchers due to its simplicity with the help of statistical software.

2 Motivation

The linear model, multiple linear regression has, to date, been applied with great success in many medical sciences research. However, in many applications including medical sciences, modelling the relationship between multiple explanatory input variables and the corresponding response variable poses a problem due to the intricate and vague associations between the predictor and response variables. The existing statistical regression techniques proved to have several limitations when dealing with an inadequate number of observations and also when distributional assumptions are not met[9, 10, 29]. The problems of vagueness in variables, small number of observations and failure to fulfill statistical assumptions, most studies related to medicine such as cancer and CHD were used artificial intelligent techniques such fuzzy regression model, fuzzy neural network model, coactive neuro-fuzzy inference system[24, 31]. The intelligent models have shown their potentials to overcome the limitation of linear model and produce a new generation of modeling which can handle high levels of uncertainties. Noor et al.[28], for example, developed a fuzzy decision support system based on neural network for the diagnosis of coronary artery disease based on evidence. Shantakumar and Kumaraswamy[25] carried out research to predict heart attack using data mining and artificial neural network. Neuro-fuzzy modeling and artificial neural networks as predictive models were also investigated by Abbod et al.[1] to predict cancer. Ritchie et al.[26] developed the optimization of neural network architecture using genetic programming to improve detection and modeling of 30 interactions genes in the studies of human diseases.

Apart from neural works, fuzzy regression models have been shown its ability in describing the relationships between medical variables. Toyoura et al.[33], for example, specifically used fuzzy linear regression method in determining oral age in relation to the number of sound teeth. Fuzzy regression systems were also used by Massad[19] in an epidemiology study of HIV infected individuals to investigate the relationship between viral load and clinical progression to AIDS among HIV positives. Fuzzy linear regression has been used as one of the feasible intelligent approaches to describe the relationship between input variables and the response variable specifically for medical fields. It seems that fuzzy regression analysis is suitable to solve the problems of human aspect that rely on subjective judgment. For the vast majority of subjective judgment, there is a need to contend with a certain level of uncertainties. Rather than the direct application of typical crisp numbers in multiple linear regressions, this paper proposes a new approach in searching the relationships between risk factors of CHD to improve the interpretation of regression coefficients by employing fuzzy numbers to a fuzzy linear regression model. The present study uses a matrix-driven multiple fuzzy linear regression model after taking into account the ambiguous relationship between the explanatory and response variables and exhibits the relationship in term of a possibility interval. The main question that sparked this research was whether there exist fuzzy regression model that can be helpful to predict the likelihood of an individual developing CHD on knowledge of their risk habits or risk factors. This research is aimed at evaluating the ability of a fuzzy regression model to investigate the risk factors that affect the CHD based on the knowledge of the individual's risk habits or risk factors. Specifically, this paper proposes a fuzzy linear regression model which can describe the relationship between risk factors and health condition of patients. Contributions of risk factors toward overall health condition were investigated using coefficient of determination of the model under different values of α-cuts. Finally, the performance of the model was also verified using receiver operating curve analysis and a comparative analysis. In other words, a matrix-driven fuzzy linear regression approach was used to observe the contributions of risk factors toward the health condition of the CHD patients, strength of relationship among the variables and performances of the linear model.

3 Preliminaries

In this section, a review of the basic definition of fuzzy sets, triangular fuzzy numbers and α-cuts are presented prior to modeling CHD risk factors.

Definition 1. *Fuzzy Sets*[35]

Let a set X be non- empty and finite. A fuzzy set A on X is an expression given by

$$A = \{< x, \mu_A(x), \gamma_A(x) > | x \in X\} \tag{1}$$

where: $\mu_A X \to [0,1]$ is the membership function of the fuzzy set A; $\mu_A X \in [0,1]$ is the membership of $x \in X$ in A.

Definition 2. *Triangular Fuzzy Number*[14]*. Let $\tilde{m} = (1, m, r)$ be a triangular fuzzy number, where the memberships function $\mu_{\tilde{m}}$ of $\tilde{m}$ is given by .*

$$\mu_m = \begin{cases} \frac{x-1}{m-1} & (1 \leq x \leq m) \\ \frac{r-x}{r-m} & (m \leq x \leq r) \end{cases} \tag{2}$$

It is easy that a triangular fuzzy number $\tilde{m} = (l, m, r)$ is reduced to a real number m can be written as a triangular fuzzy number $\tilde{m} = (m, m, m)$.

Definition 3. *Alpha-Cuts*[5]*. The α-cut, $\alpha \in [0,1]$, of a fuzzy number A is crisp set defined as*

$$A_\alpha = \{x \in R : \mu_A(x) \geq \alpha\} \tag{3}$$

Every α-cut of a fuzzy number A is a closed interval $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, where

$$A_L(\alpha) = inf\{x \in R : \mu_A(x) \geq \alpha\} \tag{4}$$
$$A_U(\alpha) = sup\{x \in R : \mu_A(x) \geq \alpha\}. \tag{5}$$

These three definitions are fully employed in implementing the matrix-driven multivariate fuzzy linear regression model.

4 Proposed method

Based on Pan et al.[23] method, nine steps of computational procedures are proposed.

Step 1. Identify the predictor matrix X.

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \tag{6}$$

Step 2. Identify the response matrix Y.

$$\tilde{Y} = \begin{bmatrix} (y_1, (1-\alpha)e_{1,L}, (1-\alpha)e_{1,R}) \\ (y_2, (1-\alpha)e_{2,L}, (1-\alpha)e_{2,R}) \\ \vdots \\ (y_n, (1-\alpha)e_{n,L}, (1-\alpha)e_{n,R}) \end{bmatrix} \tag{7}$$

where

$$(y_i, (1-\alpha)e_{i,L}, (1-\alpha)e_{i,R}) \tag{8}$$

is a symmetrical triangular fuzzy number.

Step 3. Compute the multiplication of XY_i where Y_i is the middle value of triangular fuzzy number.

Step 4. Compute the multiplication of X with left width of fuzzy numbers

$$(1-\alpha)e_{i,L}.$$

Step 5. Compute the multiplication of X with right width of fuzzy numbers

$$(1-\alpha)e_{i,R}.$$

Step 6. Compute the central value of fuzzy regression coefficients,

$$\tilde{\beta} = (X'X) - 1.X'Yi$$

Step 7. Compute the left value of fuzzy regression coefficients,

$$\tilde{\beta} = (X'X) - 1.X'$$

Step 8. Compute the right value of fuzzy regression coefficients

$$\tilde{\beta} = (X'X) - 1.X'$$

Step 9. Structure the estimated fuzzy linear regression equation.

$$\hat{Y}_i = (a_0, c_{0,L}, c_{0,R}) + (a_1, c_{1,L}, c_{1,R})X_1 + (a_2, c_{2,L}, c_{2,R})X_2 + \cdots + (a_k, c_{k,L}, c_{k,R})X_k$$

with k crisp independent variables and one fuzzy dependent variable. $(a_0, c_{0,L}, c_{0,R})$ is the fuzzy intercept coefficient; $(a_1, c_{1,L}, c_{1,R})$ is the fuzzy slope coefficient for X_1; $(a_2, c_{2,L}, c_{2,R})$ is the fuzzy slope coefficient for X_2; and $(a_k, c_{k,L}, c_{k,R})$ is the k^{th} fuzzy slope coefficient.

Fuzzy linear regression equation and coefficient of determination are the main modeling measures that would be obtained from the CHD patients.

5 Computations

This research intends to investigate the effect of four risk factors to health conditions of CHD patients. Data from a sample of hundred and thirty patients with cardiovascular disease from government funded university hospital were collected. This number of respondents are more than sufficient for intelligent models as it does not depend on linearity assumptions and hypothesis testing. This number of patients are consistent with several similar studies in the literature (see Ebadian et al.[4], Frossyniotis, et al.[11], Guler and Ubeyli[12], and Scott et al.[27]). Descriptions of the patients are summarized in Tab. 1.

Table 1: Demographic characteristics of sample

Characteristics of patients	Number of patients (%)
Gender	
Male	83 (63.8)
Female	47 (36.2)
Age (year)	
< 39	9 (6.9)
40-59	48 (36.9)
$60 >$	73 (56.2)

Four predictors of the model were identified as body mass index ($X1$) in unit of weight(kg)/[height(m)], cholesterol reading ($X2$) in unit of mg/d, systolic blood pressure ($X3$) in unit of mg/dl and serum fasting glucose level ($X4$) in unit of mg/dl.

Table 2: Mean and standard deviation of predictor

Predictors	Mean	Standard deviations
Body mass index (X1)kg/m	31.15	3.35
Cholesterol reading (X2) mg/d	210.11	5.81
Systolic blood pressure (X3) mg/dl	125.32	4.18
Serum fasting glucose level (X4) mg/dl.	147.18	3.92

Means and standard deviations of all predictors are presented in Tab. 2.

Step 1. Identify the matrix X

Data of predictors were arranged in the matrix X. Part of the matrix is given as

$$\begin{pmatrix} 1 & 36.51 & 209 & 137 & 114 \\ 1 & 31.22 & 175 & 155 & 109 \\ 1 & 26.55 & 228 & 130 & 153 \\ 1 & 26.91 & 194 & 121 & 92 \\ 1 & 23.34 & 156 & 114 & 94 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 24.26 & 190 & 126 & 99 \end{pmatrix}_{130\times 5} \tag{9}$$

Step 2. Identify the matrix Y.

The response variable refers to health condition among the patients. Triangular fuzzy numbers were introduced to define the level of healthy. Five fuzzy numbers were defined to reflect the linguistics of ‘Health’, ‘Mild’, ‘Moderate’, ‘Serious’, ‘Very Serious’. Tab. 3 shows the defined fuzzy triangular numbers for the linguistic healthy levels.

Table 3: Healthy levels and fuzzy numbers

Healthy level	Triangular Fuzzy numbers
Health	(0,0, 20)
Mild	(10,30,50)
Moderate	(40, 55,70)
Serious	(60,75,90)
Very Serious	(80, 100,100)

Based on the defined fuzzy numbers in Tab. 3, the healthy levels of the patients were arranged in matrix Y. Part of Y are given as

$$Y = \begin{pmatrix} 55 & 15 & 15 \\ 80 & 20 & 10 \\ 100 & 20 & 0 \\ 0 & 0 & 20 \\ 0 & 0 & 20 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 0 & 0 & 20 \\ 0 & 0 & 20 \end{pmatrix}_{130\times 3} \tag{10}$$

Step 3. Compute the multiplication of XYi

$$XYi = \begin{pmatrix} 1110 \\ 33051.7 \\ 237735 \\ 150775 \\ 139330 \end{pmatrix} \tag{11}$$

Step 4. Compute the multiplication of X with left width of fuzzy numbers eL

$$XeL = \begin{pmatrix} 350 \\ 10261.4 \\ 75515 \\ 47335 \\ 41340 \end{pmatrix} \tag{12}$$

Step 5. Compute the multiplication of X with right width of fuzzy numbers XeR

$$XeR = \begin{pmatrix} 480 \\ 13250.6 \\ 96835 \\ 60415 \\ 50780 \end{pmatrix} \tag{13}$$

Step 6. Compute the central value of fuzzy regression coefficients,

$$\tilde{\beta} = (X'X) - 1.X'Y_i = \begin{pmatrix} -166.821 \\ 1.654462 \\ 0.313215 \\ 0.412975 \\ 0.362876 \end{pmatrix} \tag{14}$$

Step 7. Compute the left value of fuzzy regression coefficients,

$$\tilde{\beta} = (X'X) - 1.X'eL \begin{pmatrix} -47.471 \\ 0.158911 \\ 0.134435 \\ 0.200122 \\ 0.015065 \end{pmatrix} \tag{15}$$

Step 8. Compute the right value of fuzzy regression coefficients

$$\tilde{\beta} = (X'X) - 1.X'eR = \begin{pmatrix} 38.81683 \\ -0.18479 \\ -0.01613 \\ -0.02308 \\ -0.10127 \end{pmatrix} \tag{16}$$

Step 9. Structure the estimated fuzzy linear regression equation. $\tilde{Y} = (-166.821, -47.471, 38.817) + (1.654, 0.159, -0.185)X1 + (0.313, 0.134, -0.016)X2 + (0.413, 0.200, -0.023)X3 + (0.363, 0.015, -0.101)X4$ Interpretation and discussion for the risk factors can be made from this equation.

6 Result and discussion

The result and discussion section is divided into three subsections. The subsection 6.1 discusses the fuzzy coefficient regressions and its interpretation to the risk factors. The next subsection discusses the sensitivity of α-cuts toward the fuzzy coefficient determinations. The subsection 6.3 examines the performance of the estimated model using Receiver Operating Characteristic. A comparative analysis with the conventional linear regression is made in the subsection 6.4.

6.1 Fuzzy coefficient regression and risk factors

Accordingly, the cause-and-effect relationship between the risk factors and overall coronary heart condition can be examined. It can be discovered in the above equation that the slope $X1$ is (1.654, 0.159, -0.185), indicating that regardless of cholesterol reading ($X2$), systolic blood pressure ($X3$), serum fasting glucose level ($X4$), an estimated increase of 1.654 in the condition index of the CHD (Y), an increase of 0.159 in the lower bound (left interval) and decrease -0.185 in the upper bound (right interval) occurs for each additional body mass index ($X1$). Likewise, the estimated slope coefficient of $X2$, (0.313, 0.134, -0.016), signifies that regardless of $X1$, $X3$, $X4$ and for each additional $X2$, $\tilde{Y}$ increase by 0.313, with an increase by 0.134 and decrease by -0.016 in the lower bound and upper bound, respectively. The estimated slope coefficient of X3 which is systolic blood pressure, is (0.413, 0.200, -0.023), denotes that regardless of $X1$, $X2$, $X4$ for each additional $X3$, an estimated increase of 0.413 in $\tilde{Y}$, with an increase by 0.200 in the lower bound and decrease of -0.023 in the upper bound. Similarly, the estimated slope coefficient of $X4$, (0.363, 0.015, -0.101), denotes that regardless of $X1$, $X2$, $X3$ and for each additional $X4$, an estimated increase of 0.363 in $\tilde{Y}$, with an increase of 0.015 in the lower bound and decrease of -0.101 a in the upper bound.

The fuzzy regression model indicates that body mass index, cholesterol reading, systolic blood pressure and serum fasting glucose level have positively contributed to risk of CHD. Body mass index records the highest regression coefficient among the four. The result shows that body mass index is the most influential variable contributes to CHD. The results are echoed in the research of Basir et al.[2] where thirteen risk factors were used as input variables, including body mass index, systolic blood pressure, diastolic blood pressure, blood sugar level at fasting, total cholesterol level, triglycerides level, HDL cholesterol, LDL cholesterol, smoking habit, pulse pressure, ratio of total cholesterol to HDL cholesterol, sex and age. The prediction output was the presence of any form of coronary CHD. The logistic regression analysis had identified four risk factors, namely body mass index, systolic blood pressure, total cholesterol level and age have greater impact on the outcome.

6.2 Sensitivity of α-cuts toward the model

One of the important features of fuzzy numbers is α-cuts. The flexibility of $\alpha \in [0,1]$ brings a further investigation to observe behaviour of the coefficient of determination. Using the equation

$$(HR)^2 = \frac{\sum_{i=1}^{n}(a_0+a_1X_i-\bar{Y})^2+(1-\alpha)\sum_{i=1}^{n}(c_{0,L}+c_{1,L}X_i-\bar{e}_L)^2+(1-\alpha)\sum_{i=1}^{n}(c_{0,R}+c_{1,R}X_i-\bar{e}_R)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2+(1-\alpha)\sum_{i=1}^{n}(e_{i,L}-\bar{e}_L)^2+(1-\alpha)\sum_{i=1}^{n}(e_{i,R}-\bar{e}_R)^2} \tag{17}$$

when $\alpha = 0$, the fuzzy coefficient of determination $(HR)^2$ is computed as

$$\frac{15729.739 + 1179.874 + 405.091}{34480 + 2466.667 + 970} = 0.457. \tag{18}$$

The above result shows that 45.7% of the total variation in Y can be explained by the regression line relating to the input. The fuzzy correlation coefficient thus yields 0.676 $(= \sqrt{0.457})$, signifying a positive linear relationship between CHD and the predictor variables. Therefore, these four variables provide reliable predictors for CHD.

With the similar fashion, the value of HR when $\alpha = 0.1$, 0.5, 0.8 and 1 can be obtained. The HR values yield were 0.676, 0.676, 0.675, 0.675, 0.675 when $\alpha = 0$, 0.1, 0.5, 0.8, 1, respectively. The variations of α-cuts did not give much impact on the fuzzy correlation coefficients. Therefore, it can be concluded that the estimated equation does not sensitive with the flexibility of α-cuts.

6.3 Worthiness of the model

Performance of the estimated equation was further explored using Receiver Operating Characteristic (ROC) curve analysis. ROC is defined as a plot of test sensitivity versus 1-specificity. It is an effective method of evaluating the quality or performance of diagnostic test. ROC curve is useful for assessing the accuracy of predictions. Peter[1] used the area under the ROC curve to assess the predictive accuracy of the resultant model. Therefore ROC is used to evaluate the performance of the fuzzy regression model for the prediction of risk factors of CHD. ROC plot that reflects the area under curve of the four predictors is shown in Fig. 1.

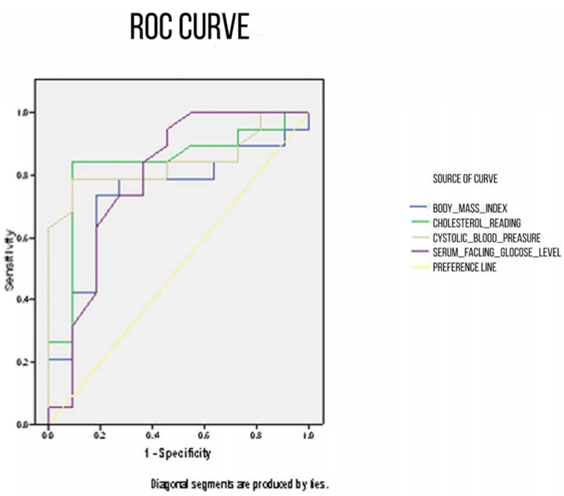

Fig. 1: ROC Plot

Area under the ROC curve represents accuracy. An area of 1 represents a perfect test while 0.5 represents a worthless test. A guide for classifying the accuracy of a diagnostic test can be retrieved from Wen et al.[36]. The ROC curve for the CHD regression equation can be translated into area and significant level. The results are shown in Tab. 4.

Table 4: Area under the ROC curve

Test Result Variable(s)	Area	Standard Error	Asymptotic Sig.
Body Mass Index	0.737	0.96	0.033
Cholesterol Reading	0.835	0.82	0.003
Systolic Blood Pressure	0.842	0.73	0.002
Serum Fasting Glucose Level	0.789	0.98	0.009

The areas under curve for all explanatory variables vary between 0.737-0.842 indicating good accuracy of the diagnostic test. Furthermore, the variables are significant at rejecting the null hypothesis when the true value 0.5 under non parametric assumption. Therefore the fuzzy regression model is robust in predicting the risk factors for CHD.

6.4 A comparative analysis

Theoretically, when the value of α-cut is one, the fuzzy regression coefficient has zero width of the fuzzy number (see Definition 3). Consequently, this particular fuzzy regression coefficient is no longer a fuzzy number and the fuzzy coefficient is now equivalent to the multiple linear regression coefficients. Therefore, the equation of multiple linear regressions can be written as, $\tilde{Y} = -166.821 + (1.654)X1 + (0.313)X2 + (0.413)X3 + (0.363)X4$. Accordingly, the cause-and-effect relationship between the risk factors and overall heart condition can be examined. It can be noticed that all central values of fuzzy coefficients are exactly the same as the multiple linear regression coefficients. The interpretation out of this multiple linear regression is similar to the fuzzy regression equation except for the level of fuzziness for each respective risk factor. The risk factors of body mass index, cholesterol reading, systolic blood pressure and serum fasting glucose level have positively contributed to heart disease since those variables have positive regression coefficients. Body mass index has been identified as the highest risk factors. This comparative analysis is indeed substantiated the results of sensitivity of α-cuts where the values of α-cuts did not sensitive to the model.

7 Conclusions

One of the main concerns in explaining the associations between risk factors of coronary heart disease and its impact to patients is by choosing the appropriate linear model which can reduce computation risk, produce better performance and giving flexible interpretations. This paper has proposed an accurate and reliable fuzzy linear regression model that can be used for the prediction of response variable when the predictors are governed by vague, ambiguous relationship and small sample size. The proposed model was to establish fitted fuzzy regression equation for estimating the cause-and-effect relationship between variables. The matrix-driven fuzzy linear regression model has been used in predicting risk factors of coronary heart disease. The inter correlated variables body mass index, cholesterol reading, systolic blood pressure and serum fasting glucose level were selected as predictors. Of the four variables, body mass index has recorded the best predictor for coronary heart disease. This investigation was further continued by determining the coefficient of determination of the model under different values of α-cuts. Despite the moderately strong positive linear correlation of the estimated fuzzy regression model, the coefficient of determinations did not sensitive to the changes of α-cuts. The performance of the model was complemented using the analysis of area under the receiver operating characteristic curve. All the predictors were giving a good accuracy of the model. The fuzzy regression equation, sensitivity of α-cuts and area under receiver operating characteristic curve were synergistically contributed in explaining the magnitude of the risk factors for deadly coronary heart disease. However, this research is far from completed. Further works need to be extended by considering more variables and patients with specific diseases. It needs to be aware that huge computations risks would anticipate if larger sample and variables were investigated.

References

[1] M. F. ABBOD, J. W. F. CATTO, et al. Artificial intelligence for the prediction of bladder cancer. *Biomedical Engineering Applications Basis & Communications*, 2004, **16**(02): 49–58.

[2] B. Abidin, R. M. Dom, et al. Use of fuzzy neural network to predict coronary heart disease in a malaysian sample. **in:** *Proceedings of the 8th Wseas international conference on Telecommunications and informatics*, 2009.

[3] Z. T. Atwa, H. H. Safar. Outcome of congenital heart diseases in egyptian children: Is there gender disparity? *Gazette of the Egyptian Paediatric Association*, 2014, **62**(2): 35–40.

[4] H. Bagher-Ebadian, H. Soltanian-Zadeh, et al. Neural network and fuzzy clustering approach for automatic diagnosis of coronary artery disease in nuclear medicine. *IEEE Transactions on Nuclear Science*, 2004, **51**(1): 184–192.

[5] A. Ban. Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval. *Fuzzy Sets & Systems*, 2008, **159**(11): 1327–1344.

[6] R. Bellazzi, B. Zupan. Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics*, 2008, **77**(2): 81–97.

[7] S. B. Bentsen, C. Miaskowski, T. Rustoen. Demographic and clinical characteristics associated with quality of life in patients with chronic obstructive pulmonary disease. *Quality of Life Research*, 2014, **23**(3): 991–998.

[8] R. A. Carels. The association between disease severity, functional status, depression and daily quality of life in congestive heart failure patients. *Quality of Life Research*, 2004, **13**(1): 63–72.

[9] J. W. Catto, M. F. Abbod, et al. Neuro-fuzzy modeling: an accurate and interpretable method for predicting bladder cancer progression. *Journal of Urology*, 2006, **175**(2): 474–479.

[10] B. Eftekhar, K. Mohammad, et al. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Medical Informatics and Decision Making*, 2005, **5**(1): 3.

[11] D. S. Frossyniotis, D. Vrettos, et al. An intelligent system for the early diagnosis of coronary artery disease. **in:** *4th Internaional Conference in Neural Networks and Expert Systems in Medicine and Healthcare (NNESMED)*, 2001.

[12] I. Guler, E. D. Ubeyli. Automated diagnostic systems with diverse and composite features for doppler ultrasound signals. *IEEE Transactions on Biomedical Engineering*, 2006, **53**(10): 1934–1942.

[13] A. J. Hanley, G. Mckeown-Eyssen, et al. Cross-sectional and prospecvtive associations between proinsulin and cardiovascular disease risk factors in a population experiencing rapid cultural transition. *Diabetic Care*, 1240-1247, **24**: 1240–1247.

[14] M. Hanss. *Applied Fuzzy Arithmetic: An Introduction with Engineering Applications*. Springer, 2004.

[15] Y. Koji, H. Tomiyama, et al. Relationship between arterial stiffness and the risk of coronary artery disease in subjects with and without metabolic syndrome. *Hypertension Research Official Journal of the Japanese Society of Hypertension*, 2007, **30**(3): 243–7.

[16] C. J. Lavie, R. V. Milani. Obesity and cardiovascular disease: the hippocrates paradox? *Journal of the American College of Cardiology*, 2003, **42**(4): 677–679.

[17] C. J. Lavie, R. V. Milani, H. O. Ventura. Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss. *Journal of the American College of Cardiology*, 2009, **53**(21): 1925.

[18] H. A. Malave, A. A. Taylor, et al. Circulating levels of tumor necrosis factor correlate with indexes of depressed heart rate variability: a study in patients with mild-to-moderate heart failure. *Chest*, 2003, **123**(3): 716.

[19] E. Massad, N. R. S. Ortega, et al. Fuzzy epidemics. *Artificial Intelligence in Medicine*, 2003, **29**(3): 241–259.

[20] M. K. Mcnabney, C. Onyike, et al. The impact of complex chronic diseases on care utilization among assisted living residents. *Geriatric Nursing*, 2013, **35**(1): 26–30.

[21] A. Menotti, P. E. Puddu, G. Maiani, G. Catasta. Lifestyle behaviour and lifetime incidence of heart diseases. *International Journal of Cardiology*, 2015, **201**(6): 293–299.

[22] N. H. A. of Malaysia. Heart diseases on the rise, second leading killer. http://www.malaysianheart.org/article.php?aid=35, 2008.

[23] N. F. Pan, T. C. Lin, N. H. Pan. Estimating bridge performance based on a matrix-driven fuzzy linear regression model. *Automation in Construction*, 2009, **18**(5): 578–586.

[24] L. Parthiban, R. Subramanian. Intelligent heart disease prediction system using canfis and genetic algorithm. *International Journal of Biological, Biomedical and Medical Sciences*, 2008, **3**(3).

[25] S. B. Patil, Y. Kumaraswamy. Intelligent and effective heart attack prediction system using data mining and artificial neural network. *European Journal of Scientific Research*, 2009, **31**(4): 642–656.

[26] M. D. Ritchie, B. C. White, et al. Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC Bioinformatics*, 2003, **4**(1): 28.

[27] J. A. Scott, K. Aziz, et al. Integration of clinical and imaging data to predict the presence of coronary artery disease with the use of neural networks. *Coronary Artery Disease*, 2004, **15**(7): 427.

[28] N. A. Setiawan, P. A. Venkatachalam, et al. Diagnosis of coronary artery disease using artificial intelligence based decision support system. *Universiti Malaysia Perlis*, 2009, **9**(5): 93C118.

[29] A. F. Shapiro. Fuzzy regression models. *Article of Penn State University*, 2005, **102**(2): 373–383.

[30] C. D. Sjöström, M. Peltonen, H. Wedel, L. Sjöström. Differentiated long-term effects of intentional weight loss on diabetes and hypertension. *Hypertension*, 2000, **36**(1): 20.

[31] P. M. Speight, A. E. Elliott, et al. The use of artificial intelligence to identify people at risk of oral cancer and precancer. *Br Dent J.*, 1995, **179**(10): 382–387.

[32] M. Tang, Y. Dai, et al. The univariation and multiple linear regression analyses for seventeen snps in thirteen cardiovascular disease-predisposing genes and blood pressure in chinese han males. *Clinical and Experimental Hypertension*, 2008, **30**(7): 648–661.

[33] Y. Toyoura, M. Kambara, et al. Evaluation of fuzzy regression analysis and its application to oral age model. *Biomedical Fuzzy & Human Sciences the Official Journal of the Biomedical Fuzzy Systems Association*, 2000, **6**: 41–49.

[34] S. Yusuf, S. Hawken, S. Ounpuu, E. Al. Interheart study investigators. effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the interheart study): case-control study. lancet 2004; 364:. *Journal of Cardiopulmonary Rehabilitation & Prevention*, 2004, **25**(1): 56–57.

[35] L. A. Zadeh. Fuzzy sets. *Information & Control*, 1965, **8**(3): 338–353.

[36] W. Zhu, N. Zeng, N. Wang. Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. **in:** *Proceedings of the SAS Conference, Baltimore, Maryland*, 2010, 9–13.