STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, May 2026, pp 3512–3525.
Published online in International Academic Press (www.IAPress.org)

Improved Risk Modeling for Concurrent Diabetes and Hypertension: A Biresponse Nonparametric Logistic Regression Approach

Marisa Rifada [1, 2, *], Nur Chamidah [1, 2], Elly Ana [1, 2], Budi Lestari [2, 3], Dursun Aydin [2, 4], Naufal Ramadhan Al Akhwal Siregar [2, 5], Muhammad Fikry Al Farizi [2, 6]

[1] *Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia*
[2] *Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia*
[3] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Jember, Jember 68121, Indonesia*
[4] *Department of Statistics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla 48000, Turkiye*
[5] *Mathematics Master Study Program, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, 60115, Indonesia*
[6] *Statistics Undergraduate Study Program, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, 60115, Indonesia*

Abstract One of the key strategic priorities in Indonesia's health development, as outlined in the Sustainable Development Goals (SDGs) agenda, is to reduce premature mortality from Non-Communicable Diseases (NCDs) by one-third. Diabetes and hypertension are two closely related NCDs that often coexist. This study aims to develop a risk model for the simultaneous incidence of diabetes and hypertension using a biresponse approach. Data were collected from 211 patients at the Internal Medicine Polyclinic of Airlangga University Hospital Surabaya. A Chi-square dependency test revealed a significant association between the incidence of diabetes and hypertension. Additionally, the relationship between each predictor variable and the observed logit of diabetes and hypertension demonstrated a non-linear pattern, suggesting that the impact of predictor variables on the risk of both diseases is not linear. A comparison of the biresponse logistic regression model with both parametric and nonparametric approaches indicated that the Biresponse Nonparametric Logistic Regression model outperformed the parametric approach in terms of performance and stability. The model's accuracy improved substantially from 0.436 to 0.626, and the Area Under the Curve (AUC) increased from 0.62 to 0.83.

Keywords Bivariate, Diabetes, Hypertension, Logit, Nonparametric

AMS 2010 subject classifications 62G08, 62H30, 62P10

DOI: 10.19139/soic-2310-5070-3118

1. Introduction

Indonesia has experienced an increase in the prevalence of Non-Communicable Diseases (NCDs) in recent years. According to the 2018 Basic Health Research (Riskesdas), the prevalence of NCDs showed an increase compared to the 2013 Riskesdas, particularly for diseases such as cancer, stroke, chronic kidney disease, diabetes mellitus, and hypertension. The 2018 Riskesdas data indicates an increase in the prevalence of diabetes mellitus from 6.9% in 2013 to 8.5% in 2018, as well as an increase in hypertension cases from 25.8% to 34.1% [7]. NCD prevention and control programs have become a focus at both the international and national levels, in line with commitments

*Correspondence to: Marisa Rifada (Email: marisa.rifada@fst.unair.ac.id). Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga. Dr. Ir. H. Soekarno Road, Surabaya, Indonesia (60115).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

under the Sustainable Development Goals (SDGs). Goal number 3 of the 17 SDGs, which is good health and well-being, includes a target to reduce premature mortality from NCDs by one-third by 2030. Therefore, NCDs have become a priority in the development agenda of various countries [33].

Diabetes and hypertension are two types of NCDs that are closely related. These diseases often occur simultaneously, making them comorbidities (diseases experienced by the same patient) [9]. Hypertension patients have a higher risk of developing diabetes, and similarly, diabetes patients are at risk of developing hypertension. This close relationship is largely due to the interrelated physiological characteristics, where the effect of one disease tends to increase the likelihood of the other disease occurring. Modeling these comorbidities jointly is advantageous as it explicitly accounts for the correlation between responses, thereby improving the efficiency of parameter estimates and providing a more realistic representation of the patient's health status.

One statistical method that can be used for disease risk modeling is logistic regression. Logistic regression is a data analysis method used to describe the relationship between a categorical response variable and one or more predictor variables that can be either categorical or continuous. In its development, studies on logistic regression have been conducted for response variables consisting of two categories (binary logistic regression) or more than two categories and ordinal scale (ordinal logistic regression), both for cases involving a single response variable (uniresponse) or two response variables (biresponse) [31, 4, 1, 6, 8, 32, 30, 5, 3, 23, 26, 27, 24, 25, 29, 28, 21, 2, 10, 19, 16, 14, 12, 17]. There are two approaches to regression modeling: the parametric approach and the nonparametric approach. The parametric approach assumes that the regression model for each individual observation has the same parameters, while the nonparametric approach assumes that not all individuals have the same parameters [31].

Several studies on diabetes and hypertension risk modeling have been conducted using logistic regression, both with parametric and nonparametric approaches for response variables consisting of two categories (binary) [31, 4, 1, 6, 8, 32, 30, 5, 3, 23]. However, these studies have not modeled the risk of diabetes and hypertension incidence simultaneously. Given the close relationship between diabetes and hypertension, it would be more realistic to model them simultaneously. Therefore, this study aims to develop a risk model for the incidence of diabetes and hypertension simultaneously using the Biresponse Nonparametric Logistic Regression (BNLR) approach.

2. Materials and Methods

2.1. Data and Data Sources

The data used in this research is primary data collected using an accidental sampling method with the study period spanned from April 2023 until December 2023. The dietary variables were measured using Semi-quantitative Food Frequency Questionnaire *SQ—FFQ*. A statement regarding ethical approval from the *Komite Etik Penelitian Kesehatan* (Health Research Ethics Committee) Universitas Airlangga Hospital with Ethical Clearance Number: 050/KEP/2023. Data collection was carried out at the Internal Medicine Polyclinic of Airlangga University Hospital, Surabaya. The study population comprised patients who fulfilled all predefined inclusion criteria, i.e. being registered as patients at the Internal Medicine Outpatient Clinic of Universitas Airlangga Hospital, Surabaya, as evidenced by possession of a medical record number; being alive at the time of data collection; and providing cooperative consent to participate as research respondents. Based on patients who met all inclusion criteria, a total sample of 211 patients was obtained for this study. This study involves two response variables: the incidence of diabetes (Y_1) and the incidence of hypertension (Y_2), as well as five predictor variables: age (X_1), Body Mass Index (BMI) (X_2), salt intake (X_3), sugar intake (X_4), and fat intake (X_5).

2.2. Data Analysis Method

Biresponse logistic regression is an extension of the logistic regression framework designed to model two correlated binary outcomes simultaneously [20]. This approach is particularly useful when the response variables exhibit dependency, since estimating them jointly yields more efficient and accurate results compared to modeling

them separately. The model allows for the incorporation of shared predictors that may influence both responses, making it suitable for complex health data.

In cases where the linearity assumption between predictors and the logit function is not satisfied, the nonparametric extension provides a flexible solution. The Biresponse Nonparametric Logistic Regression (BNLR) model utilizes smoothing techniques to capture nonlinear patterns without requiring prior specification of the functional form [22]. This flexibility enhances predictive accuracy and stability, especially in clinical and epidemiological research settings.

The BNLR analysis was implemented using the VGAM package (version 1.1-13) in the R statistical computing environment. In this study, B-splines were employed as basis functions for the nonparametric components to capture nonlinear trends. The association between the two responses was modeled using the constant odds ratio structure specified by the *binom2.or* family.

3. Results and Discussion

The descriptive statistics for the incidence of diabetes and hypertension are presented in Table 1 as follows:

Table 1. Description of Diabetes and Hypertension Incidence

Response Variable	Category	Total	Percentage	Total
Diabetes Incidence	No Diabetes	140	66.35%	211
	Diabetes	71	33.65%	(100%)
Hypertension Incidence	No Hypertension	74	35.07%	211
	Hypertension	137	64.93%	(100%)

Based on the sample size of 211 patients, it can be observed that the number of patients with hypertension is higher than those with diabetes. A total of 71 patients (33.65%) experienced diabetes, while the percentage of patients without diabetes was higher, amounting to 140 patients (66.35%). On the other hand, 137 patients (64.93%) had hypertension, and 74 patients (35.07%) did not experience hypertension.

The biresponse logistic regression model assumes that there is a significant relationship or correlation between the two response variables. Therefore, a dependency test is necessary to determine the relationship between these two categorical response variables before performing the biresponse logistic regression modeling. The dependency test that can be used is the Chi-Square test. The testing begins by constructing a contingency table between the incidence of diabetes (Y_1) and the incidence of hypertension (Y_2) as follows:

Table 2. Contingency Table between Diabetes and Hypertension Incidence

			Hypertension Incidence		Total
			No Hypertension	Hypertension	
Diabetes Incidence	No Diabetes	Total	59	81	140
		Percentage	28%	38%	66%
		Expected	49	91	140
	Diabetes	Total	15	56	71
		Percentage	7%	27%	34%
		Expected	25	46	71
		Total	74	137	211

Based on Table 2, the information shows that out of a total of 211 patients, 71 patients (34%) were diagnosed with diabetes, 56 of whom also had hypertension. When examining the hypertension status, out of 137 patients

with hypertension, 56 patients also had diabetes. To further assess the dependence between the incidence of diabetes (Y_1) and hypertension (Y_2), a statistical test was conducted using the Pearson Chi-Square test with the following hypothesis:

H_0 : Variable incidence of diabetes (Y_1) and hypertension (Y_2) are independent
H_1 : Variable incidence of diabetes (Y_1) and hypertension (Y_2) are dependent

Based on the calculations and data analysis, the values obtained are shown in Table 3 below.

Table 3. Chi-Square Test Results between Diabetes and Hypertension Incidence

Pearson Chi–Square (χ^2) **Value**	**Degrees of Freedom** (df)	***p–value***
9.1374	1	0.002504

Based on Table 3, the calculated statistical value χ^2 is 9.1374, which is greater than $\chi^2_{(0.05;1)} = 3.841$, and the $p - value$ is less than α, then the null hypothesis H_0 was rejected. This implies that there is dependency between the response variable of the incidence of diabetes (Y_1) and the incidence of hypertension (Y_2).

To assess the presence of multicollinearity or high intercorrelations among predictor variables, the variance inflation factor (VIF) was employed in this study. Multicollinearity is generally considered to be present when the VIF exceeds 5–10. The VIF values for each predictor variable are presented in Table 4 [13].

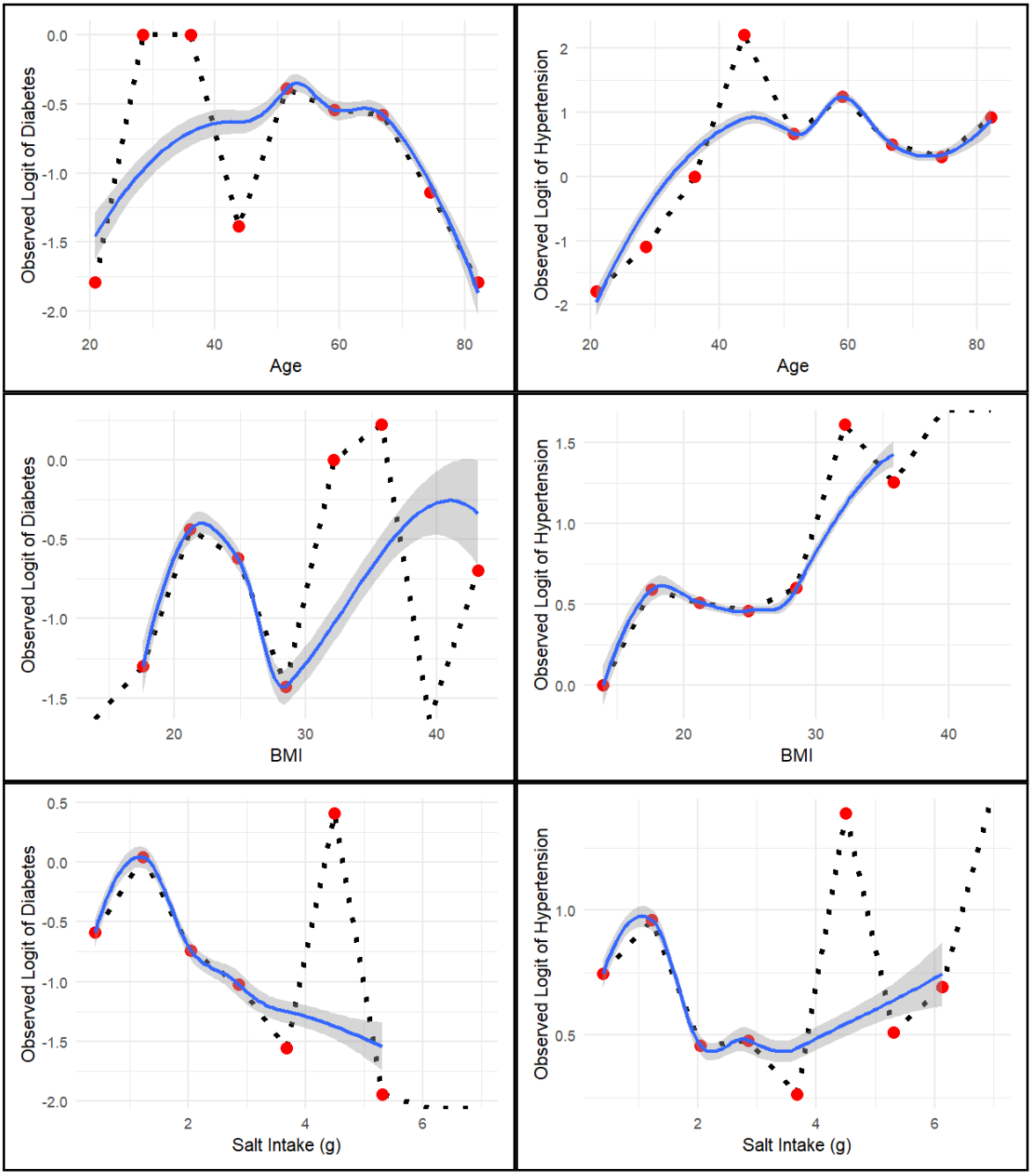
Table 4. Variance Inflation Factor Value for Each Predictor

$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$	$\boldsymbol{X_5}$
1.063936	1.027275	1.235882	1.327661	1.432385

Based on the VIF calculations presented in Table 4, the obtained values are 1.063936 for X_1, 1.027275 for X_2, 1.235882 for X_3, 1.327661 for X_4, and 1.432385 for X_5.

VIF value for each of predictor variables were found to be less than 5, indicating that there was no significant multicollinearity affecting the model estimates. Therefore, all variables can be simultaneously included in the modeling process.

In the binary logistic regression model with a nonparametric approach, the analysis starts with analyzing the linear relationship between the response variable and the predictor variable. The relationship can be analyzed visually using a scatter plot of the observed logit of the response variable and predictor variable. Below are the scatter plot results of the observed logit for each response variable and predictor variable presented in Figure 1 as follows:


Observed Logit of Diabetes
Age
Observed Logit of Hypertension
Age
Observed Logit of Diabetes
BMI
Observed Logit of Hypertension
BMI
Observed Logit of Diabetes
Salt Intake (g)
Observed Logit of Hypertension
Salt Intake (g)

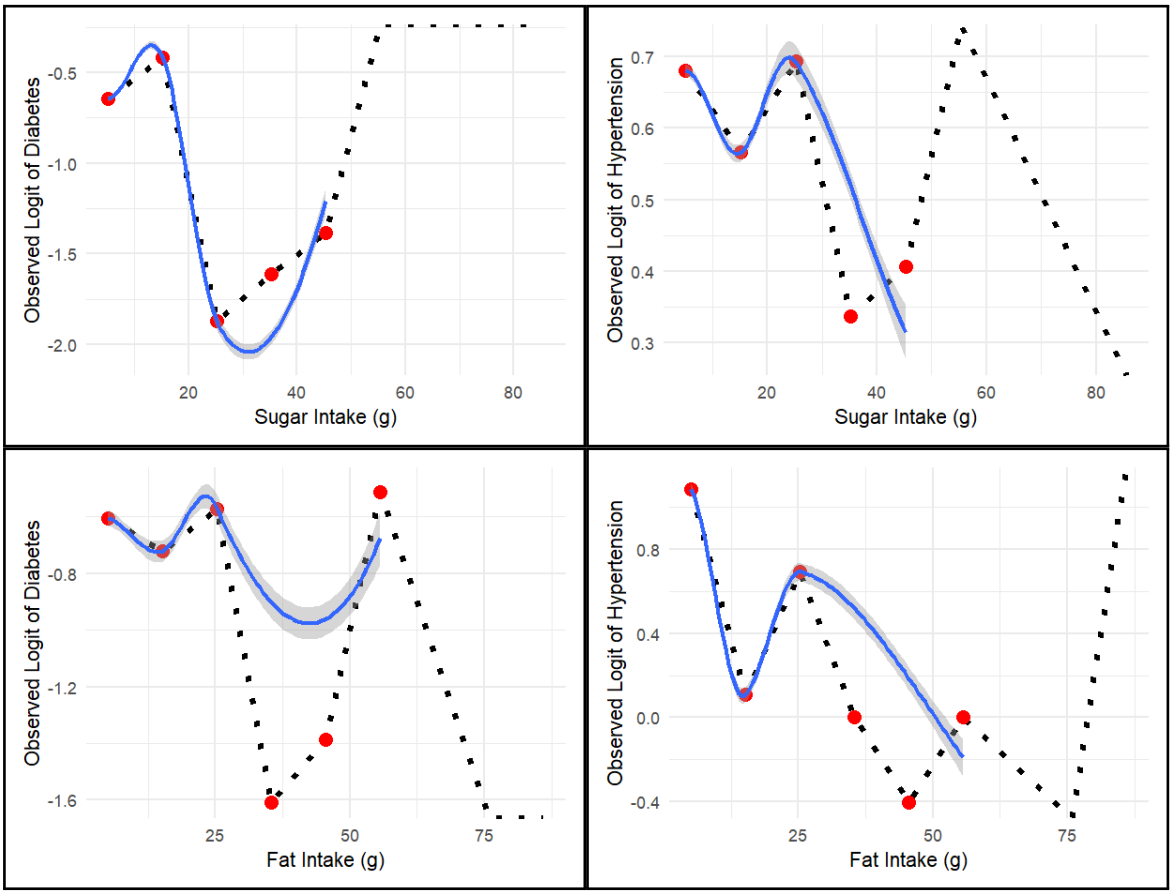


Figure 1. Scatter Plot of Observed Logit for Each Response Variable and Predictor Variable

Based on Figure 1, the plot showing the relationship between each predictor variable and the observed logit for diabetes and hypertension indicates a non-linear pattern. This non-linear pattern suggests that the effect of predictor variables on the risk of diabetes and hypertension is not always linear; rather, it depends on the level, where both low and high levels can either increase or decrease the risk. Therefore, this non-linear relationship needs to be identified, and risk modeling for diabetes and hypertension events should be performed using a nonparametric logistic regression approach.

The first step in the BNLR model building stage is to find the optimal degree of freedom (df) smoothing. To find the optimal degree of freedom (df), smoothing df values ranging from 5 to 9 will be tried, then the best model will be obtained based on the minimum Deviance value. The following are the 5 optimal smoothing df combinations based on the minimum Deviance value presented in Table 5.

Based on Table 5, from several combinations tested, it was found that the optimal combination occurred when X_1, X_2, X_3 dan X_5 each has 9 degrees of freedom, while X_4 has a degree of freedom of 6, which produces the lowest Deviance value of 404.465. This indicates that this combination provides the best performance between model flexibility and data fit, and can therefore be considered the most optimal smoothing spline configuration for modeling the nonlinear relationship between predictors and two response variables.

After obtaining the optimal smoothing df from the five predictor variables, the next step is to estimate the parameters of the bivariate response nonparametric logistic regression model using the Vector Generalized Additive Model (VGAM) approach. The estimated parameter values obtained are presented in Table 6 as follows.

Table 5. Finding the Optimal Smoothing (df) Based on the Minimum Deviance Value

Diabetes Incidence					Deviance
X_1	X_2	X_3	X_4	X_5	
9	**9**	**9**	**6**	**9**	**404.465**
9	9	9	6	8	406.773
8	9	9	6	9	407.241
9	9	9	5	9	407.316
9	8	9	6	9	407.513

Table 6. Parameter Estimation Results Using VGAM

Variable	Parameter	Estimation	Variable	Parameter	Estimation
Intercept	β_{01}	-0.103	Age (X_1)	β_{11}	-0.003
	β_{02}	0.615		β_{12}	0.003
	β_{03}	1.731			
BMI (X_2)	β_{21}	0.019	Salt Intake (X_3)	β_{31}	-0.266
	β_{22}	0.012		β_{32}	-0.135
Sugar Intake (X_4)	β_{41}	-0.031	Fat Intake (X_5)	β_{51}	0.012
	β_{42}	0.016		β_{52}	-0.03

As shown in Table 6, the positive coefficient for BMI (X_2) and Fat Intake (X_5) indicates that the higher BMI or Fat Intake is associated with an increased risk of diabetes. Clinically, for every unit increase in BMI or Fat Intake, the log-odds of the diabetes incidence increases by 0.019 and 0.012, respectively. Conversely, Age (X_1), Salt Intake (X_3), and Sugar Intake (X_4) shows a negative coefficient, reinforcing its role as a protective factor. The magnitude of the coefficient for Salt Intake is particularly large, suggesting it is the most critical determinant in this patient cohort.

In the incidence of hypertension, the positive coefficient for Age (X_1), BMI (X_2), and Sugar Intake (X_4) indicates that the higher Age, BMI, or Sugar Intake is associated with an increased risk of this disease. Clinically, for every unit increase in Age, BMI, or Sugar Intake, the log-odds of the hypertension incidence increases by 0.003, 0.012, and 0.016, respectively. Conversely, Salt Intake (X_3) and Fat Intake (X_5) shows a negative coefficient, reinforcing its role as a protective factor. The magnitude of the coefficient for Salt Intake is particularly large, suggesting it is the most critical determinant in this patient cohort.

The final model formed in the risk modeling of diabetes and hypertension based on the BNLR model is as follows:

Model logit 1:

$$\begin{aligned}\log\left(\frac{\pi_{1.}(\boldsymbol{X})}{1-\pi_{1.}(\boldsymbol{X})}\right) &= \beta_{01}+\beta_{11}X_1+\beta_{21}X_2+\beta_{31}X_3+\beta_{41}X_4+\beta_{51}X_5\\ &= -0.103-0.003X_1+0.019X_2-0.266X_3-0.031X_4+0.012X_5 \qquad (1)\end{aligned}$$

Model logit 2:

$$\begin{aligned}\log\left(\frac{\pi_{.1}(\boldsymbol{X})}{1-\pi_{.1}(\boldsymbol{X})}\right) &= \beta_{02}+\beta_{12}X_1+\beta_{22}X_2+\beta_{32}X_3+\beta_{42}X_4+\beta_{52}X_5\\ &= 0.615+0.003X_1+0.012X_2-0.135X_3+0.016X_4-0.03X_5 \qquad (2)\end{aligned}$$

Marginal probability model of Diabetes Incidence (Y_1):

$$\begin{aligned}\pi_{1.}(\boldsymbol{X}) &= \frac{\exp\left(\beta_{01}+\beta_{11}X_1+\beta_{21}X_2+\beta_{31}X_3+\beta_{41}X_4+\beta_{51}X_5\right)}{1+\exp\left(\beta_{01}+\beta_{11}X_1+\beta_{21}X_2+\beta_{31}X_3+\beta_{41}X_4+\beta_{51}X_5\right)} \\ &= \frac{\exp\left(-0.103-0.003X_1+0.019X_2-0.266X_3-0.031X_4+0.012X_5\right)}{1+\exp\left(-0.103-0.003X_1+0.019X_2-0.266X_3-0.031X_4+0.012X_5\right)}\end{aligned} \tag{3}$$

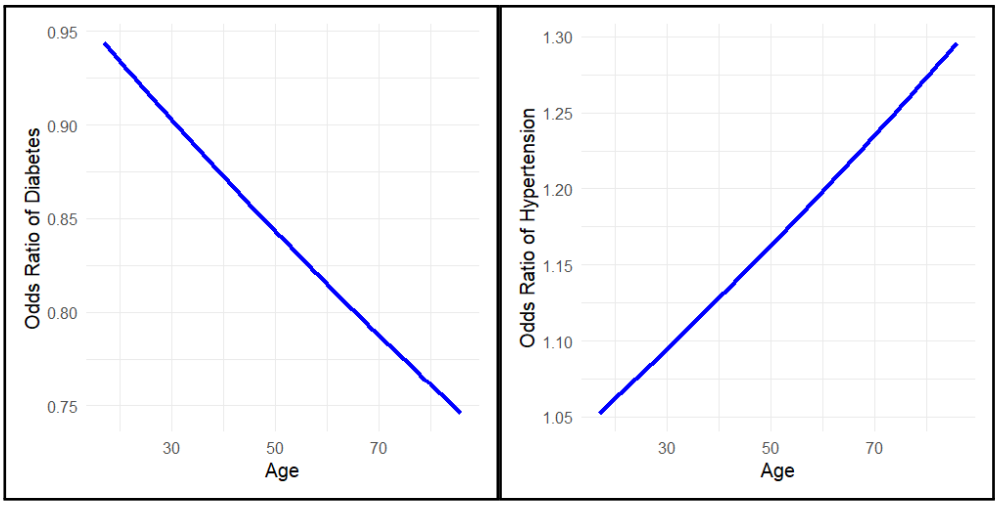
Marginal probability model of Hypertension Incidence (Y_2):

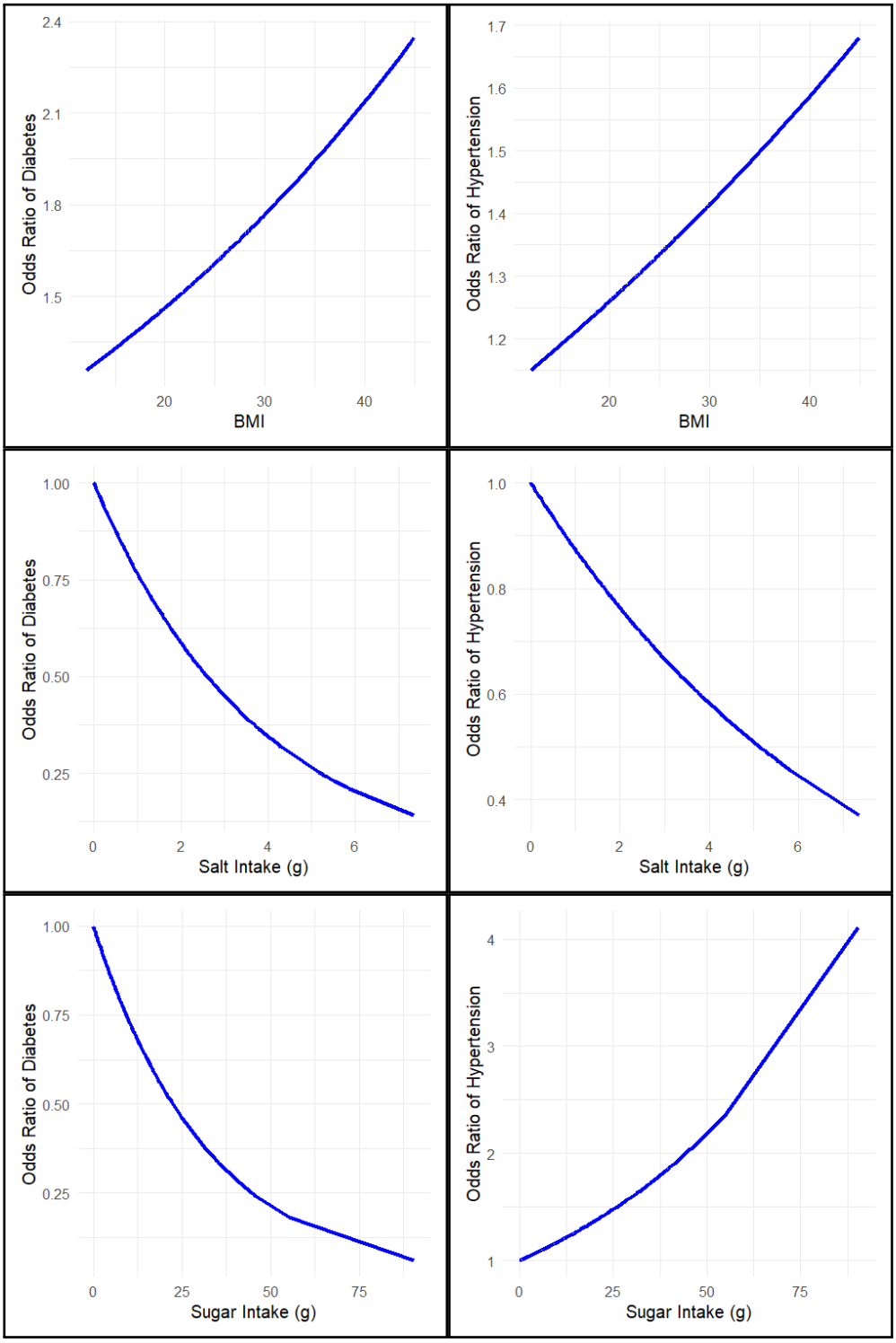
$$\begin{aligned}\pi_{.1}(\boldsymbol{X}) &= \frac{\exp\left(\beta_{02}+\beta_{12}X_1+\beta_{22}X_2+\beta_{32}X_3+\beta_{42}X_4+\beta_{52}X_5\right)}{1+\exp\left(\beta_{02}+\beta_{12}X_1+\beta_{22}X_2+\beta_{32}X_3+\beta_{42}X_4+\beta_{52}X_5\right)} \\ &= \frac{\exp\left(0.615+0.003X_1+0.012X_2-0.135X_3+0.016X_4-0.03X_5\right)}{1+\exp\left(0.615+0.003X_1+0.012X_2-0.135X_3+0.016X_4-0.03X_5\right)}\end{aligned} \tag{4}$$

To illustrate the application of the final model formulated in Equation (3) and (4), two hypothetical patients are presented below.

1. Patient A (High Risk): A 55-year-old male with hypertension and high BMI. Based on the model, his predicted probability of developing the disease is 85%.
2. Patient B (Low Risk): A 30-year-old female with normal blood pressure and active lifestyle. Her predicted probability is 12%.

This comparison demonstrates the model's capability to discriminate between risk levels in a clinical setting. Thereafter, the odds ratio for each predictor variable was computed for the entire dataset of study observations and presented in Figure 2 as follows:




Odds Ratio of Diabetes
2.4
2.1
1.8
1.5
20
30
40
BMI
Odds Ratio of Hypertension
1.7
1.6
1.5
1.4
1.3
1.2
20
30
40
BMI
Odds Ratio of Diabetes
1.00
0.75
0.50
0.25
0
2
4
6
Salt Intake (g)
Odds Ratio of Hypertension
1.0
0.8
0.6
0.4
0
2
4
6
Salt Intake (g)
Odds Ratio of Diabetes
1.00
0.75
0.50
0.25
0
25
50
75
Sugar Intake (g)
Odds Ratio of Hypertension
4
3
2
1
0
25
50
75
Sugar Intake (g)

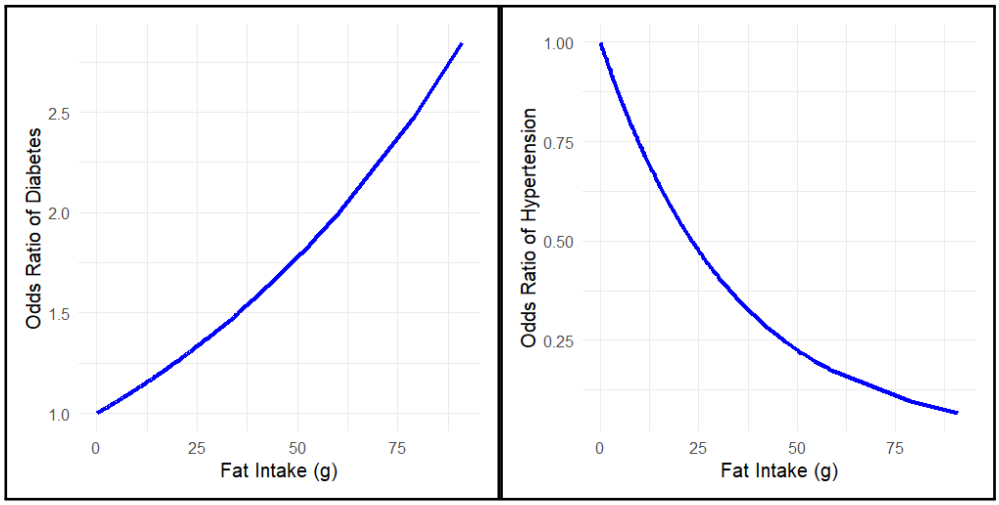


Figure 2. Changes in Odds Ratios for Each Predictor Variable with Respect to the Response

Based on Figure 2, it can be observed that each one-year increase in age is associated with a decrease in the risk of developing diabetes and an increase in the risk of hypertension. Furthermore, a one-unit increase in body mass index (BMI) consistently elevates the risk of both diabetes and hypertension. In contrast, the consumption of salt-containing foods is consistently associated with a reduction in the risk of both diabetes and hypertension. Meanwhile, the consumption of sugar-containing foods exhibits a trend that is broadly consistent with that of age. Conversely, increased consumption of fat-containing foods is associated with an increased risk of diabetes and a decreased risk of hypertension.

After estimating the probabilities for all data observations, the estimation results were then used to construct the confusion matrix, as presented in Table 7 below:

Table 7. Confusion Matrix 4×4

Actual	**Prediction**				**Total**
	Y_{00}	Y_{01}	Y_{10}	Y_{11}	
Y_{00}	38	14	0	7	59
Y_{01}	10	59	0	12	81
Y_{10}	5	2	2	6	15
Y_{11}	7	18	0	31	56
Total	**60**	**93**	**2**	**56**	**211**

where :
Y_{00} : Response variable when $Y_1 = 0$ and $Y_2 = 0$
Y_{01} : Response variable when $Y_1 = 0$ and $Y_2 = 1$
Y_{10} : Response variable when $Y_1 = 1$ and $Y_2 = 0$
Y_{11} : Response variable when $Y_1 = 1$ and $Y_2 = 1$

In general, the accuracy of a multiclass prediction model, including models with simultaneous biresponse outcomes, can be calculated using Equation (5) [34].

$$Accuracy = \sum_{i=1}^{c} \frac{n_{ii}}{n} \tag{5}$$

where :
n_{ii} : The number of observations with actual class i that are correctly classified as class i
n : The number of all observations

Based on the confusion matrix presented in Table 7, the prediction accuracies for diabetes (Y_1) and hypertension (Y_2), as defined in Equations (3) and (4), respectively, were calculated as follows.

$$\begin{aligned} Accuracy &= \sum_{i=1}^{4} \frac{n_{ii}}{n} \\ &= \frac{38 + 59 + 2 + 31}{211} \\ &= 0.61611 \end{aligned}$$

The resulting value of 0.61611 (61.611%) indicates that the bivariate prediction model correctly classified 61.611% of the observations.

The 4×4 confusion matrix was further reduced to a 3×3 and subsequently to a 2×2 matrix, as shown in Table 8 and Table 9 [15].

Table 8. Confusion Matrix 3×3

Actual	**Prediction**			**Total**
	Y_{00}	$Y_{01}Y_{10}$	Y_{11}	
Y_{00}	38	14	7	59
$Y_{01}Y_{10}$	15	63	18	96
Y_{11}	7	18	31	56
Total	**60**	**95**	**56**	**211**

Table 9. Confusion Matrix 2×2

Actual	**Prediction**	
	0	1
0	38	21
1	22	94

Next, the *Press's Q* statistic was calculated using Equation (6) to test the stability of the BNLR model with the following hypotheses [11].

H_0 : The model classification is inconsistent
H_1 : The model classification is consistent

The calculation of *Press's Q* yielded the following results.

$$Press's\text{–}Q = \frac{[(211) - (132 \times 2)]^2}{(211)(2-1)} = 13.313 \tag{6}$$

A *Press's Q* value of 13.313 was obatained, which is greater than $\chi^2_{(0.05;\,1)} = 3.841$, leading to the rejection of H_0. Thus, it can be concluded that the classification of the constructed model is consistent.

To determine which approach is more suitable for modeling the risk of diabetes and hypertension events simultaneously, a comparison was made between the parametric and nonparametric bivariate logistic regression models. The parametric model assumes a linear relationship between predictors and the response, whereas the nonparametric approach, using smoothing functions, allows for a more flexible relationship. Consequently, both models were evaluated based on Deviance value, accuracy, AUC, and model stability tested using *Press's Q*, as shown in Table 10.

Table 10. Comparison of the Performance of Biresponse Logistic Regression Model Approaches

Model Type	Deviance	Accuracy	AUC	Model Stability Test		
				Press's Q	χ^2	**Decision**
Parametric Biresponse Logistic Regression	517.318	0.436	0.62	3.455	3.841	Not Stable
Nonparametric Biresponse Logistic Regression	**404.465**	**0.626**	**0.83**	**13.313**	**3.841**	**Stable**

Based on Table 10, it is evident that the Nonparametric Biresponse Logistic Regression model demonstrates substantially better performance compared to the parametric approach. This is reflected in the Deviance value, with the nonparametric approach (404.465) being lower than the parametric approach (517.318), indicating that the nonparametric model fits the data more effectively. Furthermore, the accuracy increased substantially from 0.436 to 0.626, and the AUC also improved from 0.62 to 0.83. These results indicate that the nonparametric model has a superior classification ability in distinguishing between the two response categories simultaneously.

In terms of model stability, the nonparametric approach also demonstrated superior results. The stability test using *Press's Q* yielded a value of 13.313, which is greater than the critical value ($\chi^2 = 3.841$), indicating that the model is statistically stable. In contrast, the parametric model produced a *Press's Q* value of only 3.455, which is below the critical value $\chi^2 = 3.841$, and thus considered unstable. Therefore, it can be concluded that the biresponse nonparametric logistic regression approach is not only more accurate but also more stable, making it a better choice for modeling the risk of diabetes and hypertension events in this study.

The superior performance of the nonparametric model compared to the parametric approach can be attributed to its flexibility. Unlike parametric models that assume strict linearity, the nonparametric approach utilizes spline smoothing to adapt to the data's local structures. This allows for a more accurate representation of the non-linear fluctuations in risk factors, which are biologically plausible in medical contexts, including the diseases risk modelling. The two of highest noncommunicable diseases case occur in Indonesia, spesifically in Surabaya are diabetes and hypertension [18]. Modeling diabetes and hypertension jointly is clinically significant due to their shared pathophysiology. The bivariate analysis accounts for the correlation between these two outcomes, offering a more holistic view of patient risk compared to analyzing each disease in isolation. This suggests that interventions targeting one condition should consider the concurrent risk of the other.

4. Conclusion

The analysis shows a dependency between the incidence of these two diseases, as indicated by the Chi-square test with a calculated statistic of $\chi^2 = 9.1374$. The relationship between predictor variables and the observed logits of both diseases exhibits a non-linear pattern, suggesting that the effect of predictors is not always linear. Furthermore, model comparison results indicate that the Biresponse Nonparametric Logistic Regression approach outperforms the parametric model, with lower deviance (404.465 vs. 517.318), higher accuracy (0.626 vs. 0.436), greater AUC (0.83 vs. 0.62), and better stability (*Press's Q* $= 13.313 > \chi^2 = 3.841$, stable), demonstrating superior predictive performance and robustness.

Limitations

Several limitations of this study should be acknowledged. First, the single-center design conducted at Universitas Airlangga Hospital limits the immediate generalizability of the findings to broader Indonesian populations or other healthcare settings. This limitation highlights the need for future research to explore and validate these results in multicenter settings. Second, the development of the initial model in this study was limited to a sample of 211 patients. Therefore, future studies with larger sample sizes may help to validate and generalize these findings. Larger multi-center studies are also required to externally validate the model's predictive performance. Third, the cross-sectional nature of the data precludes the determination of causal relationships between risk factors and disease onset.

Fourth, the current performance metrics are calculated based on the training data. Nevertheless, the stability of the model in this study was evaluated using *Press's Q test*, yielding a Q statistic of 13.313, which indicates that the model is stable. This can be regarded as evidence that further external validation and cross-validation studies would be valuable for future research to confirm generalizability.

Fifth, nonparametric regression is a smoothing-based regression approach. Consequently, all predictor variables utilized in this study are measured on a ratio scale. Categorical variables, including family history, smoking history, physical activity, sex, and related factors, are irrelevant for nonparametric regression modeling due to their inability to undergo smoothing.

Acknowledgement

This work supported by Airlangga Research Fund (ARF) 2025 Batch 1 Program (Scheme: *Penelitian Dasar Unggulan*) supported by Universitas Airlangga with a contract number 1748/UN3.LPPM/PT.01.03/2025.

REFERENCES

1. T. Adiwati and N. Chamidah. Modelling of hypertension risk factors using penalized spline to prevent hypertension in Indonesia. *IOP Conference Series: Materials Science and Engineering*, 546(5):052003, June 2019. Publisher: IOP Publishing.
2. A. Agresti. *An Introduction to Categorical Data Analysis*. Wiley, July 2019. Google-Books-ID: ZX9q0QEACAAJ.
3. Z. N. Amalia, D. R. Hastuti, F. Istiqomah, and N. Chamidah. Hypertension risk modeling using penalized spline estimator approach based on consumption of salt, sugar, and fat factors. In *AIP Conference Proceedings*, volume 2264, page 030005. AIP Publishing LLC, 2020. Issue: 1.
4. E. Ana, N. Chamidah, P. Andriani, and B. Lestari. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression. In *Journal of Physics: Conference Series*, volume 1397, page 012067. IOP Publishing, 2019. Issue: 1.
5. W. A. Anam, A. Massaid, N. A. Amesya, and N. Chamidah. Modeling of diabetes mellitus risk based on consumption of salt, sugar, and fat factors using local linear estimator. In *AIP Conference Proceedings*, volume 2264, page 030009. AIP Publishing LLC, 2020. Issue: 1.
6. P. Andriani and N. Chamidah. Modelling of hypertension risk factors using logistic regression to prevent hypertension in Indonesia. In *Journal of Physics: Conference Series*, volume 1306, page 012027. IOP Publishing, 2019. Issue: 1.
7. Badan Penelitian dan Pengembangan Kesehatan. *Laporan Nasional Riskesdas 2018 [in Indonesian]*. Kementerian Kesehatan Republik Indonesia, Jakarta, June 2020.
8. N. Chamidah, E. Ana, A. Isadika, and H. Susilo. Modeling of hypertension risk based on age, body mass index, and psychological factors using local linear estimator. *AIP Conference Proceedings*, 3201(1):060018, November 2024.
9. J. H. Chi and B. J. Lee. Risk factors for hypertension and diabetes comorbidity in a Korean population: A cross-sectional study. *Plos one*, 17(1):e0262757, 2022. Publisher: Public Library of Science San Francisco, CA USA.
10. M. Enea and M. Attanasio. An association model for bivariate data with application to the analysis of university students' success. *Journal of Applied Statistics*, 43(1):46–57, January 2016. Publisher: Taylor & Francis. eprint: https://doi.org/10.1080/02664763.2014.998407.
11. S. Innassuraiya, T. Widiharih, and I. T. Utami. Analisis klasifikasi menggunakan metode regresi logistik biner dan bootstrap aggregating classification and regression trees (BAGGING CART) (Studi kasus: Nasabah koperasi simpan pinjam dan pembiayaan syariah (KSPPS)) [in Indonesian]. *Jurnal Gaussian*, 11(2):183–194, August 2022. Publisher: Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro.
12. T. M. Kaombe, J. C. Banda, G. A. Hamuza, and A. S. Muula. Bivariate logistic regression model diagnostics applied to analysis of outlier cancer patients with comorbid diabetes and hypertension in Malawi. *Scientific Reports*, 13(1):8340, May 2023. Publisher: Nature Publishing Group.
13. Jong Hae Kim. Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6):558–569, July 2019.

14. Z. Li, S. Pang, H. Qu, and W. Lian. Logistic regression prediction models and key influencing factors analysis of diabetes based on algorithm design. *Neural Computing and Applications*, 35(36):25249–25261, December 2023.
15. I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis. Multiclass confusion matrix reduction method and its application on net promoter score classification problem. *Technologies*, 9(4):81, December 2021. Publisher: Multidisciplinary Digital Publishing Institute.
16. A. C. Mathew, E. Siby, A. Tom, and S. K. R. Applications of proportional odds ordinal logistic regression models and continuation ratio models in examining the association of physical inactivity with erectile dysfunction among type 2 diabetic patients. *Physical Activity and Nutrition*, 25(1):30–34, March 2021. Publisher: Korean Society for Exercise Nutrition.
17. M. Mayawi, N. Nurhayati, T. Talib, A. W. Bustan, and N. S. Laamena. Ordinal logistic regression analysis of factors that affecting the blood sugar levels diabetes mellitus patients. *Pattimura International Journal of Mathematics (PIJMath)*, 2(1):33–42, April 2023.
18. Riski Dwi Prameswari, Natalia Christin Tiara Revita, Dwi Ayu Angraini, and Islakhil Asfarina. Epidemiological Analysis of Non-Communicable Diseases Post-Covid Era in Indonesia. *Indonesian Journal of Global Health Research*, 7(1):617–626, 2025.
19. Purhadi and M. Fathurahman. A logit model for bivariate binary responses. *Symmetry*, 13(2):326, February 2021. Publisher: Multidisciplinary Digital Publishing Institute.
20. V. Ratnasari, Purhadi, I. C. Aviantholib, and A. T. R. Dani. Parameter estimation and hypothesis testing the second order of bivariate binary logistic regression (S-BBLR) model with Berndt Hall-Hall-Hausman (BHHH) iterations. *Commun. Math. Biol. Neurosci.*, 2022(0):Article ID 35, November 2022.
21. V. Ratnasari, Purhadi, M. Rifada, and A. T. R. Dani. Explore poverty with statistical modeling: The bivariate polynomial binary logit regression (BPBLR). *MethodsX*, 14:103099, June 2025.
22. M. Rifada, D. Amelia, J. P. Setyaningrum, N. Septiandini, Y. K. Kalista, and S. N. Dwitya. Analysis of unmet need for health services based on the percentage of public health complaints with a kernel estimator approach. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 9(4):1258–1270, October 2025.
23. M. Rifada, E. Ana, C. A. Siburian, and A. G. Safitri. Risk analysis of diabetes mellitus and hypertension using biresponse binary logistic regression. *International Journal of Academic and Applied Research (IJAAR)*, 8(12):82–88, December 2024. Publisher: IJARW.
24. M. Rifada, N. Chamidah, and R. A. Ningrum. Estimation of nonparametric ordinal logistic regression model using generalized additive models (GAM) method based on local scoring algorithm. In *AIP Conference Proceedings*, volume 2668, page 070013. AIP Publishing LLC, 2022. Issue: 1.
25. M. Rifada, N. Chamidah, R. A. Ningrum, and L. Muniroh. Stunting determinants among toddlers in Probolinggo district of Indonesia using parametric and nonparametric ordinal logistic regression models. *Commun. Math. Biol. Neurosci.*, 2023(0):Article ID 8, January 2023.
26. M. Rifada, N. Chamidah, P. Nuraini, F. D. Gunawan, and L. Muniroh. Determinants of stunting among under-five years children using the ordinal logistic regression model. In *1st International Conference on Mathematics and Mathematics Education (ICMMEd 2020)*, pages 405–411. Atlantis Press, May 2021. ISSN: 2352-5398.
27. M. Rifada, N. Chamidah, V. Ratnasari, and Purhadi. Estimation of nonparametric ordinal logistic regression model using local maximum likelihood estimation. *Commun. Math. Biol. Neurosci.*, 2021(0):Article ID 28, May 2021.
28. M. Rifada, V. Ratnasari, and Purhadi. Parameter estimation and hypothesis testing of the bivariate polynomial ordinal logistic regression model. *Mathematics*, 11(3):579, January 2023. Publisher: Multidisciplinary Digital Publishing Institute.
29. M. Rifada, V. Ratnasari, and Purhadi. Parameter estimation of the bivariate polynomial ordinal logistic regression model. *AIP Conference Proceedings*, 2554(1):030009, January 2023.
30. M. Rifada and Suliyanto. The risk modeling of diabetes based on parametric and nonparametric binary logistic regression. *ARPN Journal of Engineering and Applied Sciences*, 15(20):2356–2363, 2020. Publisher: Asian Research Publishing Network (ARPN).
31. M. Rifada, Suliyanto, E. Tjahjono, and A. Kesumawati. The logistic regression analysis with nonparametric approach based on local scoring algorithm (Case study: Diabetes mellitus type II cases in Surabaya of Indonesia). *Int. J. Adv. Soft Comput. Appl*, 10:167–178, 2018.
32. Suliyanto, M. Rifada, and E. Tjahjono. Estimation of nonparametric binary logistic regression model with local likelihood logit estimation method (case study of diabetes mellitus patients at Surabaya Hajj General Hospital). In *AIP Conference Proceedings*, volume 2264, page 030007. AIP Publishing LLC, 2020. Issue: 1.
33. J. S. Thakur, R. Nangia, and S. Singh. Progress and challenges in achieving noncommunicable diseases targets for the sustainable development goals. *FASEB BioAdvances*, 3(8):563–568, 2021. eprint: https://faseb.onlinelibrary.wiley.com/doi/pdf/10.1096/fba.2020-00117.
34. Ayfer Ezgi Yilmaz and Haydar Demirhan. Weighted kappa measures for ordinal multi-class classification performance. *Applied Soft Computing*, 134:110020, February 2023.