Available online at
https://jcrinn.com/
https://crinn.conferencehunter.com/

Journal of Computing Research and Innovation 11(2) 2026

Journal of Computing Research and Innovation

Understanding Diabetes among Malaysians using Logistic Regression: What Drives the Disease?

Siti Fairusziah Khairul Anwar[1], Jaida Najihah Jamidin[2*]

[1]*AS White Global Malaysia, 59200 Kuala Lumpur, Malaysia.*
[2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia.*

ARTICLE INFO

Article history:
Received 29 June 2025
Revised 28 February 2026
Accepted 11 March 2026
Online first
Published 1 September 2026

Keywords:
Diabetes
Logistic Regression Analysis
Malaysia
Risk Factor
Risk Prediction

DOI:
10.24191/jcrinn.v11i2.558

ABSTRACT

Diabetes is one of the critical illnesses that the rate of patients is increasing day by day not only in Malaysia, but across the globe. Yet, considering Malaysians' lack of acknowledgment of the early signs, they remain unaware of how serious and dangerous this disease is. Therefore, understanding the factors that cause diabetes is very important to ensure everyone involved can develop a preventive approach. This study aims to identify the factors that contribute to diabetes among Malaysians. A survey was conducted among Malaysians, and logistic regression analysis was utilized to determine the factors contributing to diabetes. The findings highlighted four leading risk factors that contribute to diabetes, which were age, body mass index, physical exercise, and smoking. From the analysis, the Omnibus test ($p < 0.05$) indicates that the independent variables collectively have a significant relationship with the presence of diabetes, while the Hosmer–Lemeshow test ($p = 0.8799$) confirms that the model fits the data well. The model explains 32.84% of the variation in diabetes based on Cox and Snell R^2, with a Nagelkerke R^2 of 0.7877, suggesting strong explanatory power. Additionally, the model demonstrates high predictive performance, with 98.86% sensitivity, 86.21% specificity, and an overall accuracy of 97.89%. These results offer a glimpse into how lifestyle choices interfere with the likelihood of someone developing diabetes. On top of that, these findings have aided in enhancing the prevention of diabetes and offer opportunities for further research to explore other diabetes-related factors.

1. INTRODUCTION

Diabetes is a chronic metabolic illness that can harm the ability of an individual's body to process blood sugar, also known as glucose. Glucose is the main source of energy for the cells of the human body. The level of glucose in our blood is controlled by a type of hormone called insulin. If our body does not produce enough insulin or cannot use the insulin produced effectively, then it can lead to diabetes. There are two main types of diabetes disease, namely Type 1 and Type 2. Type 1 diabetes, also known as insulin

[2*] Corresponding author. *E-mail address*: jaida5698@uitm.edu.my
https://doi.org/10.24191/jcrinn.v11i2.558

dependent, occurs when the pancreas produces too little to no insulin. Hence, sugar can't enter the body's cells to be used as energy. Unfortunately, the cause of this Type 1 diabetes is still unknown and remains unpreventable (American Diabetes Association, 2024). Some studies have been done, and some strategies have been tried but there is only a little success. On the other hand, a combination of genetics and environment like maternal diet and socioeconomic condition is believed to lead to Type 1 diabetes (Abela & Fava, 2021).

While in diabetes Type 2, the pancreas produces less than normal amounts of insulin, which means the body has insulin, but it cannot be used. Among the two types of diabetes, Type 2 is the most common and is usually experienced by adults, where the body is unable to produce sufficient insulin or is immune to it. It can be said that nine out of 10 people have Type 2 diabetes (Dwivedi & Pandey, 2020). Risk factors for Type 2 diabetes are better known when compared to Type 1 diabetes. Apart from that, they in the same research stated that most Type 2 cases occur due to several contributors such as age, overweight or obesity, and physical inactivity, where a person is lacking exercise. Smoking is also another cause of increased risk of diabetes. Over time, diabetes can contribute to several other medical conditions, namely cardiovascular disease, kidney failure, blindness, and heart attack (World Health Organization, Accessed 2024).

Diabetes is also known as one of the killer diseases as almost 422 million people worldwide have diabetes, and approximately 1.5 million deaths are caused by diabetes each year (World Health Organization, 2024). Due to a lack of self-awareness and often take health-related issues for granted, many Malaysians do not realize that they have diabetes until they are experiencing its symptoms such as frequent urination, fatigue, excessive thirst and numbness in their hands or feet. Therefore, diabetes patients linger to increase in Malaysia from year to year. According to International Diabetes Federation (IDF), Malaysia was high ranked among Western Pacific region in 2018, which have large number of people diagnosed with diabetes approximately 16.8% from adult population in worldwide (Ganasegeran et al., 2020). As a result, Malaysia's healthcare delivery system faces great challenges in terms of providing quality care to diabetic patients since that. Hence, by studying the major factors that lead to diabetes and knowing its importance for the people in Malaysia, it can reduce its impact on individuals, communities and the health care system of Malaysians. Accordingly, this study desired to discover the components that are associated with the presence of diabetes.

2. LITERATURE REVIEW

The focus of this section was to review all the elements that have been selected in this study that are expected to have an impact on the root of a person suffering from diabetes. This chapter also provided some additional knowledge and information to the current study.

The first independent variable is age. Aging can be described as a decline in someone's physical function and condition that might increase the probability of developing chronic diseases such as diabetes (Burton & Faragher, 2018). According to Griffith (2024), he believed that long-term illnesses such as diabetes, particularly Type 2 diabetes, tend to be widespread among elderly people. People who are older in age are more likely to experience the deterioration of their quality of life because there was a drop in their body function, which was triggered by organ and cellular aging (Di Micco et al., 2021). Izzo et al. (2021) in their research revealed that approximately 25% of people aged 65 and above have Type 2 diabetes, which is a major health concern for this age group. However, this disease does not only strike the elderly because pre-teenagers and teenagers may also contract it if they practice poor diet and are inactive (WebMD, Accessed 2024). Frow the above reviewed literature, several researchers suggested that diabetes disease may rise due to an increasing age.

https://doi.org/10.24191/jcrinn.v11i2.558

The next variable is body mass index. Body Mass Index (BMI) can be calculated by dividing a person's weight into kilograms or pounds by their height in metres or feet. According to Soltani et al. (2021), Type 2 diabetes is extremely linked to excess body weight. The amount of body fat collected by a person through eating unhealthy and non-nutritious food can cause them to develop Type 2 diabetes. Along with the time body mass index increases, the risk of getting diabetes will likely increase linearly (Klein et al., 2022). This statement is supported by Teufel et al. (2021) in their findings that an individual with BMI of 23kg/m^2 and above are linked with a higher chance of getting diabetes, with a possibility rising to 43% for men and 41% for women versus those with a normal BMI ranging from 18.5kg/m^2 to 22kg/m^2. Next, according to Gupta and Bansal (2020), individuals who are overweight or commonly known as obese are more likely to suffer from prediabetes or diabetes compared to people who have a normal weight or underweight. Clearly, a large number of prior studies showed that body mass index is a significant predictor of whether a person has diabetes or not.

Blood pressure is the next variable. The pressure that exists when the blood flows against the walls of blood vessels while the heart pumps blood throughout the whole body is called blood pressure (Schutte et al., 2022). High blood pressure or known as hypertension occurs when the pressure of your blood vessels is too high. Besides, an individual who is overweight is related with having high blood pressure because the higher the amount of fat, the higher the oxygen and nutrient needed. Therefore, it can eventually increase the risk of having diabetes (Zhou et al., 2021). According to Nazarzadeh et al. (2021), a person with high blood pressure is at risk of developing diabetes and other cardiovascular diseases. They also believed that the prevention of diabetes can be implemented by reducing blood pressure. This can be done through exercise and staying active, as well as reducing salt intake in food. Next, based on research done by Libianto et al. (2018), they stated that hypertension is often experienced by people with diabetes, and the presence of diabetes in someone with high blood pressure indicates a high cardiovascular risk. Compared to people with normal blood pressure, individuals with uncontrolled blood pressure had a higher risk in diabetes.

Next, exercise has been proven to provide many clinical benefits, including improved insulin sensitivity and increased peak oxygen consumption (Amanat et al., 2020). As of now, exercise has been proven to be effective in preventing and treating diabetes. Active exercise plays a crucial part in treating Type 1 diabetes (Lu & Zhao, 2020). Besides, exercise can also improve blood glucose control in Type 2 diabetes, reduce cardiovascular risk, and reduce body fat. Amanat et al. (2020) in their study stated that a sedentary lifestyle, which someone who lacks significant physical movement in their daily life, is the prime factor in the expansion in Type 2 diabetes worldwide. Additionally, Talukder and Hossain (2020) in their findings revealed that individuals who do a lot of physical activities have a lower chance to diagnosed with diabetes compared to people who don't. In fact, 7% of people worldwide diagnosed with diabetes is because of them lacking in terms doing physical exercises (Thompson et al., 2020). Therefore, this study aimed to verify the findings of the previous studies which implies that diabetes is developed by lack of physical exercise.

Additionally, smoking is the practice of inhaling smoke from a burning plant material. A substance called nicotine in cigarettes works to create a sense of relaxation and calmness in the smoker's brain that makes it difficult to quit. Smoking has been found to be strongly correlated to the risk of insulin resistance and leads a person to diabetes (Lin et al., 2020). The reason smoking results in diabetes is caused by the nicotine in cigarettes. In addition, Yang et al. (2022) survey that heavy smokers show a greater risk of developing cardiovascular diseases such as diabetes, compared to people who have never smoked. Smoking may contribute to the journey from low blood sugar levels to poor glucose tolerance, thus increasing the risk of diabetes development (Durlach et al., 2022). Overall, smoking has been associated with diabetes in the population, with gender differences observed as diabetes is more likely to be developed by male heavy smokers compared to female (Kyung et al., 2024). Nonetheless, in the same study by Lin et al. (2020), they also stated that there is still insufficient evidence to indicate that smoking is the reason someone develops

https://doi.org/10.24191/jcrinn.v11i2.558

diabetes. Hence, this study aimed to examine and reveal the association between cigarette smoking and the risk of getting diabetes.

Logistic regression is a type of regression that forecasts descriptive variable based on a combination of independent variables that act like some of them are continuous variables while the other variables are discrete. A statistical method called logistic regression is suitable to measure the existence of diabetes based on several risk factors because it has categorical outcomes: having diabetes (1) and not having diabetes (0) (Rajendra & Latifi, 2021). Al Bairmani and Ismael (2021) have conducted research that study the vital factors affecting diabetes using the logistic regression method. The findings showed that among 14 independent variables, only four variables that are significant with the presence of diabetes which are cigarettes smoking, exercise, vitamin D consumption and blood pressure (Al Bairmani & Ismael, 2021). Joshi and Dhakal (2021) also utilized a logistic regression to predict the Type 2 diabetes presence among Pima Indian women. The findings for the analysis revealed that there were five major risk factors of diabetes namely level of glucose, pregnancy, body mass index, diabetes pedigree function and age. They claimed that the logistic regression model was useful and can accurately predict the development of diabetes Type 2 and it is potentially could be used to prevent diabetes from occur. Talukder and Hossain (2020) in their study seek to identify the rate of diabetes in Bangladesh and its risk variables by performing a logistic regression analysis. Based on the findings from utilizing the logistic regression, they found that diabetes was more likely to develop as the respondents age increases. Next, individuals with hypertension and is overweighed was getting higher chances of getting diabetes. However, individuals who engage in a lot of physical work was less likely to have diabetes compared to those who don't do lots of physical work. By using logistic regression to analyse the risk factors of diabetes, researchers might obtain valuable insight into the major indicators that affect diabetes and spread awareness to improve health-related habits.

3. METHODOLOGY

Methodology is the most essential aspect to determine the validity of a study. Therefore, the general outline of this research method was discussed in this section, to achieve the study's objective. This study was conducted in binary logistic regression which will determine whether the respected independent variables lead to the dependent variable, notably diabetes presence. This section also discussed the data sources and the theoretical framework that helped in giving shape and guidelines throughout the research process.

Causal research or as known as explanatory research was conducted in this study. Causal research is a research design that is focusing on defining the cause and effect of any relationships. The purpose of causal research is to comprehend which aspects that directly lead to an outcome, testing the hypothesis about the causal relationship, and forecast future outcomes under similar settings. In this study, the causal research used to determine the cause-and-effect factors that contribute to someone diagnosed with diabetes.

3.1 Theoretical framework

Theoretical framework helps in terms of guiding and explaining how the concept is applied in the study. In this research, a theoretical framework is needed as it helps to explain how different factors such as age, body mass, blood pressure, physical exercise and smoking habit contribute to diabetes risk and how these align with the existing theories. Thus, the theoretical framework of this study is shown below:

https://doi.org/10.24191/jcrinn.v11i2.558

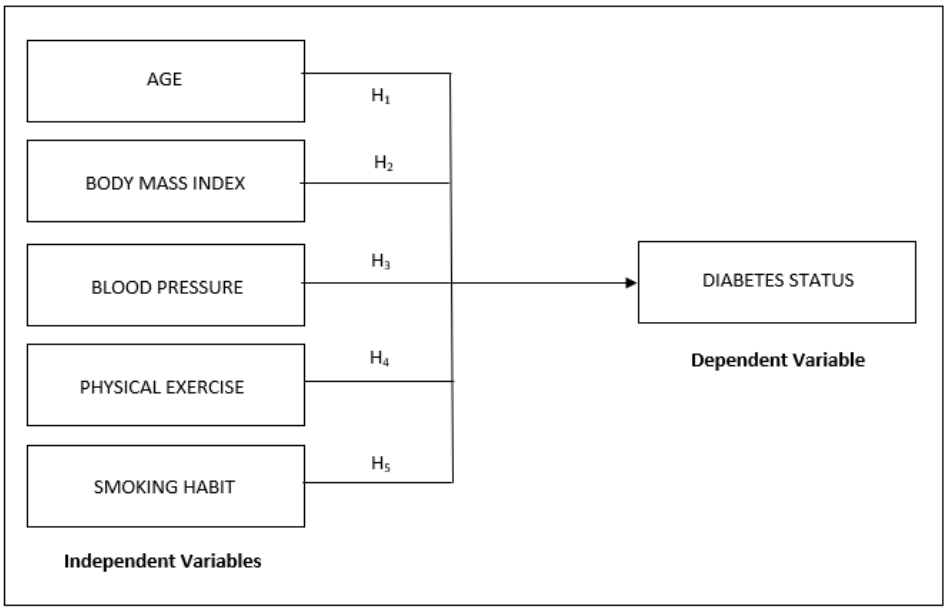

Fig. 1. The theoretical framework of risk factors that leads to diabetes

3.2 Data collection method

To obtain statistically meaningful results, a minimum of 385 respondents from Malaysia's 34,671,895 population was needed for the questionnaire. A 5% margin of error is acceptable in any social research, and a 95% confidence level indicates that 95 out of 100 samples will have true population value (Taherdoost, 2017). Hence, assuming a 95% confidence interval, a 0.05 margin of error, and an equal probability (p = 0.5), this formula offers the optimal sample size to achieve accurate data representation:

$$n = \frac{Z^2 pq}{e^2}$$

$$n = \frac{(1.6449)^2(0.5)(0.5)}{0.05^2}$$

$$n = 385$$

where,

n	:	sample size
Z	:	z-score
e	:	margin of error
p	:	probability
q	:	1-probability

This study employed a non-probability sampling technique, namely convenience sampling. The reason behind the selection of the sample was because of its simplicity and cost-effectiveness. Respondents for this study were gathering by the researcher through a direct approach from various accessible locations, such as universities, healthcare facilities, and crowded areas to answer the online questionnaire. The online platform has also been used to obtain more respondents.

https://doi.org/10.24191/jcrinn.v11i2.558

3.3 Logistic regression model

Logistic regression analysis used to estimate the probability of an event occurring, but the dependent variable is in binary form. The major purpose of using logistic regression analysis is similar to linear regression analysis, which is to estimate the association between the dependent variables and independent variables. The only difference is even though linear regression estimates continuous dependent variables, logistic regression predicts categorical dependent variables. However, before logistic regression can be utilized, there are several assumptions that need to be met. Firstly, the dependent variable must consist of only two outcomes. Secondly, the observations of the dataset must be independent of each other. Thirdly, there is no multicollinearity in the independent variables. The general form of logistic regression model is as shown:

$$\log\left[\frac{P_i}{1-P_i}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

where

P_i	:	The probability of developing diabetes (1=Yes, 0=No)
X_1	:	Age
X_2	:	Body mass index
X_3	:	Hypertension (1=Yes, 0=No)
X_4	:	Exercise
X_5	:	Smoking (1=Yes, 0=No)

3.3.1 Assumptions of logistic regression model

Logistic regression does not require many assumptions like linear regression does. However, there are still some assumptions that need to be fulfilled in logistic regression. The first assumption is the linearity between the continuous independent variables and the log-odds of dependent variable. The Box-Tidwell test was used to this assumption, with a p-value greater than 0.05 indicating a linear relationship. The last assumption is no severe multicollinearity problem. Multicollinearity denotes to the presence of high relationship between all independent variables in a regression model. A variance inflation factor (VIF) value that exceeds 10, and a low tolerance value below 0.2 show the presence of multicollinearity. With the existence of multicollinearity in the data, it will make it tough for researchers to interpret the effect of each independent variable on the presence of diabetes. Hence, this multicollinearity detection technique will be able to help researchers deal with the problems mentioned.

3.3.2 Model evaluation

Assessing the logistic regression model by using a few criteria will aids the researcher to evaluate its overall significance, predictive performance and can perceive its validity to predict diabetes status among Malaysians. The first criterion is Omnibus Test. Omnibus Test is employed to determine whether the regression coefficient is significant with other coefficients or not. If the p-value associated with the Omnibus Test attained is less than the significance level, that means there is enough evidence to reject the null hypothesis. Hence, there is a relationship between the risk factors and the presence of diabetes.

Next criterion is Wald Statistics. The purpose of Wald Statistics is to identify the significance between each coefficient in a logistic regression. In other words, the Wald Statistic shows whether the independent variable influences the prediction of diabetes existence. A p-value lower than the significance level indicates that the respected risk factors is significantly correlated with the presence of diabetes.

https://doi.org/10.24191/jcrinn.v11i2.558

The Hosmer-Lemeshow Test is the next criterion that is used to analyse and estimate whether a logistic regression model is a good fit or not. The test statistics for this test follows a chi-square distribution, and the p-value will denote the significance of the model. If the p-value exceeds the significance level, it will lead to a non-significant result which means the regression model fits the data well.

Next, the logistic regression model's predictive efficiency is used to assess how good it is to predict a new set of unknown variables. For this study, sensitivity, specificity and accuracy were utilized to predict the model's efficiency. It is a good predictive efficiency model if the percentages are more than 60%.

The last criterion is Cox and Snell R^2, also recognised as the generalised R^2 used to imply the percentage of variance explained by the logistic regression model and determine whether the model is fit or not. This test usually utilized when the outcome variable comes in a categorical variable. The range for the Cox and Snell R^2 value is between zero to one. Thus, the higher the value obtained, the better the model fits.

4. RESULT AND DISCUSSION

This section covers the results and discussion on the findings of the research. The collected dataset was used to analyse for us to obtain the best model that suits this study by going through the logistic regression model.

4.1 Assumption checking of logistic regression model

Box-Tidwell test is used to evaluate the linearity between the continuous independent variables which are age and body mass index, and the log-odds of dependent variable which is the presence of diabetes. The outcome is in Table 1.

Table 1. Box-Tidwell Test of Linearity

Variable	p-Value
Age	0.416
Body Mass Index	0.075

Table 1 showed the significance level of continuous variables, age and body mass index. Both age and body mass index variables have p-values that are more than 0.05, illustrates that both variables have a linear relationship towards the presence of diabetes. Next, the existence of multicollinearity can be detected through Variance Inflation Factor (VIF) and Tolerance values as shown in Table 2. Based on the table, all variables seem to have tolerance that is above 0.1. Also, the VIF values for all variables did not exceed 10, which means there were no multicollinearity presence.

Table 2. Box-Tidwell Test of Linearity

Variable	Tolerance	VIF
Age	0.587	1.702
Body Mass Index	0.758	1.319
High Blood Pressure	0.316	3.162
Exercise	0.299	3.340
Smoking	0.704	1.421

https://doi.org/10.24191/jcrinn.v11i2.558

4.2 Fitting logistic regression model

The first objective of this study was to identify the factors that contribute to diabetes among Malaysians. To achieve this objective, logistic regression test was employed. Therefore, the finding for this test is shown in Table 3.

Table 3. Model coefficient

Variable	Estimated Coefficient	Standard Error	p-Value
Constant	-36.7761	7.5113	0.0000
Age	0.1056	0.0455	0.0201
Body Mass Index	1.1235	0.2258	0.0000
Hypertension	-0.6897	1.1682	0.5549
Exercise	-1.3672	0.5572	0.0141
Smoking	3.8032	1.1826	0.0013

It can be determined that the risk factors contributing to diabetes were age, body mass index, exercise and smoking since all the p-values of these variables are less than 0.05.

Following is the estimated logistic regression model for the full model obtained:

$$\ln\left(\frac{p}{1-p}\right) = -36.7761 + 0.1056(Age) + 1.1235(Body\ Mass\ Index) - 0.6897(Hypertension) - 1.3672(Exercise) + 3.8032(Smoking)$$

4.3 Odd ratio

Odds ratio helps in terms of gaining better understanding on the relationship between the risk factors and the presence of diabetes. Hence, the meaning behind the odds ratio for each significant factor is as follows.

Table 4. Odd Ratio of the Estimates

Variable	Odd Ratio
Age	1.1015
Body Mass Index	3.0907
Exercise	0.3007
Smoking	44.9327

Based on the Table 4, older people have 1.10 times more chance of developing diabetes compared to youngsters. Additionally, an individual who has high body mass index has 3.09 times more chance of getting diabetes compared to people who has normal body mass index. Next, when people exercise, the odd of getting diabetes is decreased by 30.07%. This means, people who do not exercise tend to develop diabetes. Besides, a person who is smoking is 44.93 times more likely to get diabetes compared to a person who is not smoking.

4.4 Model evaluation

Omnibus Test is measured to determine whether there is a relationship between all variables involved in the model or not. Based on the Table 5, the p-value associated with Omnibus Test obtained was 0.0000 which is less than alpha value, 0.05. This suggests that there is enough evidence to reject the null hypothesis, and it can be concluded that there is a relationship between all the variables involved in both models. Next, the Hosmer-Lemeshow test was used to see whether a logistic regression model is a good fit. From the

https://doi.org/10.24191/jcrinn.v11i2.558

same table, the p-value associated with Hosmer and Lemeshow Test was 0.8799, which are extremely larger than the 5% level of significant. Hence, it reveals that the regression model fits the data very well.

Table 5. Omnibus Test and Hosmer Lemeshow Test

Model Evaluation	p-Value
Omnibus Test Model	0.0000
Hosmer and Lemeshow	0.8799

Cox and Snell R^2 was measured to indicate how much the total variation in presence of diabetes is explained by the factors. The findings in Table 6 revealed that about 32.84% of the total variation in the presence of diabetes is explained by age, body mass index, blood pressure, physical exercise, and smoking. Meanwhile, the adjusted value, Nagelkerke R^2, which helps in a better interpretation, is 0.7877. It indicates that approximately 78.77% of the total variation in the presence of diabetes is explained by age, body mass index, blood pressure, physical exercise, and smoking after considering the sample size and the number of independent variables.

Table 6. Summary of R^2 Values

R^2	Value
Cox & Snell	0.3284
Nagelkerke	0.7877

The predictive efficiency of logistic regression is utilized to measure how well the model predicts a new set of unknown variables. It includes sensitivity, specificity, and accuracy. Table 7 shows the output of the predictive efficiency of the model. The 0.9886 sensitivity value indicates that the correctly predict that several factors lead to diabetes by 98.86%. Next, the specificity value is 0.8621, which means that the model predicts respectable factors that do not lead to diabetes accurately by 86.21%. In general, the model achieves high accuracy, which is 97.89%.

Table 7. Classification Table

Predictive Efficiency	Value
Sensitivity	0.9886
Specificity	0.8621
Accuracy	0.9789

5. RESULT AND DISCUSSION

This study aimed to identify the factors that contribute to diabetes among Malaysians. This study showed that age, body mass index, physical exercise and smoking were statistically significant to the presence of diabetes. On the other hand, blood pressure did not show a correlation with the presence of the diabetes as described by prior research. The logistic regression model was found to be statistically significant and well-fitted to the data, as supported by the Omnibus Test and Hosmer–Lemeshow Test results. The model also demonstrated strong explanatory power, with Nagelkerke R^2 showing that a large proportion of variation in diabetes presence is explained by the selected factors. Additionally, the model achieved high predictive performance with strong sensitivity, specificity, and overall accuracy. Based on the findings of this study, several recommendations can be taken into consideration in order to provide more reliable and trustworthy research and serve as guidance for future research in this field. From the analysis conducted, the variable blood pressure was found to be not statistically significant to the presence of diabetes. Therefore, a large sample size that includes a diverse population is recommended in future research to investigate the relationship between these two variables in more depth. By using a larger sample size, the role of blood

https://doi.org/10.24191/jcrinn.v11i2.558

pressure towards diabetes will be more apparent. Apart from that, this study was utilizing convenience sampling as the sampling technique. Convenience sampling may limit the generalizability of the findings. Thus, the recommended sampling technique is cluster sampling. The reason behind this is if the sample size is large and includes various areas or has different diabetes risk factors, cluster sampling is very beneficial. By dividing the population into several groups according to location, health profile, or socioeconomic status, cluster sampling can provide many samples at different levels that can increase understanding of risk factors that contribute to diabetes through various groups. Simply put, using cluster sampling will increase the strength of the model in making a prediction of diabetes presence.

6. ACKNOWLEDGEMENTS/FUNDINGS

The researcher would like to express their sincere gratitude to all who contributed significantly to this project. First and foremost, to the respondents who dedicated their time to complete the survey. Finally, special thanks to the management of Universiti Teknologi MARA Seremban Branch for the invaluable assistance and support in facilitating the collaboration essential for this research project.

7. CONFLICT OF INTEREST STATEMENT

The author declare that they have no conflicts of interest to disclosure.

8. AUTHORS' CONTRIBUTIONS

All authors contributed to the development and completion of this study. The first author was responsible for conceptualising the study, developing the methodology, conducting the data analysis, interpreting the findings, and drafting the manuscript. The second author contributed to the research design, methodology, validation of the analysis, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

9. REFERENCES

Abela, A. G., & Fava, S. (2021). Why is the incidence of Type 1 diabetes increasing? *Current Diabetes Reviews*, *17*(8), 22–34. https://doi.org/10.2174/1573399817666210503133747.

Al Bairmani, Z. A. A., & Ismael, A. A. (2021). Using logistic regression model to study the most important factors which affects diabetes for the elderly in the city of Hilla/2019. *Journal of Physics: Conference Series*, *1818*(1), 012016. https://doi.org/10.1088/1742-6596/1818/1/012016.

Amanat, S., Ghahri, S., Dianatinasab, A., Fararouei, M., & Dianatinasab, M. (2020). Exercise and Type 2 diabetes. *Physical Exercise for Human Health*, 91–105. https://doi.org/10.1007/978-981-15-1792-1_6.

American Diabetes Association. (2024). Understanding Type 1 diabetes. https://diabetes.org/about-diabetes/type-1.

Burton, D. G., & Faragher, R. G. (2018). Obesity and Type-2 diabetes as inducers of premature cellular senescence and ageing. *Biogerontology*, *19*(6), 447–459. https://doi.org/10.1007/s10522-018-9763-7.

Di Micco, R., Krizhanovsky, V., Baker, D., & d'Adda di Fagagna, F. (2021). Cellular senescence in ageing:

From mechanisms to therapeutic opportunities. *Nature reviews Molecular cell biology*, *22*(2), 75–95. https://doi.org/10.1038/s41580-020-00314-w.

Durlach, V., Verge`s, B., Al-Salameh, A., Bahougne, T., Benzerouk, F., Berlin, I., Clair, C., Mansourati, J., Rouland, A., Thomas, D., et al. (2022). Smoking and diabetes interplay: A comprehensive review and joint statement. Diabetes & metabolism, *48*(6), 101370. https://doi.org/10.1016/j.diabet.2022.101370.

Dwivedi, M., & Pandey, A. R. (2020). Diabetes mellitus and its treatment: An overview. *J. Adv. Pharmacol*, *1*(1), 48–58.

Ganasegeran, K., Hor, C. P., Jamil, M. F. A., Loh, H. C., Noor, J. M., Hamid, N. A., Suppiah, P. D., Abdul Manaf, M. R., Ch'ng, A. S. H., & Looi, I. (2020). A systematic review of the economic burden of type 2 diabetes in Malaysia. *International Journal of Environmental Research and Public Health*, *17*(16), 5723. https://doi.org/10.3390/ijerph17165723.

Griffith, C. X. (2024). *Age differences of behavioral activation for comorbid depression and type 2 diabetes in low-income adults* [Master's thesis, University of Colorado]. Colorado Springs.

Gupta, S., & Bansal, S. (2020). Does a rise in BMI cause an increased risk of diabetes?: Evidence from India. *PloS one*, *15*(4), e0229716. https://doi.org/10.1371/journal.pone.0229716.

Izzo, A., Massimino, E., Riccardi, G., & Della Pepa, G. (2021). A narrative review on sarcopenia in type 2 diabetes mellitus: Prevalence and associated factors. *Nutrients*, *13*(1), 183. https://doi.org/10.3390/nu13010183.

Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health*, *18*(14), 7346. https://doi.org/10.3390/ijerph18147346.

Klein, S., Gastaldelli, A., Yki-Ja¨rvinen, H., & Scherer, P. E. (2022). Why does obesity cause diabetes? *Cell Metabolism*, *34*(1), 11–20. https://doi.org/10.1016/j.cmet.2021.12.012.

Kyung, Y., Park, Y. S., Jin, M. H., & Lee, H. J. (2024). Variability in the association of smoking status with the prevalence of diabetes mellitus in the Korean population according to different definitions of smoking status: Analysis based on the Korea National Health and nutrition examination survey (2014-2020). *International Journal of Environmental Health Research*, 1–14. https://doi.org/10.1080/09603123.2023.2300036.

Libianto, R., Batu, D., MacIsaac, R. J., Cooper, M. E., & Ekinci, E. I. (2018). Pathophysiological links between diabetes and blood pressure. *Canadian Journal of Cardiology*, *34*(5), 585–594. https://doi.org/10.1016/j.cjca.2018.01.010.

Lin, W.-Y., Liu, Y.-L., Yang, A. C., Tsai, S.-J., & Kuo, P.-H. (2020). Active cigarette smoking is associated with an exacerbation of genetic susceptibility to diabetes. *Diabetes*, *69*(12), 2819–2829. https://doi.org/10.2337/db20-0156.

Lu, X., & Zhao, C. (2020). Exercise and type 1 diabetes. *Physical Exercise for Human Health*, 107–121. https://doi.org/10.1007/978-981-15-1792-1_7.

Nazarzadeh, M., Bidel, Z., Canoy, D., Copland, E., Wamil, M., Majert, J., Byrne, K. S., Sundstro¨m, J., Teo, K., Davis, B. R., et al. (2021). Blood pressure lowering and risk of new-onset type 2 diabetes: An individual participant data meta-analysis. *The Lancet*, *398*(10313), 1803–1810.

https://doi.org/10.24191/jcrinn.v11i2.558

https://doi.org/10.1016/S0140-6736(21)01920-6.

Rajendra, P., & Latifi, S. (2021). Prediction of diabetes using logistic regression and ensemble techniques. *Computer Methods and Programs in Biomedicine Update*, 1, 100032. https://doi.org/10.1016/j.cmpbup.2021.100032.

Schutte, A. E., Kollias, A., & Stergiou, G. S. (2022). Blood pressure and its variability: Classic and novel measurement techniques. *Nature Reviews Cardiology*, *19*(10), 643–654. https://doi.org/10.1038/s41569-022-00690-0.

Soltani, S., Abdollahi, S., Aune, D., & Jayedi, A. (2021). Body mass index and cancer risk in patients with type 2 diabetes: A dose–response meta-analysis of cohort studies. *Scientific reports*, *11*(1), 2479. https://doi.org/10.1038/s41598-021-81671-0.

Taherdoost, H. (2017). Determining sample size: How to calculate survey sample size. *International Journal of Economics and Management Systems*, *2*. https://ssrn.com/abstract=3224205.

Talukder, A., & Hossain, M. Z. (2020). Prevalence of diabetes mellitus and its associated factors in Bangladesh: Application of two-level logistic regression model. *Scientific Reports*, *10*(1), 10237. https://doi.org/10.1038/s41598-020-66084-9.

Teufel, F., Seiglie, J. A., Geldsetzer, P., Theilmann, M., Marcus, M. E., Ebert, C., Arboleda, W. A. L., Agoudavi, K., Andall-Brereton, G., Aryal, K. K., Bicaba, B. W., Brian, G., Bovet, P., Dorobantu, M., Gurung, M. S., Guwatudde, D., Houehanou, C., Houinato, D., Jorgensen, J. M. A., Kagaruki, G. B., … Manne-Goehler, J. (2021). Body-mass index and diabetes risk in 57 low-income and middle-income countries: A cross-sectional study of nationally representative, individual-level data in 685 616 adults. *The Lancet*, *398*(10296), 238–248. https://doi.org/10.1016/S0140-6736(21)00844-8.

Thompson, W. R., Sallis, R., Joy, E., Jaworski, C. A., Stuhr, R. M., & Trilk, J. L. (2020). Exercise is medicine. *American Journal of Lifestyle Medicine*, *14*(5), 511–523. https://doi.org/10.1177/1559827620912192.

WebMD. (Accessed 2024). *The diabetes-old age link*. https://www.webmd.com/diabetes/diabetes-link-age.

World Health Organization. (2024). *Diabetes*. https://www.who. int/health-topics/diabetes.

Yang, Y., Peng, N., Chen, G., Wan, Q., Yan, L., Wang, G., Qin, Y., Luo, Z., Tang, X., Huo, Y., et al. (2022). Interaction between smoking and diabetes in relation to subsequent risk of cardiovascular events. *Cardiovascular Diabetology*, *21*(1), 14. https://doi.org/10.1186/s12933-022-01447-2.

Zhou, B., Perel, P., Mensah, G. A., & Ezzati, M. (2021). Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, *18*(11), 785–802. https://doi.org/10.1038/s41569-021-00559-8.

https://doi.org/10.24191/jcrinn.v11i2.558

https://doi.org/10.24191/jcrinn.v11i2.558

https://doi.org/10.24191/jcrinn.v11i2.558