MATEMATIKA, MJIAM, Volume 42, Number 2, 173–189

Logistic Regression Modelling of Complete Blood Count Parameters for Predicting Diabetes Mellitus among Sabahan Women

[1,2]Agnes Ayang Kenyang, [3]Nurliyana Juhan*, [4]Yong Zulina Zubairi, [5]Nornazirah Azizan, [1]Ho Chong Mun and [6]Abu Sayed Md. Al-Mamun

[1]Faculty of Science and Natural Resources, Universiti Malaysia Sabah, 88400, Kota Kinabalu, Sabah, Malaysia.
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 94300, Kota Samarahan, Sarawak, Malaysia.
[3]Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, 88400, Kota Kinabalu, Sabah, Malaysia.
[4]Institute of Advanced Studies, Universiti Malaya, 50603, Kuala Lumpur, Malaysia.
[5]Department of Pathology and Microbiology, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, 88400, Kota Kinabalu, Sabah, Malaysia.
[6]Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh.
*Corresponding author: liyana87@ums.edu.my

Article history
Received: 25 April 2025
Received in revised form: 27 June 2025
Accepted: 18 August 2025
Published on line: 20 July 2026

Abstract According to the National Diabetes Registry, 1.6 million Malaysians have diabetes, with women having a higher prevalence than men (57.1% versus 42.9%). These figures highlight the necessity of customized diabetes care strategies that consider gender-specific characteristics. Therefore, this study investigates the relationship between diabetes in females and complete blood count (CBC) parameters. A dataset containing the CBC test results of 537 female patients obtained from the Clinical Laboratory at the Faculty of Medicine & Health Science at Universiti Malaysia Sabah was analyzed. The univariate analysis showed all variables were significant predictors of diabetes except for ethnicity. At the multivariate analysis, the following parameters were identified as significant: age (OR = 2.77, 95% CI: 1.402–5.486, p = 0.003), hemoglobin (OR = 22.56, 95% CI: 10.556–48.178, p = 0.000), MCV (OR = 26.20, 95% CI: 12.23–56.14, p = 0.000), MCH (OR = 2.29, 95% CI: 1.17–4.51, p = 0.016), MCHC (OR = 2.81, 95% CI: 1.40–5.65, p = 0.004), HCT (OR = 3.47, 95% CI: 1.72–6.70, p = 0.001), WBC (OR = 4.72, 95% CI: 2.39–9.32, p = 0.000), and platelets (OR = 7.95, 95% CI: 3.25–19.46, p = 0.000). The Hosmer-Lemeshow goodness-of-fit test indicates that the model aligns well, accurately reflects the data and satisfies the assumptions of logistic regression. When used and interpreted appropriately, these CBC parameters can be useful indicators to support decision-making for the monitoring and treatment of diabetes mellitus-related issues in females.

Keywords Diabetes mellitus; Complete blood count; Binary logistic regression; Gender disparities; Sabahan females.

Mathematics Subject Classification 62J12, 62P10, 92C50, 62F10, 92D20.

1 Introduction

Diabetes mellitus (DM) is one of the major public health concerns worldwide. Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces [1]. According to the International Diabetes Federation 10th Edition, the number of diabetic patients from 20 to 79 years old is increasing from 151 million (in 2000) to 537 million (in 2021) globally. It is expected that this number will rise to 643 million in the year 2030 and 1.5 million deaths are directly attributed to diabetes each year [2]. Patients with diabetes face a high risk of developing other chronic diseases that shorten life expectancy, lower the quality of life and increase medical care costs [3,4]. Severe complications included nephropathy, retinopathy, neuropathy and cardiovascular diseases [5,6]. Several studies revealed that diabetes was caused by a combination of genetics and unhealthy lifestyle habits [6-8].

The Ministry of Health (MOH) Malaysia established the National Diabetes Registry in 2009 to track diabetes control and clinical outcomes of patients managed at MOH clinics and hospitals to better understand the patterns of disease and clinical management of patients managed within the MOH to develop national policies for the prevention, treatment and cure of diabetes [9]. In the latest National Diabetes Registry Report in 2020, there were 1.7 million patients in Malaysia enrolled in the National Diabetes Registry with 42.98% in males and 57.02% in females respectively. There is an alarming increase from 11.2% in 2011 to 18.3% in 2019 with a 68.3% increase in the prevalence of diabetes in Malaysia as reported by the National Institute of Health [10]. This has raised awareness of the need to detect individuals at risk of diabetes and to consider screening and intervention policies. Strategies and early interventions to treat diabetes to prevent the development of severe diabetes and its complications are crucial, particularly in Malaysia.

Complete blood count (CBC) is a basic blood test conducted to evaluate the overall health status of an individual. Besides being inexpensive, this test is useful for predicting, diagnosing, monitoring and evaluating disease treatments and routine healthcare [11,12]. Typically, the common parameters in CBC include white blood cells (WBC), WBC differential count, red blood cells (RBC), RBC indices, platelets, hematocrit and hemoglobin [12-17]. Several studies have been conducted globally to investigate the potential of CBC parameters as valuable tools for diagnosing anemia, leukemia, infectious diseases, cardiovascular, metabolic diseases and inflammatory scores due to its diagnostic capabilities [13,18-20]. In the context of COVID-19, Rahman et al., [21] developed an early scoring system (web application) for predicting disease severity in COVID-19 using CBC. Results from past studies have proven that CBC is an essential diagnostic tool that provides valuable information for a wide range of medical conditions including diabetes mellitus. The severity of diabetic patients may be lessened by considering all factors in early diabetes mellitus prognosis utilizing the results of the CBC test.

In the medical field, statistical models are crucial for clinicians and healthcare services to analyze complex data, predict clinical outcomes and assist in decision-making [22,23]. Logistic regression is widely used in medical research to predict a dependent variable with binary outcomes data (e.g., diabetic or nondiabetic) [24,25]. Several studies showed that logistic regression was as good as machine learning for predicting major chronic diseases [26]. A study conducted by Bustan & Poerwanto [27] showed the results obtained from binary logistic regression showed a significant relationship between the position and status of the breast cancer tumor with the likelihood of metastasis. Similarly, Anshori et al., [28] utilized a similar logistic regression model to predict heart disease with high accuracy. In their research, Sun et al., [29] developed an early diabetic kidney disease risk prediction model which demonstrated robust predictive capabilities. By utilizing the logistic regression model, researchers can derive valuable insights that enhance decision-making and overall patient outcomes.

The current trend of research on diabetes in Malaysia primarily focuses on the prevalence and risk factors [30-34], diabetes literacy and intervention [35-39] as well as managing financial burden [3,40]. There was a lack of research findings that predict diabetes using CBC data, particularly in Malaysia. In response to this identified gap, our research aims to utilize CBC and incorporate a logistic regression for gender-specific diabetes prevention and management. Addressing this gap is crucial as CBC has the potential to be a predictive tool and can offer valuable insights on how to improve diabetes outcomes and initiate proper treatment, especially among females in Sabah, Malaysia. This corresponds with the sustainable development goals (SDGs) of reducing gender disparities in health.

2 Research Methodology

2.1 Source of Data

The study analyses utilized secondary data from the clinical laboratory at Universiti Malaysia Sabah. In this study, the unit of analysis was the female patients who were either diabetic or non-diabetic from 2018 to 2022, with a total sample size of 537 patients.

2.2 Ethical Considerations

The study received ethical approval from the Medical Research Ethics Committee, Universiti Malaysia Sabah. Confidentiality and anonymity were maintained.

2.3 Logistic Regression and Odds Ratio (OR)

Data were cleaned, coded and entered, and all statistical analyses were performed using SPSS Version 28 and R software Version 4.4.1. The dependent variable was the patient status, which was categorized as diabetic and non-diabetic. The independent variables include demographic factors (age and ethnicity) and selected CBC parameters (hemoglobin, white blood cells, red blood cells, hematocrit, platelets, mean corpuscular hemoglobin concentration, mean corpuscular volume and mean corpuscular hemoglobin). Similarly to the dependent variable, the independent variables were divided into categorical data. The age was categorized into groups of < 60 years old and ≥ 60 years old. The ethnicity was divided into two categories, which are Malay and non-Malay. Meanwhile, the CBC parameters consisted of two categories, normal and abnormal. The CBC parameters were determined based on the data collected. If the value is within the normal range, it is classified as normal; conversely, if it is outside the normal range, it is classified as abnormal.

2.4 Model Selection Criteria

A binary logistic regression model was used to predict diabetes because of the nature of the dependent variable. The binary logistic regression model is written as Eq. (1).

$$p(x) = \frac{e^{(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}{1+e^{(\beta_0+\beta_1 x_1+\cdots+\beta_j x_j)}}. \tag{1}$$

where $p(x)$ indicates the probability of patients having diabetes and β_i are the regression coefficients and the x_i are independent variables.

For parameter estimation, Eq. (1) was transformed through logit transformation to Eq. (2).

$$y(x) = \ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n. \tag{2}$$

Probability is the ratio between the number of events favorable to some outcome and the total number of events. On the other hand, odds are the ratio between probabilities. In logistic regression, odds ratio can be conveniently calculated by exponentiated logistic regression slope coefficient ($\exp(\beta_i)$). The odds ratio indicates how much the odds of a particular outcome change for a 1-unit increase in the independent variable versus a reference group [24]. An OR=1 means that the odds of the event in one group are the same as the odds of the event in the reference group [25]. An OR>1 means that one group has a higher odd of having the event compared with the reference group [25]. Finally, an OR<1 means that one group has lower odds of having an event compared with the reference group [25].

2.5 Model Selection Criteria

At the initial stage of the analysis, all variables were tested using a collinearity test to verify the absence of collinearity among the variables. Next, univariate logistic regression using a chi-square test was conducted to evaluate the relationship between each of the independent variables and the patient status (diabetic or non-diabetic). The variables that showed significance (p-value ¡ 0.05) were included in the multivariate logistic regression analysis [41]. In the final multivariate logistic regression model, a variable was considered statistically significant if the p-value < 0.05 [41].

2.6 Model Validation and Performance

After the multivariate logistic regression model was fitted, the overall goodness of fit of the model against actual outcomes was tested using Hosmer-Lemeshow. The model is said to be better only if it exhibits p-value > 0.05 [42]. The performance of the model was shown by the receiver operating characteristic (ROC) curve and summarized by the area under the curve (AUC). Greater AUC values indicate better test performance, with AUC values that can range from 0.5 (no diagnostic ability) to 1.0 (perfect diagnostic ability) [43].

3 Results

3.1 Collinearity Statistics

The results of the collinearity test of all independent variables in analyzing the patient status are displayed in Table 1. Collinearity test results showed that there was no collinearity between the patient status and the independent variables.

Table 1: Results for the Collinearity Statistics for the Patients ($n = 537$)

Variable	**Tolerance**	**Variance Inflation Factor (VIF)**
Age	0.980	1.020
Ethnicity	0.987	1.013
Hemoglobin (Hb)	0.880	1.137
White blood cells (WBC)	0.934	1.070
Red blood cells (RBC)	0.927	1.078
Hematocrit (HCT)	0.927	1.078
Platelets	0.991	1.009
Mean corpuscular hemoglobin concentration (MCHC)	0.956	1.046
Mean corpuscular volume (MCV)	0.871	1.148
Mean corpuscular hemoglobin (MCH)	0.910	1.099

The tolerance value of all independent variables was greater than 0.10, while the VIF value for all independent variables was less than 2.00. Based on Thapa *et al.*, [44] there was no evidence of multicollinearity in the regression model if the VIF < 2.00.

3.2 Descriptive Results and Univariate Logistic Regression Analysis

Table 2 presents the descriptive analysis of the independent variables. Based on Table 2, majority of the female patients under the age of 60 were non-diabetic (51.49%). Regardless of age or patient status, Malay patients made up the majority of both diabetic (58.65%) and non-diabetic (62.13%) patients. All female patients with diabetes exhibited abnormal values in hemoglobin (78.95%), WBC (71.43%), RBC (77.44%), HCT (77.44%), platelets (85.71%), MCHC (76.69%), MCV (78.95%) and MCH (52.63%) as stated in Table 2. For non-diabetic patients, they exhibited normal values of hemoglobin (79.21%), WBC (59.90%), MCV (80.94%) and MCH (73.52%) while abnormal values of RBC (61.14%), HCT (51.73%), platelets (68.81%) and MCHC (50.49%).

The univariate logistic regression analysis in Table 2 showed ten variables were tested to find the one-to-one relationship of the independent variable with patient status. Out of ten variables, ethnicity did not differ significantly (p-value > 0.05) from the patient status. The rest of the variables were believed to have a significant relationship with the patient's status. Therefore, ethnicity would be excluded in the following multivariate logistic regression analysis.

Table 2: Descriptive Statistics Results ($n = 537$)

Variable	Patient status (Dependent variable)				p-value
	Diabetic ($n = 133$)		Non-diabetic ($n = 404$)		
	n	%	n	%	
Age					0.001
• $<$ 60 years old	46	34.59	208	51.49	
• $\geq$ 60 years old	87	65.41	196	48.52	
Ethnicity					0.475
• Malay	78	58.65	251	62.13	
• Non-Malay	55	41.35	153	37.87	
Hemoglobin (Hb)					0.000
• Normal	28	21.05	320	79.21	
• Abnormal	105	78.95	84	20.79	
White blood cells (WBC)					0.000
• Normal	38	28.57	242	59.90	
• Abnormal	95	71.43	162	40.10	
Red blood cells (RBC)					0.000
• Normal	30	22.56	157	38.86	
• Abnormal	103	77.44	247	61.14	
Hematocrit (HCT)					0.000
• Normal	30	22.56	195	48.27	
• Abnormal	103	77.44	209	51.73	
Platelets					0.000
• Normal	19	14.29	126	31.19	
• Abnormal	114	85.71	278	68.81	
Mean corpuscular hemoglobin concentration (MCHC)					0.000
• Normal	31	23.31	200	49.51	
• Abnormal	102	76.69	204	50.49	
Mean corpuscular volume (MCV)					0.000
• Normal	28	21.05	327	80.94	
• Abnormal	105	78.95	77	19.06	
Mean corpuscular hemoglobin (MCH)					0.000
• Normal	63	47.37	297	73.52	
• Abnormal	70	52.63	107	26.48	

3.3 Multivariate Logistic Regression Analysis

Table 3 shows the result of the preliminary multivariate logistic regression to identify the predictors of patient status. The preliminary multivariate logistic regression analysis identified eight variables that serve as predictors of the patient status. These variables were age, Hb, WBC, HCT, platelets, MCHC, MCV and MCH. To discover the significant predictors, the multivariate logistic regression was conducted again without RBC (p-value>0.05). The results of the final multivariate logistic analysis were presented in Table 4.

Table 3: Preliminary Multivariate Logistic Regression Analysis of Patient Status (n = 537)

Independent variable	β_i	**Standard Error**	**p-value**	**Odds Ratio (95% Confidence Interval)**
Age	0.969	0.352	0.006	2.636 (1.321, 5.257)
Hb	3.108	0.389	0.000	22.371 (10.430, 47.983)
WBC	1.606	0.353	0.000	4.981 (2.493, 9.950)
RBC	0.734	0.381	0.054	2.084 (0.988, 4.393)
HCT	1.296	0.364	0.000	3.654 (1.790, 7.460)
Platelets	2.110	0.460	0.000	8.246 (3.350, 20.297)
MCHC	1.006	0.359	0.005	2.734 (1.352, 5.527)
MCV	3.215	0.388	0.000	24.914 (11.645, 53.304)
MCH	0.715	0.352	0.042	2.043 (1.026, 4.071)

The final multivariate logistic regression found eight variables as predictors of the status of female patients, such as age, Hb, WBC, HCT, platelets, MCHC, MCV and MCH. As the result showed in Table 4, the probability of a patient having diabetes mellitus was expressed as Eq. (3).

$$p(x) = \frac{\exp(\eta)}{1+\exp(\eta)}, \qquad \begin{aligned} \eta = \beta_0 &+ 1.02\,\text{Age} + 3.116\,\text{Hb} + 1.552\,\text{WBC} \\ &+ 1.243\,\text{HCT} + 2.073\,\text{Platelets} + 1.035\,\text{MCHC} \\ &+ 3.266\,\text{MCV} + 0.829\,\text{MCH}. \end{aligned} \tag{3}$$

Table 4: Final Multivariate Logistic Regression Analysis of Patient Status ($n = 537$)

Independent variable	β_i	**Standard Error**	**p-value**	**Odds Ratio (95% Confidence Interval)**
Age	1.020	0.348	0.003	2.773 (1.402, 5.486)
Hb	3.116	0.387	0.000	22.562 (10.556, 48.178)
WBC	1.552	0.347	0.000	4.723 (2.394, 9.318)
HCT	1.243	0.358	0.001	3.467 (1.718, 6.998)
Platelets	2.073	0.457	0.000	7.946 (3.245, 19.459)
MCHC	1.035	0.356	0.004	2.814 (1.400, 5.654)
MCV	3.266	0.389	0.000	26.198 (12.225, 56.142)
MCH	0.829	0.345	0.016	2.291 (1.165, 4.505)

For parameter estimation, a multivariate logistic regression logit model was written as Eq.(4).

$$\begin{aligned} y(x) = \beta_0 &+ 1.02\,\mathrm{Age} + 3.116\,\mathrm{Hb} + 1.552\,\mathrm{WBC} + 1.243\,\mathrm{HCT} \\ &+ 2.073\,\mathrm{Platelets} + 1.035\,\mathrm{MCHC} + 3.266\,\mathrm{MCV} + 0.829\,\mathrm{MCH}. \end{aligned} \tag{4}$$

The results of the model stated in Eq.(4) showed that the female patients having diabetes mellitus were associated with age and the presence of CBC parameters including hemoglobin, white blood cells, hematocrit, platelets, mean corpuscular hemoglobin concentration, mean corpuscular volume and mean corpuscular hemoglobin (p-value < 0.05).

3.4 Model Validation and Performance

The goodness of fit based on the Hosmer-Lemeshow statistic (p-value > 0.05) in Table 5 indicates the model adequately fits the data. We can conclude that there is no difference between the observed and predicted model.

Table 5: Goodness of Fit Test

Chi-Square	**p-value**
2.577	0.958

The performance of the model was measured using ROC curve. The area under curve in Figure 1 was 0.9548 which showed the model is accurate for modelling diabetes mellitus among female patients.

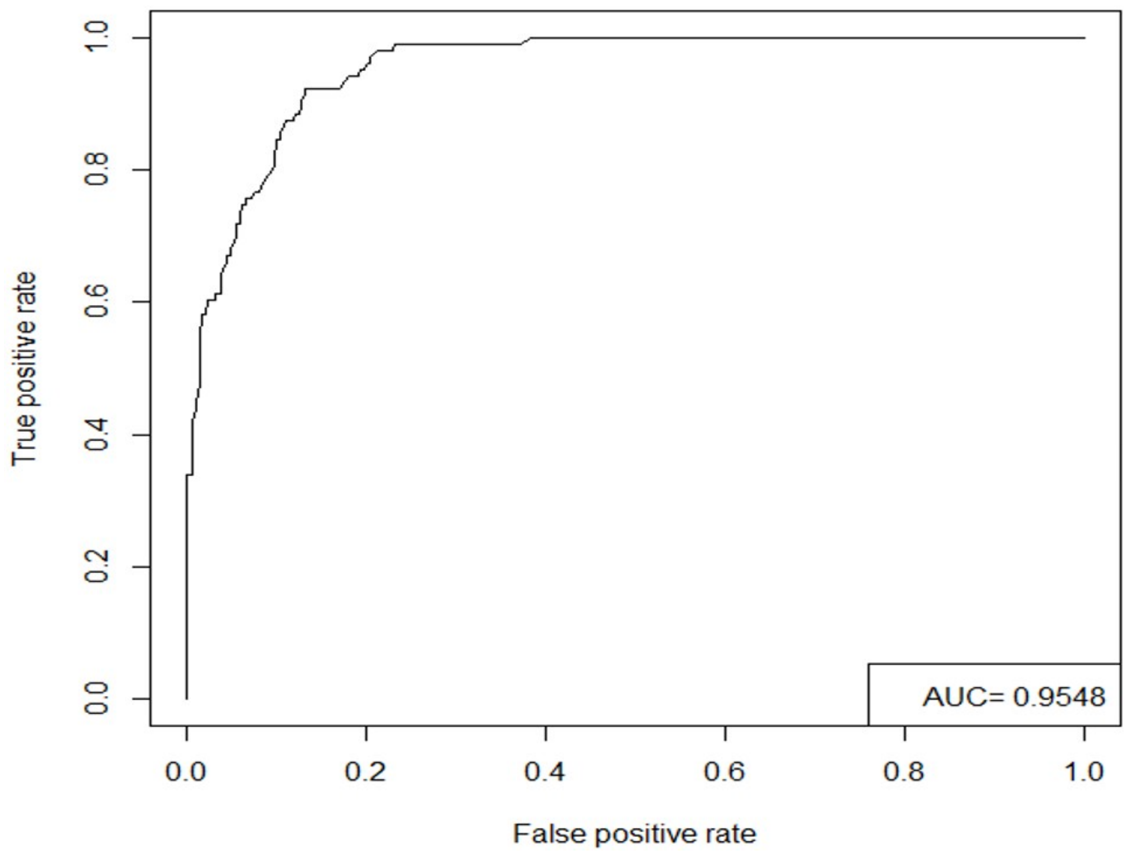

Figure 1: The area under the curve (AUC = 0.9548)

Among the CBC parameters, Hb and MCV were highly correlated with diabetes which was shown in Figure 2. A female with abnormal MCV and Hb was 26.20 times, and 22.56 times has a higher chance of having diabetes compared to those having normal MCV and Hb values. This finding supports the work of previous studies that established a connection between abnormalities in erythrocyte-related markers and diabetes. The traits revealed that MCV had a significant impact on glycated hemoglobin (HbA1c) levels, especially in relation to diabetes [15-16,45].

4 Discussions

A significant amount of recent research has been conducted on the correlation between WBC levels and diabetes. WBC levels were an essential indicator for detecting inflammatory processes, tissue damage and other degenerative disorders [46]. A diabetic patient with WBC ≥ 5.4, the chances of diabetes was 4 times more than of non-diabetics [46]. Park et al., [47] mentioned that WBC count might help to identify non-obese Korean people who were likely to develop Type 2 Diabetes Mellitus (T2DM). Another study found that a high WBC count

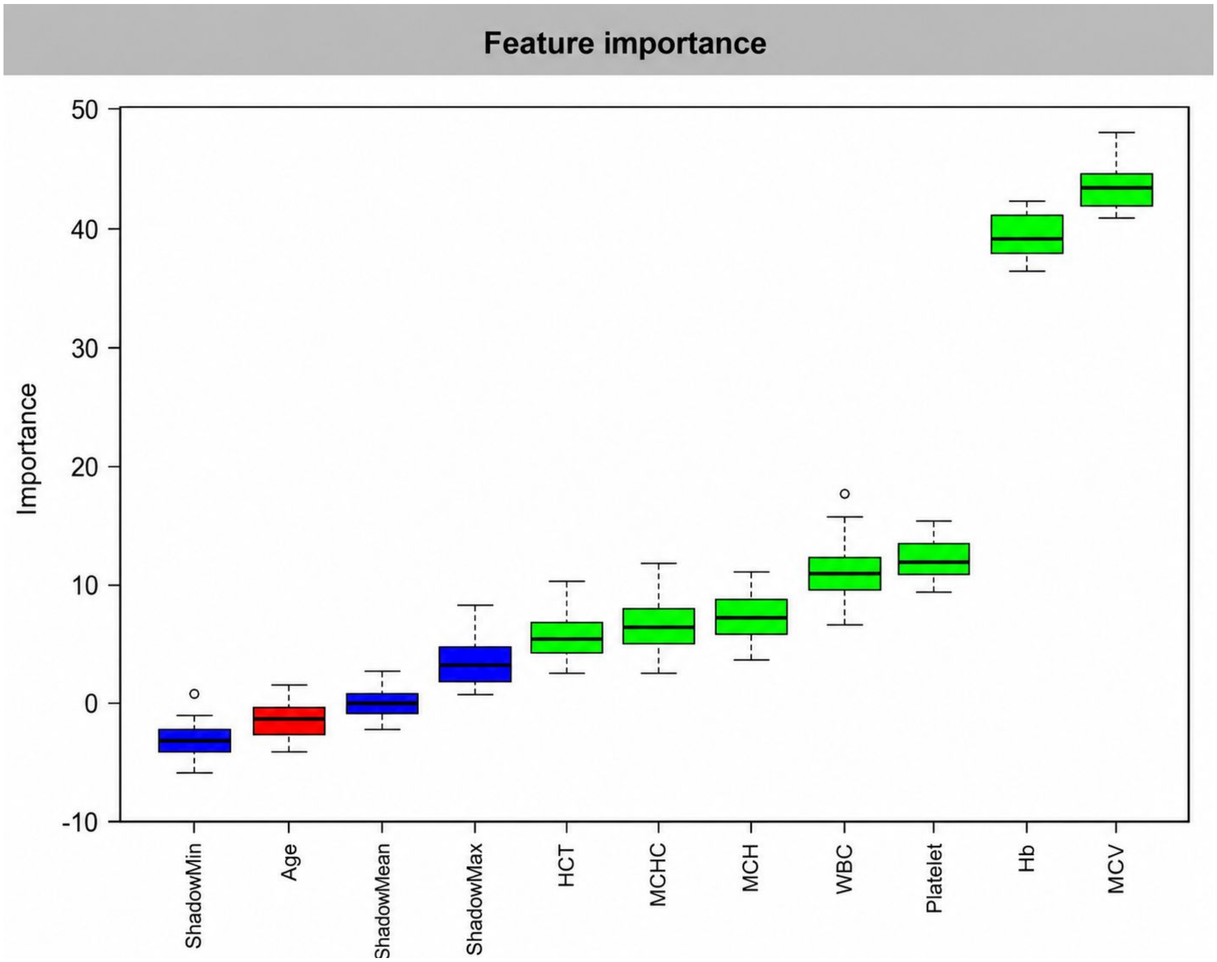

Figure 2: Feature importance. The higher the box and whisker on the chart, the more important the feature is. The green colour means the feature is important, red means it is not and blue represents variables used by Boruta to determine importance which can be ignored.

is positively related to an elevated risk of insulin resistance (IR) leading to the development of T2DM among middle-aged and older people in Taiwan [17].

Platelets have a complex function in diabetes since they contribute to both vascular problems and organ injuries caused by inflammation. Our findings indicated that a female patient with abnormal platelet levels is 7.95 times more likely to get diabetes compared to a patient with normal platelet levels. Similarly to the findings of Ilambirai et al., [48] they observed that platelet levels such as MPV and platelet distribution width (PDW) can serve as potential markers for diabetic vascular complications. Other studies have focused on the oxidative damage of platelets and the involvement of pro-oxidant enzymes in the increased activity of platelets in diabetes [49-50].

HCT, MCHC and MCH are red blood cell parameters. We also found that HCT, MCHC and MCH were positively correlated with the presence of diabetes. Similar to our results, R et al., [15] reported that diabetic patients tend to have significantly higher mean values of HCT and Hb concentration compared to non-diabetic patients. Furthermore, increased MCHC levels in diabetic patients may be linked to morphological and functional changes in red blood cells due to chronic hyperglycemia [16,46,48].

We used the female patients that were < 60 years old as the reference group [26, 51]. The value of the odds ratio was interpreted as female patients with age ≥ 60 years old were 2.77 times more likely to have diabetes than those < 60 years old (OR = 2.773, 95% CI 1.402, 5.486). An analysis using global population data revealed that the prevalence of diabetes had significantly increased in all age groups of individuals aged 65 years and older [52-53]. There was growing evidence of the association of diabetes with the older population in Malaysia. According to the results of the meta-analysis done by Akhtar et al., [32] the older population in Malaysia was more than 10 times more likely to be diagnosed with diabetes compared to the younger population. In 2019, diabetes ranked as the fourth contributor to premature mortality among older adults aged 60 and above in Malaysia, after ischemic heart disease, stroke and lower respiratory infections [54]. Given the growing number of older people in Malaysia, early and regular diabetes screening was crucial since diabetes and age were strongly associated. This would lead to elevated risks of incident heart disease, stroke, disability, cognitive impairment and mortality [55].

In this study, we examined various CBC parameters in females, some of which have not been the subject of many previous studies conducted in Malaysia, especially in Sabah. The study covered a broad age range, including adults both above and below 60 years old, and investigated the relationship among these age groups. Furthermore, our model was specifically developed for female patients, as the other available models were designed for a general population. The constraint of this study is that we did not measure HbA1c in female patients as a result of insufficient data. The resulting model applies only to Sabahan females, and it is not feasible to extrapolate the findings of this study to other regions of Malaysia. However, the results of this study can help doctors improve diabetes outcomes and initiate proper treatment, especially among females in Sabah, Malaysia using CBC parameters.

5 Conclusion

The results of our study demonstrated that the combination of CBC parameters and binary multivariate logistic regression can serve as a helpful indicator for predicting diabetes mellitus in Sabahan females. Our findings indicate a substantial correlation between age and several CBC parameters, such as Hb, WBC, HCT, platelets, MCHC, MCV, and MCH with diabetes mellitus. Among all the CBC parameters, Hb and MCV played the most crucial roles in predicting diabetes mellitus. Our research results suggest that CBC could be a valuable and cost-effective tool in the healthcare field for predicting diabetes in females from Sabah.

6 Acknowledgments

This research was supported by Malaysia Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (Project Code: FRGS/1/2023/STG06/ UMS/02/2).

References

[1] World Health Organization. *Diabetes*. https://www.who.int/health-topics/diabetes.

[2] Williams, R., Karuranga, S., Malanda, B., Saeedi, P., Basit, A., Besancon, S., Bommer, C., Estezhamati, A., Oguntova, K., Zhang, P. and Colagiuri, S. Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th edition, *Diabetes Research and Clinical Practice*, 162 (2020), 108072. DOI: https://doi.org/10.1016/j.diabres.2020.108072.

[3] Ganasegaran, 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. and Looi, I. A systematic review of the economic burden of Type 2 diabetes in Malaysia, *International Journal of Environmental Research and Public Health*, 17(16) (2020), 5723. DOI: https://doi.org/10.3390/ijerph17165723.

[4] Ganasegaran, K., Hor, C. P., Jamil, M. F. A., Suppiah, P. D., Noor, J. M., Hamid, N. A., Chuan, D. R., Manaf, M. R. A., Ch'ng, A. S. H. and Looi, I. Mapping the scientific landscape of diabetes research in Malaysia (2000–2018): A systematic scientometrics study, *International Journal of Environmental Research and Public Health*, 18(1) (2021), 318. DOI: https://doi.org/10.3390/ijerph18010318.

[5] Tomic, D., Shaw, J. E. and Magliano, D. J. The burden and risks of emerging complications of diabetes mellitus, *Nature Reviews Endocrinology*, 18(9) (2022), 525–539. DOI: https://doi.org/10.1038/s41574-022-00690-7.

[6] Bereda, G. Risk factors, complications and management of diabetes mellitus, *American Journal of Biomedical Science & Research*, 16(4) (2022), 409–412. DOI: https://doi.org/10.34297/AJBSR.2022.16.002245.

[7] Abdullah, A., Liew, S. M., Ng, C. J., Ambigapathy, S. and Paranthaman, P. V. V. Health literacy experiences of multi-ethnic patients and their health-care providers in the management of Type 2 diabetes in Malaysia: A qualitative study, *Health Expectations*, 23(5) (2020), 1166–1176. DOI: https://doi.org/10.1111/hex.13095.

[8] Akhtar, S., Nasir, J. A., Ali, A., Asghar, M., Majeed, R. and Sarwar, A. Prevalence of Type-2 diabetes and prediabetes in Malaysia: A systematic review and meta-analysis, *PLoS ONE*, 17(1) (2022), e0263139. DOI: https://doi.org/10.1371/journal.pone.0263139.

[9] Ministry of Health Malaysia. *National Diabetes Registry Report 2013–2019*, Ministry of Health Malaysia, 2020.

[10] Institute for Public Health (IPH), National Institutes of Health, Ministry of Health Malaysia. *National Health and Morbidity Survey (NHMS) 2019: Vol. I: NCDs—Non-Communicable Diseases: Risk Factors and Other Health Problems*, Ministry of Health Malaysia, 2020.

[11] Chen, D., Li, N., Zeng, S., Lv, X., Chen, L., Liu, X. and Hu, Q. Multiparameter mobile blood analysis for complex blood count using contrast-enhanced defocusing imaging and machine vision, *The Analyst*, 148(9) (2023), 2021–2034. DOI: https://doi.org/10.1039/D3AN00070B.

[12] Ahmed, M. M., Ghanavi, S. K., Jayaeged, A., Rafique, N., Hussain, W. and Khan, N. Trends of utilization of complete blood count parameters for patient management among doctors in Azad Kashmir, *Pakistan Journal of Medical Sciences*, 36(5) (2020). DOI: https://doi.org/10.12669/pjms.36.5.1885.

[13] Seo, I. and Lee, Y. Usefulness of complete blood count (CBC) to assess cardiovascular and metabolic diseases in clinical settings: A comprehensive literature review, *Biomedicines*, 10(11) (2022), 2697. DOI: https://doi.org/10.3390/biomedicines10112697.

[14] Essawi, K., Dobie, G., Shaabi, M., Hakami, W., Saboor, M., Madkhali, A. M., Hamami, A., Allallah, W., Akhter, M. S., Moharki, A. A. and Hamali, H. A. Comparative analysis of red blood cells, white blood cells, platelet count, and indices in Type 2 diabetes mellitus patients and normal controls: Association and clinical implications, *Diabetes, Metabolic Syndrome and Obesity*, 16 (2023), 3123–3132. DOI: https://doi.org/10.2147/DMSO.S422373.

[15] Rashed, E. R., Alkout, T. A., Eltomy, S. S., Etekbali, O. R. and Alkout, A. M. The effects of red blood cells parameters on HbA1c and random blood sugar levels in diabetics diagnosis, *International Journal of Diabetes and Clinical Research*, 7(3) (2020), Article 128. DOI: https://doi.org/10.23937/2377-3634/1410128.

[16] Wang, Y., Yang, P., Yan, Z., Liu, Z., Ma, Q., Zhang, Z., Wang, Y. and Su, Y. The relationship between erythrocytes and diabetes mellitus, *Journal of Diabetes Research*, (2021), 1–9. https://doi.org/10.1155/2021/6656062.

[17] Chen, J., Chen, Y., Lee, Y. and Tsou, M. The association between white blood cell count and insulin resistance in community-dwelling middle-aged and older populations in Taiwan: A community-based cross-sectional study, *Frontiers in Medicine*, 9 (2022). https://doi.org/10.3389/fmed.2022.813222.

[18] Haider, R. Z., Ujjan, I. U., Khan, N. A., Urrechaga, E. and Shamsi, T. Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias, *Diagnostics*, 12(1) (2022), 138. https://doi.org/10.3390/diagnostics12010138.

[19] Agnello, L., Giglio, R. V., Bivona, G., Scazzone, C., Gambino, C. M., Iacona, A., Ciaccio, A. M., Sasso, B. L. and Ciaccio, M. The value of a complete blood count (CBC) for sepsis diagnosis and prognosis, *Diagnostics*, 11(10) (2021), 1881. https://doi.org/10.3390/diagnostics11101881.

[20] Lin, J., Lin, J., Xie, J., Wang, J., Lu, J., Chen, Q., Cao, L., Lin, M., Tu, R., Zheng, C., Huang, C. and Li, P. Complete blood count-based inflammatory score (CBCS) is a novel prognostic marker for gastric cancer patients after curative resection, *BMC Cancer*, 20(1) (2020). https://doi.org/10.1186/s12885-019-6466-7.

[21] Rahman, T., Khandakar, A., Hoque, M. E., Ibtehaz, N., Kashem, S. B., Masud, R., Shampa, L., Islam, M. T., Maadeed, S. A., Zughaier, S. M., Badran, S., Doi, S. A. R. and Chowdhury, M. E. H. Development and validation of an early scoring system for prediction of disease severity in COVID-19 using complete blood count parameters, *IEEE Access*, 9 (2021), 120422–120441. `https://doi.org/10.1109/ACCESS.2021.3105321`.

[22] Chowdhury, M. and Turin, T. C. Variable selection strategies and its importance in clinical prediction modelling, *Family Medicine and Community Health*, 8(1) (2020), e000262. `https://doi.org/10.1136/fmch-2019-000262`.

[23] Bach, P., Wallisch, C., Klein, N., Hafermann, L., Sauerbrei, W., Steyerberg, E. W., Heinze, G. and Rauch, G. Systematic review of education and practical guidance on regression modeling for medical researchers who lack a strong statistical background: Study protocol, *PLoS One*, 15(12) (2020), e0241427. `https://doi.org/10.1371/journal.pone.0241427`.

[24] Schober, P. and Vetter, T. R. Count data in medical research: Poisson regression and negative binomial regression, *Anesthesia & Analgesia*, 132(5) (2021), 1378–1379. `https://doi.org/10.1213/ANE.0000000000005398`.

[25] Zabor, E. C., Reddy, C., Tendulkar, R. D. and Patil, S. Logistic regression in clinical studies, *International Journal of Radiation Oncology*Biology*Physics*, 112(2) (2022), 271–277. `https://doi.org/10.1016/j.ijrobp.2021.08.007`.

[26] Juhan, N., Zubairi, Y. Z., Mohd Khalid, Z. and Mahmood Zuhdi, A. S. A comparison between Bayesian and frequentist approach in the analysis of risk factors for female cardiovascular disease patients in Malaysia, *ASM Science Journal*, 10 (2020), 1–7. `https://doi.org/10.32802/asmscj.2020.sm26(1.1)`.

[27] Bustan, M. N. and Poerwanto, B. Logistic regression model of relationship between breast cancer pathology diagnosis with metastasis, *Journal of Physics: Conference Series*, 1752(1) (2021), 012026. `https://doi.org/10.1088/1742-6596/1752/1/012026`.

[28] Anshori, M. and Haris, M. S. Predicting heart disease using logistic regression, *Knowledge Engineering and Data Science*, 5(2) (2022), 188–196. `https://doi.org/10.17977/um018v5i22022p188-196`.

[29] Sun, L., Wu, Y., Hua, R. and Zou, L. Prediction models for risk of diabetic kidney disease in Chinese patients with Type 2 diabetes mellitus, *Renal Failure*, 44(1) (2022), 1455–1462. `https://doi.org/10.1080/0886022X.2022.2113797`.

[30] Bakar, F. A. A., Ismail, T. A. T., Hairon, S. M. and Yusoff, S. S. M. Risk perception survey on developing diabetes questionnaire: Translation and validation of the Malay version, *IIUM Medical Journal Malaysia*, 22(2) (2023). `https://doi.org/10.31436/iiumj.v22i2.2086`.

[31] Nasermudin, N. A., Jeffree, M. S., Kaur, N., Rahim, S. S. S. A. and Ibrahim, M. Y. Diabetic retinopathy among Type 2 diabetes mellitus patients in Sabah primary health clinics: Addressing the underlying factors, *PLoS One*, 17(1) (2022), e0261249. `https://doi.org/10.1371/journal.pone.0261249`.

[32] Akhtar, S., Nasir, J. A., Ali, A., Asghar, M., Majeed, R. and Sarwar, A. Prevalence of Type-2 diabetes and prediabetes in Malaysia: A systematic review and meta-analysis, *PLoS One*, 17(1) (2022), e0261319. https://doi.org/10.1371/journal.pone.0261319.

[33] Teng, C. L., Foo, K. Y., Ka, M. L., Shen, Y. T., Chan, X. H. and Sivalingam, N. Gestational diabetes in Malaysia: A systematic review of prevalence, risk factors and outcomes, *Sains Malaysiana*, 50(8) (2021), 2367–2377. https://doi.org/10.17576/jsm-2021-5008-19.

[34] Rahim, F. F., Abdurahman, S. A., Maideen, S. F. K. and Rashid, A. Prevalence and factors associated with prediabetes and diabetes in fishing communities in Penang, Malaysia: A cross-sectional study, *PLoS One*, 15(2) (2020), e0228570. https://doi.org/10.1371/journal.pone.0228570.

[35] Atolagbe, E. T., Siyanandy, P. and Ingle, P. V. Effectiveness of educational intervention in improving medication adherence among patients with diabetes in Klang Valley, Malaysia, *Frontiers in Clinical Diabetes and Healthcare*, 4 (2023). https://doi.org/10.3389/fcdhc.2023.1132489.

[36] Thangiah, G., Johar, H., Ismail, R., Reinghaus, U., Baminghaus, T., Thurairajasingam, S., Reidpath, D. and Su, T. T. Diabetes treatment and mental illness: A call for an integrated health care system in underserved semi-rural Malaysia, *International Journal of Environmental Research and Public Health*, 19(16) (2022), 10015. https://doi.org/10.3390/ijerph191610015.

[37] Minhah, H. S., Matt Din, H. and Raja Adnan, R. N. E. Predictors of diabetes health literacy and numeracy among urbanized older adults in Malaysia, *Malaysian Journal of Medicine and Health Sciences*, 19(5) (2023), 139–144. https://doi.org/10.47836/mjmhs.19.5.20.

[38] Khan, A. H., Iqbal, M. Z., Sulaiman, S. A. S., Ibrahim, A., Azmi, N. S. B. Y., Iqbal, M. S. and Albassam, A. A. Impact of pharmacist-led educational intervention on predictors of diabetic foot at two different hospitals of Malaysia, *Journal of Pharmacy and Bioallied Sciences*, 13(1) (2021), 108–115. https://doi.org/10.4103/jpbs.jpbs.75.20.

[39] Oluchi, S. E., Manaf, R. A., Ismail, S., Shahar, H. K., Mahmud, A. and Udeani, T. K. Health related quality of life measurements for diabetes: A systematic review, *International Journal of Environmental Research and Public Health*, 18(17) (2021), 9245. https://doi.org/10.3390/ijerph18179245.

[40] Al-Sanaani, E. A., Ismail, A., Manaf, M. R. A., Suddin, L. S., Mustafa, N., Sukor, N., Alabed, A. A. A., Alkhodary, A. A., Aljunid, S. M. Health insurance status and its determinants among patients with Type 2 diabetes mellitus in a tertiary teaching hospital in Malaysia, *PLoS One*, 17(5) (2022), e0267897. https://doi.org/10.1371/journal.pone.0267897.

[41] Hussen, H. and Alemu, Z. A. Risk of COVID-19 infection and associated factors among healthcare workers: A cross-sectional study at Eka Kotebet treatment center in Ethiopia, *International Journal of General Medicine*, 14 (2021), 1763–1772. https://doi.org/10.2147/IJGM.S301518.

[42] Fagerland, M. W. and Hosmer, D. W. A goodness-of-fit test for the proportional odds regression model, *Statistics in Medicine*, 32(13) (2012), 2235–2249. https://doi.org/10.1002/sim.5645.

[43] Elkahawy, D. M. A. S., Kiriacos, C. J. and Mansour, M. Logistic regression and other statistical tools in diagnostic biomarker studies, *Clinical and Translational Oncology*, 26(9) (2024), 2172–2180. https://doi.org/10.1007/s12094-024-03413-8.

[44] Thapa, J., Budhathoki, S. S., Niraula, S. R., Pandey, S., Thakur, N. and Pokharel, P. K. Prehypertension and its predictors among older adolescents: A cross-sectional study from Eastern Nepal, *PLoS Global Public Health*, 2(9) (2022), e0001117. https://doi.org/10.1371/journal.pgph.0001117.

[45] Sarosh, I., Ghafoor, R., Haseeb, A., Saeed, J., Shafqat, F. and Kashif, Z. Frequency of anemia and its types in Type II diabetes mellitus patients without renal insufficiency, *Pakistan Journal of Medical and Health Sciences*, 16(6) (2022), 216–218. https://doi.org/10.53350/pmjhs22166216.

[46] Mansoori, A., Sahranavard, T., Hosseini, Z., Sofjaei, S. S., Emrani, N., Nazar, E., Ghazizadeh, M., Khorasanehi, Z., Effati, S., Ghamsary, M., Ferns, G. A., Esmaili, H. and Ghayour-Mobarhan, M. Prediction of Type 2 diabetes mellitus using hematological factors based on machine learning approaches: A cohort study analysis, *Scientific Reports*, 13(1) (2023). https://doi.org/10.1038/s41598-022-27340-2.

[47] Park, J. M., Lee, H. S., Park, J., Jung, D. and Lee, J. White blood cell count as a predictor of incident Type 2 diabetes mellitus among non-obese adults: A longitudinal 10-year analysis of the Korean genome and epidemiology study, *Journal of Inflammation Research*, 14 (2021), 1235–1242. https://doi.org/10.2147/JIR.S300026.

[48] Ilambari, M. R., C. Wegerapandian, M. A., Chandrakala, K. and Rajesh, A. The role of platelet indices to predict the vascular complication in diabetic patients in South Tamil Nadu, *International Journal for Multidisciplinary Research*, 4(4) (2022), 93–103. https://doi.org/10.36948/ijfmr.2022.v04i04.909.

[49] Su, M., Fan, S., Ling, Z., Fan, X., Xia, L., Liu, Y., Li, S., Zhang, Y., Zeng, Z. and Tang, W. H. Restoring the platelet miR-223 by calpain inhibition alleviates the neointimal hyperplasia in diabetes, *Frontiers in Physiology*, 11 (2020). https://doi.org/10.3389/fphys.2020.00742.

[50] Vaidya, A. R., Wolska, N., Vara, D., Mailer, R. K., Schröder, K. and Pula, G. Diabetes and thrombosis: A central role for vascular oxidative stress, *Antioxidants*, 10(5) (2021), 706. https://doi.org/10.3390/antiox10050706.

[51] Liu, Y., Mao, B., Liang, S., Yang, J., Lu, H., Chai, Y., Wang, L., Zhang, L., Li, Q., Zhao, L., He, Y., Gu, X., Ji, X., Li, L., Jie, Z., Li, Q., Li, X., Lu, H., Zhang, W., Song, Y., Qu, J. and Xu, J. Association between age and clinical characteristics and outcomes of COVID-19, *European Respiratory Journal*, 55(5) (2020), 2001112. https://doi.org/10.1183/13993003.01112-2020.

[52] Yang, K., Yang, X., Jin, C., Ding, S., Liu, T., Ma, B., Sun, H., Zhang, J. and Li, Y. Global burden of Type 1 diabetes in adults aged 65 years and older, 1990–2019: Population based study, *BMJ*, (2024), e078432. `https://doi.org/10.1136/bmj-2023-078432`.

[53] Wong, W. J., Tan, K. M., Harrison, C., Ng, C. C., Lim, W. C. and Nguyen, T. Burden of comorbidities among older Malaysians with stroke: Associations with diabetes mellitus and frailty, *Heart, Lung and Circulation*, 33(Suppl. 4) (2024), S509. `https://doi.org/10.1016/j.hlc.2024.06.836`.

[54] Chan, Y. M., Ganapathy, S. S., Tan, L., Alias, N., Nasruddin, N. H. and Khaw, W. The burden of premature mortality among older adults: A population-based study in Malaysia, *BMC Public Health*, 22 (2022), 1181. `https://doi.org/10.1186/s12889-022-13608-9`.

[55] Gigolle, C. T., Blaum, C., Lyu, C., Ha, J., Kabeto, M. U. and Zhong, J. Associations of age at diagnosis and duration of diabetes with morbidity and mortality among older adults, *JAMA Network Open*, 5(9) (2022), e2232766. `https://doi.org/10.1001/jamanetworkopen.2022.32766`.