BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 2865-2884
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp2865-2884

REPURPOSING COMPLETE BLOOD COUNT DATA FOR MORTALITY RISK STRATIFICATION IN DIABETES: A STATEWIDE LOGISTIC REGRESSION ANALYSIS

Dg Siti Nurisya Sahirah Ag Isha [1], Nurliyana Juhan [2*], Yong Zulina Zubairi [3], Nornazirah Azizan [4], Ho Chong Mun [5], Abu Sayed Md. Al Mamun [6], Lee Qin Zhi [7]

[1,5]*Faculty of Science and Technology, Universiti Malaysia Sabah*
[2]*Preparatory Centre for Science and Technology, Universiti Malaysia Sabah*
[4]*Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah*
Jln. UMS, Kota Kinabalu, Sabah, 88400, Malaysia

[3]*Institute for Advanced Studies, University of Malaya*
Kuala Lumpur, 50603, Malaysia

[6]*Department of Statistics, University of Rajshahi*
Rajshahi, Bangladesh

[7]*Queen Elizabeth Hospital*
Locked Bag No. 2029, Kota Kinabalu, Sabah, 88586, Malaysia

*Corresponding author's e-mail: *liyana87@ums.edu.my*

Article Info

Article History:

Received: 15th September 2025
Revised: 27th March 2026
Accepted: 26th May 2026
Available online: 1st July 2026

Keywords:

Complete Blood Count (CBC);
Diabetes Mellitus (DM);
Logistic Regression;
Mortality;
Sabah.

ABSTRACT

Diabetes mellitus (DM), a disease defined by consistently high blood sugar levels, is recognized as a significant global health burden and holds the seventh position among causes of death in Malaysia. This underscores the need for affordable risk stratification. Complete blood count (CBC), a simple and widely available test, provides useful biomarkers, yet its role among Malaysian diabetic patients in Sabah remains understudied. This research aimed to examine CBC biomarkers for mortality risk stratification among diabetic patients in Sabah. This retrospective cross-sectional study utilized data from 10,672 diabetic patients retrieved from medical records at Queen Elizabeth Hospital 1, Sabah. Logistic regression (LR) was applied to estimate the probability of mortality (Alive = 0, Deceased = 1) from demographic and CBC parameters. Model performance was evaluated using calibration and discrimination metrics. Significant mortality risk factors included advanced age, inpatient status, higher neutrophil counts, platelet-to-lymphocyte ratio (PLR), red cell distribution width (RDW), neutrophil-to-lymphocyte ratio (NLR), and monocyte-to-lymphocyte ratio (MLR), as well as lower lymphocyte levels and hemoglobin. The LR model showed excellent calibration and discrimination, with non-significant Spiegelhalter Z-tests, low Brier scores, and strong performance across balanced accuracy, Matthews correlation coefficient, and F1-score. These findings highlight that CBC biomarkers can be integrated into clinical models as a low-cost approach for early detection and management of diabetes risk in resource-limited settings.

How to cite this article:

D. S. N. S. A. Isha *et al.*,"REPURPOSING COMPLETE BLOOD COUNT DATA FOR MORTALITY RISK STRATIFICATION IN DIABETES: A STATEWIDE LOGISTIC REGRESSION ANALYSIS", *BAREKENG: J. Math. & App.*, vol. 20, no. 4, pp. 2865-2882, Dec. 2026.

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

Research Article · Open Access

1. INTRODUCTION

Diabetes mellitus (DM) is defined as a chronic condition associated with consistently high blood sugar and has appeared as a significant worldwide public health challenge in recent years [1], [2]. DM is broadly divided into two major types: type 1 diabetes (T1DM) and type 2 diabetes (T2DM) [2], [3]. T1DM is an autoimmune condition characterized by immune-mediated destruction of pancreatic β-cells, leading to insulin deficiency. Conversely, T2DM arises from insulin resistance and inadequate insulin levels, in which body tissues do not respond properly to the hormone, leading to increased blood sugar levels [2], [3].

Persistent elevation of blood glucose (chronic hyperglycemia) in DM contributes to serious complications such as nephropathy, cardiovascular disease, retinopathy, and neuropathy, and progressively damages vital organs, including the kidneys, cardiovascular system, and nervous system [1], [2], [4], [5]. These processes are major drivers of the significantly increased risk of death from cardiovascular causes and from all causes combined in diabetic populations [1], [4], [5]. Globally, an estimated 537 million adults (20–79 years) live with DM, about one in ten people [6]. The trend is projected to increase to 643 million in 2030 and rise to 783 million by 2045. DM accounts for the seventh-highest cause of mortality in Malaysia, compared with its eighth position globally [7], [8].

The rising prevalence of DM and its associated mortality burden highlight the urgent need for reliable tools for risk stratification and management guidance [9]. However, current prognostic approaches in routine practice are limited by associated costs and measurement challenges [9]. Many conventional models rely on costly biomarkers or advanced imaging, which are often impractical in resource-constrained healthcare settings. This underscores the importance of developing affordable, scalable, and clinically feasible strategies for mortality risk stratification in diabetic populations [1].

Complete blood count (CBC), owing to its low cost, accessibility, minimal invasiveness, and ability to reflect underlying pathology, provides a practical alternative [1], [3], [10]. It is typically categorized into three components: platelets, red blood cells (RBC), and white blood cells (WBC) [11], [12]. Importantly, alterations in these hematological parameters are commonly observed in patients with DM, reflecting changes in cellular function, structure, and metabolism during the course of the disease [2]. Beyond its traditional diagnostic role, CBC can yield derived indices such as the monocyte-to-lymphocyte ratio (MLR), neutrophil-to-lymphocyte ratio (NLR), and platelet-to-lymphocyte ratio (PLR). These indicators provide insight into systemic inflammatory activity and have been shown to carry prognostic significance in evaluating the development of diabetes and the resulting clinical implications [3], [10].

A growing body of evidence has linked CBC parameters with mortality across a range of conditions, including sepsis in burn patients [13], coronavirus disease (COVID-19) [14], [15], and renal disease [16], [17]. Similar associations have also been reported in diverse diabetic populations, such as general cohorts [10], [18], patients with diabetic foot ulcers [19], diabetes with community-acquired pneumonia [20], cardiovascular-related mortality in diabetes [1], [9],[21], [22], diabetic kidney disease [23], cardiovascular disease with diabetes [24], and diabetes with COVID-19 [25]. However, most of these studies have been conducted outside Malaysia, and the integration of CBC-derived markers into mortality risk stratification for diabetic patients remains underexplored locally. Although these studies collectively demonstrate the prognostic relevance of CBC parameters, they predominantly focus on specific subgroups or external populations, limiting their direct applicability to Malaysian clinical settings. Furthermore, few studies have systematically translated routinely collected CBC markers into structured mortality risk stratification frameworks using real-world hospital data, particularly in resource-limited healthcare environments.

Early detection is essential in diabetes research, and logistic regression (LR) has become one of the most widely applied approaches for risk stratification [26], [27]. LR is valued for its simplicity, computational efficiency, and compatibility with standard statistical software, making it highly practical for clinical research [28], [29]. It is particularly effective for modeling binary outcomes such as survival versus death in relation to multiple risk factors, which is central to mortality stratification [30]. Given that the present study involves a large-scale, structured clinical dataset with a binary outcome and predominantly categorical predictors, LR provides a suitable and transparent modelling framework that does not require complex feature engineering or hyperparameter tuning.

Besides, evidence shows that LR often matches or outperforms more complex machine learning methods, consistently ranking among the top techniques with the added benefit of minimal training time [30], [31]. Its strong discriminatory capacity, demonstrated across metrics such as area under the precision–recall

curve (AUPRC), area under the receiver operating characteristic curve (AUROC), specificity, accuracy, sensitivity, and precision, underscores LR as a robust and reliable tool for mortality risk stratification in diabetes [32]-[38]. Moreover, LR yields interpretable odds ratios, facilitating clinical applicability and implementation in resource-limited hospital settings, which aligns with the objective of this study.

Despite the clinical value of CBC parameters, research exploring their role in mortality risk assessment among diabetic patients in Malaysia, particularly in Sabah, remains limited. To address this gap, the present study repurposed routinely collected hospital-based real-world CBC data for mortality risk stratification in patients with diabetes. To our knowledge, this is among the first studies in Sabah, and more broadly in Malaysia, to utilize large-scale real-world CBC data for mortality risk stratification in patients with diabetes. While the analytical approach applied is well established, the novelty of this study lies in its implementation at scale within a real-world hospital setting, its translational focus, and its relevance to resource-limited healthcare systems. Specifically, this study leverages data from Queen Elizabeth Hospital 1, the main tertiary referral center serving 32 hospitals across Sabah, to demonstrate that low-cost, universally available hematological parameters can be used for system-level mortality risk stratification. This provides evidence that is directly applicable to healthcare systems where access to advanced biomarkers and specialized diagnostics is limited. Therefore, this study fills a gap in understanding the prognostic value of CBC markers in diabetic patients in Sabah, Malaysia, by directly linking prior international evidence to a locally relevant, real-world clinical application.

In response, LR was employed to model mortality (alive vs. deceased) as a binary outcome. To examine the prognostic performance of CBC markers alongside demographic characteristics, univariable LR was first used to screen potential predictors, followed by multivariable LR to identify independent factors. Model performance was evaluated using calibration and discrimination metrics. The knowledge gained is expected to fill a critical gap in understanding mortality risk factors and to demonstrate the utility of an affordable, easily available diagnostic tool for the early detection of high-risk patients and timely, personalized management in resource-limited healthcare settings. Specifically, the study had two objectives: (i) to identify CBC markers significantly associated with mortality in diabetic patients, and (ii) to evaluate the performance of these parameters in mortality risk stratification.

2. RESEARCH METHODS

The research commenced with data collection, utilizing secondary patient records on diabetes obtained from Queen Elizabeth I Hospital (QEH1). QEH1 is the tertiary main referral center for 32 hospitals across Sabah, including Sabah District Hospital and several private hospitals [39], [40], [41]. This referral structure enables the dataset to capture a heterogeneous hospital-based patient population across a wide referral network, rather than a narrowly defined single-clinic cohort. Subsequently, the data underwent preprocessing, including cleaning, partitioning, and encoding procedures. In the following stage, collinearity was assessed, and potential predictors were identified. Logistic regression (LR) was then applied to identify complete blood count (CBC) parameters associated with mortality among patients with diabetes. The model's performance was evaluated and compared using calibration and discrimination metrics. All data analyses were carried out in R (v4.5.0) alongside IBM SPSS (v30.0).

2.1 Data Source and Research Design

This retrospective cross-sectional study examined 10,672 diabetic patients from QEH1. Secondary data were obtained, including demographic details, CBC results, and patient outcomes (survival status), spanning 1 January 2018 through 31 December 2022. No exclusion criteria were applied. Furthermore, ethical approval was obtained from the Universiti Malaysia Sabah Medical Research Ethics Committee and the Medical Research and Ethics Committee (MREC) of the Ministry of Health, Malaysia.

2.2 Data Preparation

The data underwent a structured cleaning process that included identifying and removing duplicate records, followed by excluding cases with missing CBC data required for analysis. A total of 1,182 records (9.97%) were excluded due to missing CBC information, resulting in a final analytic dataset comprising 10,672 patients. This final sample size represents all eligible diabetic patients with complete CBC records available within the study period at Queen Elizabeth Hospital 1 after data cleaning. As the main tertiary

referral center serving 32 hospitals across Sabah, the hospital receives referrals from multiple districts, thereby capturing a broad and heterogeneous diabetic population. Consequently, the dataset reflects a real-world hospital-based cohort rather than a selectively sampled subset, supporting its representativeness within the clinical context of Sabah.

Deletion of records with missing data is considered acceptable when the proportion of missingness is modest, particularly in large clinical datasets, and has been shown to have minimal impact on bias and estimation accuracy when missingness is below 10% [42], [43], [44]. Moreover, prior methodological studies have demonstrated that model performance degrades substantially only at high levels of missingness, whereas predictive models generally remain stable under low levels of missingness [44], [45]. Given the low proportion of missing data in the present study, this approach was considered appropriate.

In addition, the final dataset satisfied recommended sample size requirements for LR modelling, with a sufficient number of outcome events to ensure reliable estimation of regression coefficients and standard errors [46], [47]. While exclusion of records due to missing data may reduce sample size and potentially introduce selection bias [43], [45], [48], the large remaining dataset enhances statistical power, improves precision of parameter estimates, and strengthens model stability during both training and validation.

The dataset was then divided into two subsets using stratified random sampling: 70% for training and 30% for testing [37]. The training dataset was used to develop the predictive model and to conduct internal validation via 10-fold cross-validation (CV). Model performance in terms of discrimination and calibration was determined by averaging results across all folds [27]. The testing dataset was subsequently applied to externally validate the model's performance and generalizability [27].

The demographic variables involved were age, gender, and ethnicity. Age was dichotomized into <60 years and ≥60 years. Ethnicity was categorized into Bajau, Bruneian-Malay, Chinese, Kadazan-Dusun, and others, representing the four major ethnic groups in Sabah [49], [50]. Hospitalization status was defined as outpatient or inpatient. Each CBC parameter was categorized as low, normal, or high, based on reference ranges reported in multiple authoritative guidelines [51]-[57], thereby simplifying the model [58]. This categorization enhanced interpretability, facilitated communication of findings, and provided practical decision-making criteria for diagnosis, treatment, and prognosis [59]-[62].

The mean platelet volume-to-platelet ratio (MPVPL) and monocyte-to-lymphocyte ratio (MLR) were derived from existing CBC parameters and lacked standardized reference ranges. To address this, median cutoff values of 0.037 and 0.36 were determined in this study and used to classify MPVPL and MLR into low and high levels [61], [62]. All categorical independent variables were encoded as dummy variables for the LR analysis. The designated reference groups were: age <60 years, male sex, Bajau ethnicity, outpatient status, and normal CBC levels; for MPVPL and MLR, the low category was used as the reference group.

2.3 Identification of Potential Risk Factors

At this stage, the objective was to identify factors that showed a significant association with the outcome [63]. For this purpose, Pearson's chi-square test of independence was applied. This test is applicable to contingency tables of varying sizes [64] and is considered appropriate when fewer than 20% of the expected cell counts are less than five [65]. Variables demonstrating a statistically significant association with patient outcomes ($p < 0.05$) were retained as candidate factors for further analysis.

2.4 Assessment of Multicollinearity

Collinearity, or multicollinearity, arises when two or more independent variables are highly correlated, which can compromise the stability and interpretability of regression models [66], [67]. To address this, both pre- and post-analysis assessments were conducted to ensure that variables contributed independently to outcome prediction, thereby improving model accuracy and interpretability [66], [67]. In the pre-analysis stage, correlations among categorical variables were examined using Cramer's V, as shown in Eq. (1). This measure, applicable to contingency tables of any size, evaluates the strength of associations between categorical variables, with scores ranging from 0 (no correlation) to 1 (perfect correlation) [65], [68], [69]. For 2×2 tables, Cramer's V is equal to the phi coefficient [70]. When strong correlations were detected, one of the variables was excluded in accordance with established recommendations [65], [71], [72].

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}, \tag{1}$$

where, χ^2 is chi-square value, N is number of samples, k is the number of rows or columns (the smaller of the two).

In the post-analysis stage, multicollinearity in the regression model was evaluated using the variance inflation factor (VIF) and tolerance (the denominator of the VIF formula), as shown in Eq. (2). A VIF greater than 10 or a tolerance below 0.1 was considered indicative of problematic collinearity [68], [73]. Variables exceeding these thresholds were excluded from the final model.

$$VIF = \frac{1}{1-R^2}, \tag{2}$$

where, R^2 is coefficient determination.

2.5 Logistic Regression

Univariate and multivariate logistic regression (LR) were applied to examine factors associated with patient outcomes. A significance criterion of $p < 0.05$ (two-tailed) was applied to all statistical tests. Prior to modelling, the key LR assumptions were reviewed and satisfied, including the specification of a binary outcome, assessment of multicollinearity, adequacy of sample size, and independence of errors **[74]**,**[75]**. Assumptions regarding outliers and the linearity of continuous predictors were not evaluated, as only categorical variables were used.

Univariate LR was first conducted to assess the effect of each variable independently on patient outcomes. Variables with $p < 0.05$ were retained for the initial multivariate model. In the multivariate stage, non-significant variables were sequentially removed, yielding a final optimized model comprising only statistically significant predictors. Results from the final model are presented as adjusted odds ratios (OR) along with 95% confidence intervals and associated *p*-values. In this framework, the dependent variable (Y) was binary, coded as 0 for alive and 1 for deceased, with independent variables denoted as (X). The general specification of the multivariate LR model is provided in Eq. (3).

$$y_i = logit\,(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij}, \tag{3}$$

where, $i = 1,2,3,\ldots,n;\ j = 1,2,3,\ldots,k$; y_i is the observed value for the dependent variable (patient outcomes) for the i-th observation, $\beta_0,\ldots,\beta_j$ is the regression coefficient, X_{ij} are independent variables (demographic, hospitalization status, CBC parameters) for the i-th observation, n number of observations, and k number of predictors.

Equivalently, the probability of mortality for the i-th observation can be expressed as:

$$\pi_i = \frac{e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_j X_{ij}}}{1+e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_j X_{ij}}}. \tag{4}$$

In this formulation, the regression coefficients (β_j) represent the change in the log-odds of mortality associated with a one-unit change in the predictor (X_{ij}), holding other variables constant. Exponentiating these coefficients yields the odds ratios (e^{β_j}), which quantify the multiplicative effect of each predictor on the odds of mortality.

2.6 Evaluation of Model Performance

Calibration and discrimination were used as the key measures to evaluate the model's performance. Calibration assesses how closely the predicted probabilities match the observed event rates [76]. It was assessed using the Brier score, defined as the average squared difference between observed outcomes and predicted probabilities, as shown in Eq. (5). Lower Brier scores, closer to zero, signify superior calibration [27], [35], [77].

$$Brier\ Score = \frac{1}{n}\sum_{i=1}^{n}(E_i - O_i)^2, \tag{5}$$

where N represents the total number of cases, E_i is the predicted outcome, O_i is the outcome (1, occurs, 0 otherwise).

Spiegelhalter's Z-test, as written in Eq. (6), was applied to further assess the calibration component of the Brier score by comparing observed versus expected probabilities [78], [79].

$$z(E_i, O_i) = \frac{\sum_{i=1}^{n}(O_i - E_i)\,(1 - 2E_i)}{\sqrt{\sum_{i=1}^{n}(1 - 2E_i)^2\, E_i(1 - E_i)}}. \tag{6}$$

If $|z(E_i, O_i)| > q_{1-\alpha/2}$, where $q_{1-\alpha/2}$ is the critical value from the standard normal distribution at a significance level of $\alpha = 0.05$, the result is considered significant, indicating poor calibration. A non-significant result is preferable, as it suggests adequate calibration [78], [79]. In all cases, larger absolute Z-values reflect greater miscalibration.

Calibration was further evaluated using binning-based approaches, as applied in recent studies to estimate the expected calibration error (ECE) and the maximum calibration error (MCE), as shown in Eqs. (7) and (8), respectively [78]. To calculate these metrics, predictions are sorted and divided into K bins, each containing approximately the same number of cases. The ECE quantifies the average calibration error across all bins; meanwhile, the MCE captures the largest single-bin calibration error [78], [80].

$$ECE = \sum_{i=1}^{n} P(i) \cdot |o_i - e_i|\,, \tag{7}$$

$$MCE = \max_{i\,=\,1,\,\ldots\,,\,K} (|o_i - e_i|), \tag{8}$$

where $P(i)$ is the proportion of all cases in bin i, o_i is the proportion of positive cases in bin i, and e_i is the average predicted probability for cases in bin i. The number of bins, K, is set to 10.

Due to the unequal class representation in the datasets, model discrimination was assessed using the F1 score (Eq. (9)), balanced accuracy (BAcc) (Eq. (10)), Matthews Correlation Coefficient (MCC) (Eq. (11)), and the area under the receiver operating characteristic curve (AUROC). Discrimination refers to a model's ability to distinguish between negative and positive cases (e.g., alive vs. deceased) [81]. BAcc is robust for imbalanced data [82], while the F1 score provides a balanced evaluation of binary classifiers [38]. AUROC summarizes overall classification performance across all possible decision thresholds. MCC incorporates all four components of the confusion matrix, yielding values from –1 to 1, and is widely regarded as an unbiased measure for imbalanced datasets [83]. In contrast, F1, Bacc, and AUROC range from 0 to 1, with higher values signifying superior performance.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}, \tag{9}$$

$$BAcc = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right), \tag{10}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{11}$$

3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis and Relationships Among Factors

The distribution of patient characteristics for a study sample of 10,672 patients is presented in Table 1. The data was divided into training (7,471 patients) and testing (3,201 patients) subsets. In the overall sample, 74.9% of patients were classified as "Alive". This proportion was consistent across both the training and testing sets. Similarly, 25.1% of patients were classified as "Deceased", with the same proportion across all subsets. Females slightly outnumber males across the dataset, with a consistent distribution between subsets. Although not shown, this study demonstrated consistent proportions across the overall, training, and testing groups for all CBC parameters, indicating a balanced representation across subsets.

Table 1. Distributions of Patient Demographics Across Sample Subsets

Demographic Profiles	Overall Sample (n =10,672)	Training (n =7,471)	Testing (n =3,201)
1. Patient Status			
Alive	7997 (74.9)	5598 (74.9)	2399 (74.9)
Deceased	2675 (25.1)	1873 (25.1)	802 (25.1)
2. Sex			
Male	5278 (49.5)	3700 (49.5)	1578 (49.3)
Female	5394 (50.5)	3771 (50.5)	1623 (50.7)
3. Hospitalization Status			
Outpatient	4355 (40.8)	3058 (40.9)	1297 (40.5)
Inpatient	6317 (59.2)	4413 (59.1)	1904 (59.5)
4. Age			
< 60	6340 (59.4)	4440 (59.4)	1900 (59.4)
≥ 60	4332 (40.6)	3031 (40.6)	1301 (40.6)
5. Race			
Bajau	1869 (17.5)	1311 (17.5)	558 (17.4)
Bruneian-Malay	1104 (10.3)	750 (10)	354 (11.1)
Chinese	1821 (17.1)	1274 (17.1)	547 (17.1)
Kadazan-Dusun	3284 (30.8)	2301 (30.8)	983 (30.7)
Others	2594 (24.3)	1835 (24.6)	759 (23.7)

The associations between patient status and clinical factors in the training subset, evaluated using the chi-square test is summarized in Table 2. Significant associations were observed for age, hospitalization status, race, WBC, neutrophil, lymphocyte, monocyte, eosinophil, IG, MCV, MCH, MCHC, RDW, platelet, PLCR, NRBC, PDW, HGB, RBC, HCT, NLR, PLR, PCT, MPVPL, and MLR. Variables that did not reach statistical significance (MPV, basophil, and sex) were excluded from the univariate LR analysis, leaving 25 retained predictors.

Table 2. Associations of Factors with Patient Status (Training Sample, n = 7,471)

No.	Variable	χ^2	No.	Variable	χ^2
1.	Age	214.037***	15.	Red cell distribution width (RDW)	436.2025***
2.	Sex	0.8630$^+$	16.	Platelet	102.1817***
3.	Hospitalization Status	206.4117***	17.	Platelet-large cell ratio (PLCR)	12.5412***
4.	Race	17.0973**	18.	Nucleated red blood cell (NRBC)	91.8478***
5.	White blood cell (WBC)	584.549***	19.	Mean platelet volume (MPV)	3.7373$^+$
6.	Neutrophil	1222.002***	20.	Platelet distribution width (PDW)	49.1785***
7.	Lymphocyte	1301.9854***	21.	Hemoglobin (HGB)	655.6079***
8.	Monocyte	187.2971***	22.	Red blood cell (RBC)	455.4517***
9.	Eosinophil	21.2291***	23.	Hematocrit (HCT)	617.459***
10.	Basophil	2.91$^+$	24.	Neutrophil-to-lymphocyte ratio (NLR)	2265.974***
11.	Immature granulocyte (IG)	581.2432***	25.	Platelet-to-lymphocyte ratio (PLR)	1528.0933***
12.	Mean corpuscular volume (MCV)	59.9510***	26.	Plateletcrit (PCT)	10.9860***
13.	Mean corpuscular hemoglobin (MCH)	56.9966***	27.	MPV-to-platelet ratio (MPVPL)	4.1997*
14.	Mean corpuscular hemoglobin concentration (MCHC)	46.4867***	28.	Monocyte-to-lymphocyte ratio (MLR)	1566.5709***

*Note: ***p<0.001, **p<0.01, *p<0.05, $^+$p>0.05*

3.2 Collinearity Assessment and Logistic Regression Analyses

Although not shown, collinearity was assessed in the training sample using Cramer's V, with thresholds of 0.350 for df = 2 and 0.500 for df = 1. Several variable pairs exceeded these thresholds, including HGB–HCT (0.747, df = 2), neutrophil–WBC (0.732, df = 2), PDW–PLCR (0.698, df = 1), and PCT–MPVPL (0.675, df = 1). Additional strong correlations were observed between MCH and MCV, NLR and PLR, HCT and RBC, PLT and MPVPL, neutrophil and NLR, HGB and RBC, IG and neutrophil, PLR with NLR, PLR with lymphocyte, PLR with PLT, WBC and monocyte, and platelet with PCT. To reduce collinearity, monocyte, HCT, WBC, RBC, PCT, PDW, IG, MPVPL, and MCH were excluded, leaving 16 variables for univariate analysis. All 16 were significant and entered into the multivariate model. In the initial multivariate

LR, race, eosinophil, MCV, MCHC, PLCR, and NRBC were not significantly associated with mortality and were sequentially removed. The final model identified several clinical variables significantly associated with mortality risk, which can be seen in Table 3.

Table 3. Results of the Finalized Multivariate LR Model (Training Sample, n = 7,471)

Parameter	Estimate	*p*-value	Adjusted Odd Ratio, OR [95% C.I]
Intercept	-8.9849	<0.001	-
Hospitalization Status			
Outpatient		Reference	
Inpatient	**1.3765**	**<0.001**	**3.9611 [3.3171, 4.7430]**
Age			
< 60		Reference	
≥ 60	**1.1898**	**<0.001**	**3.2863 [2.7693, 3.9082]**
Neutrophil			
Normal		Reference	
Low	0.1298	0.7375	1.1386 [0.5258, 2.4098]
High	**1.4652**	**<0.001**	**4.3283 [3.6018, 5.2164]**
Lymphocyte			
Normal		Reference	
Low	**1.1024**	**<0.001**	**3.0115 [2.4272, 3.7444]**
High	-0.3035	0.5217	0.7382 [0.2811, 1.8113]
Red cell distribution width (RDW)			
Normal		Reference	
Low	0.6706	0.1060	1.9553 [0.8571, 4.3764]
High	**1.1898**	**<0.001**	**3.2863 [2.7473, 3.9381]**
Platelet			
Normal		Reference	
Low	**1.4680**	**<0.001**	**4.3406 [3.1626, 5.9890]**
High	-0.2117	0.0796	0.8092 [0.6384, 1.0252]
Hemoglobin (HGB)			
Normal		Reference	
Low	**1.0332**	**<0.001**	**2.8102 [2.3370, 3.3845]**
High	-0.0759	0.8716	0.9269 [0.3534, 2.2575]
Neutrophil-to-lymphocyte ratio (NLR)			
Normal		Reference	
Low	-0.8871	0.1799	0.4119 [0.0911, 1.3175]
High	**3.2956**	**<0.001**	**26.9940 [18.1159, 41.6789]**
Platelet-to-lymphocyte ratio (PLR)			
Normal		Reference	
Low	-0.0680	0.7046	0.9343 [0.6558, 1.3248]
High	**1.4000**	**<0.001**	**4.0554 [3.2876, 5.0133]**
Monocyte-to-lymphocyte ratio (MLR)			
Low		Reference	
High	**1.3730**	**<0.001**	**3.9472 [3.1856, 4.9104]**

As shown in Table 3, hospitalization status was a significant predictor of mortality, with inpatients demonstrating a markedly higher risk (OR = 3.9611, $p < 0.001$). This is consistent with prior studies showing increased mortality among hospitalized diabetes patients. Gao et al. [84] reported higher in-hospital mortality among diabetic inpatients compared with non-diabetics, while Eyob Tediso et al. [85] highlighted that most deceased patients were older, known diabetics with multiple comorbidities and complications. Similarly, Jang et al. [86] found elevated mortality and longer hospital stays in both newly diagnosed and known diabetes cases compared with non-diabetic individuals.

Among demographic variables, only age was significantly associated with mortality, with individuals aged ≥60 years showing nearly a threefold higher risk (OR = 3.2863, $p < 0.001$). This finding is consistent with prior evidence highlighting advanced age as a critical risk factor in diabetes outcomes. Yan et al. [87] emphasized that older adults are more likely to experience diabetes and its associated complications, while Forbes [88] demonstrated increased mortality risk among those diagnosed between 60 and 70 years. Vega-López and González-Pérez [89] further showed that diabetes substantially reduced life expectancy among older adults in Mexico. Supporting this, Eyob Tediso et al. [85] found that more than half of the deceased

inpatients with diabetes were aged over 60 years. Collectively, these findings confirm that advanced age amplifies vulnerability to diabetes-related mortality.

Complete blood count (CBC) parameters provide essential insights into both inflammatory status [10], [25], [90], [91] and anemia-related conditions [2], [4], [92]-[95], which are regarded as important determinants of death in patients with diabetes. Regarding inflammatory markers, higher neutrophil levels were significantly associated with greater mortality risk (OR = 4.3283, $p < 0.001$), suggesting that patients with elevated neutrophil counts had more than four times the odds of death compared with those with normal levels. This finding is consistent with earlier studies showing that higher neutrophil levels are linked with cardiovascular (CVD) and all-cause death in individuals with diabetic kidney disease (DKD) [90], broader diabetic populations [96], and diabetic patients with COVID-19 [25], [97], [47]. Higher neutrophil levels reflect ongoing inflammation that can damage blood vessels and accelerate the buildup of arterial plaque (atherosclerosis), increasing cardiovascular and overall mortality risk in diabetes [90]. These activated neutrophils can also promote clot formation and contribute to declining kidney function, further raising the risk of poor outcomes [90].

Additionally, lower lymphocyte counts were significantly associated with a higher risk of death among diabetic patients (OR = 3.0115, $p < 0.001$), indicating that patients with lymphopenia had approximately threefold higher odds of mortality compared with those with normal lymphocyte levels. This observation supports prior evidence that lymphopenia contributes to greater mortality in diabetes patients affected by COVID-19 [25], [98] and is also linked with higher all-cause and CVD death among those with prediabetes or diabetes [21], [91]. A possible mechanism is that reduced lymphocyte levels reflect weakened immune defense capacity, resulting in a poorer ability to respond to severe illness and increasing the risk of adverse outcomes [21].

Among anemia-related markers, both RDW and hemoglobin emerged as significant risk factors for mortality in this cohort. Elevated RDW was significantly associated with diabetes-related mortality (OR = 3.2863, $p < 0.001$), indicating more than threefold higher odds of death compared with normal RDW levels. This finding is consistent with previous studies demonstrating that high RDW levels markedly increase mortality risk in patients with diabetic ketoacidosis [92], diabetes with acute kidney injury [4], diabetic kidney disease [93], coexisting congestive heart failure and diabetes [95], and diabetic foot amputation [94]. High RDW reflects variability in red cell size [2], [4], resulting from chronic inflammation and oxidative stress that impair normal red cell production (erythropoiesis) and may occur alongside nutritional deficiencies [4], [92], [93]. These abnormalities, which are common in advanced diabetes, can impair blood vessel function (endothelial dysfunction) and cause reduced oxygen delivery, representing cumulative metabolic stress that worsens clinical outcomes in this patient cohort [4], [92], [93].

Similarly, low hemoglobin, indicative of anemia, was significantly associated with higher mortality risk (OR = 2.81, $p < 0.001$), indicating nearly threefold higher odds of death compared with normal hemoglobin levels. In line with these results, earlier research has shown that reduced hemoglobin levels are linked to all-cause mortality in adults with diabetes [99], death among diabetic patients with COVID-19 [97], [100], mortality related to diabetic foot in individuals with diabetes mellitus [5], and mortality in the broader diabetic population [101]. Reduced hemoglobin levels lower oxygen-carrying capacity, promoting tissue hypoxia and increased cardiovascular strain, which may contribute to vascular and organ dysfunction, and increased mortality risk in patients with diabetes [2].

Platelet count was another significant CBC marker associated with diabetes-related mortality. Lower platelet counts were significantly associated with higher mortality risk (OR = 4.3406, $p < 0.001$), indicating more than fourfold higher odds of death compared with normal platelet levels. Other studies have reported comparable results, showing that thrombocytopenia is linked with higher mortality in diabetic patients with COVID-19 [25], [47], [102], [103] and with increased CVD and all-cause death in people with prediabetes or diabetes [21], [91]. In diabetes, hyperglycemia accelerates platelet turnover by activating platelets and shortening their lifespan, leading to ongoing platelet consumption at sites of vascular injury and inflammation, which may worsen vascular damage and increase the risk of cardiovascular events and mortality [20], [104].

This research demonstrated that inflammation-related blood markers, including the monocyte-to-lymphocyte ratio (MLR), neutrophil-to-lymphocyte ratio (NLR), and platelet-to-lymphocyte ratio (PLR), were significantly associated with mortality in individuals with diabetes. Higher MLR (OR = 3.9472, $p < 0.001$) indicates that patients with elevated MLR had nearly four times the odds of death compared with those

with lower MLR levels. This finding aligns with previous studies linking higher MLR to increased mortality in diabetic populations with chronic kidney disease [105], CVD [106], and in general diabetes cohorts [10]. An elevated MLR reflects increased inflammatory activity driven by monocytes, together with reduced immune regulation from lymphocytes. This imbalance may contribute to endothelial injury, worsen insulin resistance, and increase vulnerability to infection, potentially leading to poor outcomes in people with diabetes [10].

Similarly, high PLR (OR = 4.0554, $p < 0.001$) implies that patients with elevated PLR had approximately four times the odds of death compared with those with normal PLR levels. High PLR has been reported as a marker of mortality risk in patients with diabetic foot ulcers [19], diabetes with community-acquired pneumonia [20], and broader diabetic populations [10]. Elevated PLR reflects platelet hyperactivity alongside reduced lymphocyte levels. This condition promotes blood clot formation and vascular inflammation, accelerating atherosclerosis, while reduced lymphocyte responses impair immune control and tissue repair, heightening the risk of adverse outcomes and death [19].

In the present cohort, NLR demonstrated the strongest association with mortality (OR = 26.994, $p < 0.001$), suggesting that patients with elevated NLR had nearly 27 times the odds of death compared with those with normal NLR levels. Previous studies have consistently reported higher mortality among diabetic patients with elevated NLR, including general diabetes cohorts [10], [18], diabetic foot ulcers [19], diabetes with community-acquired pneumonia [20], cardiovascular–related death [1], [9], [22] diabetic kidney disease [23], cardiovascular disease with diabetes [24], and diabetes with COVID-19 [25]. A high NLR represents excessive neutrophil-driven inflammation and reduced immune regulation, leading to oxidative stress, endothelial dysfunction, and impaired wound healing, mechanisms that are amplified in diabetes and contribute to increased mortality [19], [20], [24].

Overall, MLR, PLR, and NLR serve as practical indicators of systemic inflammation and immune imbalance in diabetes [19], [20], [105], [106]. Chronic inflammation, procoagulant imbalance, and endothelial dysfunction play key roles in the progression of diabetes and its complications [10], [19], [105]. These processes contribute to tissue injury, microvascular and macrovascular complications, end-organ damage, and cardiovascular events, which remain a leading cause of death among individuals with diabetes [19]. Taken together, these findings support the prognostic value of inflammation-based blood markers for identifying diabetic patients at higher risk of adverse outcomes and mortality.

Furthermore, the variance inflation factor (VIF = 1.0150–1.7731) and tolerance values (0.5640–0.9852) for all predictors fell well within the acceptable thresholds (VIF < 10; tolerance > 0) [68], [73], indicating minimal multicollinearity and supporting the reliability of the parameter estimates. Collectively, these results underscore the critical role of clinical markers in assessing mortality, with inpatient status, advanced age, elevated neutrophil counts, reduced lymphocyte levels, increased RDW, decreased HGB, and higher NLR, PLR, and MLR emerging as significant risk factors in this patient cohort. The final regression model is presented in Eq. (12).

$$\begin{aligned} log\left(\frac{P(Deceased)}{1-P(Deceased)}\right) = {} & -8.9849 + 1.3765 \cdot x_{HS_{In}} + 1.1898 \cdot x_{Age\geq 60} + 1.4652 \cdot x_{Neut_H} \\ & +1.1024 \cdot x_{Lymph_L} + 1.1898 \cdot x_{RDW_H} + 1.4680 \cdot x_{PLT_L} + 1.0332 \cdot x_{HB_L} \\ & +3.2956 \cdot x_{NLR_H} + 1.4 \cdot x_{PLR_H} + 1.373 \cdot x_{MLR_H}. \end{aligned} \quad (12)$$

These findings are in line with earlier studies that have identified hospitalization status, advanced age, neutrophil-lymphocyte imbalance, red cell distribution width, hemoglobin level, and platelet-related indices as important factors associated with mortality among individuals with diabetes [4], [10], [84], [87], [91], [96], [99]. Validation of these associations in a large cohort reinforces existing evidence and extends it to an underrepresented population in Sabah. This alignment with earlier work strengthens confidence in the robustness and generalizability of the finalized model.

3.3 Evaluation of the Final Model's Performance

The performance of the final model for assessing mortality risk in diabetic patients was evaluated using calibration and discrimination metrics across both training and testing datasets. To enhance robustness and mitigate the risk of overfitting, cross-validation was applied to these evaluation measures. Model calibration was evaluated using 10-fold cross-validation on the training dataset (internal validation) and a separate testing dataset (external validation). Metrics included the Brier score, Spiegelhalter's Z-test, ECE, and MCE, as

presented in Table 4. Brier scores were close to zero in both internal and external validations, indicating strong calibration and reliable predictions for alive and deceased outcomes [27], [35]. Spiegelhalter's Z-test produced *p*-values > 0.05, suggesting no significant difference between predicted and observed outcomes, consistent with good calibration [78], [79]. ECE values were 0.0641 (internal) and 0.0319 (external), showing minimal average deviation between predicted and actual outcomes. MCE values were 0.1902 and 0.0807, reflecting the largest discrepancies observed within any bin.

Table 4. Model Calibration Assessment

No.	Measure	Internal Validation	External Validation
1.	Brier Score	0.0776	0.0767
2.	Spiegelhalter's Z-score	0.0812, *p*-value: 0.4506	-0.4638, *p*-value: 0.6427
3.	Expected Calibration Error (ECE)	0.0641	0.0319
4.	Maximum Calibration Error (MCE)	0.1902	0.0807

Model discrimination was evaluated using the same 10-fold cross-validation on the training dataset for internal validation and the independent testing dataset for external validation. Discrimination metrics included F1-score, balanced accuracy, the Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUROC), which together provide a comprehensive evaluation of the model's capability to correctly distinguish between alive and deceased outcomes.

Across both internal and external validations, balanced accuracy exceeded 0.84, indicating high overall classification performance (Table 5). F1-scores were above 0.77, demonstrating a reliable balance between sensitivity and precision. MCC values approached 0.70, suggesting strong agreement between predicted and actual outcomes beyond chance. In addition, the model demonstrated excellent discriminatory ability, with AUROC values of 0.9501 for internal validation and 0.9529 for external validation, indicating a strong, consistent capacity to differentiate between survival and mortality. Collectively, these results indicate that the model achieved consistent and robust discriminatory performance across datasets [27].

Table 5. Model Discrimination Assessment

No.	Measure	Internal Validation	External Validation
1.	Balanced accuracy	0.8530	0.8407
2.	F1-Score	0.7827	0.7703
3.	Matthews correlation coefficient (MCC)	0.7114	0.6978
4.	Area under the receiver operating characteristic curve (AUROC)	0.9501	0.9529

A contextual comparison of predictive models reported in the literature for diabetes-related clinical outcomes is presented in Table 6. Owing to substantial heterogeneity in outcome definitions, predictor sets, populations, and validation strategies, the reported performance metrics are provided for contextual reference only and should not be interpreted as direct head-to-head comparisons.

Table 6. Contextual Comparison of Logistic Regression–Based Mortality Risk Stratification with Alternative Predictive Models Reported in the Literature

Study	Population	Outcome (DV)	Predictor Set (IVs)	Model(s)	AUROC	Key Notes
This study	Diabetes (Sabah, n =10,672)	Mortality	CBC + patients' profile	LR	0.9529	CBC, age, hospitalization status
[10]	T2DM in ICU (MIMIC-IV, n =4783)	Mortality	CBC, AISI, SOFA	LR	0.741	Inflammatory markers
[20]	T2DM (China, n =1360)	In-hospital mortality	T2DMCAP	LR	0.858	More complex features with NLR
[47]	DM with COVID-19 (China, n = 325)	Mortality	CBC + clinical characteristics + comorbidities	LR	0.724	CBC, dyspnea, CVD, CKMD

Study	Population	Outcome (DV)	Predictor Set (IVs)	Model(s)	AUROC	Key Notes
[100]	DM with COVID-19 (Iran, n=455)	In-hospital mortality	CBC + patients' profile + comorbidities	LR	0.760	Neutrophil has the highest diagnostic accuracy
[107]	DKD (Singapore, n=6066)	Disease status	Imaging + labs + patients' profile	LR, NN	0.828	Hybrid (LR +NN) is slightly better, but LR is the second highest
[108]	CVD, CKD, DM, HTN (Singapore, n=6,762)	Disease status	Multiple predictors	NN, SVM, RF, GBM, KNN, LR	0.768 (DM)	LR has the highest performance for CKD and DM
[109]	DM, HTN (Taiwan, n=58,618)	Mortality	Hospital cost, vital signs and symptoms, comorbidities, and demographic	LR, RC, GBM, RF, SVM	0.9700	LR has the highest performance

Note. LR, logistic regression; NN, neural network; SVM, support vector machine; RF, random forest; GBM, gradient boosting machine; KNN, k-nearest neighbors; RC, ridge classifier; AUROC, area under the receiver operating characteristic curve; ICU, intensive care unit; DKD, diabetic kidney disease; DM, diabetes mellitus; HTN, hypertension; CVD, cardiovascular disease; CKD, chronic kidney disease; SOFA, Sequential Organ Failure Assessment score; AISI, aggregate index of systemic inflammation. Reported AUROC values correspond to external or testing datasets where available. Differences in outcome definitions, predictor availability, population characteristics, and validation frameworks preclude direct performance benchmarking.

As shown in Table 6, externally validated AUROC values reported for alternative machine learning and hybrid approaches are generally comparable to those observed in the present study. However, many of these models rely on expanded predictor sets, including clinical scores, imaging data, comorbidities, healthcare utilization metrics, or multi-modal electronic health record features. In contrast, the present study prioritizes interpretability and feasibility by using routinely available CBC parameters combined with basic patient characteristics, aligning with the goal of transparent and scalable mortality risk stratification in real-world hospital settings rather than algorithmic performance optimization.

3.4 Limitations and Future Directions

This research has some limitations that must be recognized. First, it relied on secondary data, restricting variable selection to those already available. As a result, the analysis focused primarily on CBC parameters and a limited set of demographic characteristics. Important variables such as comorbidities, socioeconomic factors, lifestyle factors, family and medication history were excluded due to data unavailability, limiting the ability to account for potential confounders and reducing the depth of clinical interpretation. Second, the study's observational design limits the extent to which causal inferences can be made, and residual confounding cannot be ruled out. Third, although the dataset originates from a single hospital, QEH1 is Sabah's major referral center for 32 hospitals, enhancing the sample's representativeness within the state. However, generalizability beyond Sabah remains limited and should be interpreted accordingly.

Future research should consider prospective data collection or the use of more comprehensive, multimodal datasets that integrate laboratory, clinical, behavioral, comorbidity, medication, socioeconomic, and family history variables. Incorporating such data would allow for a more holistic and accurate understanding of patient outcomes. In particular, future model extensions may incorporate non-CBC biomarkers, clinical severity indices, or imaging-based parameters to evaluate whether predictive performance and risk stratification accuracy can be further improved while maintaining interpretability. At the institutional level, strengthening hospital data infrastructure by systematically capturing these variables is recommended to enhance the quality and scope of future investigations.

3.5 Novelty of the Present Study

Rather than proposing a novel statistical or machine learning methodology, the originality of this study lies in its contextual application, the scale of the real-world data, and its translational relevance. This work is among the first to utilize routinely collected complete blood count (CBC) data for mortality risk stratification in patients with diabetes in Sabah. Although CBC is low-cost and widely available, its prognostic utility in diabetes outcomes remains underexplored, particularly in resource-limited healthcare settings. By applying an established, clinically interpretable logistic regression framework to large-scale, real-world hospital data,

this study demonstrates the feasibility of repurposing routinely collected hematological parameters for prognostic use at the population level.

The proposed model demonstrated excellent calibration and discrimination in both internal and external validations, with non-significant Spiegelhalter Z-tests, low Brier scores, and strong performance across balanced accuracy, F1-score, and Matthews correlation coefficient (MCC). These results confirm the model's robustness, generalizability, and reliable ability to stratify mortality risk, even in the presence of imbalanced outcomes. Importantly, the findings highlight the feasibility of integrating CBC-based markers into clinical risk assessment models, offering a cost-effective and scalable approach for early identification of high-risk patients. This contextual contribution is particularly valuable for low- and middle-income healthcare systems, where access to advanced biomarkers, longitudinal monitoring, and specialized diagnostic tools is often limited.

From a practical standpoint, the model is readily implementable in real-world hospital settings, as it relies solely on routinely available CBC parameters and basic patient information. In resource-limited regions such as Sabah, this framework may be integrated into electronic health record systems or clinical decision-support tools to assist clinicians in early identification of high-risk patients. Future implementation studies are warranted to evaluate its operational feasibility, usability, and impact on patient outcomes in routine clinical practice.

4. CONCLUSION

This research was conducted to evaluate the prognostic role of complete blood count (CBC) parameters in mortality risk stratification among 10,672 diabetic patients from Queen Elizabeth Hospital 1 (QEH1), Sabah, using logistic regression. The findings demonstrated that inpatient status, advanced age, and elevated neutrophil counts, neutrophil-to-lymphocyte ratio (NLR), red cell distribution width (RDW), platelet-to-lymphocyte ratio (PLR), and monocyte-to-lymphocyte ratio (MLR), together with lower lymphocyte counts and hemoglobin levels, were strongly associated with mortality. The logistic regression model showed excellent calibration and discrimination across both internal and external validations, underscoring its robustness and generalizability for clinical application. Importantly, these results demonstrate that meaningful mortality risk stratification can be achieved by repurposing routinely collected laboratory data, without reliance on costly or specialized biomarkers. By translating established analytical methods into a scalable, real-world risk stratification approach, this study highlights the feasibility of integrating CBC-based markers and demographic factors into clinical practice. This approach enables automated identification of high-risk patients and supports timely, personalized interventions. Such approaches are particularly valuable in resource-limited healthcare settings, where advanced diagnostic facilities are scarce. Beyond their clinical implications, these findings provide a foundation for advancing healthcare strategies in Sabah by promoting earlier detection, optimizing resource allocation, and improving patient management. Future research should explore additional factors such as comorbidities, lifestyle factors, and family history to further enhance model performance and understanding of mortality risk in diabetes. Policymakers could also leverage these insights to strengthen public health strategies, ultimately contributing to more efficient, equitable, and patient-centered diabetes care across Malaysia. By integrating accessible CBC markers into structured mortality risk stratification frameworks, healthcare providers in resource-limited settings such as Sabah can substantially improve early identification and timely management of high-risk diabetic patients, thereby reducing preventable mortality and enabling more sustainable healthcare delivery.

Author Contributions

Dg Siti Nurisya Sahirah Ag Isha: Data Curation, Methodology, Analysis, Software, and Writing-Original Draft. Nurliyana Juhan: Data Curation, Funding Acquisition, Project Administration, Supervision, Review and Editing. Yong Zulina Zubairi: Conceptualization, Methodology, Supervision, Review. Nornazirah Azizan: Data Curation, Writing and Review. Ho Chong Mun: Methodology, Supervision, Funding Acquisition. Abu Sayed Md. Al Mamun: Validation, Writing and Review. Lee Qin Zhi: Data Curation and Methodology. All authors reviewed the findings and participated in preparing the final manuscript.

Funding Statement

Gratitude is extended to the Malaysia Ministry of Higher Education for supporting this research under the Fundamental Research Grant Scheme (FRGS) (Project Code: FRGS/1/2023/STG06/UMS/02/2).

Acknowledgment

The authors wish to acknowledge the Malaysia Ministry of Higher Education for funding through the Fundamental Research Grant Scheme (FRGS) (Project Code: FRGS/1/2023/STG06/UMS/02/2). Appreciation is further extended to Queen Elizabeth Hospital for their invaluable collaboration and provision of the data that made this study possible.

Declarations

The authors affirm that there are no potential conflicts of interest to disclose.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement, including grammar, spelling, and clarity. The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. All final text was reviewed and approved by the authors.

REFERENCES

[1] G. Dong, M. Gan, S. Xu, Y. Xie, M. Zhou, and L. Wu, "THE NEUTROPHIL–LYMPHOCYTE RATIO AS A RISK FACTOR FOR ALL-CAUSE AND CARDIOVASCULAR MORTALITY AMONG INDIVIDUALS WITH DIABETES: EVIDENCE FROM THE NHANES 2003–2016," *Cardiovasc Diabetol*, vol. 22, no. 1, Sep. 2023, doi: https://doi.org/10.1186/s12933-023-01998-y

[2] A. Williams *et al.*, "PATHOPHYSIOLOGY OF RED BLOOD CELL DYSFUNCTION IN DIABETES AND ITS COMPLICATIONS," *Pathophysiology*, vol. 30, no. 3, pp. 327–345, Aug. 2023, doi: https://doi.org/10.3390/pathophysiology30030026.

[3] M. Aghaei, S. S. Bahreiny, Z. D. Zayeri, N. Davari, M. M. Abolhasani, and N. Saki, "EVALUATION OF COMPLETE BLOOD COUNT PARAMETERS IN PATIENTS WITH DIABETES MELLITUS: A SYSTEMATIC REVIEW," *Health Sci Rep*, vol. 8, no. 2, Feb. 2025, doi: https://doi.org/10.1002/hsr2.70488.

[4] M. Mo *et al.*, "INFLUENCE OF RED BLOOD CELL DISTRIBUTION WIDTH ON ALL-CAUSE DEATH IN CRITICAL DIABETIC PATIENTS WITH ACUTE KIDNEY INJURY," *Diabetes Metab Syndr Obes*, vol. Volume 15, pp. 2301–2309, Aug. 2022, doi: https://doi.org/10.2147/DMSO.S377650.

[5] J. Li *et al.*, "ASSOCIATION BETWEEN ANEMIA AND THE RISK AND OUTCOMES OF DIABETIC FOOT IN PATIENTS WITH TYPE 2 DIABETES MELLITUS," *Exp Ther Med*, vol. 26, no. 2, p. 384, Jun. 2023, doi: https://doi.org/10.3892/etm.2023.12083.

[6] Dianna. Magliano and E. J. . Boyko, *IDF diabetes Atlas*, 10th ed. BRUSSELS: INTERNATIONAL DIABETES FEDERATION, 2021. Accessed: Jul. 29, 2025. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK581934

[7] World Health Organization, "DATA – MALAYSIA." [Online]. Available: https://data.who.int/countries/458 [Accessed: 28 July 2025].

[8] World Health Organization, "THE TOP 10 CAUSES OF DEATH." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Accessed: 26 July 2025]

[9] G. Chen *et al.*, "ASSOCIATION OF NEUTROPHIL-LYMPHOCYTE RATIO WITH ALL-CAUSE AND CARDIOVASCULAR MORTALITY IN US ADULTS WITH DIABETES AND PREDIABETES: A PROSPECTIVE COHORT STUDY," *BMC Endocr Disord*, vol. 24, no. 1, p. 64, May 2024, doi: https://doi.org/10.1186/s12902-024-01592-7

[10] L. Liu, Y.-B. Zhao, Z.-T. Cheng, Y.-H. Li, and Y. Liu, "DEVELOPMENT AND VALIDATION OF A PROGNOSTIC MODEL FOR CRITICALLY ILL TYPE 2 DIABETES PATIENTS IN ICU BASED ON COMPOSITE INFLAMMATORY INDICATORS," *Sci Rep*, vol. 15, no. 1, p. 3627, 2025, doi: https://doi.org/10.1038/s41598-025-87731-z.

[11] L. Agnello *et al.*, "THE VALUE OF A COMPLETE BLOOD COUNT (CBC) FOR SEPSIS DIAGNOSIS AND PROGNOSIS," *Diagnostics*, vol. 11, no. 10, p. 1881, Oct. 2021, doi: https://doi.org/10.3390/diagnostics11101881.

[12] I.-H. Seo and Y.-J. Lee, "USEFULNESS OF COMPLETE BLOOD COUNT (CBC) TO ASSESS CARDIOVASCULAR AND METABOLIC DISEASES IN CLINICAL SETTINGS: A COMPREHENSIVE LITERATURE REVIEW," *Biomedicines*, vol. 10, no. 11, p. 2697, Oct. 2022, doi: https://doi.org/10.3390/biomedicines10112697.

[13] M. Kim *et al.*, "BIG DATA INSIGHTS INTO THE DIAGNOSTIC VALUES OF CBC PARAMETERS FOR SEPSIS AND SEPTIC SHOCK IN BURN PATIENTS: A RETROSPECTIVE STUDY," *Sci Rep*, vol. 14, no. 1, p. 800, Jan. 2024, doi: https://doi.org/10.1038/s41598-023-50695-z.

[14] Mohd. S. F. Omar, S. N. Mohd Tajuddin, S. Md Noor, and Z. Seman, "FULL BLOOD COUNT PARAMETERS IN COVID-19 PATIENTS WITH DISEASE SEVERITY, PATIENT OUTCOME AND VACCINATION STATUS," *Malaysian Journal of Medicine and Health Sciences*, vol. 19, no. s16, pp. 16–23, Dec. 2023, doi: https://doi.org/10.47836/mjmhs.19.s16.4.

[15] R. Yadav, V. Yadav, S. Pokhriyal, U. Zahid, and A. Gandhi, "RETROSPECTIVE OBSERVATIONAL STUDY OF COMPLETE BLOOD COUNT (CBC) PARAMETERS AND ICU MORTALITY OF COVID-19 DISEASE IN DELTA VARIANT AND OMICRON VARIANT IN A COMMUNITY-BASED HOSPITAL IN NEW YORK CITY," *Cureus*, vol. 15, no. 2, Feb. 2023, doi: https://doi.org/10.7759/cureus.34894.
[16] M. J. Balasa, "ACCURACY OF PLATELET INDICES AS PREDICTOR OF IN-HOSPITAL MORTALITY IN PATIENTS WITH ACUTE KIDNEY INJURY REQUIRING RENAL REPLACEMENT THERAPY," *Nephrology Dialysis Transplantation*, vol. 38, no. Supplement_1, Jun. 2023, doi: https://doi.org/10.1093/ndt/gfad063d_5304.
[17] G. E. Umeres-Francia, M. V. Rojas-Fernández, P. Herrera-Añazco, and V. A. Benites-Zapata, "NEUTROPHIL-TO-LYMPHOCYTE RATIO AND PLATELET-TO-LYMPHOCYTE RATIO AS RISK FACTORS FOR MORTALITY IN PERUVIAN ADULTS WITH CHRONIC KIDNEY DISEASE," *Ren Replace Ther*, vol. 8, Dec. 2022, doi: https://doi.org/10.1186/s41100-022-00420-9.
[18] L. L. F. Hoes *et al.*, "RELATIONSHIP OF NEUTROPHIL-TO-LYMPHOCYTE RATIO, IN ADDITION TO C-REACTIVE PROTEIN, WITH CARDIOVASCULAR EVENTS IN PATIENTS WITH TYPE 2 DIABETES," *Diabetes Res Clin Pract*, vol. 213, p. 111727, Jul. 2024, doi: https://doi.org/10.1016/j.diabres.2024.111727.
[19] W. Chen *et al.*, "NEUTROPHIL-TO-LYMPHOCYTE RATIO AND PLATELET-TO-LYMPHOCYTE RATIO PREDICT MORTALITY IN PATIENTS WITH DIABETIC FOOT ULCERS UNDERGOING AMPUTATIONS," *Diabetes Metab Syndr Obes*, vol. 14, pp. 821–829, Feb. 2021, doi: https://doi.org/10.2147/DMSO.S284583.
[20] S. Cheng *et al.*, "RISK PREDICTION OF IN-HOSPITAL MORTALITY AMONG PATIENTS WITH TYPE 2 DIABETES MELLITUS AND CONCOMITANT COMMUNITY-ACQUIRED PNEUMONIA," *Ann Palliat Med*, vol. 9, no. 5, pp. 3313–3325, Sep. 2020, doi: https://doi.org/10.21037/apm-20-1489.
[21] Y. He *et al.*, "ASSOCIATION BETWEEN HEMOGLOBIN, ALBUMIN, LYMPHOCYTE, AND PLATELET SCORE AND ALL-CAUSE AND CARDIOVASCULAR MORTALITY AMONG POPULATION WITH DIABETES: EVIDENCE FROM THE NHANES 2003–2016," *Diabetes Res Clin Pract*, vol. 224, p. 112212, Jun. 2025, doi: https://doi.org/10.1016/j.diabres.2025.112212.
[22] J.-M. He and Y. Yang, "ASSOCIATION BETWEEN NEUTROPHIL-LYMPHOCYTE RATIO AND ALL-CAUSE AND CARDIOVASCULAR MORTALITY IN PATIENTS WITH DIABETES OR PREDIABETES WITH COMORBID OBSTRUCTIVE SLEEP APNEA SYMPTOMS: EVIDENCE FROM NHANES 2005-2008 AND 2015-2018," *Front Endocrinol (Lausanne)*, vol. 16, Apr. 2025, doi: https://doi.org/10.3389/fendo.2025.1512621.
[23] Y. Bai *et al.*, "NEUTROPHIL/LYMPHOCYTE RATIO AND ALL-CAUSE MORTALITY IN DIABETIC KIDNEY DISEASE: A RETROSPECTIVE COHORT STUDY," *J Inflamm Res*, vol. 17, pp. 10739–10753, Dec. 2024, doi: https://doi.org/10.2147/JIR.S483637.
[24] B. Zhu *et al.*, "ASSOCIATION OF NEUTROPHIL-TO-LYMPHOCYTE RATIO WITH ALL-CAUSE AND CARDIOVASCULAR MORTALITY IN CVD PATIENTS WITH DIABETES OR PRE-DIABETES," *Sci Rep*, vol. 14, no. 1, p. 24324, Oct. 2024, doi: https://doi.org/10.1038/s41598-024-74642-8.
[25] S. Madaschi *et al.*, "PREDICTIVE MARKERS FOR CLINICAL OUTCOMES IN A COHORT OF DIABETIC PATIENTS HOSPITALIZED FOR COVID-19," *Diabetol Metab Syndr*, vol. 14, no. 1, p. 168, Nov. 2022, doi: https://doi.org/10.1186/s13098-022-00941-7.
[26] M. M. Ahsan, S. A. Luna, and Z. Siddique, "MACHINE-LEARNING-BASED DISEASE DIAGNOSIS: A COMPREHENSIVE REVIEW," *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, doi: https://doi.org/10.3390/healthcare10030541.
[27] L. P. Nguyen *et al.*, "THE UTILIZATION OF MACHINE LEARNING ALGORITHMS FOR ASSISTING PHYSICIANS IN THE DIAGNOSIS OF DIABETES," *Diagnostics*, vol. 13, no. 12, p. 2087, Jun. 2023, doi: https://doi.org/10.3390/diagnostics13122087.
[28] P. Arora, D. Boyne, J. J. Slater, A. Gupta, D. R. Brenner, and M. J. Druzdzel, "BAYESIAN NETWORKS FOR RISK PREDICTION USING REAL-WORLD DATA: A TOOL FOR PRECISION MEDICINE," *Value in Health*, vol. 22, no. 4, pp. 439–445, Apr. 2019, doi: https://doi.org/10.1016/j.jval.2019.01.006.
[29] N. Juhan, Y. Z. Zubairi, Z. Mohd Khalid, and A. S. Mahmood Zuhdi, "A COMPARISON BETWEEN BAYESIAN AND FREQUENTIST APPROACH IN THE ANALYSIS OF RISK FACTORS FOR FEMALE CARDIOVASCULAR DISEASE PATIENTS IN MALAYSIA," *ASM Science Journal*, vol. 13, pp. 1–7, Apr. 2020, doi: https://doi.org/10.32802/asmscj.2020.sm26(1.1)
[30] A. S. Mohd Faizal, W. Y. Hon, T. M. Thevarajah, S. M. Khor, and S.-W. Chang, "A BIOMARKER DISCOVERY OF ACUTE MYOCARDIAL INFARCTION USING FEATURE SELECTION AND MACHINE LEARNING," *Med Biol Eng Comput*, vol. 61, pp. 2527–2541, Oct. 2023, doi: https://doi.org/10.1007/s11517-023-02841-y.
[31] H. Sufriyana, Y.-W. Wu, and E. C.-Y. Su, "ARTIFICIAL INTELLIGENCE-ASSISTED PREDICTION OF PREECLAMPSIA: DEVELOPMENT AND EXTERNAL VALIDATION OF A NATIONWIDE HEALTH INSURANCE DATASET OF THE BPJS KESEHATAN IN INDONESIA," *EBioMedicine*, vol. 54, p. 102710, Apr. 2020, doi: https://doi.org/10.1016/j.ebiom.2020.102710.
[32] B. K. Betzler *et al.*, "DEEP LEARNING ALGORITHMS TO DETECT DIABETIC KIDNEY DISEASE FROM RETINAL PHOTOGRAPHS IN MULTIETHNIC POPULATIONS WITH DIABETES," *Journal of the American Medical Informatics Association*, vol. 30, no. 12, pp. 1904–1914, Nov. 2023, doi: https://doi.org/10.1093/jamia/ocad179.
[33] K. Y. Ngew, H. Z. Tay, and A. K. M. Yusof, "DEVELOPMENT AND VALIDATION OF A PREDICTIVE MODELS FOR PREDICTING THE CARDIAC EVENTS WITHIN ONE YEAR FOR PATIENTS UNDERWENT PERCUTANEOUS CORONARY INTERVENTION PROCEDURE AT IJN," *BMC Cardiovasc Disord*, vol. 23, Nov. 2023, doi: https://doi.org/10.1186/s12872-023-03536-w.
[34] K. Phirom, P. Charoenkwan, W. Shoombuatong, P. Charoenkwan, S. Sirichotiyakul, and T. Tongsong, "DEEPTHAL: A DEEP LEARNING-BASED FRAMEWORK FOR THE LARGE-SCALE PREDICTION OF THE A+-THALASSEMIA TRAIT USING RED BLOOD CELL PARAMETERS," *J Clin Med*, vol. 11, no. 21, Oct. 2022, doi: https://doi.org/10.3390/jcm11216305.

[35] C. Sang *et al.*, "DIAGNOSIS OF FIBROSIS USING BLOOD MARKERS AND LOGISTIC REGRESSION IN SOUTHEAST ASIAN PATIENTS WITH NON-ALCOHOLIC FATTY LIVER DISEASE," *Front Med (Lausanne)*, vol. 8, Feb. 2021, doi: https://doi.org/10.3389/fmed.2021.637652

[36] K. Soonklang, B. Siribumrungwong, B. Siripongpreeda, and C. Auewarakul, "COMPARISON OF MULTIPLE STATISTICAL MODELS FOR THE DEVELOPMENT OF CLINICAL PREDICTION SCORES TO DETECT ADVANCED COLORECTAL NEOPLASMS IN ASYMPTOMATIC THAI PATIENTS," *Medicine*, vol. 100, no. 20, p. e26065, May 2021, doi: https://doi.org/10.1097/MD.0000000000026065.

[37] S. Nusinovici *et al.*, "LOGISTIC REGRESSION WAS AS GOOD AS MACHINE LEARNING FOR PREDICTING MAJOR CHRONIC DISEASES," *J Clin Epidemiol*, vol. 122, pp. 56–69, Jun. 2020, doi: https://doi.org/10.1016/j.jclinepi.2020.03.002.

[38] C.-Y. Ting, H. Zakariah, and Y. Z. M. Yusri, "PROBABILISTIC RISK ASSESSMENT OF COVID-19 PATIENTS AT COVID-19 ASSESSMENT CENTRE," *International Journal of Technology*, vol. 13, no. 6, pp. 1193–1201, Nov. 2022, doi: https://doi.org/10.14716/ijtech.v13i6.5882.

[39] K. Ng, A. K. Y. Chow, sooth lu, kok wong, and N. Sulaiman, "A RETROSPECTIVE STUDY ON THE INCIDENCE AND OUTCOMES OF ACUTE KIDNEY INJURY AMONG PATIENTS DIAGNOSED WITH MALARIA INFECTION IN SABAH – A TERTIARY CENTRE EXPERIENCE," *Journal of Clinical and Translational Nephrology (JCTN)*, vol. 1, Jan. 2021.

[40] F. F. Lem, M. Y. Lim, and W. Z. Teong, "HOW DEPRESSING CAN CHRONIC PAIN BE?: A CROSS-SECTIONAL STUDY IN QUEEN ELIZABETH HOSPITAL, SABAH," *Malaysian Journal of Medicine & Health Sciences*, vol. 17, no. 3, pp. 82–87, 2021.

[41] Clinical Research Malaysia, "HOSPITAL QUEEN ELIZABETH I." [Online]. Available: https://clinicalresearch.my/portfolio/hospital-queen-elizabeth-i/ [Accessed: 08 Dec. 2025].

[42] J. Rashid and Khalid Waheed, "MISSING VALUES AND OUTLIERS IN RESEARCH DATA," *Pakistan Postgraduate Medical Journal*, vol. 31, no. 04, p. 167, Jun. 2021, doi: https://doi.org/10.51642/ppmj.v31i04.404.

[43] L. Ren, T. Wang, A. Sekhari Seklouli, H. Zhang, and A. Bouras, "A REVIEW ON MISSING VALUES FOR MAIN CHALLENGES AND METHODS," *Inf Syst*, vol. 119, p. 102268, Oct. 2023, doi: https://doi.org/10.1016/j.is.2023.102268.

[44] K. P. Junaid, T. Kiran, M. Gupta, K. Kishore, and S. Siwatch, "HOW MUCH MISSING DATA IS TOO MUCH TO IMPUTE FOR LONGITUDINAL HEALTH INDICATORS? A PRELIMINARY GUIDELINE FOR THE CHOICE OF THE EXTENT OF MISSING PROPORTION TO IMPUTE WITH MULTIPLE IMPUTATION BY CHAINED EQUATIONS," *Popul Health Metr*, vol. 23, no. 1, p. 2, Feb. 2025, doi: https://doi.org/10.1186/s12963-025-00364-2.

[45] P. A. Popoola, J.-R. Tapamo, and A. G. H. Assounga, "EFFECTIVE AND EFFICIENT HANDLING OF MISSING DATA IN SUPERVISED MACHINE LEARNING," *Data Science and Management*, vol. 8, no. 3, pp. 361–373, Sep. 2025, doi: https://doi.org/10.1016/j.dsm.2024.12.002.

[46] P. C. Austin and S. van Buuren, "THE EFFECT OF HIGH PREVALENCE OF MISSING DATA ON ESTIMATION OF THE COEFFICIENTS OF A LOGISTIC REGRESSION MODEL WHEN USING MULTIPLE IMPUTATION," *BMC Med Res Methodol*, vol. 22, no. 1, p. 196, Dec. 2022, doi: https://doi.org/10.1186/s12874-022-01671-0.

[47] Y.-F. Xiao *et al.*, "MAJOR CHARACTERISTICS OF SEVERITY AND MORTALITY IN DIABETIC PATIENTS WITH COVID-19 AND ESTABLISHMENT OF SEVERITY RISK SCORE," *Front Med (Lausanne)*, vol. 8, p. 655604, Jun. 2021, doi: https://doi.org/10.3389/fmed.2021.655604.

[48] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "AN INVESTIGATION OF THE IMPUTATION TECHNIQUES FOR MISSING VALUES IN ORDINAL DATA ENHANCING CLUSTERING AND CLASSIFICATION ANALYSIS VALIDITY," *Decision Analytics Journal*, vol. 9, p. 100341, Dec. 2023, doi: https://doi.org/10.1016/j.dajour.2023.100341.

[49] D. K. Ag Budin and S. A. Wafa, "THE RELATIONSHIP BETWEEN GENDER AND ETHNICITY UPON HOFSTEDE'S CULTURAL DIMENSIONS AMONG SABAH ETHNICITIES," *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 10, pp. 55–58, 2013, [Online]. Available: https://api.semanticscholar.org/CorpusID:86867059. doi: https://doi.org/10.9790/487X-1064649

[50] Sabah State Government, "PEOPLE & HISTORY." [Online]. Available: https://sabah.gov.my/en/content/people-history [Accessed: 23 July 2025].

[51] J. El Brihi and S. Pathak, *NORMAL AND ABNORMAL COMPLETE BLOOD COUNT WITH DIFFERENTIAL*. 2025.

[52] Cleveland Clinic, "COMPLETE BLOOD COUNT." Accessed: Jul. 24, 2025. [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/4053-complete-blood-count

[53] L. C. Dean, *Blood groups and red cell antigens*. Bethesda: National Center for Biotechnology Information (US), 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:88933786

[54] A. N. Drăgoescu *et al.*, "NEUTROPHIL TO LYMPHOCYTE RATIO (NLR) - A USEFUL TOOL FOR THE PROGNOSIS OF SEPSIS IN THE ICU," *Biomedicines*, vol. 10, no. 1, p. 75, Dec. 2021, doi: https://doi.org/10.3390/biomedicines10010075.

[55] M. I. Khan and I. Ullah, "DIAGNOSTIC IMPORTANCE OF MEAN PLATELET VOLUME, PLATELET DISTRIBUTION WIDTH AND PLATELET LARGE CELL RATIO AS SCREENING TOOL IN IMMUNE THROMBOCYTOPENIA," *Porto Biomed J*, vol. 5, no. 6, p. e094, Nov. 2020, doi: https://doi.org/10.1097/j.pbj.0000000000000094.

[56] K. Pogorzelska, A. Krętowska, M. Krawczuk-Rybak, and M. Sawicka-Żukowska, "CHARACTERISTICS OF PLATELET INDICES AND THEIR PROGNOSTIC SIGNIFICANCE IN SELECTED MEDICAL CONDITION – A SYSTEMATIC REVIEW," *Adv Med Sci*, vol. 65, no. 2, pp. 310–315, Sep. 2020, doi: https://doi.org/10.1016/j.advms.2020.05.002.

[57] D. Thombare, A. Bhalerao, P. Dixit, S. Joshi, and P. Dapkekar, "NEUTROPHIL-TO-LYMPHOCYTE RATIO AND PLATELET-TO-LYMPHOCYTE RATIO IN ANTENATAL WOMEN WITH PRE-ECLAMPSIA: A CASE-CONTROL STUDY," *Cureus*, vol. 15, no. 6, Jun. 2023, doi: https://doi.org/10.7759/cureus.40338.

[58] N. Mahachai *et al.*, "CLINICAL PREDICTIVE SCORE FOR IDENTIFYING METABOLIC DYSFUNCTION-ASSOCIATED STEATOTIC LIVER DISEASE IN INDIVIDUALS WITH PREDIABETES USING TRANSIENT ELASTOGRAPHY," *J Clin Med*, vol. 12, no. 24, p. 7617, Dec. 2023, doi: https://doi.org/10.3390/jcm12247617.

[59] E. L. Busch, "CUT POINTS AND CONTEXTS," *Cancer*, vol. 127, no. 23, pp. 4348–4355, Dec. 2021, doi: https://doi.org/10.1002/cncr.33838.

[60] Y. Chen *et al.*, "A NOVEL APPROACH TO DETERMINE TWO OPTIMAL CUT-POINTS OF A CONTINUOUS PREDICTOR WITH A U-SHAPED RELATIONSHIP TO HAZARD RATIO IN SURVIVAL DATA: SIMULATION AND APPLICATION," *BMC Med Res Methodol*, vol. 19, Dec. 2019, doi: https://doi.org/10.1186/s12874-019-0738-4.
[61] Y. Jhan, W. Li, S. Ruan, J. Sie, and I. Lian, "OPTIMAL DICHOTOMIZATION OF BIMODAL GAUSSIAN MIXTURES," *Statistical Papers*, vol. 65, pp. 3285–3301, Jul. 2024, doi: https://doi.org/10.1007/s00362-023-01521-1.
[62] F. Tustumi, "CHOOSING THE MOST APPROPRIATE CUT-POINT FOR CONTINUOUS VARIABLES," *Rev Col Bras Cir*, vol. 49, 2022, doi: https://doi.org/10.1590/0100-6991e-20223346-en.
[63] C. Partanapat, C. Jaruskulchai, and C. Jandaeng, "A HYBRID MODEL FOR CORONARY HEART DISEASE PREDICTION IN THAI POPULATION," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 414–425, 2020, doi: https://doi.org/10.25046/aj050552.
[64] IBM Corporation, "CROSSTABS STATISTICS." [Online]. Available: https://www.ibm.com/docs/en/spss-statistics/31.0.0?topic=crosstabs-statistics [Accessed: 23 July 2025].
[65] S. H. Abu-Bader, *USING STATISTICAL METHODS IN SOCIAL SCIENCE RESEARCH : WITH A COMPLETE SPSS GUIDE*, 3rd ed. Oxford: Oxford University Press, 2021.
[66] M. Gregorich, S. Strohmaier, D. Dunkler, and G. Heinze, "REGRESSION WITH HIGHLY CORRELATED PREDICTORS: VARIABLE OMISSION IS NOT THE SOLUTION," *Int J Environ Res Public Health*, vol. 18, no. 8, Apr. 2021, doi: https://doi.org/10.3390/ijerph18084259.
[67] K. I. Sundus, B. H. Hammo, M. B. Al-Zoubi, and A. Al-Omari, "SOLVING THE MULTICOLLINEARITY PROBLEM TO IMPROVE THE STABILITY OF MACHINE LEARNING ALGORITHMS APPLIED TO A FULLY ANNOTATED BREAST CANCER DATASET," *Inform Med Unlocked*, vol. 33, p. 101088, 2022, doi: https://doi.org/10.1016/j.imu.2022.101088.
[68] B. M. Benito, "AUTOMATED MULTICOLLINEARITY MANAGEMENT - PACKAGE 'COLLINEAR," 2024. [Online]. Available: https://cran.r-project.org/web/packages/collinear/collinear.pdf [Accessed: 25 July 2025].
[69] IBM Corporation, "CRAMER'S V." [Online]. Available: https://www.ibm.com/docs/en/cognos-analytics/12.1.0?topic=terms-cramrs-v [Accessed: 25 July 2025].
[70] O. Charles C., "ALTERNATIVE METHODS OF SOLVING BIASEDNESS IN CHI - SQUARE CONTINGENCY TABLE," *Academic Journal of Applied Mathematical Sciences*, vol. 5, no. 1, pp. 1–6, Jan. 2019, doi: https://doi.org/10.32861/ajams.51.1.6.
[71] C. C. Serdar, M. Cihan, D. Yücel, and M. A. Serdar, "SAMPLE SIZE, POWER AND EFFECT SIZE REVISITED: SIMPLIFIED AND PRACTICAL APPROACHES IN PRE-CLINICAL, CLINICAL AND LABORATORY STUDIES," *Biochem Med (Zagreb)*, vol. 31, no. 1, Feb. 2021, doi: https://doi.org/10.11613/BM.2021.010502.
[72] S. Sugathan and L. Jacob, "USE OF EFFECT SIZE MEASURES ALONG WITH P-VALUE IN SCIENTIFIC PUBLICATIONS," *Borneo Epidemiology Journal*, vol. 2, no. 2, pp. 89–97, Dec. 2021, doi: https://doi.org/10.51200/bej.v2i2.3629.
[73] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *MULTIVARIATE DATA ANALYSIS*, 8TH ED. United Kingdom: Cengage, 2019.
[74] A. K. Nørskov *et al.*, "ASSESSMENT OF ASSUMPTIONS OF STATISTICAL ANALYSIS METHODS IN RANDOMISED CLINICAL TRIALS: THE WHAT AND HOW," *BMJ Evid Based Med*, vol. 26, no. 3, pp. 121–126, Jun. 2021, doi: https://doi.org/10.1136/bmjebm-2019-111268.
[75] R. R. Ramalu, R. Kandasamy, A. Raffiq, Z. Idris, and J. S. Adnan, "RISK FACTORS ASSOCIATED WITH OUTCOME OF SPONTANEOUS INTRACEREBRAL HAEMORRHAGE: HOSPITAL KUALA LUMPUR EXPERIENCE," *Malaysian Journal of Medical Sciences*, vol. 29, no. 1, pp. 76–90, Feb. 2022, doi: https://doi.org/10.21315/mjms2022.29.1.8.
[76] P. Mekanimitdee, T. Morasert, J. Patumanond, and P. Phinyo, "THE MAGENTA MODEL FOR INDIVIDUAL PREDICTION OF IN-HOSPITAL MORTALITY IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE WITH ACUTE EXACERBATION IN RESOURCE-LIMITED COUNTRIES: A DEVELOPMENT STUDY," *PLoS One*, vol. 16, no. 8, p. e0256866, Aug. 2021, doi: https://doi.org/10.1371/journal.pone.0256866.
[77] W. Han, N. A. Allameen, I. Ibrahim, P. Dhanasekaran, F. Mengling, and M. Lahiri, "HOSPITAL ADMISSION RISK STRATIFICATION OF PATIENTS WITH GOUT PRESENTING TO THE EMERGENCY DEPARTMENT," *Clin Rheumatol*, vol. 41, pp. 1801–1807, Jun. 2022, doi: https://doi.org/10.1007/s10067-021-05902-5.
[78] Y. Huang, W. Li, F. Macheret, R. A. Gabriel, and L. Ohno-Machado, "A TUTORIAL ON CALIBRATION MEASUREMENTS AND CALIBRATION MODELS FOR CLINICAL PREDICTION MODELS," *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 621–633, Apr. 2020, doi: https://doi.org/10.1093/jamia/ocz228.
[79] O. Lindhiem, I. T. Petersen, L. K. Mentch, and E. A. Youngstrom, "THE IMPORTANCE OF CALIBRATION IN CLINICAL PSYCHOLOGY," *Assessment*, vol. 27, no. 4, pp. 840–854, Jun. 2020, doi: https://doi.org/10.1177/1073191117752055.
[80] T. Dawood *et al.*, "UNCERTAINTY AWARE TRAINING TO IMPROVE DEEP LEARNING MODEL CALIBRATION FOR CLASSIFICATION OF CARDIAC MR IMAGES," *Med Image Anal*, vol. 88, p. 102861, Aug. 2023, doi: https://doi.org/10.1016/j.media.2023.102861.
[81] N. Juhan, Y. Z. Zubairi, A. S. Mahmood Zuhdi, and Z. Mohd Khalid, "PREDICTORS ON OUTCOMES OF CARDIOVASCULAR DISEASE OF MALE PATIENTS IN MALAYSIA USING BAYESIAN NETWORK ANALYSIS," *BMJ Open*, vol. 13, no. 11, p. e066748, Nov. 2023, doi: https://doi.org/10.1136/bmjopen-2022-066748
[82] P. Thölke *et al.*, "CLASS IMBALANCE SHOULD NOT THROW YOU OFF BALANCE: CHOOSING THE RIGHT CLASSIFIERS AND PERFORMANCE METRICS FOR BRAIN DECODING WITH IMBALANCED DATA," *Neuroimage*, vol. 277, p. 120253, Aug. 2023, doi: https://doi.org/10.1016/j.neuroimage.2023.120253.
[83] S. J. J. Guesné, T. Hanser, S. Werner, S. Boobier, and S. Scott, "MIND YOUR PREVALENCE!," *J Cheminform*, vol. 16, no. 1, p. 43, Apr. 2024, doi: https://doi.org/10.1186/s13321-024-00837-w.
[84] T. Gao, K. E. Agho, M. K. Piya, D. Simmons, and U. L. Osuagwu, "ANALYSIS OF IN-HOSPITAL MORTALITY AMONG PEOPLE WITH AND WITHOUT DIABETES IN SOUTH WESTERN SYDNEY PUBLIC HOSPITALS (2014–2017)," *BMC Public Health*, vol. 21, no. 1, p. 1991, Dec. 2021, doi: https://doi.org/10.1186/s12889-021-12120-w.

[85] D. Eyob Tediso, F. Bekele Daba, and T. Ayele Mega, "IN-HOSPITAL MORTALITY AND ITS PREDICTORS AMONG HOSPITALIZED DIABETES PATIENTS: A PROSPECTIVE OBSERVATIONAL STUDY," *Int J Clin Pract*, vol. 2023, p. 9367483, Mar. 2023, doi: https://doi.org/10.1155/2023/9367483.

[86] S. A. Jang, K. Min Kim, H. Jin Kang, S.-J. Heo, C. Sik Kim, and S. Won Park, "HIGHER MORTALITY AND LONGER LENGTH OF STAY IN HOSPITALIZED PATIENTS WITH NEWLY DIAGNOSED DIABETES," *Diabetes Res Clin Pract*, vol. 210, p. 111601, Apr. 2024, doi: https://doi.org/10.1016/j.diabres.2024.111601.

[87] Z. Yan, M. Cai, X. Han, Q. Chen, and H. Lu, "THE INTERACTION BETWEEN AGE AND RISK FACTORS FOR DIABETES AND PREDIABETES: A COMMUNITY-BASED CROSS-SECTIONAL STUDY," *Diabetes, Metabolic Syndrome and Obesity*, vol. 16, pp. 85–93, Jan. 2023, doi: https://doi.org/10.2147/DMSO.S390857.

[88] A. Forbes, "REDUCING THE BURDEN OF MORTALITY IN OLDER PEOPLE WITH DIABETES: A REVIEW OF CURRENT RESEARCH," *Front Endocrinol (Lausanne)*, vol. 11, p. 133, Mar. 2020, doi: https://doi.org/10.3389/fendo.2020.00133.

[89] M. G. Vega-López and G. J. González-Pérez, "MORTALITY FROM DIABETES MELLITUS AND ITS IMPACT ON LIFE EXPECTANCY AT 60 YEARS OF AGE IN MEXICO," *Rev Saude Publica*, vol. 55, p. 61, Oct. 2021, doi: https://doi.org/10.11606/s1518-8787.2021055003292 9.

[90] R. Xie, D. M. Bishai, D. T. W. Lui, P. C. H. Lee, and D. Y. H. Yap, "HIGHER CIRCULATING NEUTROPHIL COUNTS IS ASSOCIATED WITH INCREASED RISK OF ALL-CAUSE MORTALITY AND CARDIOVASCULAR DISEASE IN PATIENTS WITH DIABETIC KIDNEY DISEASE," *Biomedicines*, vol. 12, no. 8, p. 1907, Aug. 2024, doi: https://doi.org/10.3390/biomedicines12081907.

[91] P. Zhao, Z. Zhang, M. Li, J. Hao, and Y. Wang, "ASSOCIATION BETWEEN HEMOGLOBIN–ALBUMIN–LYMPHOCYTE–PLATELET SCORE AND ALL-CAUSE OR CARDIOVASCULAR MORTALITY IN PATIENTS WITH DIABETES OR PREDIABETES: MEDIATED EFFECTS OF RENAL FUNCTION," *BMC Cardiovasc Disord*, vol. 25, no. 1, p. 331, Apr. 2025, doi: https://doi.org/10.1186/s12872-025-04791-9.

[92] H. Dai, X. Su, H. Li, and L. Zhu, "ASSOCIATION BETWEEN RED BLOOD CELL DISTRIBUTION WIDTH AND MORTALITY IN DIABETIC KETOACIDOSIS," *Journal of International Medical Research*, vol. 48, no. 3, p. 00060520911494, Mar. 2020, doi: https://doi.org/10.1177/0300060520911494.

[93] S. Roumeliotis *et al.*, "RED BLOOD CELL DISTRIBUTION WIDTH IS ASSOCIATED WITH DETERIORATION OF RENAL FUNCTION AND CARDIOVASCULAR MORBIDITY AND MORTALITY IN PATIENTS WITH DIABETIC KIDNEY DISEASE," *Life*, vol. 10, no. 11, p. 301, Nov. 2020, doi: https://doi.org/10.3390/life10110301.

[94] Y. K. Yoon *et al.*, "RED BLOOD CELL DISTRIBUTION WIDTH IS AN INDEPENDENT PREDICTOR OF MORTALITY FOLLOWING AMPUTATION FOR DIABETIC FOOT," *Sci Rep*, vol. 15, no. 1, p. 4441, Feb. 2025, doi: https://doi.org/10.1038/s41598-025-85684-x.

[95] K. Zhang *et al.*, "ASSOCIATION BETWEEN RED BLOOD CELL DISTRIBUTION WIDTH AND IN-HOSPITAL MORTALITY AMONG CONGESTIVE HEART FAILURE PATIENTS WITH DIABETES AMONG PATIENTS IN THE INTENSIVE CARE UNIT: A RETROSPECTIVE COHORT STUDY," *Crit Care Res Pract*, vol. 2024, p. 9562200, Jan. 2024, doi: https://doi.org/10.1155/2024/9562200.

[96] L. Yang *et al.*, "PERIPHERAL MONOCYTE COUNT IS AN INDEPENDENT PREDICTOR OF ALL-CAUSE MORTALITY IN TYPE 2 DIABETES WITH MACRO-VASCULAR COMPLICATIONS," *Medicine*, vol. 99, no. 4, p. e18876, Jan. 2020, doi: https://doi.org/10.1097/MD.0000000000018876.

[97] J. Carbajal *et al.*, "RISK FACTORS ASSOCIATED WITH THE MORTALITY OF COVID-19 IN PATIENTS WITH TYPE 2 DIABETES MELLITUS," *World Acad Sci J*, vol. 6, no. 6, p. 62, Sep. 2024, doi: https://doi.org/10.3892/wasj.2024.277.

[98] Y. Deng *et al.*, "ASSOCIATION OF DIABETES WITH SEVERITY AND MORTALITY IN HOSPITALIZED PATIENTS WITH COVID-19 IN WUHAN, CHINA: A SINGLE-CENTERED, RETROSPECTIVE STUDY," *Arch Endocrinol Metab*, vol. 65, no. 5, pp. 596–608, Jul. 2021, doi: https://doi.org/10.20945/2359-3997000000384

[99] A. K. Koyama, E. Lundeen, K. McKeever Bullard, and M. E. Pavkov, "PREVALENCE OF ANEMIA AND ASSOCIATED ALL-CAUSE MORTALITY AMONG ADULTS WITH DIABETES: THE ROLE OF CHRONIC KIDNEY DISEASE," *Diabetes Res Clin Pract*, vol. 200, p. 110695, Jun. 2023, doi: https://doi.org/10.1016/j.diabres.2023.110695.

[100] H. Rastad *et al.*, "FACTORS ASSOCIATED WITH THE POOR OUTCOMES IN DIABETIC PATIENTS WITH COVID-19," *J Diabetes Metab Disord*, vol. 19, no. 2, pp. 1293–1302, Dec. 2020, doi: https://doi.org/10.1007/s40200-020-00646-6.

[101] D. Erez *et al.*, "ANEMIA IN PATIENTS WITH DIABETES AND PREDIABETES WITH NORMAL KIDNEY FUNCTION: PREVALENCE AND CLINICAL OUTCOMES," *Endocrine Practice*, vol. 28, no. 2, pp. 129–134, Feb. 2022, doi: https://doi.org/10.1016/j.eprac.2021.10.005

[102] M. Faisal, F. Alzahraa Al-Hattab, A. Mohammed Al-boinin, B. Mahmoud Al-Qudah, M. A. Waheed, and M. Danjuma, "THE EFFECTS OF THROMBOCYTOPENIA, TYPE 2 DIABETES MELLITUS, AND ENDOTHELIAL DYSFUNCTION ON CLINICAL OUTCOMES IN PATIENTS WITH COVID-19," *Qatar Med J*, vol. 2023, no. 1, p. 3, May 2023, doi: https://doi.org/10.5339/qmj.2023.3.

[103] C. Izzi-Engbeaya *et al.*, "ADVERSE OUTCOMES IN COVID-19 AND DIABETES: A RETROSPECTIVE COHORT STUDY FROM THREE LONDON TEACHING HOSPITALS," *BMJ Open Diabetes Res Care*, vol. 9, no. 1, p. e001858, Jan. 2021, doi: https://doi.org/10.1136/bmjdrc-2020-001858.

[104] S. Jena, N. Raizada, S. Kalra, and S. Bhattacharya, "PLATELET INDICES AS PREDICTORS OF GLYCAEMIC STATUS AND COMPLICATIONS IN DIABETES MELLITUS," *Apollo Medicine*, vol. 22, no. 3, pp. 214–219, May 2025, doi: https://doi.org/10.1177/09760016251317265.

[105] C. Qiu, S. Liu, X. Li, W. Li, G. Hu, and F. Liu, "PROGNOSTIC VALUE OF MONOCYTE-TO-LYMPHOCYTE RATIO FOR 90-DAY ALL-CAUSE MORTALITY IN TYPE 2 DIABETES MELLITUS PATIENTS WITH CHRONIC KIDNEY DISEASE," *Sci Rep*, vol. 13, no. 1, p. 13136, Aug. 2023, doi: https://doi.org/10.1038/s41598-023-40429-6.

[106] H. Li *et al.*, "MONOCYTE–LYMPHOCYTE RATIO PREDICTS CARDIOVASCULAR DISEASES DEATH IN INDIVIDUALS WITH TYPE 2 DIABETES," *J Diabetes Investig*, vol. 16, no. 1, pp. 137–145, Jan. 2025, doi: https://doi.org/10.1111/jdi.14329.

[107] B. K. Betzler *et al.*, "DEEP LEARNING ALGORITHMS TO DETECT DIABETIC KIDNEY DISEASE FROM RETINAL PHOTOGRAPHS IN MULTIETHNIC POPULATIONS WITH DIABETES," *Journal of the American Medical Informatics Association*, vol. 30, no. 12, pp. 1904–1914, Nov. 2023, doi: https://doi.org/10.1093/jamia/ocad179.

[108] S. Nusinovici *et al.*, "LOGISTIC REGRESSION WAS AS GOOD AS MACHINE LEARNING FOR PREDICTING MAJOR CHRONIC DISEASES," *J Clin Epidemiol*, vol. 122, pp. 56–69, Jun. 2020, doi: https://doi.org/10.1016/j.jclinepi.2020.03.002.

[109] D. Barsasella *et al.*, "A MACHINE LEARNING MODEL TO PREDICT LENGTH OF STAY AND MORTALITY AMONG DIABETES AND HYPERTENSION INPATIENTS," *Medicina (B Aires)*, vol. 58, no. 11, p. 1568, Oct. 2022, doi: https://doi.org/10.3390/medicina58111568.