Malaysian Journal of Computing, 10 (2): 2188-2202, 2025

eISSN: 2600-8238

PREVALENCE OF RISK FACTORS AND SEVERITY LEVEL OF DIABETIC RETINOPATHY IN UITM MEDICAL SPECIALIST CENTRE:AN ORDINAL ANALYSIS

Aniza Ayuna Che Jani[1], Zuraida Khairudin[2*], Nornadiah Mohd Razali[3], Nor Azimah Abd Aziz[4], Hezlin Aryani Abd Rahman[5], and Nur Syamimi Haji Abu Bakar[6]

[1,2,5,6]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia*

[3] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia

[4]Faculty of Medicine, Universiti Teknologi Mara Sg Buloh, Selangor, Malaysia

[1]aniza.ayuna97@gmail.com, [2*]zuraida_k@fskm.uitm.edu.my, [3]nornadiah@uitm.edu.my, [4]azimah80@uitm.edu.my, [5]hezlin@fskm.uitm.edu.my, [6]syamimi@uitm.edu.my

ABSTRACT

Diabetic Retinopathy (DR) has become a major concern as the number of diabetic patients who are diagnosed with DR has been on the rise each year. Hence, the adoption of a strategic approach is needed to determine the factors that contribute to the level of severity in DR patients. This research aims to determine the significant risk factors influencing the severity level of DR as well as to evaluate the classification rate between the severity level of 157 patients at UiTM Private Specialist Centre by employing the Ordinal Logistic Regression (OLR) analysis. The study revealed that there were 60 (38%) patients with Mild Non-Proliferative Diabetic Retinopathy (NPDR), 61 (39%) patients with moderate NPDR, 9 (6%) patients with severe NPDR and 27 (17%) patients with Proliferative Diabetic Retinopathy (PDR). On top of that, the severity level of DR was found to be influenced by the duration of diabetes melitus (DM) (p=0.005), nephropathy (p=0.011) and dyslipidemia (p=0.035). Patients who did not have nephropathy were 0.6 times less likely and patients who did not have the dyslipidemia were 0.7 times more likely to have the highest severity level of DR compared to lower severity levels. For the duration of DM, an increase in the duration was associated with an increase in the odds of having the highest severity level. Finally, the results showed that severity level 2 (Mild NPDR), severity level 3 (Moderate NPDR) and severity level 4 (Severe NPDR) were the most accurate categories predicted by the model. This study can contribute to the improvement of health among DR patients and provides alternatives to the hospital in giving treatments to the patients.

Keywords: *Diabetes, Diabetic Retinopathy (DR), Risk Factors, Ordinal Logistic Regression (OLR), Severity Level.*

Received for review: 02-09-2025; Accepted: 18-09-2025; Published: 01-10-2025

DOI: 10.24191/mjoc.v10i1.7135

1. Introduction

Diabetic Retinopathy (DR) is a diabetes complication marked by issues with the small blood vessels within the retina. These issues can manifest as microvascular abnormalities, cotton wool spots, increased vascular permeability, neovascularization, hemorrhages, and microaneurysms (Sultan et al., 2024). Caused by hyperglycemia and other metabolic disorders, DR results from the weakening of these blood vessels, which can lead to the leakage of fatty substances and capillary blockages. Researchers have found it difficult to precisely determine the role of hyperglycemia in the development of DR in patients (Ansari et al., 2022).

A significant study by Khairudin (2020) found that several factors are closely linked to the presence of diabetic retinopathy (DR) in diabetic patients. These include the duration of diabetes, the patient's HbA1C level, and the presence of other diabetes-related complications such as diabetic foot ulcers, nephropathy, and neuropathy. To some extent, DR may affect and limit daily activities including driving, reading, walking, and working, hence potentially impacting the psychological well-being of patients (Zhao et al., 2020). Diabetic eye disease is a leading cause of preventable vision loss and is negatively associated with health-related quality of life. Globally, 43.3 million individuals were predicted to be blind in 2020, with Asia having the greatest population (GBD 2021 Blindness and Vision Impairment Collaborators, 2021).

Based on the international clinical disease severity scale for diabetic retinopathy (DR), a widely accepted framework classifies the condition into six distinct stages (Ministry of Health Malaysia, 2011). This scale ranges from "No apparent retinopathy" to "Advanced Diabetic Eye Disease (ADED)," with four stages that include Mild, Moderate, Severe Non-Proliferative DR (NPDR) and Proliferative DR (PDR). Each of these stages presents with its own set of symptoms, indicating the varying levels of disease severity. While existing studies of DR have established, the lifestyle-related risk factors across different severity levels of DR remains insufficiently explored, particularly within local populations. Therefore, this paper seeks to investigate the risk factors of these different levels of DR. This paper is organized to start with medical data mining, methodology, results and findings, and conclusions.

2. Medical Data Mining

Data mining techniques are becoming essential for predicting the severity of diabetic retinopathy (DR) by extracting valuable, previously unseen knowledge from patient data. This approach has already demonstrated significant progress in the medical industry for diagnosing and managing a variety of diseases, including liver disorders (Ahn et al., 2021; Razali et al., 2020), heart disease (Khan et al., 2020), diabetes (Rastogi et al., 2023; Khan et al., 2021), brain injuries (Rodger, 2015), and chronic obstructive pulmonary disease (Abugabah et al., 2020; Brockway & Ahmed, 2024). The results from these methods can be applied in both scientific research and real-world clinical practice to improve the quality of healthcare for diabetic patients (Khairudin et al., 2020; Bora et al., 2021: Saxena et al.,2021).

There are quite a handful of studies on the prediction of diabetes mellitus (DM) using logistic regression techniques. Komi et al. (2017) used Support Vector Machine (SVM), Extreme Learning Machine (ELM), Gaussian Mixture Model (GMM), logistic regression, and Artificial Neural Network (ANN) in their study. Gholipour et al. (2018) performed both logistic regression and ANN in identifying Type 2 Diabetes Melitus (T2DM) risk factors; it turned out that the logistic regression model was better than ANN and clinically more comprehensible. Numerous studies used data mining techniques such as ordinal logistic regression (OLR) in predicting severity levels of DR (Essuman et al., 2021; Fitria et al., 2021; Khairudin et al., 2020).

Though data mining techniques have been applied widely (Aimran et al., 2022; Hashad et al., 2024; Ibrahim et al., 2024; Sunardi et al., 2023; Pala & Camurcu, 2014), the focus of this study is narrowed down to predicting the severity level of diabetic patients only. An analysis of the data could yield a more effective strategy, producing a simpler, more easily interpretable model for determining the severity of diabetic retinopathy (DR).

3. Methodology

3.1 Scope of Study

This study was conducted from January 2014 to December 2018 on patients who were diagnosed with Diabetes Mellitus (DM) at UiTM Medical Specialist Centre. The data comprised 157 diabetic retinopathy (DR) patients, with ages ranging from 28 to 79 years. The dependent variable was the severity level of DR. The independent variables were divided into three categories: the patients' demographic profile (age in years, gender, BMI in kg/m2 and ethnicity), patient's diabetic history (stroke, duration of diabetes, Systolic Blood Pressure (SBP) in mmHg, Diastolic Blood Pressure (DBP) in mmHg, the presence of hypertension, nephropathy, neuropathy, Coronary Artery Disease (IHD), Diabetic ketoacidosis (DKA) and diabetic foot ulcer) and patient's blood result (Glycosylated haemoglobin level in percent (HbA1C) and Dyslipidaemia). Diabetic Retinopathy (DR) has several severity stages, ranging from level 1 to level 6. For this study, patients with severity level 1 were not included since severity level 1 indicates "no apparent DR," where the retina is normal and free from any signs of the disease (Ministry of Health Malaysia, 2011). Similarly, severity level 6 was also excluded from the study because data for patients at this level were not available.For the OLR analysis, 70% of the data were used as the training set and the remaining 30% were used for model validation. Figure 1 below shows the theoretical framework of this study.

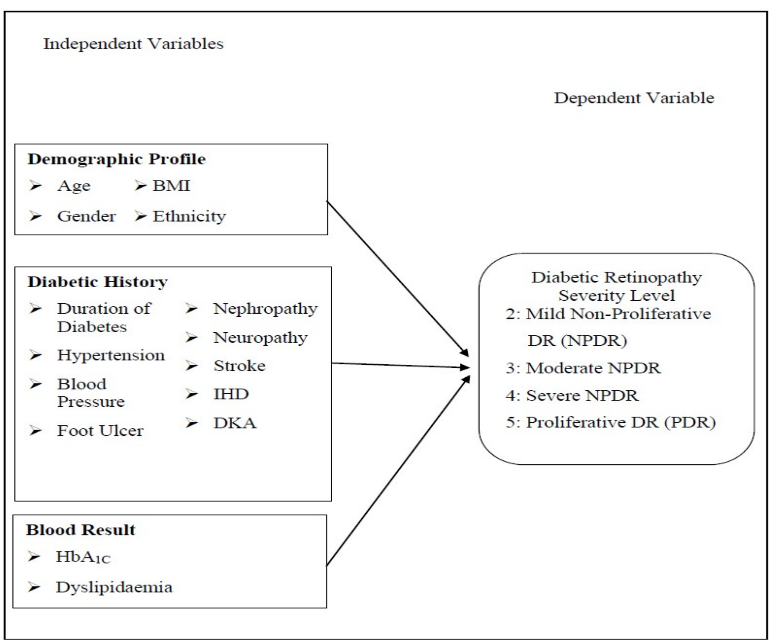

Figure 1. Research framework

3.2 Ordinal Logistic Regression (OLR)

This study utilized an Ordinal Logistic Regression (OLR) model, which is well-suited for analyzing data with an ordered, categorical dependent variable. This approach was appropriate for our research as the primary objective was to identify risk factors for diabetic retinopathy (DR) by using the ranked severity levels of the disease as the response variable. Below are the assumptions of OLR from Abreu et al. (2008):

- The dependent variable should be measured in ordinal level.
- Independent variables should be continuous, ordinal or categorical.
- No multicollinearity should be present among variables.
- The proportional odds assumption, where the logit surfaces are parallel and the odds ratio is constant across all possible cut points of the outcome, must be met.

If one of the above assumptions is violated, proportional odds model or multinomial logistic regression can be considered. The steps in OLR analysis involve Model Fitting Information, Goodness of Fit, Pseudo R-Square, Parameter Estimates, and Test of Parallel Lines (Ari & Yildiz, 2014). The parameter estimates produce coefficients that describe the relationship between the dependent and independent variables. Through a Wald test, the significance of the independent variable is determined by looking at the Wald statistic and its associated p-value. The odds ratio is also helpful in describing the relationship. The Wald statistics are the square of the ratio of the coefficient to the standard error. A high value of Wald statistic indicates that the corresponding independent variable is significant. The OLR collapsed the response categories into two sets, known as dichotomies, one for each possible cut-off of the dependent variables. In this study, the cut-off was:

- Severity level 2 versus (Severity level 3, Severity level 4 or Severity level 5).
- (Severity level 2, Severity level 3 or Severity level 4) versus severity level 5.

Therefore, the binary logistic model is defined for the log-odds of each cut-off, and the Ordinal Logistic Regression (OLR) model, also known as the cumulative probabilities model, can be derived from the parameter estimates (Harrell, 2015).

$$P(Y \leq j) = \frac{\exp\left[\alpha_j - (\beta_1 X_1 + \beta_2 X_2 + ..\beta_k X_k)\right]}{1 + \exp\left[\alpha_j - (\beta_1 X_1 + \beta_2 X_2 + ..\beta_k X_k)\right]} \tag{1}$$

where:

α_j is the threshold for the *j*th category, *j*=2,3,4 and 5

$\beta_1 \ldots \beta_k$ are the regression coefficients

$X_1 \ldots X_k$ are the independent variables

k is the number of independent variables

j is the category of dependent variables

3.3 Evaluation of Model Performance

The performance of the predictive models for severity level of DR patients was evaluated based on their misclassification rate, accuracy, sensitivity, specificity and also receiver operating curves (ROC). The main goal of the matrix is to determine which model is the most accurate and has the highest performances classification.

3.4 Confusion Matrix

The performance of the OLR model can be further investigated by using the confusion matrix. The number of correct and incorrect predictions made by the model compared with the actual classifications in the validation data is displayed in the confusion matrix (Peng, Lee, & Ingersoll, 2002). The accuracy, sensitivity, specificity, and misclassification rate are examined to identify the classification performance of the model. The matrix contains True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) as shown in (4), (5), and (6).

In this study, the severity levels were classified into four classes ω2, ω3, ω4 and ω5. Each class ωi has a conditional distribution p(x | ωi), and a prior probability P(ωi). Class assigned is based on Bayes Rule, which assigns membership to the highest posterior output as shown in (2):

$$P(\omega_1 \mid x) = \frac{P(\omega i)p(x \mid \omega i)}{P(\omega 2)p(x \mid \omega 2) + P(\omega 3)p(x \mid \omega 3) + P(\omega 4)p(x \mid \omega 4) + P(\omega 5)p(x \mid \omega 5)} \quad (2)$$

Thus, x was assigned according to (i=2,3,4,5):

$$\text{argmaxci=1P}(\omega i \mid x) \quad (3)$$

Accuracy is the probability of cases that are correctly classified which is calculated as the ratio between the number of cases correctly classified and the total number of cases. The formula is shown in equation (4):

$$\text{Accuracy} = \frac{TP_2 + TP_3 + TP_4 + TP_5}{TP + TN + FN + FP} \quad (4)$$

Sensitivity is defined as the proportion of true positive results that are correctly predicted by the model:

$$\text{Sensitivity (Severity Level } i\text{)} = \frac{TP_i i}{TPi + \varepsilon i_2 + \varepsilon i_3 + \varepsilon i_4 + \varepsilon i_5} \quad (5)$$

On the other hand, Specificity is defined as the proportion of true negative results that are correctly predicted by the model:

$$\text{Specificity (Severity Level } i\text{)} = \frac{TN_1}{TNi + \varepsilon_2 i + \varepsilon_{31} + \varepsilon_{41} + \varepsilon_{51}} \quad (6)$$

Misclassification rate refers to the error indicating the probability of cases being incorrectly classified. This error is calculated as shown in equation (7) (Jenness & Wynne, 2005).

$$\text{Misclassification Rate (Severity Level } i) = \frac{\varepsilon_{12}+\varepsilon_{13}+\varepsilon_{14}+\varepsilon_{21}+\varepsilon_{31}+\varepsilon_{41}+\varepsilon_{51}}{TP_1+TN_1+\varepsilon_{12}+\varepsilon_{13}+\varepsilon_{14}+\varepsilon_{21}+\varepsilon_{31}+\varepsilon_{41}+\varepsilon_{51}} \quad (7)$$

3.4.1 Receiver Operating Curves (ROC)

According to Fawcett (2006), the ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1-specificity) at various cut-off points. When the curve is closer to the left-hand border and the top border of the ROC space, the test is more accurate. When the curve gets closer to 45-degree diagonal of the ROC curve, the test is less accurate. Accuracy is measured based on area under the ROC curve (AUC) where perfect classifier gives AUC of 1.

4. Results and Findings

4.1 Descriptive Analysis

The distribution of diabetic patients by their severity levels was summarized in Figure 2. The figure indicated that there were 60 (38%) patients with mild NPDR, 61 (39%) patients with moderate NPDR, 9 (6%) patients with severe NPDR and 27 (17%) patients with PDR.

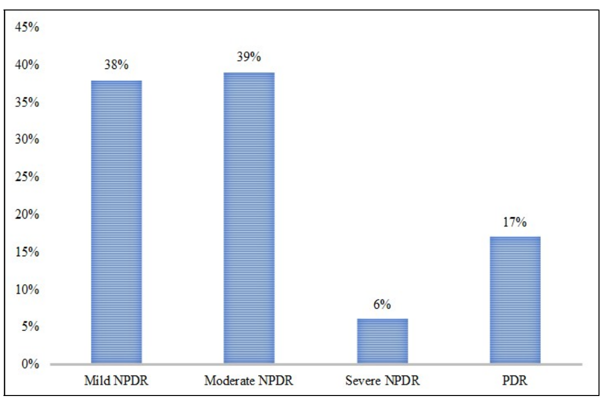

Figure 1. Four Levels of Diabetic Retinopathy

Severity level 3 or moderate NPDR showed the highest prevalence among all severity levels. Referring to the Table 1, the study sample consisted predominantly of males (59.9%) and Malays (78.3%), with the majority being married (98.1%). Dyslipidemia was the most common comorbidity, affecting 83.4% of participants, followed by ischemic heart disease (IHD) (43.3%), nephropathy (35.7%), and neuropathy (27.4%). Stroke (5.7%) and foot ulcer (12.1%) were less frequent. Across all stages of diabetic retinopathy, males and Malays remained the largest groups, while comorbidities such as dyslipidemia, ischemic heart disease, nephropathy, and neuropathy were consistently observed. Stroke and foot ulcer were relatively uncommon but appeared more often in moderate NPDR and proliferative DR cases.

Table 1. Summary Statistics for Qualitative Variables

Variable	n (%)	Mild NPDR n (%)	Moderate NPDR n (%)	Severe NPDR n (%)	PDR n (%)
Gender					
Male	94 (59.9)	36 (22.9)	38 (24.2)	5 (3.2)	15 (9.6)
Female	63 (40.1)	24 (15.3)	23 (14.6)	4 (2.5)	12 (7.6)
Ethnicity					
Malay	123 (78.3)	46 (29.3)	50 (31.8)	8 (5.1)	19 (12.1)
Chinese	16 (10.2)	7 (4.5)	4 (2.5)	1 (0.01)	4 (2.5)
Indian	16 (10.2)	6 (3.8)	7 (4.5)	0 (0.0)	3 (1.9)
Others	2 (1.3)	1 (1.3)	0 (0.0)	0 (0.0)	1 (0.01)
Marital Status					
Single	2 (1.3)	1 (0.01)	0 (0.0)	0 (0.0)	1 (0.01)
Married	154 (98.1)	59 (37.6)	60 (38.2)	9 (5.7)	26 (16.6)
Widowed	1 (0.6)	0 (0.0)	1 (0.01)	0 (0.0)	0 (0.0)
Neprophathy					
Yes	56 (35.7)	14 (8.9)	25 (15.9)	4 (2.5)	13 (8.3)
No	101 (64.3)	46 (29.3)	36 (22.9)	5 (3.2)	14 (8.9)
Neurophathy					
Yes	43 (27.4)	15 (9.6)	19 (12.1)	1 (0.6)	8 (5.1)
No	114 (72.6)	45 (28.7)	42 (26.8)	8 (5.1)	19 (12.1)
Stroke					
Yes	9 (5.7)	2 (1.3)	6 (3.8)	0 (0.0)	1 (0.6)
No	148 (94.3)	58 (36.9)	55 (35.0)	9 (5.7)	26 (16.6)
IHD					
Yes	68 (43.3)	23 (14.6)	24 (15.3)	7 (4.5)	14 (8.9)
No	89 (56.7)	37 (23.6)	37 (23.6)	2 (1.3)	13 (8.3)
Foot Ulcer					
Yes	19 (12.1)	6 (3.8)	8 (5.1)	0 (0.0)	5 (3.2)
No	138 (87.9)	54 (34.4)	53 (33.8)	9 (5.7)	22 (14.0)
Dyslipidemia					
Yes	131 (83.4)	54 (34.4)	50 (31.8)	7 (4.5)	20 (12.7)
No	26 (16.6)	6 (3.8)	11 (7.0)	2 (1.3)	7 (4.5)

From Table 2, the average Body Mass Index (BMI) was 28.74 (SD = 5.12), ranging from 20 to 49, indicating that most participants were overweight or obese. The mean age of participants was 57.85 years (SD = 9.60), with ages ranging between 28 and 79 years. This suggests that the sample mainly consisted of middle-aged to older adults. The duration of diabetes mellitus (DM) averaged 12.17 years (SD = 7.41), with a minimum of 1 year and a maximum of 40 years, showing that participants had varied lengths of living with diabetes. For blood pressure, the mean systolic BP was 140.06 mmHg (SD = 18.67) and ranged from 88 to 198 mmHg, while the mean diastolic BP was 77.96 mmHg (SD = 12.49) with a range of 46 to 123 mmHg. This indicates that many participants were hypertensive. The mean HbA1c level was 9.18% (SD = 2.15), with values between 5.5% and 16.1%. This suggests poor glycemic control on average among participants.

Table 2. Summary Statistics for Quantitative Variables

Variable	n	Minimum	Maximum	Mean	Std. Deviation
BMI	157	20	49	28.74	5.124
Age	157	28	79	57.85	9.598
Duration_of DM	157	1	40	12.17	7.406
SBP	157	88	198	140.06	18.665
HBA1C	157	5.5	16.1	9.181	2.1495
DBP	157	46	123	77.96	12.494

Table 3 recorded a negative value for the skewness of age which indicated that the distribution of age was skewed to the left. Therefore, the normality assumption was not met. Hence, the median and inter quartile range *(IQR)* were used to describe the variable. The distribution of DR patients according to their age group was shown in Table 4. The table shows that most of the DR cases (52.9%) were contributed by the patients who were less than 59 years old.

Table 3. Summary Statistics

Summary Statistics for Age and Severity Level		
Variable	**Skewness**	**Median (SD)**
Age	-0.698	59.00 (9.598)
Severity Level	0.911	

Table 4. Percentage Of DR Patients by Age Group

Age (years)	Percentage
<59	52.9
≥59	47.1

4.2 Ordinal Logistic Regression (OLR)

The negative log-log link function was employed since the distribution of ordinal outcomes was positively skewed (0.911) as shown in Table 5. This section presents the results of the estimated model. In Table 5, a statistically significant chi-square test ($p = 0.042$) indicated that the model with independent variables gives a significant improvement compared to the model with intercept only.

Table 5. Model Fitting Information

Model	-2LL	Chi-Square	df	Sig.
Intercept Only	362.683			
Final	334.658	28.026	18	.042

In addition, a nonsignificant chi-square statistic ($p = 0.540$), as shown in Table 6, indicated that the observed data were consistent with the estimated values in the fitted model. This means that, the negative log-log link model obtained fits the data adequately.

Table 6. Goodness Of Fit Statistic

Goodness of Fit	Chi-Square	df	Sig.
Pearson	438.816	435	0.540

Meanwhile in Table 7, the Negelkerke *R*-Square and McFadden values of 1.000 indicated that 100% of the variability in the outcome was explained by the independent variables.

Table 7. Pseudo R-Square

Cox and Snell	Negel kerke	McFa dden
0.917	1.000	1.000

4.3 Evaluation of Model

The actual and classified response was categorized using the classification table. Table 8 shows the classification result of accuracy for the severity level of DR using the negative log-log link. The results pointed out that the prediction accuracy was 100% for each severity level of DR when the three significant factors which were Duration of DM, nephropathy and dyslipidemia were included in the model. Hence, the overall fit of the model yielded 100.0% correct classifications**.**

Table 8. Classification Table

Actual	Classification Result			
	Predicted			
	Severity Level 2	Severity Level 3	Severity Level 4	Severity Level 5
Severity Level 2 (%)	20 (100.0)	0 (33.3)	0 (0)	0 (0)
Severity Level 3 (%)	0 (0)	14 (100.0)	0 (0)	0 (0)
Severity Level 4 (%)	0 (0)	0 (0)	3 (100.0)	0 (0)
Severity Level 5 (%)	0 (0)	0 (0)	0 (0)	6 (100.0)

Table 9 displays the results of the parameter estimates. At 5% significance level, the severity level of DR was found to be significantly associated with the duration of DM (p=0.005), and the presence of nephropathy (p=0.011) and dyslipidemia (p=0.035). A positive parameter estimate obtained for the duration of DM indicated that an increase in the duration of DM was associated with an increase in the odds of having the highest severity level (severity level 5) compared to the lower severity levels (severity level 2, 3 and 4). In addition, the odds ratio for the presence of nephropathy ($e^{-0.537} = 0.6$) indicated that patients who did not have the nephropathy were 0.6 times less likely to have the highest severity level of DR compared to lower severity levels (severity levels 2, 3 and 4). For the presence of dyslipidemia, patients who did not have the dyslipidemia were 1.73 times ($e^{0.549} = 1.73$) more likely to have the highest severity level of DR compared to lower severity levels (severity levels 2, 3 and 4).

Table 9. Parameter Estimates

Variable	**Parameters Estimate**				
	Estimate	**Std. Error**	**Wald**	**df**	**Sig.**
[Severity Level = 2]	0.245	0.250	0.958	1	0.000
[Severity Level = 3]	1.617	0.289	31.414	1	0.000
[Severity Level = 4]	1.952	0.305	40.952	1	0.000
Duration of DM	0.038	0.014	7.775	1	0.005
Nephropathy (No)	-0.537	0.212	6.430	1	0.011
Dyslipidemia (No)	0.549	0.261	4.432	1	0.035

Based on the parameter estimates, the estimated model can be written as follows,

$$Z_2 = 0.245 + 0.038(\text{Duration of Diabetes Melitus}) - 0.537(\text{Neprophaty}) + 0.549(\text{Dyslipidemia})$$
$$Z_3 = 1.617 + 0.038(\text{Duration of Diabetes Melitus}) - 0.537(\text{Neprophaty}) + 0.549(\text{Dyslipidemia})$$
$$Z_4 = 1.952 + 0.038(\text{Duration of Diabetes Melitus}) - 0.537(\text{Neprophaty}) + 0.549(\text{Dyslipidemia})$$

where Z_2 = Severity Level 2, Z_3 = Severity Level 3, Z_4 = Severity Level 4.

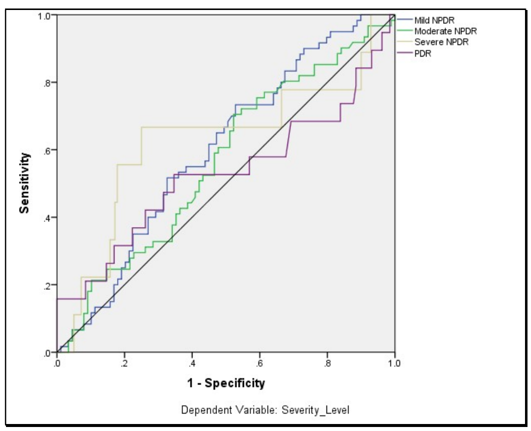

Figure 2. ROC Curve for Severity Level of Diabetic Retinopathy

As illustrated in Figure 3, the OLR model exhibited superior discriminatory ability for severity level 2 (Mild NPDR), severity level 3 (Moderate NPDR) and severity level 4 (Severe NPDR). The ROC curves for these categories were positioned closer to the upper-left boundary of the plot, indicating higher sensitivity and specificity. Conversely, the discriminatory performance for severity level 5 was comparatively weaker, as their ROC curves aligned more closely with the 45° reference line, reflecting limited predictive accuracy.

5. Conclusion

A study by Khan et al. (2017) indicated that there was a statistically significant difference between race and diabetic status. A study from one of the Asian countries where the participants were Malay, Indian, and Chinese aged 40 years and above found that Indian ethnicity had a higher prevalence of diabetic retinopathy compared to Chinese and Malay (Tan et al., 2018). This study, on the other hand, showed that the Malay community had the highest prevalence of diabetic retinopathy compared to other major races in Malaysia. The results obtained in the study seem to be consistent with the previous study. The Non-Communicable Disease (NCD) Risk Factor Collaboration (2016) reported that females have more chances of having DR compared to males. However, based on this study, the number of male patients (59.9%) and female patients (40.1%) has caused the influence of male patients to be more detectable with a ratio of 1.49 to 1.

From the Ordinal Logistic Regression (OLR) model, the parameter estimates showed that the duration of diabetes mellitus, nephropathy and dyslipidemia were found to be significant predictors of the severity level of diabetic retinopathy. Worldwide studies have revealed the same outcomes where the severity level of diabetes mellitus was strongly associated with nephropathy (Patel et al., 2020; Ali et al., 2020). A study using data from the 2008–2011 Korea National Health and Nutrition Examination Survey discovered that diabetic retinopathy was significantly associated with the duration of diabetes mellitus. The risks were significantly higher in patients with diabetes less than 10 years. The risk of developing diabetic retinopathy was higher among patients with more than 5 hours of daily sunlight exposure (Lee, Kim, & Lee, 2020). Meanwhile, dyslipidemia is a characteristic feature of diabetes complications involving high plasma triglyceride concentration, increased concentrations of Low-Density lipoprotein (LDL) cholesterol particles, and High Density Lipoprotein (HDL) cholesterol.

The result of parameter estimate (Table 9) shows that threshold severity level 2 is (1/0.245) ~ 4.082 times less occurring than severity level 5. While, threshold level 3 is 1.617 times more likely to occur as compared to severity level 5. Threshold severity level 4 is 1.952 more likely to occur compared to severity level 5.

A study involving DR and severity is a retrospective study that relies on medical records. Thus, some of the patients' details were not available or recorded in the medical records. In addition, majority of diabetic patients in this study were Malays and therefore, other ethnic groups were not equally covered. Thus, this study may not be representative to a larger community. For future studies, it is suggested to incorporate more risk factors to the model such as family history, smoking habits, anemia and asthma and fasting blood sugar so that it can provide a better overview of the disease. More samples should also be added to the study by considering appropriate size of ethnicity. In the future study, Artificial Neural Network (ANN) is recommended for analysis. Model comparison could be considered in the study to find the best model for predicting the severity level of DR.

The analysis provided effective strategy on monitoring the factors contributed to the DR. Based on the OLR model, it shows that the only significant factors associated with severity level of DR were the duration of DM, nephropathy and dyslipidemia. This means that the DR patients at UiTM Medical Specialist Centre are prone to develop retinopathy disease when they have been diagnosed with a long-term DM, with the presence of nephropathy and dyslipidemia. It is advisable for those who work in Ophthalmology Clinics to set up frequent treatment sessions from time to time to monitor patients' level of illness in order to decrease the number of patients who suffer from a more serious retinopathy. Help and support from patient's family and friends can also be beneficial for the patients. Giving the patients some advice to follow treatments that was set up by the hospital or clinic could help in controlling the patient's severity level and avoiding it from getting worse.

Acknowledgement

The authors would like to thank Faculty of Medicine, UiTM Sungai Buloh for providing the data. The authors declare no competing interest. Aniza collected and analysed the data for the study. Zuraida Khairudin and Hezlin Aryani Abd Rahman drafted the manuscript. Nornadiah Mohd Razali and Nur Syamimi Haji Abu Bakar have participated in revision of the intellectual content. Also, thanks to Dr Azimah from Faculty of Medicine, UiTM for her expert opinion. All authors have read and approved the final manuscript.

Funding

Research NEXUS Universiti Teknologi MARA (ReNeU), Shah Alam, Selangor.

Author Contribution

Aniza collected and analysed the data for the study. Zuraida Khairudin and Hezlin Aryani Abd Rahman drafted the manuscript. Nornadiah Mohd Razali and Nur Syamimi Haji Abu Bakar have participated in revision of the intellectual content. Lastly, Dr Azimah from Faculty of Medicine, UiTM for her expert opinion.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

Abreu, M. N. S., Siqueira, A. L., Cardoso, C. S., & Caiaffa, W. T. (2008). Ordinal logistic regression models: Application in quality of life studies. Cadernos de Saúde Pública, 24(4).

Abugabah, A., AlZubi, A. A., Al-Obeidat, F., Alarifi, A., & Alwadain, A. (2020). Data mining techniques for analyzing healthcare conditions of urban space-person lung using meta-heuristic optimized neural networks. *Cluster Computing*, *23*, 1781-1794.

Ahn, J. C., Connell, A., Simonetto, D. A., Hughes, C., & Shah, V. H. (2021). Application of artificial intelligence for the diagnosis and treatment of liver diseases. *Hepatology*, *73*(6), 2546-2563.

Aimran, N., Rambli, A., Afthanorhan, A., Mahmud, A., Sapri, A., & Aireen, A. A. (2022). Prediction of Malaysian women divorce using machine learning techniques. Malaysian Journal of Computing (MJoC), 7(2), 1067–1081.

Ali, M., Ejaz, A., Solangi, S. A., et al. (2020). Vitamin D deficiency in end-stage renal disease patients with diabetes mellitus undergoing hemodialysis. Cureus, 12, e11668. https://doi.org/10.7759/cureus.11668

Ansari, P., Tabasumma, N., Snigdha, N. N., Siam, N. H., Panduru, R. V., Azam, S., & Abdel-Wahab, Y. H. (2022). Diabetic retinopathy: an overview on mechanisms, pathophysiology and pharmacotherapy. *Diabetology*, *3*(1), 159-175.

Ari, E., & Yildiz, Z. (2014). Parallel lines assumption in ordinal logistic regression and analysis approaches. International Interdisciplinary Journal of Scientific Research, 1(3), 8–23.

Bora, A., Balasubramanian, S., Babenko, B., Virmani, S., Venugopalan, S., Mitani, A., ... & Bavishi, P. (2021). Predicting the risk of developing diabetic retinopathy using deep learning. *The Lancet Digital Health*, *3*(1), e10-e19.

Brockway, K., & Ahmed, S. (2024). Aligning Recommendations for Pain Management in Patients with Chronic Obstructive Pulmonary Disease (COPD): A Systematic Review of Clinical Practice Guidelines. *Archives of Physical Medicine and Rehabilitation*, *105*(4), e153-e154.

Essuman, V. A., Tagoe, N. N., Akpalu, J., et al. (2021). Morbidity and complications of diabetes mellitus in children and adolescents in Ghana: Protocol for a longitudinal study. JMIR Research Protocols, 10(1).

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861–874.

Fitria, R., Yulisda, D., & Ula, M. (2021). Data Mining Classification Algorithms For Diabetes Dataset Using Weka Tool. *Sisfo: Jurnal Ilmiah Sistem Informasi*, *5*(2).

GBD 2021 Blindness and Vision Impairment Collaborators. (2021). Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study. The Lancet Global Health, 9, e130–e143.

Gholipour, K., Jafarabadi, M. A., Iezadi, S., Jannati, S., & Keshavarz, S. (2018). Modeling the prevalence of diabetes mellitus risk factors based on artificial neural network and multiple regression. Eastern Mediterranean Health Journal, 24(8), 770–777.

Harrell, F. (2015). Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

Hashad, A. A., Wah, K. K., Alnoor, A., & Chew, X. (2024). Exploratory analysis with association rule mining algorithms in the retail industry. Malaysian Journal of Computing (MJoC), 9(1), 1746–1758. https://doi.org/10.24191/mjoc.v9i1.21433

Ibrahim, N., Ishak, U. M., Ali, N. N. A., & Shaadan, N. (2024). Machine learning-based approaches for credit card debt prediction. Malaysian Journal of Computing (MJoC), 9(1), 1722–1751.

Jenness, J., & Wynne, J. J. (2005). Cohen's Kappa and classification table metrics 2.0: An ArcView 3x extension for accuracy assessment of spatially explicit models. Science for a Changing World, 1363, 1–86.

Khairudin, Z., Abdul Razak, N. A., Abd Rahman, H. A., Kamarudin, N., & Abd Aziz, N. A. (2020). Prediction of diabetic retinopathy among type II diabetic patients using data mining techniques. *Malaysian Journal of Computing (MJoC)*, *5*(2), 572-586.

Khan, A. R., Nisha, D. S. S., & Sathik, D. M. M. (2020). Researcher reviews on data mining techniques for medical dataset: A survey. Mukt Shabd Journal, 9(6), 7646–7651.

Khan, F. A., Zeb, K., Al-Rakhami, M., Derhab, A., & Bukhari, S. A. C. (2021). Detection and prediction of diabetes using data mining: a comprehensive review. *IEEE Access*, *9*, 43711-43735.

Khan, H. M. R., Mende, S., Rafiq, A., Gabbidon, K., & Reddy, P. H. (2017). Methods needed to measure predictive accuracy: A study of diabetic patients. Biochimica et Biophysica Acta.

Komi, M., Li, J., Zhai, Y., & Zhang, X. (2017). Application of data mining methods in diabetes prediction. In 2nd International Conference on Image, Vision and Computing (ICIVC) (pp. 1006–1010).

Lee, H. J., Kim, C. O., & Lee, D. C. (2020). Association between daily sunlight exposure duration and diabetic retinopathy in Korean adults with diabetes. PLoS ONE, 15, e0237149.

Awang Nawi, M.A., Ahmad, W. M. A. W., Mat Yudin, Z., & Abdul Rohim, R. A. (2019). Determining the associated factors related to diabetes mellitus type II by using multiple logistic regression in Malaysia. International Journal of Public Health and Clinical Sciences, 6(2), 172–179.

Ministry of Health Malaysia. (2011). Clinical Practice Guidelines: Screening for diabetic retinopathy (Vol. 19, No. 3).

NCD Risk Factor Collaboration. (2016). Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis. The Lancet, 387(10026), 1377–1396.

Pala, T., & Camurcu, A. Y. (2014). Evaluation of data mining classification and clustering techniques for diabetes. Malaysian Journal of Computing (MJoC), 2(1), 37–45.

Patel, Z., Rathod, A., Kothari, R. N., Uyas, A., & Puthak, M. (2020). Study of clinical correlation of diabetic retinopathy with diabetic nephropathy and diabetic neuropathy in patients with type 2 diabetes mellitus. Indian Journal of Clinical and Experimental Ophthalmology, 6(2), 295–299. https://www.ipinnovative.com/media/journals/IndianJClinExpOphthalmol-6-2-295-299.pdf

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. ResearchGate, 96(1).

Rastogi, R., & Bansal, M. (2023). Diabetes prediction model using data mining techniques. *Measurement: Sensors*, *25*, 100605.

Razali, N., Mustapha, A., Abd Wahab, M. H., Mostafa, S. A., & Rostam, S. K. (2020). A data mining approach to prediction of liver diseases. In *Journal of Physics: Conference Series* (Vol. 1529, No. 3, p. 032002). IOP Publishing.

Rodger, J. A. (2015). Discovery of medical big data analytics: Improving the prediction of traumatic brain injury survival rates by data mining Patient Informatics Processing Software Hybrid Hadoop Hive. Informatics in Medicine Unlocked, 1, 17–26.

Saxena, A., Jain, M., & Shrivastava, P. (2021). Data mining techniques based diabetes prediction. Indian Journal of Artificial Intelligence and Neural Networking, 1(2), 29–35. https://doi.org/10.35940/ijainn.b1012.041221.

Sultan,S., Riaz, M. (2024). Chapter 8 - Microvascular complications: Diabetic eye disease. BIDE's Diabetes Desk Book. 181-199.Elsevier.

Sunardi, F., Fadlil, A., & Kusuma, N. M. P. (2023). Cyber fraud profiling with routine activity theory using data mining techniques. Malaysian Journal of Computing (MJoC), 8(2), 1517–1533.

Tan, G. S., Gan, A., Sabanayagam, C., Tham, Y. C., Neelam, K., & Wang, J. J. (2018). Ethnic differences in the prevalence and risk factors of diabetic retinopathy: The Singapore Epidemiology of Eye Diseases Study. Ophthalmology.

Zhao, L., Ren, H., Zhang, J., Cao, Y., Wang, Y., Meng, D., Wu, Y., & Zhang, R. (2020). Diabetic retinopathy, classified using the lesion-aware deep learning system, predicts diabetic end-stage renal disease in Chinese patients. Rapid Electronic Articles in Press.