SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Application of Intermediate Level Data Fusion to Improve Classification of Diabetic Retinopathy

Syafawati Ab Saad[1,*], Maz Jamilah Masnan[1,2], Safwati Ibrahim[1], Karniza Khalid[3], Dodi Devianto[4]

[1] Institute of Engineering Mathematics, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[2] Centre of Excellence for Advanced Sensor Technology (CEASTech), Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia
[3] Special Protein Unit, Specialized Diagnostic Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, 50588 Kuala Lumpur, Malaysia
[4] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Andalas University, Limau Manis Campus, Padang 25163, Indonesia

ABSTRACT

This study highlights the application of intermediate level data fusion to improve the classification of diabetic retinopathy stages among type 2 diabetes mellitus patients. Intermediate level data fusion was applied to analyse the demographic factors, clinical predictors, and risk factors for diabetic retinopathy and diabetic nephropathy, independently. The investigation focuses on the two diseases due to their inter-relation implication towards diabetes patients after certain period. Two models namely baseline model and mean model for the clinical predictors were applied in modelling the classification rules using ordinal logistic regression. The aim of the study is to evaluate the performance of the selected classification rule based on different sets of significant predictors from diabetic retinopathy, diabetic nephropathy, and the fusion of both predictors. The developed classification models with different combinations of predictors were tested to confirm the best model to classify diabetic retinopathy stages among type 2 diabetic patients who are at risk of retinopathy. In conclusion, intermediate level data fusion based on the baseline model shows better classification performance in classifying the stages of diabetic retinopathy.

Keywords:
Diabetic retinopathy; Diabetic nephropathy; Intermediate level data fusion; Ordinal logistic regression

1. Introduction

One of the earliest and mostly applied data fusion frameworks was the JDL process model, developed in the mid-1980s by the U.S. Joint Directors of Laboratory (JDL) Data Fusion Working Group (DFS). Later, the JDL-DFS group was established in 1986 to unify the terminology and procedure in data fusion where eventually the JDL process model was developed and has gained attention from multidisciplinary researchers [1]. The JDL process model is a conceptual model which identifies the process, functions, categories of techniques, and specific techniques applicable to data fusion through different function levels. To implement the data fusion system, one of the key issues is to decide where in the data flow does the fusion takes place. Usually, fusion could occur either at

* *Corresponding author.*
E-mail address: syafawatisaad@unimap.edu.my

https://doi.org/10.37934/araset.59.2.1019

the raw data level (prior to feature extraction), at the feature vector level (prior to identity declaration), or [2] at the decision level (after each input has made an independent declaration of identity) [1]. Simply, these fusion processes can be mapped to low level data fusion or data level fusion, followed by intermediate level data fusion or feature level fusion, and the final is high level data fusion or decision level fusion. Recently, data fusion framework has become an important mechanism to boost robustness and improved classification performance. One of the areas that gain interest is in medical health management such as data fusion techniques and technologies for wearable health monitoring [3] clinical data fusion for smart healthcare [4] review in human activity detection and health monitoring [5] and managing clinical data using data fusion approach [6].

Recent decades have witnessed the rise in the prevalence of type 2 diabetes mellitus globally. More than 463 million people worldwide have been diagnosed with diabetes, and the number is predicted to escalate to 700 million by 2045 [7]. Unfortunately, Malaysia is one of the countries facing this threat and has recorded the highest prevalence of diabetes among Western Pacific regions [8]. Diabetes mellitus among adults may lead to disabilities and reduce quality of life and may prolong microvascular and macrovascular complications if it is not treated well. Microvascular complications include diabetic retinopathy, diabetic nephropathy, and diabetic neuropathy, whereas macrovascular diseases include coronary heart disease, cardiomyopathy, arrhythmias-sudden death, cerebrovascular disease, and peripheral arterial disease [9]. Diabetic retinopathy refers to the vascular changes in the retina occurring as a complication from poorly controlled diabetes [10]. On the other hand, diabetic nephropathy is characterized by albuminuria, hypertension and progressive renal failure [11].

For these reasons, the risk of vision loss for diabetes patients with diabetic retinopathy and nephropathy increases with the progression of these diseases. Kim *et al.,* [12] revealed that diabetic retinopathy had a 4.37-times greater chance of having diabetic nephropathy than people without diabetic retinopathy. Previous studies have suggested a hypothetical relationship between diabetic retinopathy and nephropathy [13,14]. They share similar predictors, such as duration of diabetes, age of onset of diabetes, gender and HbA1c readings [15-17]. Patients with diabetic retinopathy are more likely to acquire diabetic nephropathy than those with diabetic nephropathy, and vice versa [11]. Even though many studies have begun to relate predictors and risk factors of diabetic retinopathy and nephropathy among diabetes patients, less research discusses these issues in the direction of the linkage between the two diseases of diabetic retinopathy and nephropathy simultaneously [11,17-19]. Interactions among diabetes complications should be identified and early diagnosis.

Since diabetes patients may suffer from diabetic complications such as diabetic retinopathy and nephropathy, diagnosis using factors or predictors is one of the classification problems [20]. The development of classification models for diabetes complications were recently investigated [20-23]. Many researchers have developed various classification models for diabetes complications using different techniques such as the Naives' Bayes Algorithm, Decision Tree, K-Nearest Neighbor, neural network, logistic regression [23-26]. Among those techniques, logistic regression model consistently gives higher accuracy than the other classifier in diabetes complications diagnosis [26].

Even though classification modelling for diabetes complications has recently recognizable, very rare research that relates data fusion specifically intermediate level data fusion with the classification of diabetic retinopathy's stages. Since diabetic retinopathy and diabetic nephropathy are linked to each other, this study adapted intermediate level data fusion to capture significant demographic factors, clinical predictors and risk factors of diabetic retinopathy and nephropathy to improve the classification performance of diabetic retinopathy's stages using Ordinal Logistic Regression.

2. Methodology

2.1 Data Collection

This study involved type 2 diabetes mellitus (T2DM) patients who have developed either diabetic retinopathy or nephropathy, or those who have already developed diabetic retinopathy and nephropathy, regardless of which one occurred first. A total sample of 377 T2DM patients were selected. Data were collected from outpatient medical records beginning from 1st January 2019 until 31st December 2020 from two hospitals in the northern region of Malaysia. The variables included in the data collection were partly accustomed according to the hospital format in the data recording which can be divided into three categories namely demographic factors, clinical predictors, and risk factors.

This study focuses on two types of clinical predictors data which are the Baseline data and Mean data. Baseline data refers to the medical result of each clinical predictor diagnosed during patients' first visit at the respective health clinic or outpatient department. At baseline, medical history in detailed were taken including the previous treatment of diabetes, hypertension and smoking habit [12]. A comprehensive medical examination will be taken during the first visit for Baseline data. The Baseline data is considered important to this study because at the first visit, a medical examination will be taken comprehensively before diagnosing the stages either diabetic retinopathy or nephropathy to the patients. While Mean data refers to the clinical data that is based on the average value calculated from the clinical data records from patients' overall visits at the respective health clinic.

2.2 Classification Modelling

The aim of this study is to develop a classification model using Ordinal Logistic Regression to classify stages of diabetic retinopathy using fused subset predictors from diabetic retinopathy and nephropathy. Due to limited previous studies on data fusion related to diabetes complication, this study begins by implementing independent predictor selection for each disease just to make sure the process of finding the best subset of predictors from a large pool of predictors is achieved. Let $X_{DR} = \left[x_1, x_2, ..., x_P\right]$ be the pool of all predictors from diabetic retinopathy denoted and $X_{DN} = \left[x_1, x_2, ..., x_Q\right]$ be the pool of predictors for diabetic nephropathy. Let the dependent variable is the stages of diabetic retinopathy, $\left(Y_{DR}\right)$ that are no diabetic retinopathy (stage 0), mild NPDR (stage 1), moderate NPDR (stage 2), severe NPDR (stage 3), PDR (stage 4) and ADED (stage 5), $Y_{DR} = \left[0,1,2,3,4,5\right]$. The process of selecting and evaluating the pool of predictors from both diseases was performed independently, $X_{DR} = \left[x_1, x_2, ..., x_P\right]$ and $X_{DN} = \left[x_1, x_2, ..., x_Q\right]$, as well as using the intermediate level data fusion method, $X_{DR+DN} = [x_1, x_2, ..., x_p] + [x_1, x_2, ..., x_q]$.

Three different sets of predictors involved in this study are; distinct predictors for diabetic retinopathy, set B, denotes as $X_{DR} = [x_1, x_2, ..., x_p]$, distinct predictors for diabetic nephropathy, Set C, denotes as $X_{DN} = [x_1, x_2, ..., x_q]$; and the shared predictors of diabetic retinopathy and nephropathy, Set A, represents as $X_{DR+DN} = [x_i]$ where $i = 1,2,...,r$, with the condition that $r < p$ and $r < q$. Thus, the development of classification model is based on the ILDF with different feeder combination such as [A + B], [A + C] and [A + B + C]. The case of combination [A + B] is the fused subset predictors for shared predictors from both diseases with subset predictors of diabetic retinopathy. While the combination of [A + C] reflects the fused subset predictors for shared predictors from both diseases

with the subset predictors of diabetic nephropathy. Finally, for the combination of [A + B + C], it represents the fused subset predictors for shared predictors from both diseases and the subset predictors of diabetic retinopathy and nephropathy. The details of set predictors are illustrated in Figure 1.

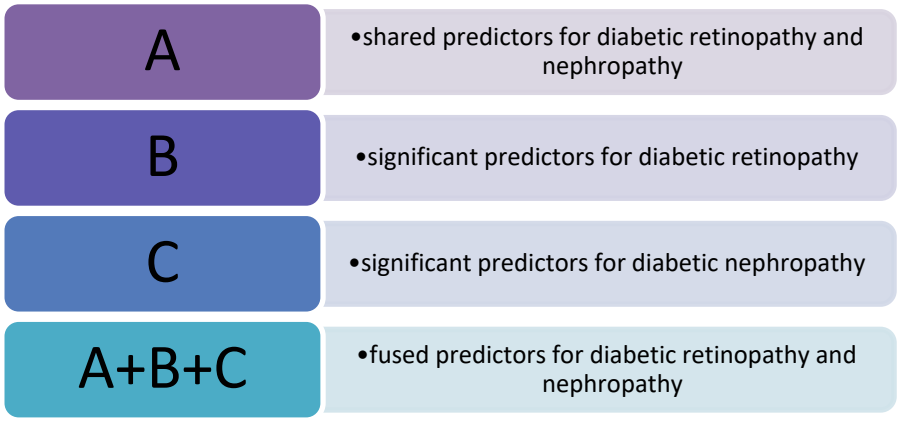


Fig. 1. Sets of predictors for ILDF

2.2.1 Intermediate level data fusion (ILDF)

ILDF is one of the methods from data fusion. Data fusion can be classified into three levels which are the low-level data fusion (LLDF), intermediate level data fusion (ILDF) and high-level data fusion (HLDF). However, for the purpose of this research, the ILDF method was considered suitable due to its ability to let each disease recommend the subset of useful predictors independently. This study implemented the ILDF to fuse the significant predictors from diabetic retinopathy and nephropathy. Bear in mind that for our data collection from the two different hospitals, the same demographic factors, clinical predictors, and risk factors were recorded as variables for each disease. Thus, the implementation of ILDF would clarify which of these predictors are associated to diabetic retinopathy and nephropathy, similarly and differently. There are possibilities that these diseases share similar predictors, and this leads to a certain explanation. There are also high chance that each disease may lead to dissimilar subset of predictors which explained the disease further detail. But what most important for this study is the contribution of significant subset of predictors identified from diabetic nephropathy that may increase the accuracy. Ordinal logistic regression is extended from binary logistic regression which dependent variable is in ordered variable. In this study, the ordinal logistic regression classifies the stages of diabetic retinopathy. Stages are denoted as stage 0, stage 1, stage 2, stage 3, stage 4 and stage 5. The linear relationship of the dependent variable to the fused predictors from diabetic retinopathy and nephropathy is formulated as follows $Y_{DR} = Z_{DR+DN} = X_{DR} + X_{DN}$, where only the evaluated subset predictors $X_{DR} = \left[x_1, x_2, ..., x_p \right]$ and $X_{DN} = \left[x_1, x_2, ..., x_q \right]$ were included in the final classification model. The search and evaluation methods will be further discussed in the next section. The ILDF methodology can be illustrated using Figure 2.

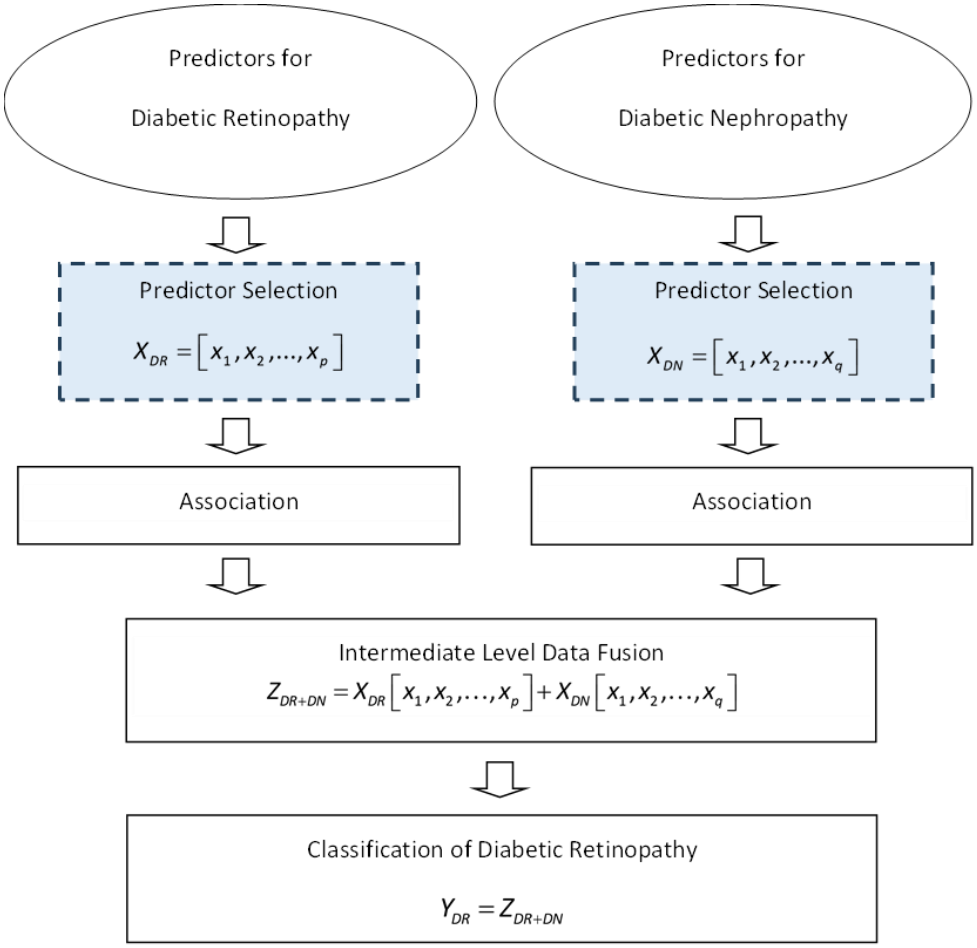


Fig. 2. The ILDF methodology for classification of stages of diabetic retinopathy

2.2.2 Criterion in choosing the classification model

This study used several criteria in choosing the best classification model. The criteria are the classification rate (accuracy), Akaike's Information Criteria (AIC) and Nagelkerke's R-Square. Particularly, classification rate, also known as classification accuracy, is a metric used to measure the performance of a classification model. It represents the percentage of correctly classified instances out of the total instances in the dataset. The classification rate formula as follows:

$$\text{Classification Rate} = \frac{\text{Corrected Classification Values}}{\text{Total}} x\ 100\% \quad (1)$$

AIC is one of the tools in comparing the model. Model with AIC minimum value indicates that the best model [27]. Furthermore, AIC does not provide information about the model's absolute quality; rather, it merely provides information about a model's quality in relation to other models.

$$AIC = n\log\left(\hat{\sigma}^2\right) + 2K \quad (2)$$

Nagelkerke R-squared or pseudo-R squared is a measure of the proportion of variance explained by a logistic regression model. While Nagelkerke R-squared ranges from 0 to 1, with higher values indicating a better fit of the logistic regression model.

$$R_N^2 = \frac{R_{cs}^2}{1 - L\left(B^{(0)}\right)^{2/n}} \tag{3}$$

3. Results and Discussion

This study involved two types of clinical data which are the baseline and mean data. The classification model for both baseline and mean clinical data was performed separately. Common significant factors and predictors for both diseases were fused using five different data fusion strategies (A, B, A+B, A+C, A+B+C) before it was fed to ordinal logistic regression to form the classification models. Detail of set of predictors A, B and C can refer previously in Figure 1.

Subset of predictors A is referring to shared significant predictors for diabetic retinopathy and nephropathy. It was found that the stages at diagnosed diabetic nephropathy the only predictors that significant to both model; baseline and mean clinical data. While for subset of predictors B, there were five factors and predictors that found significant to baseline clinical model and only three factors and predictors found significant to mean clinical model. Subset C which represented the factors and predictors that significant to diabetic nephropathy only, found that different factors and predictors for both baseline and mean model. Fused predictors (A+B+C) resulted most significant predictors among the other subset of predictors. These common significant factors and predictors for baseline and mean model were shown in Table 1.

Table 1
Common significant predictors for diabetic retinopathy and nephropathy for baseline and mean model

Subset of Predictors	Common factors and Predictors for Diabetic Retinopathy and Nephropathy	
	Baseline Model	Mean Model
A	Stages at Diagnosed Diabetic Nephropathy	Stages at Diagnosed Diabetic Nephropathy
B	Duration of Diabetic Retinopathy Stages at Diagnosed Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia Baseline Readings of Cholesterol	Stages at Diagnosed Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia
C	Age Race Baseline Readings of Urea Baseline Readings of Creatinine Baseline Readings of LDL	Mean Readings of eGFR Mean Readings of Creatinine
A+B	Stages at Diagnosed Diabetic Nephropathy Duration of Diabetic Retinopathy Stages at Diagnosed Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia Baseline Readings of Cholesterol	Stages at Diagnosed Diabetic Nephropathy Stages at Diagnosed Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia
A+C	Stages at Diagnosed Diabetic Nephropathy Age Race Baseline Readings of Urea Baseline Readings of Creatinine Baseline Readings of LDL	Stages at Diagnosed Diabetic Nephropathy Mean Readings of eGFR Mean Readings of Creatinine

A+B+C	Stages at Diagnosed Diabetic Nephropathy Stages at Diagnosed Diabetic Retinopathy Duration of Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia Baseline Readings of Cholesterol Age Race Baseline Readings of Urea Baseline Readings of Creatinine Baseline Readings of LDL	Stages at Diagnosed Diabetic Nephropathy Stages at Diagnosed Diabetic Retinopathy Risk of Right Eye Diabetic Retinopathy Dyslipidaemia Mean Readings of eGFR Mean Readings of Creatinine

Table 2 provides the results from classification model developed using ILDF using ordinal logistic regression. The classification rate (accuracy), AIC and Nagelkerke R-squared were used to indicate the best model. Each of the criterion were implemented for each subset predictors; A, B, A+B, A+C and A+B+C. Highest value for classification rate indicates the best accuracy through the group of subset predictors. As in Table 2, the highest value of classification rate for baseline and mean model is obtained in fused predictors (A+B+C). In the case of AIC, smallest value indicates the better fit model through the three link functions. However, the smallest AIC can be found in subset predictors A for both baseline and mean model. Another criterion that had to consider is Nagelkerke R-squared. For Nagelkerke R-squared, the highest value is in fused predictors for both baseline and mean model. Thus, these results indicated that the fused predictors; A+B+C for both baseline and mean model was the best subset predictors in conjunctions to predict the stages of diabetic retinopathy. The details of the result were represented in Table 2.

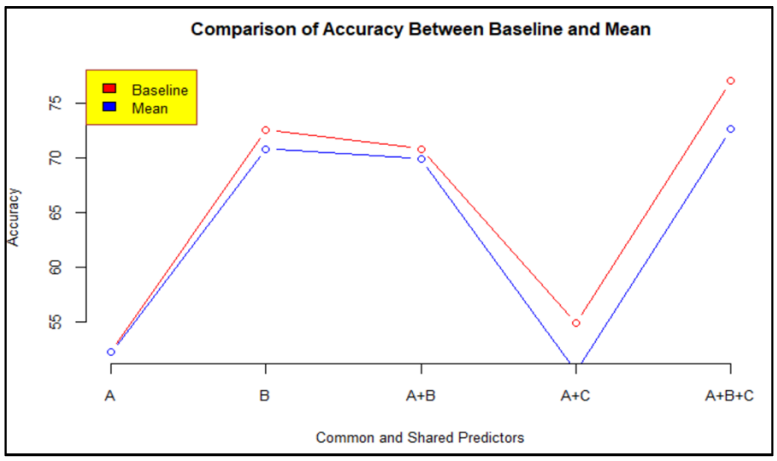
Table 2
The results of classification rate, AIC and Nagelkerke R-squared for baseline and mean model

Common, Shared and Fused Predictors	Baseline Model			Mean Model		
	Classification Rate	AIC	Nagelkerke R-squared	Classification Rate	AIC	Nagelkerke R-squared
A	52.21	77.31	0.316	67.76	77.13	0.316
B	72.56	197.69	0.788	70.79	192.36	0.811
A+B	70.79	202.09	0.811	69.02	192.35	0.811
A+C	54.86	321.75	0.42	71.68	316.16	0.321
A+B+C	76.99	197.33	0.868	71.68	201.62	0.817

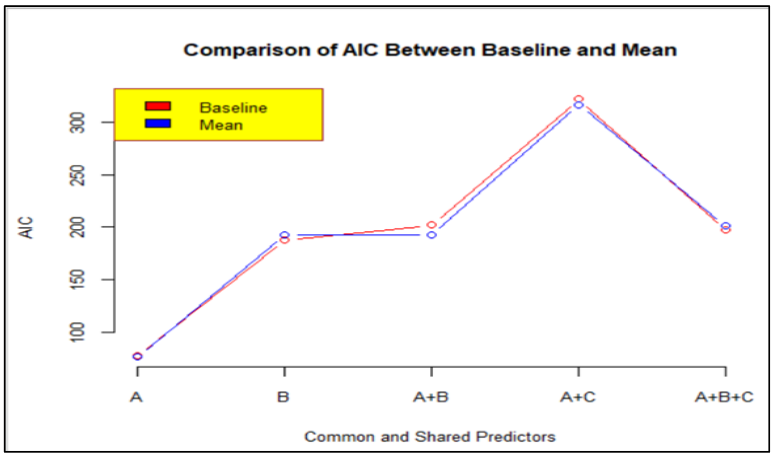
In comparing the best model between baseline model and mean model, it was easier to compare using Figure 3. Figure 3 shown the comparison between accuracy, AIC and Nagelkerke R-squared for baseline and mean model. Figure 3(a) shown that baseline model had higher classification rate for all subset predictors compare to mean model. But not for results in AIC and Nagelkerke R-squared, (Figure 3(b) and 3(c)) as some of the set predictors, mean model recorded higher and same value with the baseline model. Thus, based on the results shown in Table 2 and Figure 3, the fused predictors; A+B+C is the best model compared to the other of subset predictors or single predictors. This result shown that with the fused predictors help to increase the accuracy of the classification model. Thus, the fused predictors give the important role in classification model.

In comparison between the baseline and mean model, baseline model found to be the best model. The baseline model genuinely provides the full picture for a patient in clinical readings and accurately depicts their health when they were first diagnosed. Furthermore, this study also included the factors that have timely period characteristics such as duration of diabetes, duration of diabetic retinopathy and duration of diabetic nephropathy which is better in baseline model rather than mean

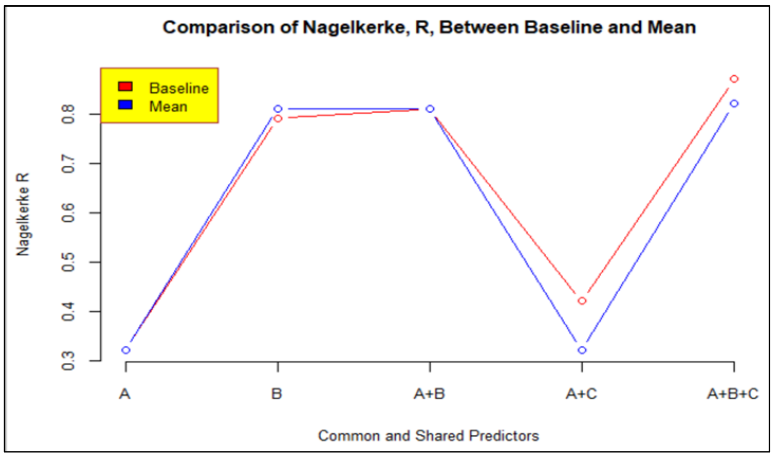
model. In addition, in mean model, its only depend on one value which represented for the whole clinical readings where the readings might be fluctuation readings and subjective. Thus, this data for this study, it was suitable to use baseline model.



a) Classification rate (Accuracy)



b) AIC



c) Nagelkerke R-squared

Fig. 3. Comparison of Accuracy, AIC and Nagelkerke R-squared between Baseline and Mean Model

4. Conclusion

This study emphasizes the formulation of a model with various combinations of risk factors and predictors selected using the ILDF to classify the stages of diabetic retinopathy. Based on the results of the analyses, the best model for this investigation was determined to be the fused predictors of [A+B+C] which consists of age, race, stages at diagnosed diabetic nephropathy, stages at diagnosed diabetic retinopathy, duration of diabetic retinopathy, risk of right eye diabetic retinopathy, dyslipidaemia, baseline readings of cholesterol, baseline readings of urea, baseline readings of creatinine, baseline readings of LDL. The fused predictors can improve the accuracy of both the Mean and Baseline models. While for this study the Baseline model is superior to the Mean model. ILDF as the basis for the formulation of classification model using the Ordinal Logistic Regression is appropriate to help classifies the stages of diabetic retinopathy using predictors and risk factors from both diabetic retinopathy and nephropathy. Identifying the best classification model to classify the stages of diabetic retinopathy could give some clinical decision's information that could assist medical officer in diabetes management.

Acknowledgement

This research was not funded by any grant. This study has received ethical clearance from the Medical Research and Ethics Committee of the Ministry of Health Malaysia (KKM/NIHSEC/P20-376) and was registered with the National Medical Research Register (NMRR-19-NMRR-19-3896-47105 IIR).

References

[1] Hall, David Lee, and Sonya AH McMullen. *Mathematical techniques in multisensor data fusion*. Artech House, 2004.
[2] Hall, David L., and James Llinas. "An introduction to multisensor data fusion." *Proceedings of the IEEE* 85, no. 1 (1997): 6-23. https://doi.org/10.1109/5.554205
[3] King, Rachel C., Emma Villeneuve, Ruth J. White, R. Simon Sherratt, William Holderbaum, and William S. Harwin. "Application of data fusion techniques and technologies for wearable health monitoring." *Medical engineering & physics* 42 (2017): 1-12. https://doi.org/10.1016/j.medengphy.2016.12.011
[4] Kalamkar, Shrida. "Clinical data fusion and machine learning techniques for smart healthcare." In *2020 International Conference on Industry 4.0 Technology (I4Tech)*, pp. 211-216. IEEE, 2020. https://doi.org/10.1109/I4Tech48345.2020.9102706
[5] Nweke, Henry Friday, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions." *Information Fusion* 46 (2019): 147-170. https://doi.org/10.1016/j.inffus.2018.06.002
[6] Kulkarni, Mrunalini Harish, Chaitanya Kulkarni, K. Suresh Babu, Saima Ahmed Rahin, Shweta Singh, and D. Dinesh Kumar. "Data Fusion Approach for Managing Clinical Data in an Industrial Environment using IoT." *Scientific Programming* 2022, no. 1 (2022): 3603238. https://doi.org/10.1155/2022/3603238
[7] Webber, Sara. "International Diabetes Federation." *Diabetes Research and Clinical Practice*, Vol. 102. (2021).
[8] Akhtar, Sohail, Jamal Abdul Nasir, Aqsa Ali, Mubeen Asghar, Rizwana Majeed, and Aqsa Sarwar. "Prevalence of type-2 diabetes and prediabetes in Malaysia: A systematic review and meta-analysis." *PloS one* 17, no. 1 (2022): e0263139. https://doi.org/10.1371/journal.pone.0263139
[9] Harreiter, Jürgen, and Michael Roden. "Diabetes mellitus—Definition, classification, diagnosis, screening and prevention (Update 2019)." *Wiener Klinische Wochenschrift* 131 (2019): 6-15. https://doi.org/10.1007/s00508-019-1450-4
[10] Altomare, Filiberto, Amin Kherani, Julie Lovshin, and Diabetes Canada Clinical Practice Guidelines Expert Committee. "Retinopathy." *Canadian journal of diabetes* 42 (2018): S210-S216. https://doi.org/10.1016/j.jcjd.2017.10.027
[11] Zhang, Junlin, Yiting Wang, Li Li, Rui Zhang, Ruikun Guo, Hanyu Li, Qianqian Han, Geer Teng, and Fang Liu. "Diabetic retinopathy may predict the renal outcomes of patients with diabetic nephropathy." *Renal failure* 40, no. 1 (2018): 243-251. https://doi.org/10.1080/0886022X.2018.1456453
[12] Kim, Yoon Jeon, June-Gone Kim, Joo Yong Lee, Kyoung Sub Lee, Soo Geun Joe, Joong-Yeol Park, Min-Seon Kim, and Young Hee Yoon. "Development and progression of diabetic retinopathy and associated risk factors in Korean

patients with type 2 diabetes: the experience of a tertiary center." *Journal of Korean medical science* 29, no. 12 (2014): 1699. https://doi.org/10.3346/jkms.2014.29.12.1699

[13] Pavkov, Meda E., Jessica L. Harding, Chiu-Fang Chou, and Jinan B. Saaddine. "Prevalence of diabetic retinopathy and associated mortality among diabetic adults with and without chronic kidney disease." *American journal of ophthalmology* 198 (2019): 200-208. https://doi.org/10.1016/j.ajo.2018.10.019

[14] Jawa, Ali, Juanita Kcomt, and Vivian A. Fonseca. "Diabetic nephropathy and retinopathy." *Medical Clinics* 88, no. 4 (2004): 1001-1036. https://doi.org/10.1016/j.mcna.2004.04.012

[15] Pedro, Romero-Aroca, Sagarra-Alamo Ramon, Baget-Bernaldiz Marc, Fernández-Ballart Juan, and Méndez-Marin Isabel. "Prevalence and relationship between diabetic retinopathy and nephropathy, and its risk factors in the North-East of Spain, a population-based study." *Ophthalmic epidemiology* 17, no. 4 (2010): 251-265. https://doi.org/10.3109/09286586.2010.498661

[16] Romero-Aroca, Pedro, Marc Baget-Bernaldiz, Javier Reyes-Torres, Juan Fernandez-Ballart, Nuria Plana-Gil, Isabel Mendez-Marin, and Alicia Pareja-Rios. "Relationship between diabetic retinopathy, microalbuminuria and overt nephropathy, and twenty-year incidence follow-up of a sample of type 1 diabetic patients." *Journal of Diabetes and its Complications* 26, no. 6 (2012): 506-512. https://doi.org/10.1016/j.jdiacomp.2012.06.010

[17] Ahmed, Mohamed H., Einas S. Elwali, Heitham Awadalla, and Ahmed O. Almobarak. "The relationship between diabetic retinopathy and nephropathy in Sudanese adult with diabetes: population based study." *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 11 (2017): S333-S336. https://doi.org/10.1016/j.dsx.2017.03.011

[18] Kaewput, Wisit, Charat Thongprayoon, Ram Rangsin, Prajej Ruangkanchanasetr, Michael A. Mao, and Wisit Cheungpasitporn. "Associations of renal function with diabetic retinopathy and visual impairment in type 2 diabetes: A multicenter nationwide cross-sectional study." *World journal of nephrology* 8, no. 2 (2019): 33. https://doi.org/10.5527/wjn.v8.i2.33

[19] Sabanayagam, Charumathi, Miao Li Chee, Riswana Banu, Ching-Yu Cheng, Su Chi Lim, E. Shyong Tai, Thomas Coffman, and Tien Y. Wong. "Association of diabetic retinopathy and diabetic kidney disease with all-cause and cardiovascular mortality in a multiethnic Asian population." *JAMA network open* 2, no. 3 (2019): e191540-e191540. https://doi.org/10.1001/jamanetworkopen.2019.1540

[20] Zhou, Liang, Xiaoyuan Zheng, Di Yang, Ying Wang, Xuesong Bai, and Xinhua Ye. "Application of multi-label classification models for the diagnosis of diabetic complications." *BMC medical informatics and decision making* 21, no. 1 (2021): 182. https://doi.org/10.1186/s12911-021-01525-7

[21] Sisodia, Deepti, and Dilip Singh Sisodia. "Prediction of diabetes using classification algorithms." *Procedia computer science* 132 (2018): 1578-1585. https://doi.org/10.1016/j.procs.2018.05.122

[22] Dudkina, Tetiana, Ievgen Meniailov, Kseniia Bazilevych, Serhii Krivtsov, and Anton Tkachenko. "Classification and Prediction of Diabetes Disease using Decision Tree Method." In *IT&AS*, pp. 163-172. 2021.

[23] Phongying, Methaporn, and Sasiprapa Hiriote. "Diabetes classification using machine learning techniques." *Computation* 11, no. 5 (2023): 96. https://doi.org/10.3390/computation11050096

[24] Karegowda, Asha Gowda, M. A. Jayaram, and A. S. Manjunath. "Cascading k-means clustering and k-nearest neighbor classifier for categorization of diabetic patients." *International Journal of Engineering and Advanced Technology* 1, no. 3 (2012): 147-151.

[25] Hang, Ong Yee, Wiwied Virgiyanti, and Rosly Rosaida. "Diabetes Prediction using Machine Learning Ensemble Model." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 37, no. 1 (2024): 82-98. https://doi.org/10.37934/araset.37.1.8298

[26] Brieva, Jorge. "Datamining and its applications." *Journal of Artificial Intelligence and Technology* 2, no. 3 (2022): 77-79. https://doi.org/10.37965/jait.2022.0125

[27] Chowdhury, Mohammad Ziaul Islam, and Tanvir C. Turin. "Variable selection strategies and its importance in clinical prediction modelling." *Family medicine and community health* 8, no. 1 (2020). https://doi.org/10.1136/fmch-2019-000262