SCIENTIFIC PAPER

A deep learning nomogram of continuous glucose monitoring data for the risk prediction of diabetic retinopathy in type 2 diabetes

Rui Tao[1] · Xia Yu[1] · Jingyi Lu[2] · Yaxin Wang[2] · Wei Lu[2] · Zhanhu Zhang[1] · Hongru Li[1] · Jian Zhou[2]

Abstract

Continuous glucose monitoring (CGM) data analysis will provide a new perspective to analyze factors related to diabetic retinopathy (DR). However, the problem of visualizing CGM data and automatically predicting the incidence of DR from CGM is still controversial. Here, we explored the feasibility of using CGM profiles to predict DR in type 2 diabetes (T2D) by deep learning approach. This study fused deep learning with a regularized nomogram to construct a novel deep learning nomogram from CGM profiles to identify patients at high risk of DR. Specifically, a deep learning network was employed to mine the nonlinear relationship between CGM profiles and DR. Moreover, a novel nomogram combining CGM deep factors with basic information was established to score the patients' DR risk. This dataset consists of 788 patients belonging to two cohorts: 494 in the training cohort and 294 in the testing cohort. The area under the curve (AUC) values of our deep learning nomogram were 0.82 and 0.80 in the training cohort and testing cohort, respectively. By incorporating basic clinical factors, the deep learning nomogram achieved an AUC of 0.86 in the training cohort and 0.85 in the testing cohort. The calibration plot and decision curve showed that the deep learning nomogram had the potential for clinical application. This analysis method of CGM profiles can be extended to other diabetic complications by further investigation.

Introduction

Diabetic retinopathy (DR) is the most common microvascular complication of diabetes, and it greatly damages visual impairment and affects the quality of life [1–3]. Although management methods, such as laser photocoagulation, drugs and blood pressure control, have been demonstrated to be effective for treating DR [4, 5], the prevalence of DR in type 2 diabetes (T2D) patients is still increasing. Therefore, it is critical to promptly identify patients at a high risk of DR and promote early intervention. Currently, fundus imaging is widely used to diagnose DR in T2D patients [6]. However, the accuracy is greatly influenced by the equipment [7]. On

the other hand, only a small percentage of people are able to undergo routine fundus examinations [8, 9]. According to the American Diabetes Association, no more than half of patients received any examination for detecting DR [9]. Although glycosylated hemoglobin (HbA_{1c}) is considered to be a significant risk factor for DR, other metrics of glycemic control also have been demonstrated to have the potential to contribute to the development of DR [4]. These studies suggested that there are risk indicators other than HbA1c associated with DR.

The continuous glucose monitoring (CGM) system can facilitate a deeper understanding of the changing trend of blood glucose levels, including the direction, amplitude, duration, and frequency of abnormal blood glucose [10]. CGM is also convenient for patients and healthcare providers to monitor their glucose levels dynamically with a portable device [11]. Analysis of CGM profiles could be used for subtype classification of patients with type 2 diabetes and high/low blood glucose alarms [12–14]. Several recent studies have explored the risk factors associated with CGM and DR [15–17]. Sartore et al. [15] explored the significant correlation between the standard deviation of blood glucose

✉ Jian Zhou
zhoujian@sjtu.edu.cn

1 College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, China

2 Department of Endocrinology and Metabolism, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai Clinical Center for Diabetes, 600 Yishan Road, Shanghai 200233, China

and DR in diabetes patients. A study by Lu et al. [16] found that time in range (TIR) is an independent risk factor for DR in T2D patients. Hirsch et al. [17] demonstrated that glycemic variability obtained by CGM is an important risk factor involved in the pathogenesis of DR. Therefore, we postulate that the analysis of DR from CGM profiles will provide new perspectives and valuable information.

Despite these advantages, CGM profiles have not been widely used to predict DR due to the following challenges. First, a large amount of data leads to difficulty in interpreting the CGM relationship with DR [11, 18]. Various metrics, such as TIR and mean blood glucose (MBG), were designed based on CGM. Individual CGM metrics are one-sided and do not accurately reflect the connection between complications and CGM. Second, the risk of complex clinical events, such as DR, is a nonlinear combination of the patients' covariates [19]. Although the traditional risk prediction model nomogram is commonly used in clinical practice, it does not sufficiently resolve the effect of nonlinearity [20,

21]. To address the shortcomings of CGM analysis in constructing a risk prediction model, deep learning should be considered. As an artificial intelligence algorithm, deep learning can solve the problems above and has been successfully applied in pattern recognition [22, 23]. Deep learning has also been successfully applied in many medical fields [24, 25].

Therefore, we constructed a deep learning nomogram for the prediction of DR in patients with T2D from CGM profiles. The objective of this study was to promote T2D patients with a high risk of DR to receive more timely examinations and thereby improve long-term outcomes.

Methodology

The framework of the proposed nomogram is presented in Fig. 1. The preprocessing block cleans and interpolates the CGM data to regularize it. The feature extraction and
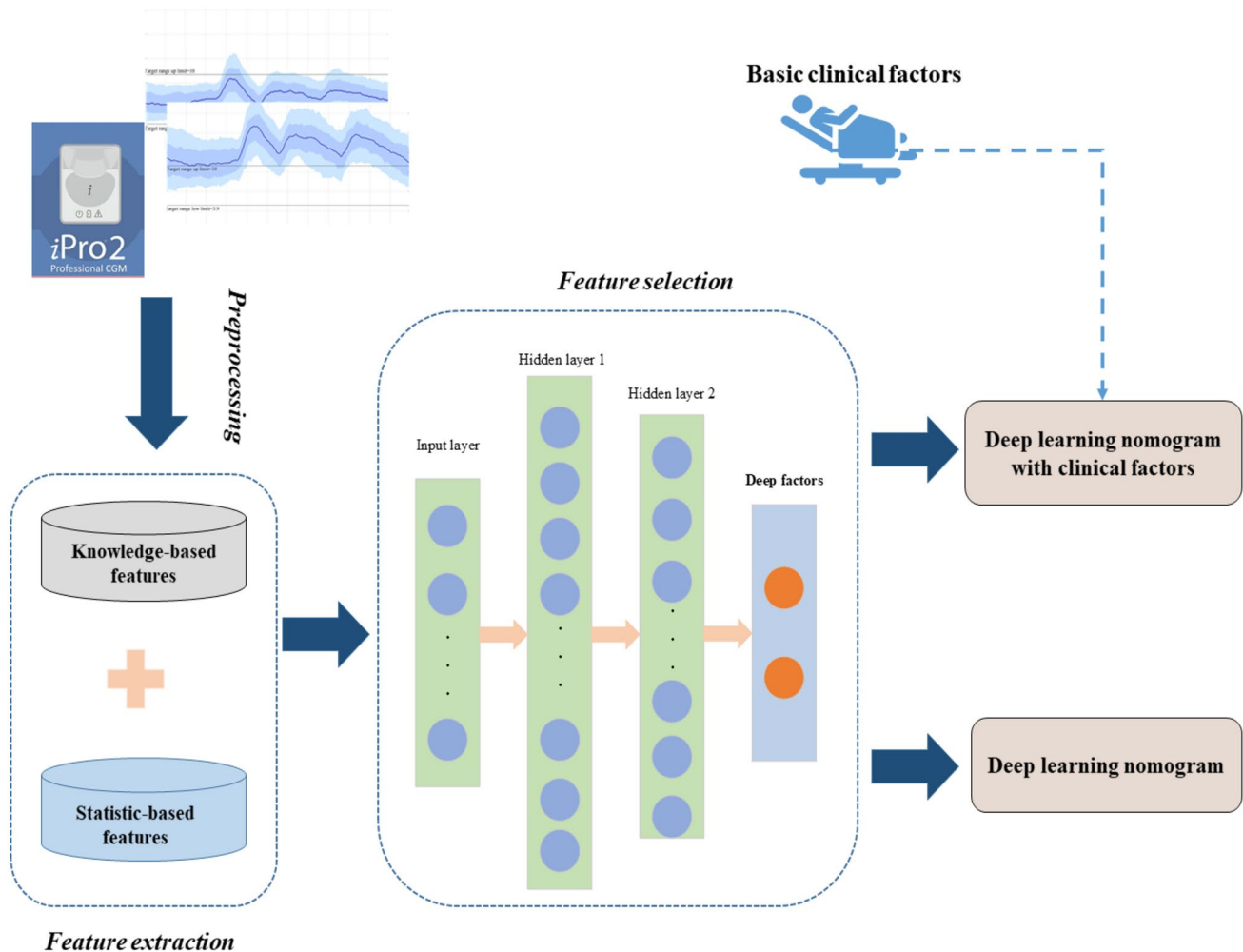

Fig. 1 The framework of the proposed deep learning nomogram for DR prediction from CGM time series data

selection blocks determine efficient patient representations. The model construction block generates the nomogram model based on the efficient representations from feature selection blocks.

Preprocessing

CGM profiles are usually high-dimensional, sparse, and often have missing values due to the transmission of the device. This irregularity in the CGM profiles leads to a shortage of structures that complicates feature learning and risk prediction model construction. This study addressed these problems by propensity score matching (PSM) and spline interpolation.

PSM is a logistic regression with the intervention factor (group) as the dependent variable and all observed non-research factors as independent variables. After PSM, we can select the experimental and control groups so that the non-research type factors in the distinct groups achieve equilibrium. Thus we can better dig into the relationship between the research factors and the dependent variable. The Propensity score itself does not control confounding, but by PSM, weighting, stratification, and constantly adjusting the manner of confounding to varying degrees to improve equilibrium, we can attenuate or balance the effect of covariates on the results [26, 27]. Spline interpolation is one of the most common methods in data filling. This method fills in the "missing" measurements of the patient CGM data for feature extraction using piecewise polynomials. More information about spline interpolation can be found in [28].

This study focused on exploring the correlation between CGM profiles and DR in patients with T2D. In order to overcome gender-specific data bias. We applied the PSM method to eliminate the effects of gender and age in the data on the DR event before the train/test split. Indeed, many other non-CGM factors may contribute to an increased risk in DR [29]. The high-lipoprotein cholesterol (HDL) and low-density lipoprotein cholesterol (LDL) were no significant differences between the two groups patients with DR and patients with no DR (nDR). HbA_{1c} and Duration were also risk factors, which be added to the subsequent construction of the mixed model. For the data after PSM, we screened the data by examining the CGM for missing conditions. If the patient had no missing data on CGM, we included it in our patient database. If the patient's CGM data had less than 15 min (less than 3 points) of consecutive missing data, we included it in our patient database after spline interpolation. In other cases, we will delete the data of patients with excessive missing data.

Feature extraction

Despite the many advantages of CGM profiles, the enormous amount of sparse data prevented the development of a perfect approach for extracting CGM features. To better interpret the CGM profiles, this study extracted the well-established metrics from CGM according to [30]. The effective feature extraction in advance could initially mine the valid information and avoid the low successful utilization of deep learning methods in CGM analysis [9].

The metrics used for CGM analysis can be divided into two perspectives: the knowledge-based metrics pool and the statistical-based metrics pool. The knowledge-based metrics pool includes the percentages of CGM values in the target range (TIR, [3.9–10] mmol/L), the low and high blood glucose indices (LBGI and HBGI), J index, M value, glycemic risk assessment diabetes equation (GRADE) score, liability index (LI), average daily risk range (ADRR) and area under the curve (AUC). The statistical-based metrics pool contains the following features: MBG), mean amplitude of glycemic excursions (MAGE), the largest amplitude of glycemic excursions (LAGE), coefficient of variation (CV), and standard deviation of blood glucose (SDBG).

Feature selection

Feature selection can be performed to remove redundant information, making it easier to retrieve the data structure for high-dimensional data, especially for CGM profiles. It could also solve the pattern mining errors caused by covariance. Finding particular connections between features can enhance data information, which will improve the accuracy of the model [31, 32].

In the regularized nomogram, the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm was applied in feature selection [33]. However, the main problem in constructing the nomogram is that the regularized nomogram can only mine the linear combination of the patients' covariates. Deep learning is a typical machine learning method and can overcome the difficulty of mining nonlinear relationships. The continuous development of computing resources and the application of large amounts of data has led to the widespread use of deep learning in a variety of medical situations [34, 35].

In this block, a deep feedforward neural network for feature selection was constructed from CGM profiles. The DR complications of T2D were a target in the training cohort of patients. As shown in supplementary Fig. 1, there were four fully connected layers which include three hidden layers and one output layer. The penultimate layer of the network can be extracted as the deep factors. The deep factors contained the nonlinear relationships between the input CGM metrics and DR events. The loss of the network is defined as follows:

$$Loss(\theta) = -\frac{1}{N}\sum_i y_i * \log(P(y_i)) + (1-y_i) * \log(1-P(y_i))$$

(1)

where y_i means the presence of DR which 1 means DR and 0 means nDR. $P(y_i)$ indicates the probability of having DR, which was calculated by softmax function. The penultimate layer of the network could be extracted respectively as the deep factors. We randomly divided 20% of the train sets into validation sets. An early stop was set when the accuracy of the validation set does not increase more than 0.01% for 200 epochs. The loss and accuracy curve could be seen in Supplementary Fig. 2.

The deep factors can be influenced by the architecture of the deep learning network. The parameters of the network were optimized by Bayesian Hyper-parameters Optimization in the training cohort. The final network architecture can be seen in Supplementary Table 1. Finally, two higher-order deep factors were obtained from the input CGM indicators, named deep factor 1 and deep factor 2. We then performed a nomogram on the same training cohort described above.

Nomogram

The nomogram includes three types of lines. A point line is used to identify the score level of the risk factors. Each risk factor also has a point line for indicating the number of points. A total points line represents the sum of points of each risk factor. A probability line indicates the probability corresponding to the total points line [36].

Point lines

The LP_{ip} values for each deep factor based on feature learning were calculated by the regression coefficient in the regression model. The LP_{ip} values and point values were calculated using the following formula:

$$LP_{ip} = \beta_i x_{ip}$$

(2)

$$Points_{ip} = \frac{LP_{ip} - \min_{ip} LP_{ip}}{\max_p LP_{*p} - \min_p LP_{*p}} \times 100$$

(3)

where i = 1,2 is the number of deep factors. LP_{ip} is represented as the product of the regression coefficient β_i and the category of each risk factor x_{ip}. The point values ranged from 0 to 100. LP_{*p} is the largest LP_{ip} value with the estimated regression coefficient.

Total points line

The total points line can be obtained by simply summing the points of each risk factor:

$$Total\,points = \sum_i Points_{ip}$$

(4)

Probability line

This line represents the probability value. It corresponds to the total points by summarizing the LP value from the total points line.

Model validation

For the regularized nomogram and proposed deep learning nomogram, the model was trained on the training cohort. Furthermore, it was evaluated in both the training cohort and the testing cohort. This study mainly compares and evaluates the models in terms of both *discrimination* and *calibration*. Besides, other evaluation metrics like sensitivity and specificity also be considered in this research.

Discrimination

Discrimination of the model means that the model correctly distinguishes whether an outcome event will occur in an individual. The most common metric for assessing the discrimination ability of a model is the area under the receiver operating curve (AUC-ROC). The ROC curve can reveal the relationship between sensitivity and specificity. The larger the AUC-ROC, the better the discriminative ability of the prediction model. Generally AUC < 0.60 is considered poor discriminative ability, 0.60–0.75 is considered that the model has the moderate discriminative ability, and > 0.75 is considered a good discriminative ability [37, 38].

Calibration

The calibration of a prediction model is an important indicator to evaluate the accuracy of a disease risk model in classifying the probability of an outcome event in an individual. The probability reflects the degree of agreement between the risk judged by the model and the actual risk of occurrence, so it can also be called consistency. A good calibration indicates a high accuracy of the model, while a poor calibration indicates that the model may overestimate or underestimate the risk of the actual outcome event. The calibration plot can be used to determine the connection between the true probability and the predicted probability. If the calibration plot line is close to the 45-point line, the better the calibration of the model [39].

Evaluation metrics

For a more comprehensive and detailed evaluation of the model. The following evaluation metrics were used to quantify the performance of our proposed deep learning nomogram:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$SPE = \frac{TN}{TN + FP} \tag{7}$$

$$PRE = \frac{TP}{TP + FP} \tag{8}$$

$$F1-score = \frac{2 \times PRE \times SEN}{PRE + SEN} \tag{9}$$

where ACC, SEN, SPE, PRE, and F1-score represent the accuracy, sensitivity, Specificity, precision, and F1-score. And TP, TN, FP, and FN mean the true positive, true negative, false positive, and false negative, respectively.

Results

Patient selection

The secondary study protocol was approved by the Ethics Committee of Shanghai Jiao Tong University Affiliated Sixth People's Hospital. Every patient provided informed consent. For the 1556 patients, we recorded basic information such as disease duration, age and gender. We also documented several clinical measures of admission testing like HbA1c, triglyceride, etc. Each patient wore a CGM system (iPro2, Medtronic, USA) on the day of hospitalization, and it was removed after the third day. The CGM system of iPro2 will collect the patient's blood glucose value every 5 min. The CGM sensors were calibrated by finger stick blood glucose collection every 12 h. CGM data (0:00–24:00) for each patient on Day 2 or Day 3 were extracted. Patient inclusion criteria for this study were as follows: hospitalized patients with stable treatment for at least 3 months, age ≥ 18 years, and no missing clinical characteristics data. Exclusion criteria included a hyperglycemic hyperosmolar state or severe and recurrent events within the previous 3 months, diabetic ketoacidosis and a history of malignancy or mental disorders. The patient selection process is illustrated in Fig. 2.

Preprocessing

A total of 788 patients were included in this study after processing by spline interpolation and propensity score matching. 2017–2018 data were used as the training cohort, and 2019 data was used as the testing cohort. The CGM metrics pool, as shown in the Methodology, was calculated in these patients based on their CGM profiles. The clinical characteristics are shown in Table 1 and Fig. 3. It could be found that the extracted CGM metrics have a significant difference between groups of DR and nDR on both the training cohort and testing cohort (except LBGI [p > 0.05 in the training cohort and testing cohort]). These differences could prove that diabetic patients with DR and nDR are distinguishable in terms of CGM metrics.

Variables were presented as the mean ± standard deviation. The difference between groups was examined by the Mann–Whitney U test. A 2-tailed P value of less than 0.05 was considered statistically significant. Abbreviations: HDL-high density lipoprotein; LDL-low density lipoprotein.

Performance of the model

The training cohort included 478 patients and 247 developed DR. The testing cohort included 294 patients and 147 developed DR. The model was trained with the entire training cohort and tested on the testing cohort. We set a regularized nomogram model with LASSO as the baseline. Three features were identified through LASSO by the lambda with minimum partial likelihood deviance value (Supplementary Fig. 3, Supplementary Table 2). These were SDBG, TIR, and GRADE. The ROC curves for the regularized nomogram and deep learning nomogram are presented in Fig. 4, and the corresponding ROC-AUC, specificity, sensitivity, and other metrics are shown in Table 2.

The performance metrics AUC of the regularized nomogram were superior to those features obtained by LASSO. In particular, the regularized nomogram achieved AUCs of 0.79 and 0.75 in the training and testing cohort, respectively. Moreover, the AUC values of the deep learning nomogram were 0.82 and 0.80 in the training and testing cohort, respectively, showing that the quality of the positive predictions of DR was improved by nonlinear feature mining by the deep network. Compared with regularized nomogram, the deep learning nomogram is also superior in Sensitivity and F1-score. The deep learning nomogram gets the value of 0.76 in the training cohort and 0.72 in the testing cohort at F1-score, which means that the proposed nomogram could not only provide better screening of DR patients but also have a lower rate of false negative and false positive. The calibration plot in Fig. 4 also indicated good agreement between the nomogram prediction and actual observation for T2D in
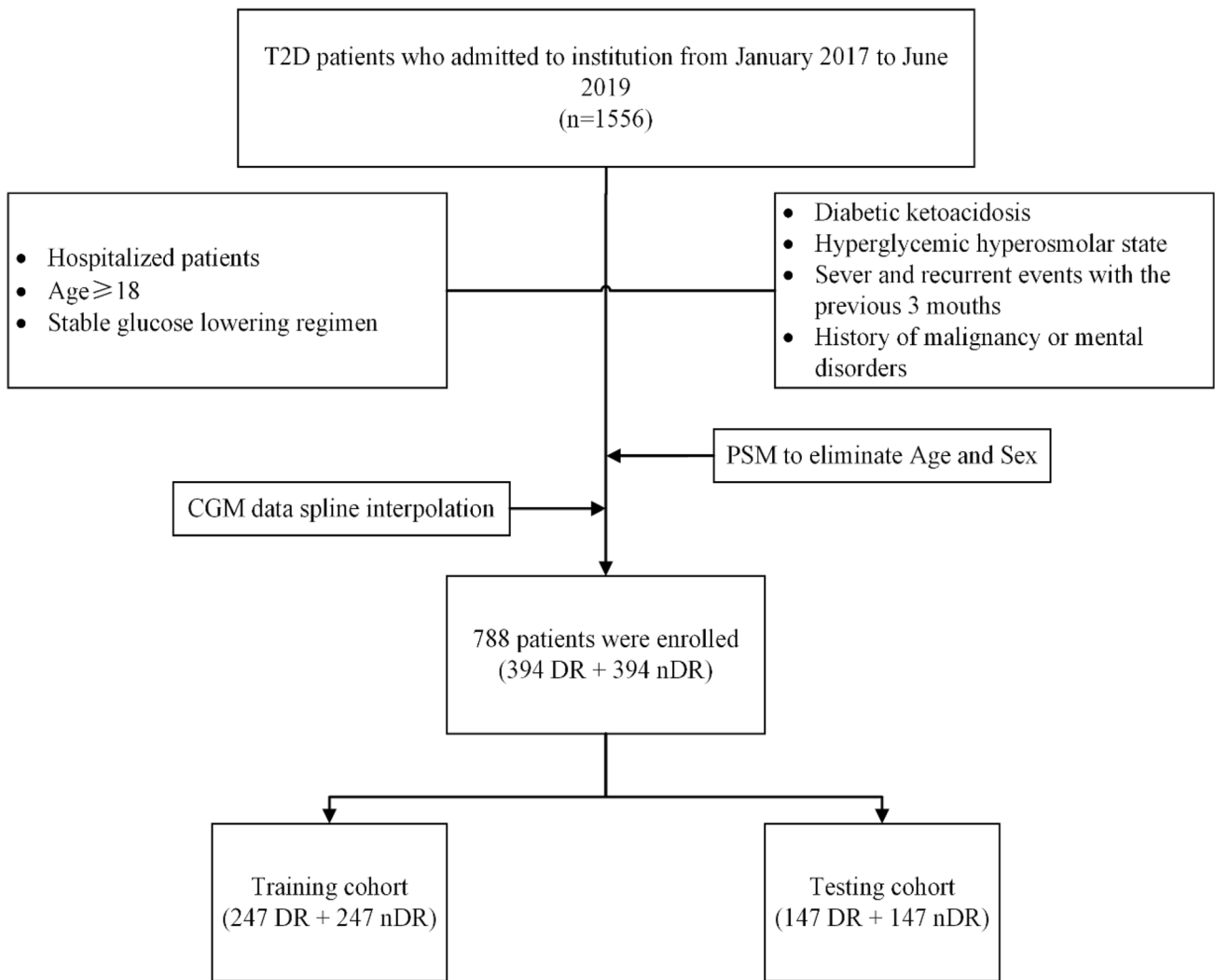
Fig. 2 Patients selection flowchart

Table 1 Baseline Characteristics of the patients in the DR and nDR groups

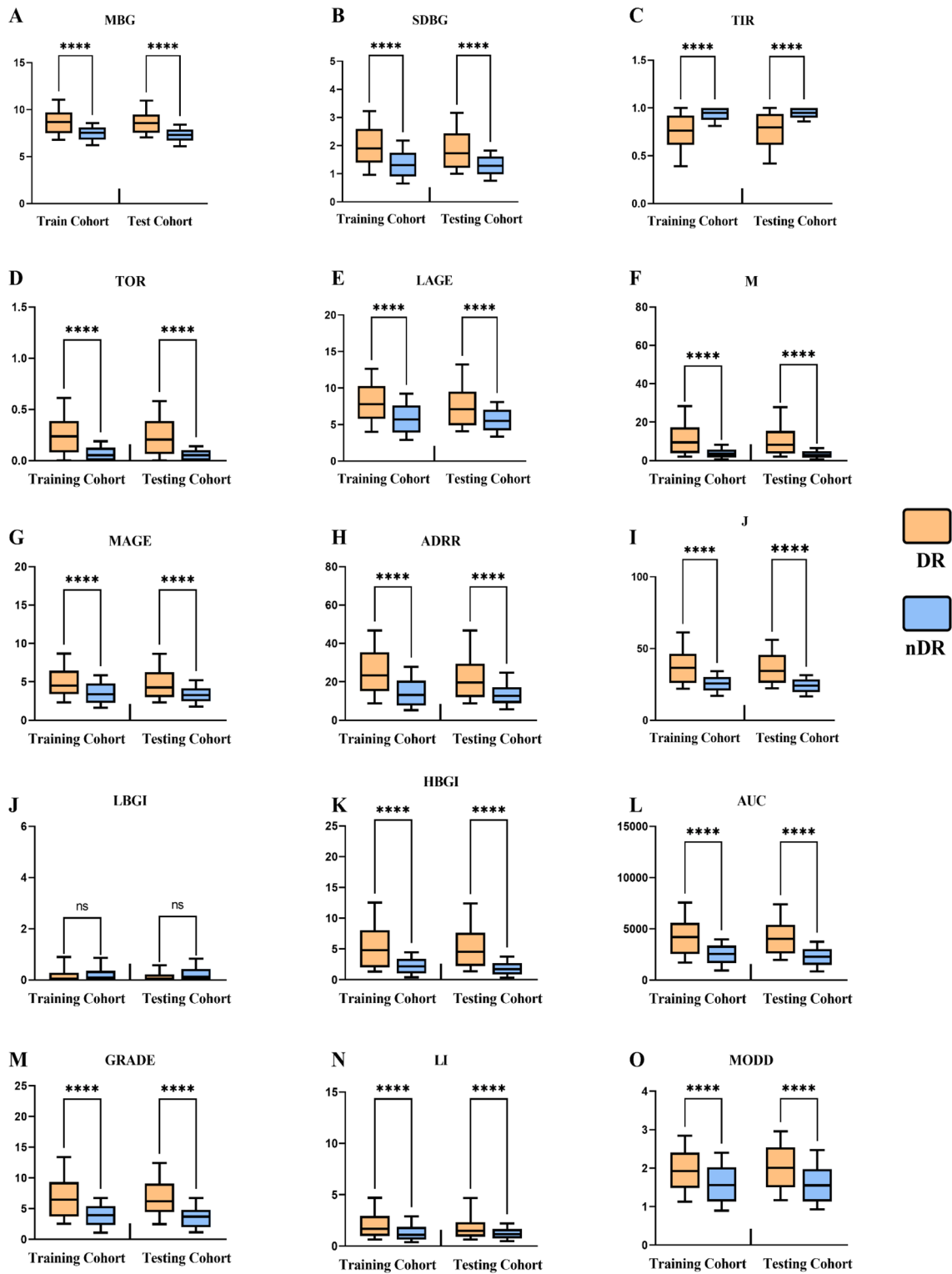|	Training cohort			Testing cohort		
	nDR (N = 247)	DR (N = 247)	P value	nDR (N = 147)	DR (N = 147)	P value
Basic information						
Duration (years)	10.28 ± 6.81	14.17 ± 7.21	< 0.01	9.54 ± 7.18	14.13 ± 7.13	< 0.01
Sex = Female (%)	37.2	38.1	0.99	38.1	38.1	0.99
Age (years)	58.99 ± 11.44	58.52 ± 11.65	0.61	58.41 ± 11.28	58.62 ± 11.28	0.67
Clinical measures						
HbA1c (%)	7.23 ± 1.47	8.71 ± 1.93	< 0.01	7.32 ± 1.57	8.73 ± 1.72	< 0.01
Triglyceride (mmol/L)	1.67 ± 1.39	2.04 ± 2.23	0.03	1.57 ± 1.37	2.01 ± 2.18	0.03
Total cholesterol (mmol/L)	4.30 ± 1.12	4.57 ± 1.39	0.01	4.34 ± 0.98	4.64 ± 1.21	0.02
HDL (mmol/L)	1.07 ± 0.31	1.07 ± 0.30	0.99	1.11 ± 0.35	1.08 ± 0.28	0.44
LDL (mmol/L)	2.53 ± 0.92	2.58 ± 0.91	0.51	2.51 ± 0.80	2.67 ± 0.82	0.18
Uric acid (UMO)	344.24 ± 88.61	333.41 ± 91.73	0.19	333.42 ± 79.79	335.51 ± 96.12	0.84

Fig. 3 Box plots of CGM metrics between the Training Cohort and Testing Cohort. The difference between groups was examined by the Mann–Whitney U test. **** means the P value of less than 0.01 and *ns* means there is no significant difference between the two groups. Abbreviations: MBG-mean blood glucose; TIR-percentages of values within the target range ([3.9–10] mmol/L); SDBG-standard devia-tion of the blood glucose, CV-coefficient of variation; LAGE-largest amplitude of glycemic excursions; MAGE-mean amplitude of gly-cemic excursions; HBGI(LBGI)-high (low) blood glucose indices; GRADE-glycemic risk assessment diabetes equation score; LI-lia-bility index; ADRR-average daily risk range; M-M value; J-J index [Color needed]
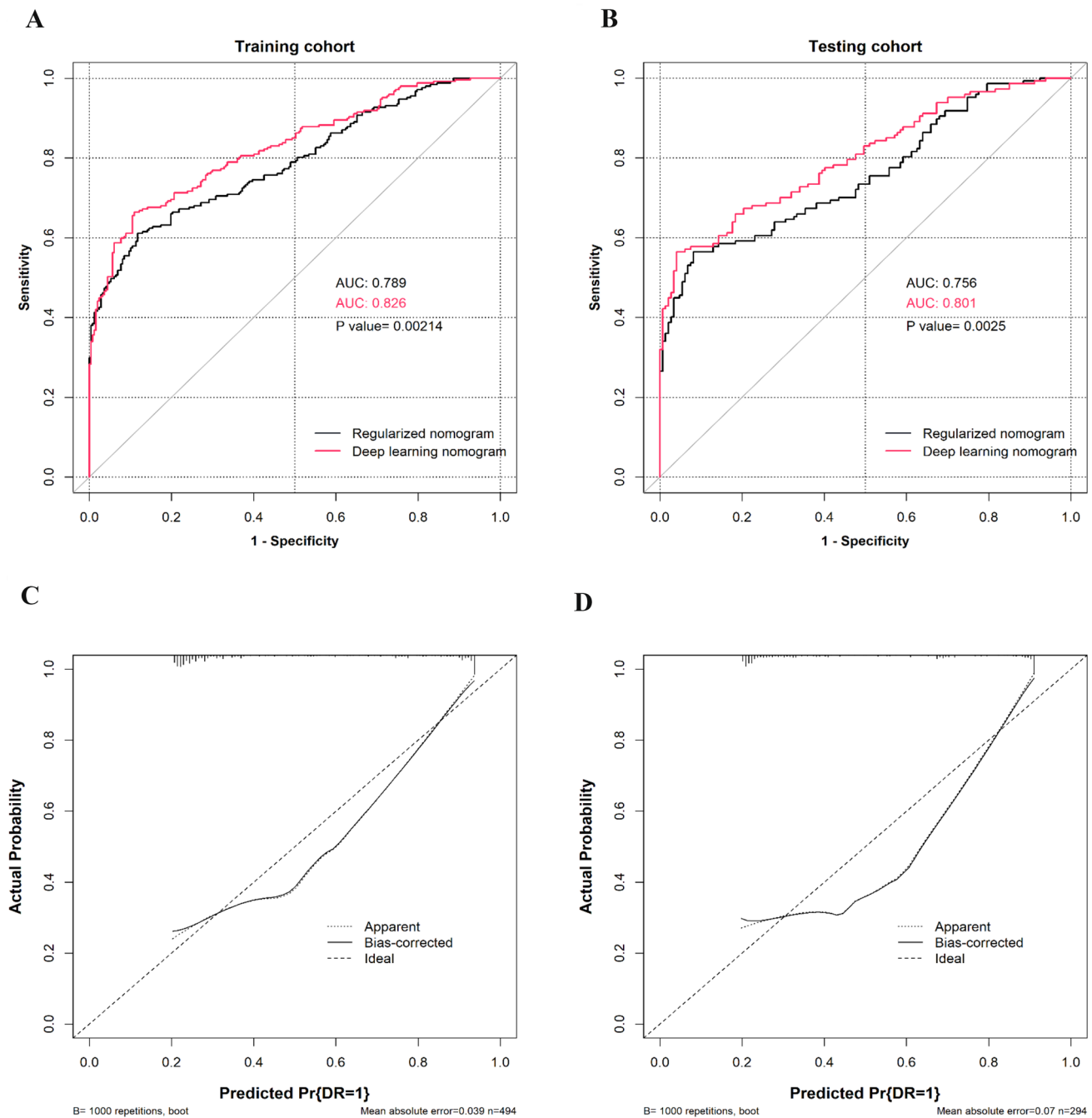
Fig. 4 The performance of regularized nomogram and the deep learning nomogram. **a** ROC curve of the regularized nomogram and the deep learning nomogram in training and **b** testing cohorts, respectively. **c, d** The calibration curve of the deep learning nomogram [Color needed]

both the training and testing cohorts. These comparisons illustrated that the deep learning nomogram could exploit the nonlinear relationships between CGM metrics. And the proposed deep learning approach could improve the effectiveness of the nomogram.

Clinical utility of the deep nomogram

The decision curve analysis for the deep learning nomogram is presented in supplementary Fig. 4. The decision curve showed that if the threshold probability of a patient was

Table 2 The performance of regularized nomogram and deep learning nomogram in all metrics

	Regularize nomogram		Deep learning nomogram	
	Training	Testing	Training	Testing
AUC	0.79 (0.74–0.82)	0.75 (0.70–0.81)	0.82 (0.78–0.85)	0.80 (0.75–0.85)
ACC	0.72 (0.68–0.76)	0.70 (0.64–0.76)	0.76 (0.71–0.79)	0.72 (0.66–0.77)
SEN	0.58 (0.51–0.64)	0.59 (0.50–0.67)	0.72 (0.66–0.77)	0.70 (0.62–0.77)
SPE	0.89 (0.84–0.93)	0.80 (0.73–0.87)	0.80 (0.74–0.84)	0.77 (0.69–0.83)
PRE	0.84 (0.79–0.89)	0.75 (0.66–0.82)	0.78 (0.72–0.83)	0.75 (0.67–0.81)
F1-score	0.67 (0.62–0.74)	0.66 (0.57–0.74)	0.76 (0.68–0.80)	0.72 (0.64–0.79)

0.20, using the deep learning nomogram to predict DR adds more benefit. Within this range, the deep learning nomogram showed more predictive value than the extreme curves. This result showed that most patients may benefit from this proposed model.

Discussion

This study fused deep learning with a nomogram model and constructed a deep learning nomogram to predict the risk of DR in T2D based on CGM profiles. The results demonstrated that the proposed deep learning nomogram could easily use CGM data to identify patients at high risk of DR. To our knowledge, this is the first study to develop a risk prediction model from CGM profiles. Based on our study, the deep learning nomogram could be embedded into the CGM device to provide a daily score assessment of patient glucose values. This deep learning nomogram could also help to select patients with T2D who should receive fundus examination, which could increase the effectiveness of DR screening. Meanwhile, our findings suggest that this deep learning nomogram developed from regularized logistic regression may be useful for developing personalized medicine for other diabetic complications, such as diabetic nephropathy, diabetic neuropathy, and microvascular diseases.

A nomogram is a statistical model used in clinical practice. The construction of mathematical equations allows for the exploration of the impact of indicators and disease-related risk factors to easily obtain risk probability. The strength of the deep learning nomogram is the use of deep learning in feature selection [22]. Although numerous metrics based on CGM have been proposed, such as TIR, glucose coefficient of variation (CV), and MBG [40], those metrics were developed to assess the quality of glucose control, not for the prediction of diabetes-related adverse outcomes, including DR. The regularized nomogram model is a visualization model that assumes a patient's risk of disease is a linear combination of the patient's clinical factors. Deep learning can synthesize linear and nonlinear associations of multiple indicators with target variables, explaining the higher-order information for predicting tasks. We have

shown that this deep learning nomogram efficiently triages DR patients with T2D with high precision.

To improve the prediction efficiency of the above model based on deep factors, we further integrated the common patient characteristics (diabetes duration and HbA1c) into the deep learning nomogram (Fig. 5). To demonstrate superior discriminating ability, we also computed the *Akaike Information Criterion* (AIC) values of the three models. The performance of the deep learning nomogram was improved significantly by including clinical factors (Fig. 6). The performance of the combined model was significantly lower than the deep learning nomogram and regularized nomogram. The results show that the new model has the best adaptive properties. The AUC of the promoted model in the training cohort was 0.86, and in the testing cohort, it was 0.85. The calibration plot indicated good calibration based on the promoted nomogram model in both the training and testing cohort.

The majority of the literature [41–44] built diagnostic models for DR in T2D based on fundus photographs and electronic health records (EHR). In [41], Omolola constructed a machine-learning approach from EHR with an AUC of 0.72 in the validation cohort. Our proposed deep learning nomogram with clinical factors could achieve 0.84 in the validation cohorts. Without a multitude of laboratory variables, the CGM-based model led to the economical and convenient prediction of DR. In [42], M.T.Esfahan built a CNN network to detect DR from fundus images. They reported an accuracy of 85% and a sensitivity of 86%, which was significantly higher than ours. It was worth noting that the CGM-based DR prediction model was not intended to replace the current eye examination, which was the gold standard for DR diagnosis. Although the comparison of results directly was one-sided due to the different samples. However, the results of the comparison could indicate that the CGM-based model could be a useful screening method to identify those patients at higher risk. This will help these patients to receive further inspection in time. Therefore, the role of the CGM-based models in the development of DR should be investigated.

There are several limitations to this study that should be mentioned. First, the datasets are limited, and more samples

Fig. 5 The nomogram and calibration curves of the model based on the deep factors and clinical factors. **a**, **b** The calibration curves between the nomogram prediction and actual observation for nDR and DR patients in the training and test datasets, respectively. **c**, **d** The diagonal dotted line represents an ideal evaluation, while the solid lines and dashed lines represent the performance of the corrected and apparent bias, respectively. The closer the fit is to the diagonal dotted line, the better the evaluation. **e** The deep learning nomogram integrated the deep factors with the Duration and HbA1c in the training datasets. The probability value of each T2D patient with DR is marked on each axis. **f** The table of the performance of deep learning nomogram with clinical factors
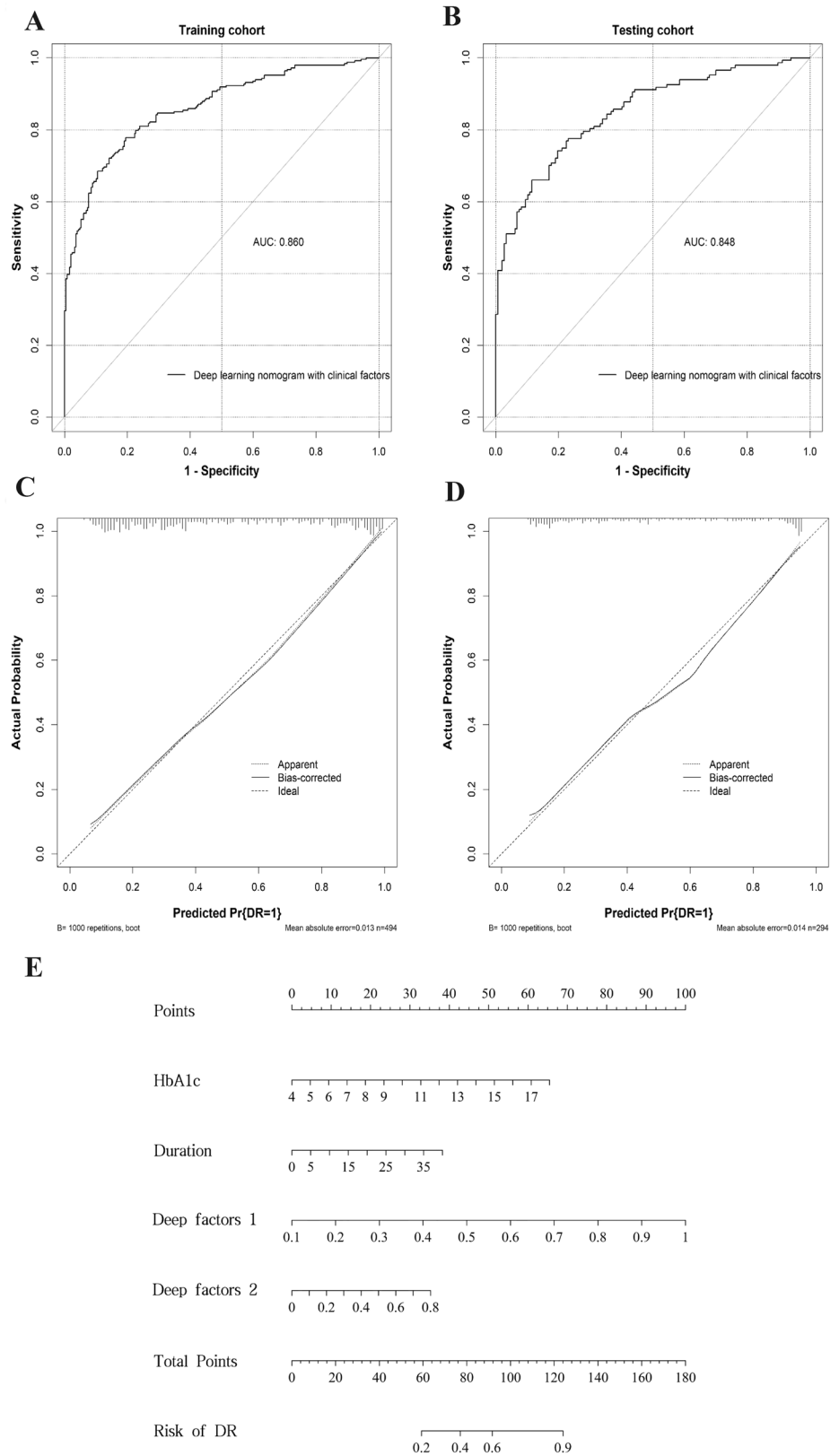
Fig. 5 (continued)

F	Training cohort	Testing cohort
AUC	0.86	0.85
	(0.83-0.89)	(0.81-0.89)
ACC	0.79	0.77
	(0.75-0.82)	(0.72-0.82)
Sensitivity	0.77	0.75
	(0.71-0.82)	(0.67-0.81)
Specificity	0.81	0.80
	(0.75-0.86)	(0.73-0.86)
Precision	0.80	0.79
	(0.74-0.85)	(0.71-0.85)
F1-score	0.78	0.76
	(0.72-0.83)	(0.69-0.83)

Fig. 6 The histogram of the AIC values between the three models. The lower the AIC value, the more accurate the model fit. *** indicates that the AIC values of the two models are significantly different
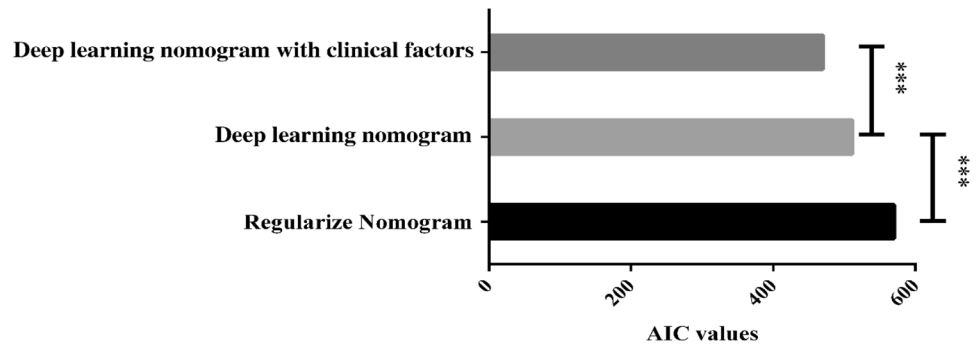
should be considered in the future to enhance the model's scalability. Second, multi-center data will improve the generality of our funding. In addition, the one-day CGM data may not be fully representative of patients' historical glycemic control, and the results should be interpreted with caution. Moreover, although this paper solves some of the data bias through data collection and PSM, further construction of a bias model will make the CGM-based DR prediction model more clinically universal [45, 46].

Conclusion

The developments of CGM sensors provide a fertile environment for powerful analysis in precession medicine in T2D management. In this study, we established a novel CGM-based DR prediction model with an AUC of 0.86 and a specificity of 0.81. The novelty of this work lies in the combination of deep learning and nomogram to analyze the un-linear relationship between CGM profiles and DR. The method could be embedded in CGM sensors to provide daily DR risk prediction for patients, which could screen patients and recommend high-risk groups for more comprehensive examinations. There are still other complications that exist in the actual situation. Hence, future research will consider the application of the CGM-based deep learning nomogram to other diabetic complications. The use of such a method can assist in the treatment of precision medicine for T2D patients.

Data availability The data that support the studying are not publicly available due to privacy or ethical restrictions.

Declarations

Conflict of interest The authors declare no conflict of interest relevant to this work.

Ethical approval Research Ethics Committees of Shanghai Jiao Tong University School of Medicine Affiliated Sixth People's Hospital (Date 2017-04-27/No.2017-047).

Consent to participate Informed consent was obtained from individual participants included in the study.

Consent to publication Patients signed informed consent regarding publishing their data.

References

1. Aiello L (1998) Diabetic retinopathy (technical review). Diabetes Care 21(1):143–156
2. Frederick L, Macular E (1984) A complication diabetic retinopathy. Survey of Ophthalmol 28:452–461. https://doi.org/10.1016/0039-6257(84)90227-3
3. Yau JWY, Rogers SL, Kawasaki R et al (2012) Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 35(3):556–564
4. Hendarto H, Pramono A, Saksono D et al (2016) The relationship of glycemic exposure (HbAlc) to the risk of subclinical hypothyroidism in type 2 diabetes mellitus patients. Diabetes Res Clin Pract 120(1):S204–S204
5. Mohamed Q, Gillies MC, Wong TY (2007) Management of diabetic retinopathy: a systematic review. JAMA 298(1):902–916
6. Pei X, Yao X, Yang Y et al (2022) Efficacy of artificial intelligence-based screening for diabetic retinopathy in type 2 diabetes mellitus patients. Diabetes Res Clin Pract 184:109190
7. Maistry A, Pillay A, Jembere E (2020) (2020) Improving the accuracy of diabetes retinopathy image classification using augmentation. Conference of the South African Institute of Computer Scientists and Information Technologists 1:134–140
8. Shin KH, Chi MJ (2009) Fundus examination rate in diabetics and the public health factors associated with fundus examination rate. J Korean Ophthalmol Soc 50(1):1319
9. Abràmoff MD, Niemeijer M, Suttorp-Schulten MSA et al (2008) Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. Diabetes Care 31:193–198
10. Bolinder J, Antuna R, Geelhoed-Duijvestijn P et al (2016) Novel glucose-sensing technology and hypoglycaemia in type 1 diabetes: a multicentre, non-masked, randomised controlled trial. The Lancet 388(10057):2254–2263
11. Li L, Sun J, Liemin R et al (2021) Time-series analysis of continuous glucose monitoring data to predict treatment efficacy in patients with T2DM. J Clin Endocrinol Metab 8(8):1–18
12. Zahedani AD, Torbaghan SS, Rahili S et al (2021) Improvement in glucose regulation using a digital tracker and continuous glucose monitoring in healthy adults and those with type 2 diabetes. Diabetes Therapy 1:1–16
13. Tao R, Yu X, Lu J et al (2021) Multilevel clustering approach driven by continuous glucose monitoring data for further classification of type 2 diabetes. BMJ Open Diabetes Res Care 9(1):e001869
14. Yu X, Ma N, Yang T et al (2021) A multi-level hypoglycemia early alarm system based on sequence pattern mining. BMC Med Inform Decis Mak 21(1):1–11
15. Sartore G, Chilelli NC, Burlina S et al (2013) Association between glucose variability as assessed by continuous glucose monitoring (CGM) and diabetic retinopathy in type 1 and type 2 diabetes. Acta Diabetol 50(1):437–442
16. Lu J, Ma X, Zhou J et al (2018) Association of time in range, as assessed by continuous glucose monitoring, with diabetic retinopathy in type 2 diabetes. Diabetes Care 41(11):2370–2376
17. Hirsch IB (2015) Glycemic variability and diabetes complications: does it matter? of course it does! Diabetes Care 38(8):1610–1614
18. De Nicolao G, Magni L, Dalla Man C et al (2011) Modeling and control of diabetes: towards the artificial pancreas. IFAC Proceedings Volumes 44(1):7092–7101
19. Fox KAA, FitzGerald G, Puymirat E et al (2014) Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. BMJ Open 4(2):e004425
20. Sperandei S (2014) Understanding logistic regression analysis. Biochemia medica 24(1):12–18
21. Legge BJ, Banister EW (1986) The Astrand-Ryhming nomogram revisited. J Appl Physiol 61(3):1203–1209
22. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press, Cambridge
23. Carin L, Pencina MJ (2018) On deep learning for medical image analysis. JAMA 320(11):1192–1193
24. Katzman JL, Shaham U, Cloninger A et al (2018) DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 18(1):1–12
25. Cai W, He B, Hu M et al (2019) A radiomics-based nomogram for the preoperative prediction of posthepatectomy liver failure in patients with hepatocellular carcinoma. Surg Oncol 28:78–85
26. Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. Journal of economic surveys 22(1):31–72
27. Randolph JJ, Falbe K (2014) A step-by-step guide to propensity score matching in R. Pract Assess Res Eval 19(18):6
28. Prenter PM (2008) Splines and variational methods. Dover Publicatio, New York
29. Flaxel CJ, Adelman RA, Bailey ST et al (2020) Diabetic retinopathy preferred practice pattern. Ophthalmology 127(1):P66–P145
30. J. Z. WeIPing J, Yuqian B. (2019) Continuous glucose monitoring. Shanghai Science and Technology Pass, Shanghai.
31. Remeseiro B, Bolon-Canedo V (2019) A review of feature selection methods in medical applications. Comput Biol Med 112:103375
32. Raj RJS, Shobana SJ, Pustokhina IV et al (2020) Optimal feature selection-based medical image classification using deep learning model in internet of medical things. IEEE Access 8:58006–58017
33. Roth V (2004) The generalized LASSO. IEEE Trans Neural Networks 15(1):16–28
34. Jiang C, Kong Z, Liu S et al (2019) Fusion radiomics features from conventional MRI predict MGMT promoter methylation status in lower grade gliomas. Eur J Radiol 121:108714
35. Huang Y, Li H, Yu X (2021) A multiview feature fusion model for heartbeat classification. Physiol Meas 42(6):065003
36. Seo JH, Kim HJ, Lee JY (2020) Nomogram construction to predict dyslipidemia based on a logistic regression analysis. J Appl Stat 47(5):914–926
37. Zhou XH, Higgs RE (1998) COMPROC and CHECKNORM: computer programs for comparing accuracies of diagnostic tests using ROC curves in the presence of verification bias. Comput Methods Programs Biomed 57(3):179–186
38. Obuchowski N A, Bullen J A. (2018) Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Physics in Medicine & Biology 63(7): 07TR01.
39. Osterhoudt H W, Ray Jr L N. (1967) Gel permeation chromatography: On the shape of the calibration plot and molecular size separability. Journal of Polymer Science Part A-2: Polymer Physics 5(3): 569–581.
40. Bailey R (2019) The relationship between continuous glucose monitor (CGM) derived metrics and indices of glycemic control. University of South Florida, South Florida

41. Ogunyemi O, Kermah D (2015) Machine learning approaches for detecting diabetic retinopathy from clinical and public health records. AMIA Ann Symp Proc 11:983–990

42. Wang R, Miao Z, Liu T, Liu M, Grdinovac K, Song X, Liang Y, Delen D, Paiva W (2021) Derivation and validation of essential predictors and risk index for early detection of diabetic retinopathy using electronic health records. J Clin Med. https://doi.org/10.3390/jcm10071473

43. Esfahani MT, Ghaderi M, Kafiyeh R (2018) Classification of diabetic and normal fundus images using new deep learning method. Leonardo Electron J Pract Technol 17(32):233–248

44. Quellec G, Charriere K, Boudi Y, Cochener B, Lamard M (2017) Deep image mining for diabetic retinopathy screening. Med Image Anal 39:178–193

45. Rashid MM, Askari MR, Chen C et al (2022) Artificial intelligence algorithms for treatment of diabetes. Algorithms 15(9):299

46. Askari MR et al (2022) Detection and classification of unannounced physical activities and acute psychological stress events for interventions in diabetes treatment. Algorithms 15(10):352