Status of glycosylated hemoglobin and prediction of glycemic control among patients with insulin-treated type 2 diabetes in North China: a multicenter observational study

Jiao Wang[1], Meng-Yang Wang[1], Hui Wang[1], Hong-Wei Liu[1], Rui Lu[1], Tong-Qing Duan[1], Chang-Ping Li[1], Zhuang Cui[1], Yuan-Yuan Liu[1], Yuan-Jun Lyu[2], Jun Ma[1]

[1]Department of Health Statistics, College of Public Health, Tianjin Medical University, Tianjin 300070, China;
[2]Department of Endocrinology, Tianjin Hospital, Tianjin 300211, China.

Abstract

Background: Blood glucose control is closely related to type 2 diabetes mellitus (T2DM) prognosis. This multicenter study aimed to investigate blood glucose control among patients with insulin-treated T2DM in North China and explore the application value of combining an elastic network (EN) with a machine-learning algorithm to predict glycemic control.

Methods: Basic information, biochemical indices, and diabetes-related data were collected via questionnaire from 2787 consecutive participants recruited from 27 centers in six cities between January 2016 and December 2017. An EN regression was used to address variable collinearity. Then, three common machine learning algorithms (random forest [RF], support vector machine [SVM], and back propagation artificial neural network [BP-ANN]) were used to simulate and predict blood glucose status. Additionally, a stepwise logistic regression was performed to compare the machine learning models.

Results: The well-controlled blood glucose rate was 45.82% in North China. The multivariable analysis found that hypertension history, atherosclerotic cardiovascular disease history, exercise, and total cholesterol were protective factors in glycosylated hemoglobin (HbA1c) control, while central adiposity, family history, T2DM duration, complications, insulin dose, blood pressure, and hypertension were risk factors for elevated HbA1c. Before the dimensional reduction in the EN, the areas under the curve of RF, SVM, and BP were 0.73, 0.61, and 0.70, respectively, while these figures increased to 0.75, 0.72, and 0.72, respectively, after dimensional reduction. Moreover, the EN and machine learning models had higher sensitivity and accuracy than the logistic regression models (the sensitivity and accuracy of logistic were 0.52 and 0.56; RF: 0.79, 0.70; SVM: 0.84, 0.73; BP-ANN: 0.78, 0.73, respectively).

Conclusions: More than half of T2DM patients in North China had poor glycemic control and were at a higher risk of developing diabetic complications. The EN and machine learning algorithms are alternative choices, in addition to the traditional logistic model, for building predictive models of blood glucose control in patients with T2DM.

Keywords: Type 2 diabetes; Blood glucose; HbA1c; Elastic network; Machine learning

Introduction

It was estimated that in 2017, there were 451 million people with diabetes worldwide; moreover, this figure is expected to increase to 693 million by 2045.[1] According to the report of the International Diabetes Federation in 2017, the type 2 diabetes mellitus (T2DM) epidemic in China is the largest in the world, with 11.44 million people living with diabetes, which was expected to increase to 11.98 million by 2045.[2] The healthcare expenditure due to T2DM was USD 630 billion in 2017 in China.[2] Despite strong financial support, the blood glucose control rate among individuals with diabetes in China is still very low.[3]

The prognosis of diabetes is mainly referred to based on diabetic complications and comorbidities, including hypertension, cardiovascular diseases, neuropathy, nephropathy, and retinopathy, which can result in poor quality of life and reduced life expectancy.[4-5] However, the World Health Organization has endorsed the use of glycosylated hemoglobin (HbA1c) as a screening test for people at high risk of diabetes and, more importantly, as a test for predicting the risk of microvascular complications.[6]

Access this article online	
Quick Response Code:	Website: www.cmj.org
	DOI: 10.1097/CM9.0000000000000585

Correspondence to: Dr. Jun Ma, Department of Health Statistics, College of Public Health, Tianjin Medical University, No. 22 Qixiangtai Road, Heping District, Tianjin 300070, China
E-Mail: junma@tmu.edu.cn

Chinese Medical Journal 2020;133(1)
Received: 08-06-2019 **Edited by:** Yuan-Yuan Ji

Hence, as the gold standard measurement for long-term blood glucose control, HbA1c is a good index of the prognosis of diabetes.

Previous studies have examined the incidence of diabetes and blood glucose control in the Chinese population. For example, Xu *et al*[3] showed that the overall blood glucose control level of the Chinese population in 2010 was 39.7%, but diabetes is related to lifestyle and economic development. Therefore, the data may have changed. Moreover, an increasing number of studies have focused on diabetes in North China. Gao *et al*[7] found that sedentary civil servants had high rates of diabetes in Xinjiang, China. A multicenter study in Shanxi, China described the simultaneous control rate of blood pressure, blood sugar and blood lipids among drug-treated T2DM patients.[8] Unfortunately, most of these studies focused on the situation in a single region or special population that rarely touch upon the general situation in North China.[9-12]

In recent years, with the development of machine learning, random forest (RF), support vector machine (SVM), and back propagation artificial neural network (BP-ANN) have been widely used in the field of diabetes prediction.[13,14] In addition, studies have shown that combining machine learning algorithms with dimensionality reduction can improve their performance.[15-17] However, few studies have introduced the method of combining dimensionality reduction with machine learning in the field of diabetes.[18]

Hence, the primary objective of this study was to investigate the general status of diabetic blood glucose control and its influencing factors in insulin-treated T2DM outpatients in North China using multicenter data from various provinces and cities. Moreover, this study explored the application value of combining an elastic network (EN) with a machine-learning algorithm to predict diabetic blood glucose control.

Methods

Ethical approval

The study was approved by the Ethics Committees of Tianjin Medical University (No. TMUHMEC2013032). All patients gave their informed consent.

Subjects

According to the pre-calculated sample size [Supplementary File 1, http://links.lww.com/CM9/A160], 27 centers (including secondary hospitals and tertiary hospitals) from six cities in China (Tianjin, Tangshan, Datong, Qinhuangdao, Cangzhou, and Taiyuan) that agreed to participate in the experiment were selected in North China [Supplementary File 2, http://links.lww.com/CM9/A161]. From January 2016 to December 2017, data from the first five outpatients were collected daily, and 2787 consecutive patients who met the inclusion and exclusion criteria were eventually included. The inclusion criteria were as follows: (1) T2DM diagnosis, (2) ≥18 years old, and (3) basal insulin use ≥3 months. The exclusion criteria were as follows: (1) refusal to sign informed consent, (2) history of drug allergy or allergies, (3) preparation to get pregnant or current pregnancy or lactation, or (4) psychiatric conditions.

Data collection

Basic information (sex, age, smoking status, alcohol consumption, marital status, etc) and diabetes-related information (typical characteristics of diabetes, duration of diabetes, exercise, diet, oral medications, complications, hypoglycemia, insulin injection time and injection dosage, etc) were collected by questionnaire after the subjects were recruited. The physical examination included height, weight, waist circumference, hip circumference, and blood pressure. The laboratory tests included fasting blood glucose (FBG), 1-h blood glucose (1HBG) and 2-h blood glucose (2HBG), which were measured by the glucose oxidase method; HbA1c, which was determined by high-performance liquid chromatography; total cholesterol (TC) and triglycerides (TGs), which were determined by an enzymatic method; and high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C), which were measured by a homogeneous method.

Quality control

Uniform training for all investigators was provided using standardized procedures. The structured interview method was used to conduct the questionnaire survey to avoid differences in the inquiry process. All data from the questionnaire were thoroughly scrutinized, and missing or illogical options were reassessed and modified.

Statistical analysis

Statistical analyses were performed using the statistical software SAS 9.4 (SAS Inc., North Carolina, USA) and R 3.5.1 (https://www.r-project.org/). The continuous variables were expressed as the median (P_{25}, P_{75}). The categorical variables were represented by the frequency and composition ratio. According to the "Guidelines for the Prevention and Treatment of Type 2 Diabetes in China (2017 Edition),"[19] HbA1c <7.0% was defined as the standard of glycemic control. The continuous data and categorical data were compared by the Wilcoxon rank-sum test and Chi-square test, respectively, between the controlled and uncontrolled groups.

Multicollinearity usually exists in the data of chronic diseases,[20] so we choose an EN with machine learning validation to obtain a smaller prediction error. It is important to note that ENs eliminate certain predictors to avoid overfitting.[21] The EN algorithm is a combination of least absolute shrinkage and selection operator (LASSO) and ridge regression. When there are multiple collinear predictors, LASSO selects only one, ignores others, or zeroes some regression coefficients. The ridge method counteracts collinearity and variance inflation by shrinking the regression coefficients towards zero but without reaching zero. EN combines the penalties of the LASSO and ridge approaches.[22] This algorithm achieves sparse coefficient estimates by minimizing the sum of squared errors by adding an ℓ 1 and a squared ℓ 2 penalty simultaneously to the coefficient β.[23] The EN removes the limitation of the number of selected variables of LASSO and encourages the grouping effect.

In this study, RF, SVM, and BP-ANN were selected to model and predict diabetes-related variables before and after dimensional reduction by EN. The data were divided into training sets and testing sets on a 7:3 scale. The optimal parameters of each model were found using ten-fold cross-validation, and the area under the receiver operating characteristics curve (AUC) was used to select the optimal model. Then, we estimated the performance of the training model by predicting HbA1c control in the test set. Sensitivity, specificity, accuracy, and AUC were selected to compare the predictive performance of each model. Classification accuracy is a common method in pattern recognition and refers to the ratio of the number of correctly classified samples to the total number of samples. Sensitivity and specificity are the statistical indicators of the performance of classification tests. Sensitivity measures the proportion of actual positive cases that were correctly identified. Specificity refers to the proportion of negativities that were correctly identified. Accuracy, sensitivity, and specificity are expressed as follows, where TP means number of positive cases those are correctly classified; TN means number of negative cases those are correctly classified; FP means number of positive cases those are misclassified; and FN means number of negative cases those are misclassified:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\% \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100\% \tag{3}$$

For each algorithm, the final result of the predictive performance is taken as the average value of ten cycles. To compare with machine learning models, this study also established a stepwise logistic regression model.

The RF classifier is composed of multiple tree classifiers, in which each classifier is generated by a random vector independent of the input vector, and the input vector is classified by unit voting.[24] The RF classifier used in this study was created by creating a tree using randomly selected features or a combination of features on each node. As a supervised classifier, the basic principle of SVM is to map the input vector to a high-dimensional feature space through the pre-selected non-linear relation and find an optimal classification hyperplane in this space to maximize the classification interval between the two classes.[25] BP-ANN is a multilayer feedforward neural network trained by an error back-propagation algorithm, which is one of the most widely used neural network models. The learning rule of BP-ANN is to use the steepest descent method and continuously adjust the weight and threshold of the network through the back-propagation algorithm to minimize the error square sum of the network to identify an optimal model.[26]

Results

General information of the subjects

According to the inclusion and exclusion criteria, 2787 T2DM patients were enrolled in the study, including 1407 males (50.48%) and 1380 females (49.52%) aged from 19 to 91 years old. Almost all participants (98.49%) were married, and only a few were single (1.51%). Table 1 shows the general information of all subjects in the two groups.

Only 1277 (45.82%) subjects met the HbA1c control standard. The percentage of central adiposity in the controlled group was 46.28% compared to 57.02% in the uncontrolled group ($\chi^2 = 31.9762$, $P < 0.0001$). Almost 90% of the subjects in the controlled group exercised, but the rate in the uncontrolled group was only 76.49% ($\chi^2 = 77.7556$, $P < 0.0001$).

The disease conditions and drug treatments of the subjects are shown in Table 2. The overall duration of T2DM was 4 (2.0, 7.5) years and 7 (3.0, 10.0) years in the two groups ($Z = -10.5209$, $P < 0.0001$). Over 50% of participants had typical disease characteristics (polydipsia, polyphagia, polyuria, and emaciation) in the uncontrolled group compared to 40.33% in the controlled group. In the controlled group, 33.59% of the subjects had complications, while 53.18% had complications in the non-controlled group ($\chi^2 = 107.5936$, $P < 0.0001$). The rate of hypoglycemia was lower in the controlled group than in the uncontrolled group (14.41% and 23.38%, $\chi^2 = 35.7774$, $P < 0.0001$). The time to the initiation of basal insulin between the two groups was significantly different ($\chi^2 = 23.3311$, $P < 0.0001$). Approximately 9.4% of the subjects were not taking oral hypoglycemic drugs (OHA) in the controlled group, compared to only 6.75% in the uncontrolled group ($\chi^2 = 14.1481$, $P = 0.0068$).

Table 3 shows the serum biochemical indices in the two groups. All biochemical indicators were standardized according to the median values. The differences in FBG, 2HBG, blood pressure, TC, total TGs, HDL-C, and LDL-C between the two groups were statistically significant.

Results of dimension reduction by elastic net

Figure 1 shows the process of EN variable filtering. Table 4 shows the results of dimension reduction by elastic net regression. A total of 42 variables with statistical significance in single factor analysis were entered into the model, and then 19 variables were selected. The variance analysis of the model showed that the model is valid [Table 5] ($F = 190.55$, $P < 0.0001$). Hypertensive history, atherosclerotic cardiovascular disease history, nocturnal hypoglycemia, exercise, and TC were protective factors for HbA1c control. Central adiposity, family

Table 1: General information of patients with insulin-treated type 2 diabetes in HbA1c controlled and uncontrolled group.

Characteristics	Controlled group ($n = 1277$)	Uncontrolled group ($n = 1510$)	χ^2/Z	P
Age (years)	57 (48, 63)	58 (50, 65)	–3.6010*	0.0003
Gender			6.5609†	0.0104
Male	611 (47.85)	796 (52.72)		
Female	666 (52.15)	714 (47.28)		
Married	1263 (98.90)	1482 (98.15)	2.6781†	0.1017
Height (cm)	167 (160, 173)	168 (160, 173)	–0.8608*	0.3894
Weight (kg)	70 (63, 78)	70 (63, 79)	–1.4624*	0.1436
BMI (kg/m^2)	25.31 (23.67, 27.04)	25.39 (23.48, 27.34)	–0.8008*	0.4233
BMI			10.7287†	0.0133
Thin (BMI <18.5 kg/m^2)	7 (0.55)	15 (0.99)		
Normal (18.5≤ BMI <24 kg/m^2)	380 (29.76)	462 (30.60)		
Overweight (24≤ BMI <28 kg/m^2)	701 (54.89)	754 (49.93)		
Obesity (BMI ≥28 kg/m^2)	189 (14.80)	279 (18.48)		
Waist (cm)	86 (80, 92)	89 (81, 96)	5.8483*	<0.0001
Central adiposity	591 (46.28)	861 (57.02)	31.9762†	<0.0001
Smoking history	393 (30.78)	526 (34.83)	5.1581†	0.0231
Drinking history	387 (30.31)	504 (33.38)	3.0023†	0.0831
Family history	309 (24.20)	550 (36.42)	48.5075†	<0.0001
Exercise	1140 (89.27)	1155 (76.49)	77.7556†	<0.0001
Exercise frequency			79.2510†	<0.0001
0 time/week	137 (10.73)	355 (23.51)		
≤2 times/week	424 (33.20)	432 (28.61)		
3–4 times/week	579 (45.34)	566 (37.48)		
≥5 times/week	137 (10.73)	157 (10.40)		
Diet adjustment	1195 (93.58)	1299 (86.03)	41.9458†	<0.0001
Staple food	1170 (91.62)	1264 (83.71)	39.1583†	<0.0001
Vegetable	319 (24.98)	329 (21.79)	3.9513†	0.0468
Vegetable oil	796 (62.33)	720 (47.68)	59.8719†	<0.0001
Salt	468 (36.65)	540 (35.76)	0.2357†	0.6273
Anamnesis	261 (20.44)	456 (30.20)	34.4926†	<0.0001
Hypertensive	179 (14.02)	314 (20.79)	21.8273†	<0.0001
ASCVD	79 (6.19)	146 (9.67)	11.3064†	0.0008

Values were expressed as median (P_{25}, P_{75}) or n (%).Central adiposity was designed as waist size greater than 85 cm for women and 90 cm for men. Diet adjustment is one or more that controls the intake of staple foods, vegetables, vegetable oils and salt. * Wilcoxon rank-sum test. † The Chi-square test. HbA1c: Glycosylated hemoglobin; BMI: Body mass index; ASCVD: Atherosclerotic cardiovascular disease.

history, duration of T2DM, typical disease characteristics, complications, insulin dose, OHA, FBG, 2HBG, blood pressure, HDL-C, LDL-C, and hypertension were risk factors for HbA1c control.

Results of machine learning

Table 6 shows the prediction results of RF, SVM, BP-ANN, and logistic regression. To increase the reliability of the prediction, ten cycles were performed for each model in this study, and the average value of the ten cycles was used for the final predictive performance evaluation index.

According to the results, the sensitivity, specificity, accuracy, and AUC of the RF algorithm without EN dimension reduction for variables were 0.80, 0.71, 0.74, and 0.73, respectively. Although the sensitivity of the reduced dimensional model decreased to 0.79, the specificity, accuracy, and AUC were improved overall. After using an EN to reduce the dimensions of the variables, the sensitivity of the SVM algorithm was improved by 37.70%, specificity by 7.94%, accuracy by 17.74%, and AUC by 18.03%, which were the most improved among the three methods. When the RF algorithm was combined with the variables of dimensionality reduction through EN, the predictive performance was generally improved; the sensitivity was increased from 0.75 to 0.78, the specificity from 0.69 to 0.70, the accuracy from 0.71 to 0.73, and the AUC from 0.70 to 0.72. Among the three machine learning algorithms, RF performed better than SVM and BP-ANN in predicting blood glucose control, both before and after the dimensionality reduction. Generally, after variable screening using the EN first, the three machine learning algorithms were used to predict the long-term blood glucose control in T2DM patients, and this method significantly improved the predictive performance of the abovementioned machine learning algorithm.

Compared with the results of traditional logistic regression, although the logistic specificity was slightly higher

Table 2: Diabetes condition and treatment of patients with insulin-treated type 2 diabetes in HbA1c controlled and uncontrolled groups.

Items	Controlled group ($n = 1277$)	Uncontrolled group ($n = 1510$)	χ^2/Z	P
Duration of T2DM (years)	4.0 (2.0, 7.5)	7.0 (3.0, 10.0)	–10.5209*	<0.0001
Insulin dose (U)	16.0 (12.0, 20.0)	17.0 (13.0, 22.0)	–5.8229*	<0.0001
Typical disease characteristics	515 (40.33)	792 (52.45)	40.8202†	<0.0001
Polydipsia	345 (27.04)	624 (41.38)	62.6501†	<0.0001
Polyphagia	240 (18.81)	383 (25.40)	17.2755†	<0.0001
Polyuria	323 (25.31)	579 (38.40)	54.0041†	<0.0001
Emaciation	324 (25.41)	455 (30.17)	7.7682†	0.0053
Complication	429 (33.59)	803 (53.18)	107.5936†	<0.0001
DN	101 (8.22)	281 (19.30)	67.0639†	<0.0001
Diabetic retinopathy	175 (14.03)	391 (26.24)	61.6787†	<0.0001
DPN	291 (23.87)	523 (36.19)	47.3201†	<0.0001
Diabetic foot	13 (1.07)	28 (1.94)	3.2627†	0.0709
LEAD	116 (9.53)	231 (15.93)	23.9410†	<0.0001
Comorbidity	328 (25.69)	531 (35.17)	29.1645†	<0.0001
Hypertension	173 (13.55)	314 (20.79)	25.2002†	<0.0001
Hyperlipidemia	94 (7.36)	122 (8.08)	0.4995†	0.4797
ASCVD	83 (6.50)	139 (9.21)	6.9091†	0.0086
Previous hypoglycemia	184 (14.41)	353 (23.38)	35.7774†	<0.0001
Hypoglycemia			35.7816†	<0.0001
Mild hypoglycemia	180 (97.83)	345 (97.73)		
Severe hypoglycemia	4 (2.17)	8 (2.27)		
Nocturnal hypoglycemia	37 (2.90)	80 (5.30)	9.9139†	0.0016
Time to start using basic insulin			23.3311†	0.0001
<6 months	407 (31.87)	388 (25.70)		
6–11 months	411 (32.18)	470 (31.13)		
12–35 months	326 (25.53)	422 (27.95)		
≥36 months	133 (10.42)	230 (15.23)		
Oral hypoglycemic drugs			14.1481†	0.0068
None	120 (9.40)	102 (6.75)		
1 kind	566 (44.32)	628 (41.59)		
2 kinds	536 (41.97)	685 (45.36)		
3 kinds	51 (3.99)	88 (5.83)		
4 kinds	4 (0.31)	7 (0.46)		
Biguanides	769 (60.27)	957 (63.42)	2.9166†	0.0877
Sulfonylureas prolactin	258 (20.20)	303 (20.08)	0.0066†	0.9351
Glinides	200 (15.67)	243 (16.11)	0.1000†	0.7518
DPP-4 inhibitors	54 (4.23)	105 (6.96)	9.5763†	0.0020
Follow medical advice‡	1221 (96.61)	1404 (92.98)	8.7718†	0.0031

Values were expressed as median (P_{25}, P_{75}) or n (%). * Wilcoxon rank-sum test. † The Chi-square test. ‡ Follow medical advice includes oral hypoglycemic drugs and insulin treatment. HbA1c: Glycosylated hemoglobin; T2DM: Type 2 diabetes; DN: Diabetic nephropathy; DPN: Diabetic peripheral neuropathy; LEAD: Diabetic lower-extremity arterial disease; ASCVD: Atherosclerotic cardiovascular disease; DPP-4: Dipeptidyl peptidase.

than that of the EN and machine learning model, the sensitivity and accuracy (sensitivity = 0.51, accuracy = 0.55) were much lower than those of the EN and machine learning model. In summary, the EN and machine learning model is a good alternative to the traditional logistic regression for blood glucose prediction in individuals with T2DM.

Discussion

In this study, a combination of the EN and machine learning algorithms was used to analyze the factors affecting the standard level of blood glucose among individuals with diabetes, and good results were obtained. On the basis of the traditional Logistic regression model, a new idea on prediction of glycemic control was provided.

Blood glucose control is closely related to complications and prognosis,[27,28] so effective blood glucose control is considered the basis of diabetes treatment. The overall glycemic control rate in our study was 45.82%, which was higher than the 32.60% obtained in a study assessing individuals with HbA1c <7%.[29] The results showed that central adiposity, family history, duration of diabetes, blood pressure, and hypertension were risk factors for high HbA1c, which was consistent with previous research conclusions.[30,31] Exercise and modifying vegetable oil

Table 3: Serum biochemical indices of patients with insulin-treated type 2 diabetes in HbA1c controlled and uncontrolled group.

Indices	Controlled group (n = 1277)	Uncontrolled group (n = 1510)	χ^2	P
FBG			291.6056	<0.0001
Normal	582 (45.58)	241 (15.96)		
Abnormal	695 (54.42)	1269 (84.04)		
2HBG			581.2302	<0.0001
Normal	566 (44.32)	84 (5.56)		
Abnormal	711 (55.68)	1426 (94.44)		
BP			39.9767	<0.0001
Normal	1006 (76.78)	1035 (68.54)		
Abnormal	271 (21.22)	475 (31.46)		
TC			13.0408	0.0003
Normal	979 (76.66)	1066 (70.60)		
Abnormal	298 (23.34)	444 (29.40)		
TGs			17.5275	<0.0001
Normal	860 (67.35)	901 (59.67)		
Abnormal	417 (32.65)	609 (40.33)		
HDL-C			8.5490	0.0035
Normal	884 (69.22)	966 (63.97)		
Abnormal	393 (30.78)	544 (36.03)		
LDL-C			33.1003	<0.0001
Normal	867 (67.89)	865 (52.28)		
Abnormal	410 (32.11)	645 (42.72)		

Values were expressed as n (%). HbA1c: Glycosylated hemoglobin; FBG: Fasting blood glucose (normal range: 3.8–6.1 mmol/L); 2HBG: 2-h Blood glucose (normal range: <7.8 mmol/L); BP: Blood pressure (normal range: systolic BP <140 mmHg and diastolic BP <90 mmHg); TC: Total cholesterol (normal range: 2.5–5.7 mmol/L); TGs: Triglycerides (normal range: 0.3–1.7 mmol/L); HDL-C: High-density lipoprotein cholesterol (normal range: 0.9–2.0 mmol/L); LDL-C: Low-density lipoprotein cholesterol (normal range: 1.5–3.3 mmol/L).

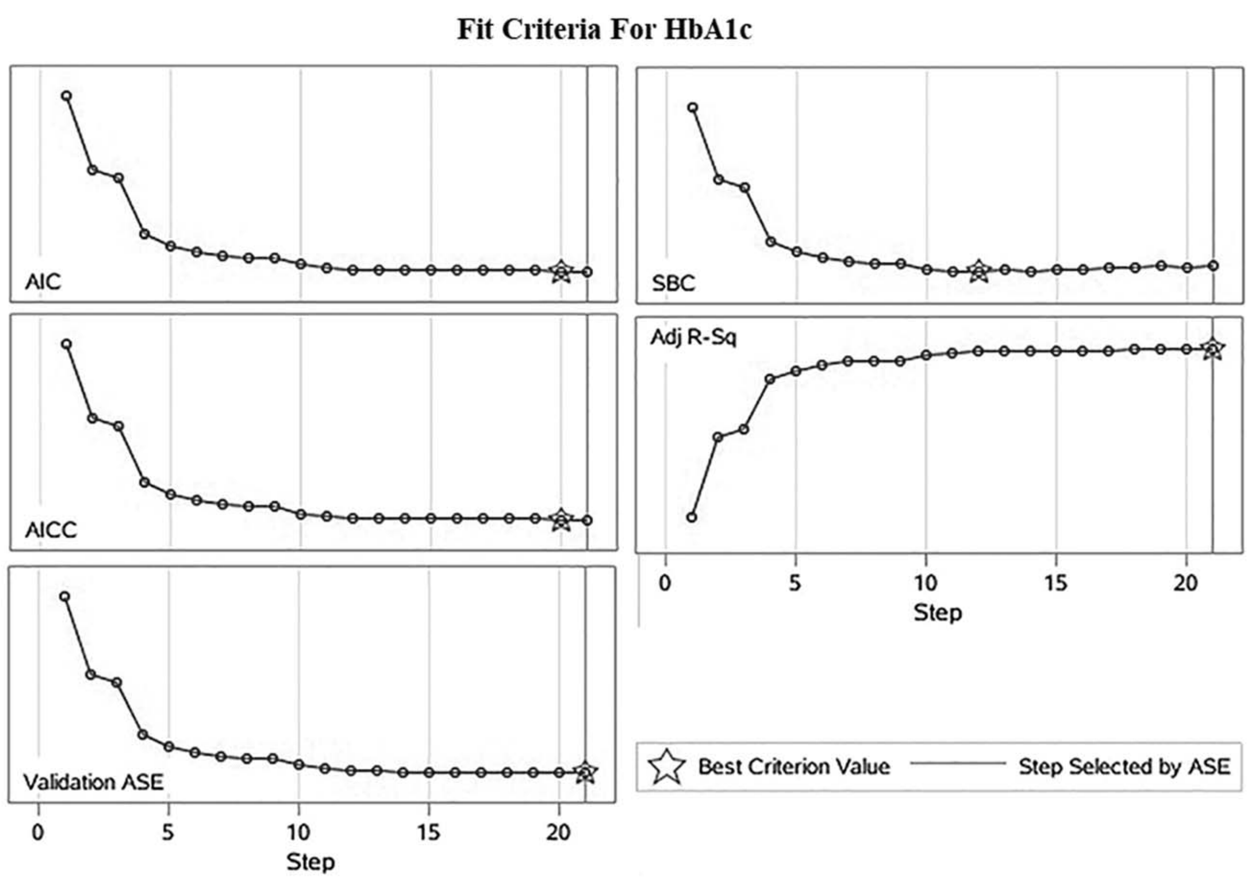

Figure 1: Elastic network model fitting process. The abscissa represents the variable selection step and the circle represents the fitting process for each step. ☆ means the best criteria value. Adj R-Sq: Adjusted R^2; ASE: The average squared error; AIC: Akaike information criterion; AICC: Akaike information criterion based on Kullback-Laible (KL) information improvement; SBC: Schwartz Bayes criterion.

Table 4: Results of dimension reduction by elastic net.

Parameters	DF	Estimate	Standardized estimate
Central adiposity	1	0.002	0.002
Family history	1	0.047	0.036
Hypertensive history	1	–0.018	–0.010
ASCVD history	1	–0.018	–0.007
Duration of T2DM	1	0.004	0.041
Typical disease characteristics	1	0.056	0.053
Complication	1	0.034	0.031
Insulin dose	1	0.003	0.087
Nocturnal hypoglycemia	1	–0.040	–0.011
Oral hypoglycemic drugs	1	0.007	0.016
Exercise	1	–0.053	–0.066
Vegetable oil intake	1	–0.007	–0.007
FBG	1	0.154	0.178
2HBG	1	0.374	0.448
Blood pressure	1	0.035	0.025
TC	1	–0.010	–0.007
HDL-C	1	0.008	0.006
LDL-C	1	0.061	0.051
Hypertension	1	0.002	0.001

DF: Degree of freedom; T2DM: Type 2 diabetes; ASCVD: Atherosclerotic cardiovascular disease; FBG: Fasting blood glucose; 2HBG: 2-h Blood glucose; TC: Total cholesterol; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol.

intake were protective factors for HbA1c control, which is also consistent with previous studies.[32,33]

In this study, the dose of insulin and the use of OHA were risk factors for elevated HbA1c, which may be because the patients taking the high dosage of insulin were mostly patients with severe disease or complications, and their blood glucose control level was relatively poor.[34] Unlike previous studies,[32] TC was a protective factor for HbA1c control, possibly because a decreased serum TC level can produce an elevated serum interleukin-6 level,[35] which is an inflammatory protein related to insulin resistance.[36]

The results of this study showed that the predictive performance of a machine learning algorithm could be improved by using dimensionality reduction to solve the variable collinearity. It was shown that the EN algorithm has high application value in influencing factor analysis in the field of diabetes and can be used to reduce the dimension of a large number of covariates of diabetes. Future studies should apply the approach of this study to the analysis of diabetes and other chronic diseases.

This study also had some limitations. First, the study did not use rigorous random sampling, and the results may be biased. However, various confounding factors were adjusted to make the results more reliable. Second, this study included only outpatients and did not consider community patients, which may lead to an overestimation of glycemic control. However, the diagnosis and treatment of diabetic patients in China are carried out in secondary and tertiary hospitals. Community hospitals have no specialists and are responsible for only daily medications. Therefore, the research object has a certain representativeness. Third, the study utilized an observational study design and; therefore, failed to produce a causal conclusion that could provide a scientific basis for predicting glycemic control. In future studies, the sample size can be increased on the basis of this study to further explore the analytical ideas.

Acknowledgement

The authors thank all the medical staff who participated in this study for their contributions to this study.

Table 5: Variance analysis of elastic net.

Source	DF	Sum of squares	Mean square	*F*	*P*
Model	19	680.17	35.80	190.55	<0.0001
Error	1942	364.83	0.19		
Uncorrected total	1961	1045			

DF: Degree of freedom.

Table 6: Comparison of predictive performance.

Models	Variable selection	Sensitivity	Specificity	Accuracy	AUC
Logistic	Stepwise	0.52	0.75	0.56	0.74
RF	Non-dimensionality reduction	0.80	0.71	0.74	0.73
	Dimensionality reduction	0.79	0.73	0.75	0.75
SVM	Non-dimensionality reduction	0.61	0.63	0.62	0.61
	Dimensionality reduction	0.84	0.68	0.73	0.72
BP-ANN	Non-dimensionality reduction	0.75	0.69	0.71	0.70
	Dimensionality reduction	0.78	0.70	0.73	0.72

AUC: Area under the operating curve; RF: Random forest; SVM: Support vector machine; BP-ANN: Back propagation artificial neural network.

Funding

This study was supported by grants from the Ministry of Education of the Humanities and Social Science Project (No. 17YJAZH048) and the National Natural Science Foundation of China (No. 81803333).

Conflicts of interest

None.

References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, *et al*. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018;138:271–281. doi: 10.1016/j.diabres.2018.02.023.
2. The IDF Diabetes Atlas: 8th Edition. China Country Report 2017 & 2045. Avaliable at: https://reports.instantatlas.com/report/view/704ee0e6475b4af885051bcec15f0e2c/CHN. [Accessed September 02, 2018]
3. Xu Y, Wang L, He J, Bi Y, Li M, Wang T, *et al*. Prevalence and control of diabetes in Chinese adults. JAMA 2013;310:948–958. doi: 10.1001/jama.2013.168118.
4. Braunwald E. Harrison's Principles of Internal, Medicine. Beijing: People's Medical Publishing House; 2001:287–300.
5. UK Prospective Diabetes Study (UKPDS 33) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837–853. doi: 10.1016/S0140-6736(98)07019-6.
6. John WG, Hillson R, Alberti SG. Use of hemoglobin A1c (HbA1c) in the diagnosis of diabetes mellitus. The implementation of World Health Organization (WHO) guidance 2011. Pract Diabetes Int 2012;29:12–112. doi: 10.1002/pdi.1648.
7. Gao Y, Xie X, Wang SX, Li H, Tang HZ, Zhang J, *et al*. Effects of sedentary occupations on type 2 diabetes and hypertension in different ethnic groups in North West China. Diab Vasc Dis Res 2017;14:372–375. doi: 10.1177/1479164117696050.
8. Xu S, Sun F, Xu W, Jiao K, Shi B, Xie X, *et al*. Simultaneous control of blood glucose, blood pressure, and lipid among drug-treated type 2 diabetes patients from Shaanxi province, North-Western China: a multicenter study. Niger J Clin Pract 2016;19:784–792. doi: 10.4103/1119-3077.
9. Liu L, Zhou C, Du H, Zhang K, Huang D, Wu JY. The prevalence of impaired fasting glucose and diabetes mellitus in working age men of North China: Anshan Worker Health Survey. Sci Rep 2014;4:4835. doi: 10.1038/srep04835.
10. Wei W, Xin X, Shao B, Zeng FF, Love EJ, Wang BY. The relationship between anthropometric indices and type 2 diabetes mellitus among adults in north-east China. Public Health Nutr 2015;18:1675–1683. doi: 10.1017/S1368980014002250.
11. Hu J, Gruber KJ, Liu H, Zhao H, Garcia AA. Diabetes knowledge among older adults with diabetes in Beijing, China. J Clin Nurs 2013;22:51–60. doi: 10.1111/j.1365-2702.2012.04273.x.
12. Fu JF, Liang L, Gong CX, Xiong F, Luo FH, Liu GL, *et al*. Status and trends of diabetes in Chinese children: analysis of data from 14 medical centers. World J Pediatr 2013;9:127–134. doi: 10.1007/s12519-013-0414-4.
13. Kang S, Kang P, Ko T, Cho S, Rhee SJ, Kyung-Sang Yu. An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction. Expert Syst Appl 2015;42:4265–4273. doi: 10.1016/j.eswa.2015.01.042.
14. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104–116. doi: 10.1016/j.csbj.2016.12.005.
15. Shan P, Zhao Y, Sha PX, Wang YQ, Lv YX, Peng SL, *et al*. Interval lasso regression based extreme learning machine for nonlinear multivariate calibration of near infrared spectroscopic datasets. Anal Methods 2018;10:3011–3022. doi: 10.1039/c8ay00466h.
16. Sun Z, Qiao Y, Lelieveldt BPF, Staring M. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. Neuroimage 2018;178:445–460. doi: 10.1016/j.neuroimage.2018.05.051.
17. Niaf E, Flamary R, Rakotomamonjy A, Rouviere O, Lartizien C. SVM with feature selection and smooth prediction in images: application to CAD of prostate cancer. 2014 IEEE International Conference on Image Processing (ICIP), Paris, France 2015;2246–2250. doi: 10.1109/icip.2014.7025455.
18. Sunghwan B, Taesung P. Risk prediction of type 2 diabetes using common and rare variants. Int J Data Min Bioinform 2018;20:77. doi: 10.1109/bibm.2017.8217926.
19. The diabetes society of the Chinese medical association. Guidelines for the prevention and treatment of type 2 diabetes in China (2017 edition) (In Chinese). Chin J Diabetes Mellit 2018;1:4–67.
20. Zheng W. Factor analysis of diabetic retinopathy in Chinese patients. Diabetes Res Clin Pract 2011;92:244–252. doi: 10.1016/j.diabres.2011.02.007.
21. Wu Y. Elastic net for cox's proportional hazards model with a solution path algorithm. Stat Sin 2012;22:27–294. doi: 10.5705/ss.2010.107.
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1–22. doi: 10.18637/jss.v033.i01.
23. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996;58:267–288. doi: 10.1111/j.1467-9868.2011.00771.x.
24. Breiman L. Random forests. Mach Learn 2001;45:5–32. doi: 10.1023/A:1010933404324.
25. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9:293–300. doi: 10.1023/A:1018628609742.
26. Wang J, Wang J. Forecasting stochastic neural network based on financial empirical mode decomposition. Neural Netw 2017;90:8–20. doi: 10.1016/j.neunet.2017.03.004.
27. Yoo DE, Park JT, Oh HJ, Kim SJ, Lee MJ, Dong HS, *et al*. Good glycemic control is associated with better survival in diabetic patients on peritoneal dialysis: a prospective observational study. PLoS One 2012;7:e30072. doi: 10.1371/journal.pone.0030072.
28. Poolsup N, Suksomboon N, Rattanasookchit S. Meta-analysis of the benefits of self-monitoring of blood glucose on glycemic control in type 2 diabetes patients: an update. Diabetes Technol Ther 2009;11:775–784. doi: 10.1089/dia.2009.0091.
29. Chen R, Ji L, Chen L, Chen L, Cai D, Feng B, *et al*. Glycemic control rate of T2DM outpatients in China: a multi-center survey. Med Sci Monit 2015;21:1440–1446. doi: 10.12659/MSM.892246.
30. Liu L, Lou Q, Guo X, Yuan L, Shen L, Sun ZL, *et al*. Management status and its predictive factors in patients with type 2 diabetes in China: a nationwide multicenter study. Diabetes Metab Res Rev 2015;31:811–816. doi: 10.1002/dmrr.2757.
31. Shan S, Gu L, Lou Q, Ouyang X, Yu Y, Wu HD, *et al*. Evaluation of glycemic control in patients with type 2 diabetes mellitus in Chinese communities: a cross-sectional study. Clin Exp Med 2017;17:79–84. doi: 10.1007/s10238-015-0406-x.
32. Omar SM, Musa IR, Osman OE, Adam I. Assessment of glycemic control in type 2 diabetes in the Eastern Sudan. BMC Res Notes 2018;11:373. doi: 10.1186/s13104-018-3480-9.
33. Martha AB, Rachel MC, Deena BS, Banh T, Angelotti A. Linoleic acid, glycemic control and Type 2 diabetes. Prostaglandins Leukot Essent Fatty Acids 2018;132:30–33. doi: 10.1016/j.plefa.2018.03.001.
34. Gao LL, Ji LN, Lu JM, Guo XH, Yang WY, Wen JP, *et al*. Current status of blood glucose control and treatment of type 2 diabetes in China 2009-2012 (In Chinese). Chin J Diabetes 2014;4:397–401. doi: 10.3969/j.issn.1006-6187.2014.07.005.
35. Zhou P, Li B, Liu B, Chen T, Xiao J. Prognostic role of serum total cholesterol and high-density lipoprotein cholesterol in cancer survivors: a systematic review and meta-analysis. Clin Chim Acta 2017;477:94–104. doi: 10.1016/j.cca.2017.11.039.
36. Bhartia NP, Archna G, Shraddha S, Sandeep B, Bushra I, Tiwari S, *et al*. Circulating interleukins (IL6)-an early predictor of insulin resistance. Asian J Med Sci 2016;7:20–25. doi: 10.3126/ajms.v7i6.15250.

How to cite this article: Wang J, Wang MY, Wang H, Liu HW, Lu R, Duan TQ, Li CP, Cui Z, Liu YY, Lyu YJ, Ma J. Status of glycosylated hemoglobin and prediction of glycemic control among patients with insulin-treated type 2 diabetes in North China: a multicenter observational study. Chin Med J 2020;133:17–24. doi: 10.1097/CM9.0000000000000585