Diabetes, Metabolic Syndrome and Obesity

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Health Management of Type 2 Diabetes Mellitus and Its Complications: A Machine Learning Algorithm-Based Retrospective Study in Chinese Communities

Xin Luo[1],*, Jingming Liang[1],*, Hong Pan[1], Dian Zhou[1], Hong Ye[2], Ying Zhao[2], Jijia Sun[2], An Zhang[1]

[1]Department of Health Management, School of Public Health, Shanghai University of Traditional Chinese Medicine, Shanghai, People's Republic of China; [2]Department of Mathematical Sciences and Computational Intelligence, School of Traditional Chinese Medicine and Artificial Intelligence, Shanghai University of Traditional Chinese Medicine, Shanghai, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jijia Sun, Department of Mathematical Sciences and Computational Intelligence, School of Traditional Chinese Medicine and Artificial Intelligence, Shanghai University of Traditional Chinese Medicine, No. 1200 Cailun Road, Pudong New Area, Shanghai, 201203, People's Republic of China, Tel +86-021-51322185, Email jijiasun@163.com; An Zhang, Department of Health Management, School of Public Health, Shanghai University of Traditional Chinese Medicine, No. 1200 Cailun Road, Pudong New Area, Shanghai, 201203, People's Republic of China, Tel +86-13651673015, Email 13052289046@163.com

Background: The global burden of diabetes mellitus (DM) and its complications is a major global public health challenge. This study aimed to improve community capacity for DM management by developing a risk prediction model for complications and providing health management recommendations using machine learning (ML).

Methods: A retrospective analysis was conducted of 4916 type 2 diabetes (T2DM) patients from Shanghai communities. Model I was developed and compared by using the least absolute shrinkage and selection operator (Lasso) regression, support vector machine (SVM), decision tree (DT) and logistic regression (LR). A Bayesian Network (BN) model to uncover potential causal relationships. Model I was evaluated and adjusted using the receiver operating characteristic (ROC) curve, area under the curve (AUC), calibration curve, and decision curve analysis (DCA). The BN model was assessed using AUC, accuracy, specificity, and sensitivity.

Results: Five consistent predictors were identified: disease course, diastolic blood pressure, HbA1c, urinary creatinine, and urinary microalbumin. Model I achieved AUCs of 0.695 (training) and 0.676 (validation), with decision curve analysis showing risk thresholds of 12–92% and 20–92% respectively. The calibration curves showed good calibration. The tree-augmented BN model achieved the AUC of 0.755, accuracy of 0.733, specificity of 0.802 and sensitivity of 0.519.

Conclusion: Effective models for predicting complication risk in T2DM patients were developed. T2DM patients with chronic comorbidities, higher income, and longer disease duration as key targets for community management. We recommend prioritizing UMA as a key monitoring indicator and strengthening comprehensive interventions, including health education, dietary self-management, and family-community support.

Keywords: T2DM, machine learning, health management

Introduction

The global burden of diabetes is escalating, with an age-standardized mortality rate of 19.8 deaths per 100000 population,[1] ranking it among the top ten causes of death worldwide and positioning it as a critical public health challenge. As one of the most common forms of DM,[2] Type 2 diabetes mellitus (T2DM) is a metabolic disorder characterized by hyperglycemia caused by insulin resistance.[3] Currently, the prevalence of T2DM is increasing yearly due to the sedentary lifestyle and population aging.[4] In 2015, an estimated 415 million people worldwide had DM, with over 90% having T2DM; this number is projected to rise to 642 million by 2040.[5] The associated medical expenditure

Received: 5 August 2025
Accepted: 10 November 2025
Published: 27 November 2025

may be about $770 billion.[6] A study revealed that nearly half of the world's patients with T2DM are from India and China.[7] According to the International Diabetes Federation,[8] as of 2021, China had approximately 140 million patients with DM and spent $165 billion on diabetes-related health expenditure. Currently, more than half of patients with DM do not know that they have the disease. Therefore, their continuing hyperglycemic state may lead to the occurrence of DM complications,[9] increased mortality, reduced patients' quality of life, and increased burden of health care.[10–12] A previous study indicated that the treatment of patients with at least one DM complication is more expensive to treat than that of those with DM alone.[13]

Consequently, early screening of the population at risk and effective interventions have proven to be effective.[14,15] Identifying multidimensional risk factors is of great importance. Public health interventions targeting modifiable risk factors can effectively facilitate the precision prevention of type 2 diabetes and its complications. Non-clinical factors such as diet, physical activity, smoking, and alcohol consumption also significantly influence the progression of metabolic diseases.[16] For instance, diet impacts gut microbiota homeostasis, and specific dietary components can modulate the interaction between gut flora and metabolism1.[17,18] Currently, with the development of artificial intelligence (AI) and the increase in big data availability, machine learning (ML) is increasingly used in research.[19,20] In the field of medical research, ML is mainly used for the prediction, diagnosis, treatment, and prognosis of diseases, and to a certain extent, it can be used to help in the screening of diseases and the development of interventions,[21] due to its capability to more accurately identify disease influencing factors and infer the risk of disease outcomes, thereby providing an effective early warning tool for the health management of community-based T2DM patients. Common methods include least absolute shrinkage and selection operator (Lasso) regression, support vector machine (SVM), classification regression tree, and Bayesian network (BN), all of which have their own advantages and disadvantages, and several different methods are often used in most studies to construct multiple models.[22] BN are capable of inferring uncertainties and probabilities of events, thereby clearly revealing the proximal and distal contributions of various factors to disease outcomes.[23] As such, this modeling approach has been employed in disease prediction research to supplement and explain the complex interplay between diseases and their risk factors.[24]

The Healthy China Action Plan (2019–2030) states that the burden from four types of chronic diseases (DM, cardiovascular diseases, cancer, and chronic respiratory diseases) accounts for more than 70% of the total disease burden, which is an important factor limiting the improvement of healthy life expectancy in China. With the orderly implementation of graded diagnosis and treatment, the patients' primary community is gradually becoming the main institution for managing and preventing chronic diseases. Moreover, the concept of health management is highly valued by the society and government. However, there are few studies on the community health management of DM complications. In our previous study, we developed a prediction model for a specific high-risk complication combination (lower extremity vascular disease, diabetic foot, and diabetic retinopathy) in T2DM patients, using a Shanghai community cohort, association rule mining, and machine learning.[25] Although this model provides valuable clinical insights, it is inherently designed for a predetermined combination of pathologies and does not incorporate non-clinical factors. This limitation may restrict its utility for universal risk screening and comprehensive health management planning at the community level. Therefore, this study expands upon our previous work by leveraging the same community cohort from Shanghai's Pudong New Area and employing machine learning to develop a more comprehensive and generalizable risk prediction model for diabetic complications. The key advancements are as follows: First, we shifted the prediction target from a specific complication cluster to the overall risk of developing any complication, and developed a user-friendly web application based on the model to facilitate its deployment in primary healthcare settings. Second, we integrated a wide array of non-clinical influencing factors (eg, education, income, history of chronic diseases, dietary habits, health management advice provided by community physicians, family support, and social support) with key clinical indicators from the predictive model. Furthermore, a Bayesian network model was employed to elucidate the complex probabilistic dependencies and potential causal pathways among these multi-faceted factors, thereby constructing a more robust and interpretable risk prediction framework. In summary, this study is expected to improve the level of chronic disease management in the community, reduce the medical burden of DM complications for patients, families, and society, and contribute to the advancement of the "Healthy China" initiative.

Methods

Data Collection

This retrospective study, which analyzed data collected from community residents in a T2DM cohort project (the "Construction of Key Disciplines of Health Education and Promotion in the Fourth Round of Three-Year Action Plan of Public Health"), was conducted in the Shanghai University of Traditional Chinese Medicine. The project included participants from six community health service centers (Huamu, Jin Yang, Yin Xing, Siping, Sanlin, and Daqiao Community Health Service Centers) in Shanghai, in collaboration with the institution, who voluntarily participated in the study between December 2015 and April 2016. The study followed the basic principles of the Declaration of Helsinki, and the study protocol was approved by the Ethics Committee of Shanghai University of Traditional Chinese Medicine. Written informed consent was obtained from all study participants.

To ensure the clinical relevance of our findings for community health practice, this study focused on the middle-aged and elderly population (≥40 years), which bears the highest burden of T2DM and is the primary target of community-level management strategies.[26] Inclusion criteria for study subjects: (1) patients diagnosed with type 2 diabetes mellitus according to standard diagnostic criteria[27] or those with diabetes recorded in community health service center archives; (3) Age ≥ 40 years; (4) Other physical examination information is complete and accurate. Exclusion criteria: (1) Patients with type 1 diabetes or other special types of diabetes; (2) Pregnant women; (3) Refusal to participate in the questionnaire survey or failure to complete the questionnaire. The primary outcome was the presence of at least one diabetic complication, ascertained either by patient self-report in the questionnaire or by a recorded diagnosis in the community health records, including diabetic retinopathy, nephropathy, cardio-cerebrovascular disease, peripheral neuropathy, lower extremity vascular disease, and diabetic foot.

The project obtained the baseline data on 49 underlying variables from 5078 patients with DM who volunteered to participate in the cohort project, using questionnaires, physical examinations, and biochemical tests (The questionnaire can be found in the supplementary material). We imputed missing values using the overall mean of each variable, a common and effective single imputation method,[28] given the low missing rate (<5%) across all variables. Outliers were excluded using the interquartile range (IQR) method (where values below Q_1–1.5×IQR or above Q_3 + 1.5×IQR were considered outliers).

Finally, the study data of 4916 participants on 49 variables were finally included. The 49 variables included one dependent variable (presence [n=1684] or absence [n=3232] of DM complications) and 48 independent variables, 20 clinical indicator and 28 non-clinical variables. To avoid overfitting, 4916 cases were randomized into a training set (n=3936) and verification set (n=980) according to the ratio of 8:2, using R software (version 4.1.0, https://www.r-project.org/).

The 20 clinical indicator were: sex, abdominal obesity, age, systolic blood pressure (SBP), diastolic blood pressure (DBP), waist circumference, body mass index (BMI), disease course of T2DM, glucose (GLU), 2-hour postprandial plasma glucose (GLU2h), glycosylated hemoglobin (HbA1c), total cholesterol (TC), triglycerides (TG), low-density lipoprotein (LDL), high-density lipoprotein (HDL), blood urea nitrogen (BUN), serum creatinine (SCR), uric acid (UA), urinary creatinine (UCR), and urinary microalbumin (UMA). The area and violin plots of 20 clinical indicators were plotted by using Microsoft Office Excel software (Microsoft Corp., USA) and the plotrix[29] package in R software.

The 28 non-clinical variables were: education, income, history of T2DM, hypertension, hyperlipidemia, coronary heart disease (CHD), cerebral apoplexy (CA), malignant tumor, dietary control, exercise, smoking, drinking, drinking tea, dietary advice by a community doctor, exercise advice by a community doctor, written health advice by a community doctor, regular outpatient visits, regular follow-up service, whether the participant had a contracted general practitioner and knew the general practitioner's name and telephone number, family support, community support, staple food intake, meat intake, bean products intake, vegetable intake, and fruit intake.

The following 10 DM complications were included through the questionnaire survey results of 4619 participants in this study: ketoacidosis, diabetic ketoacidosis, diabetic nephropathy (DN), diabetic cardio–cerebrovascular complications, hyper-glycemic hyperosmolar state, diabetic retinopathy (DR), lower extremity vascular disease, lactic acidosis, diabetic neuropathy, and diabetic foot.

Statistical Analysis and Modeling Methods

Variable screening and model building were performed for the two groups of variables separately. The first screening focused on the 20 clinical indicator affecting DM complications; thereafter, a DM complication prediction model was established for screening of the at-risk population. To reduce the influence of irrelevant clinical indicator on other variables and explore the causal relationship between the clinical indicator, health management status of individuals and communities and DM complication, the characteristic clinical indicator from the first screening were incorporated as independent variables in the second screening and model building.

Four common ML methods in medical research include: supervised, unsupervised, semi-supervised, and reinforcement learning.[30,31] In disease prediction and modeling research, supervised learning algorithms and related derivatives are the most common, including regularization, lasso regression, SVM, BN and decision tree (DT).[32] Lasso regression simplifies models by shrinking minor coefficients to zero; it then computes individual scores from the non-zero coefficients for use as categorical predictors.[22] Support Vector Machines (SVM) are effective in high-dimensional spaces and are capable of delivering precise classification and prediction for complex datasets.[33] DT, specifically the C5.0 algorithm, was used to segment data based on the greatest information gain, providing variable importance rankings. Boosting was employed to enhance accuracy and efficiency.[34] The DT based on C5.0 algorithm works by segmenting samples according to a field that provides the greatest information gain, by continuous segmenting of each subsample from the previous segmentation according to different fields until it is impossible to continue segmenting the subsamples, and finally, removing and pruning segments that have no significant impact on the model. When there is covariance in the data and high correlation between two groups, using Lasso regression may lead to overfitting and loss of model predictive power, while SVM can compensate for this deficiency and avoid overfitting. DT has high predictive power and show the importance of variables.

Therefore, by using the glmnet,[35] e1071,[36] kernlab[37] and caret[38] packages in R software and IBM SPSS Modeler (version 18.0), all models were run with default parameters; this study used Lasso regression based on ten-fold cross-validation, linear kernel function-based SVM analysis and DT based on the C5.0 algorithm for the first screening of the 20 clinical indicators to establish a predictive model I by logistic regression (LR) analysis. In this study, forest plot and nomogram plot created using the forestplot[39] and rms[40] packages were created for the characteristic clinical indicators included in model I to demonstrate the degree of influence of the variables on the occurrence of related complications in patients with T2DM. As a visualization result of LR, the line length of each predicted variable in the nomogram plot corresponds to the size of its regression coefficient. By assigning scores to the value level of each predicted variable, the total score is obtained by adding each score. Finally, the predicted value of the individual's outcome event is calculated by converting the function between the total score and the probability of the target outcome event occurring.[22] And further utilize the DynNom[41] package to develop a web-based calculator for diabetes complication risk.

BN, a probabilistic predictive classifier based on the Bayes' theorem, can visualize the hidden correlation and causality between variables through indicators and methods such as conditional probability table and directed acyclic graph (DAG). It is more interpretable than the general prediction algorithms and can avoid the overfitting problem to a certain extent.[42] In the causality between variables, BN has better performance in terms of interpretability and intuitiveness.[43] Therefore, the second variable screening and model construction were performed using simple BN, created by Tree Augmented Naive Bayes with maximum likelihood method, as a parametric learning method. In addition, to explore the effect of the important variables screened by BN model on DM complications and the causality between the variables, the study drew a DAG of the variables by using Netica (NORSYS software corp, USA) with default parameters, which employs the Hill-Climbing algorithm for the TAN structure learning. Finally, the important variables included in the BN model were analyzed and discussed from both individual and community perspectives.

Model Evaluation and Comparison

The receiver operating characteristic (ROC) curve and the area under curve (AUC) are two important metrics used to evaluate the discrimination of a binary classification model. The ROC curves provide a visual representation of the classifier's performance at different thresholds; while the AUC values provide a quantitative assessment of the classifier's overall performance. In this study, ROC curve and AUC were used to evaluate and optimize the model I by using rms[44] and pROC[45] packages. In addition, calibration curves were used to assess the calibration capability of the model I, and estimating the uncertainty of the curve by bootstrap-resampling. The calibration performance of the model I can be evaluated by comparing the Apparent curve with the Bias-corrected curve. If the two curves are very close and closely follow the diagonal (ideal curve), then the model I has good calibration performance. In addition, Decision Curve Analysis (DCA) were used to assess the value of the model I for clinical application at different thresholds, specifically by weighing the potential clinical benefits gained or lost by patients through comparing the magnitude of net benefit (NB) values. In BN model, AUC, accuracy, specificity and sensitivity are often used to evaluate model performance.[41] Among them, accuracy is the correct prediction rate of BN model. Specificity is the percentage of samples correctly determined as negative, while sensitivity is the percentage of samples correctly determined as positive.

The basic statistical analysis of the 48 variables was performed by using IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., NY, Armonk, USA). The quantitative variables were described as the mean ± standard deviation, while the qualitative variables were described as the number of cases (percentage). Finally, the univariate analysis was performed using two independent samples *t*-test, four-grid table, χ^2 test, and rank sum test as appropriate. All statistical tests used were two-sided, and statistical significance was set at $p < 0.05$.

The study design is depicted in Figure 1.

Results

Basic Statistical Description

The area and violin plots of the 20 clinical indicators are shown in Figure 2. The statistical descriptions of the 28 non-clinical variables are shown in Table 1. The baseline characteristics of training set and validation set are shown in Tables S1 and S2.

Model Construction and Prediction Results

Variable screening was performed on the training set using Lasso regression, SVM, and DT. The results for each of the 20 clinical indicators are reported below. First, Lasso regression with 10-fold cross-validation identified 9 variables at the optimal lambda value ($\lambda = 0.018543$), as shown in Figure 3a. In contrast, the results of the linear kernel function-based SVM analysis, showed that the highest accuracy was 0.7129 when the SVM was filtered to 6 variables (Figure 3b). A total of 16 variables with importance greater than 0 were screened by the DT, and the importance ranking of the variables is shown in Figure 3c. Five variables were common to all three methods: disease course, DBP, HbA1c, UCR, and UMA. These five clinical indicators were used as independent variables, with "presence or absence of DM complications" as the dependent variable. A risk prediction model (Model I) was then developed using LR for T2DM patients:

$$y_{model\,I} = -1.877 + 0.065 \times DiseaseCourse - 0.012 \times DBP + 0.201 \times HbA1c - 0.034 \times UCR + 0.009 \times UMA \quad (1)$$

The result of model I corresponded to an AUC of 0.695 (95% confidence interval [CI] = 0.677–0.712) in training set and AUC of 0.670 (95% CI = 0.633–0.707) in validation set (Figure 4a and b). The calibration curves showed moderate agreement (Figure 4c and d). The risk thresholds of DCA were 12–92% n the training set and 20–92% in the validation set, which showed good clinical application of model I (Figure 4e and f). These results are consistent with the assessment results obtained in the standardization data. For model I, forest plot and nomogram plot are shown in Figure 5. In addition, a web-based application (APP, at https://studentluo.shinyapps.io/T2DM_DM/) for predicting the risk probability of DM complications in patients with T2DM was created. The application can facilitate the screening of people at risk for DM complications and commencing management as early as possible by community health workers (CHWs).

A total of 33 variables (5 clinical and 28 non-clinical indicator variables) were included in a second round of screening and were analyzed using the BN on the overall dataset. In the constructed BN model, the results of the top 20 variables in importance and evaluation are shown in Figure 6. Finally, the DAG plots was constructed with the occurrence of DM complications as the parent node and the top 20 variables as children nodes, are shown in Figure 7.

As shown in the results of the variable importance ranking of the BN model (Figure 6) and the corresponding DAG plots of the variables (Figure 7), (1) CHD and CA were risk factors for DM complications; (2) the higher the degree of

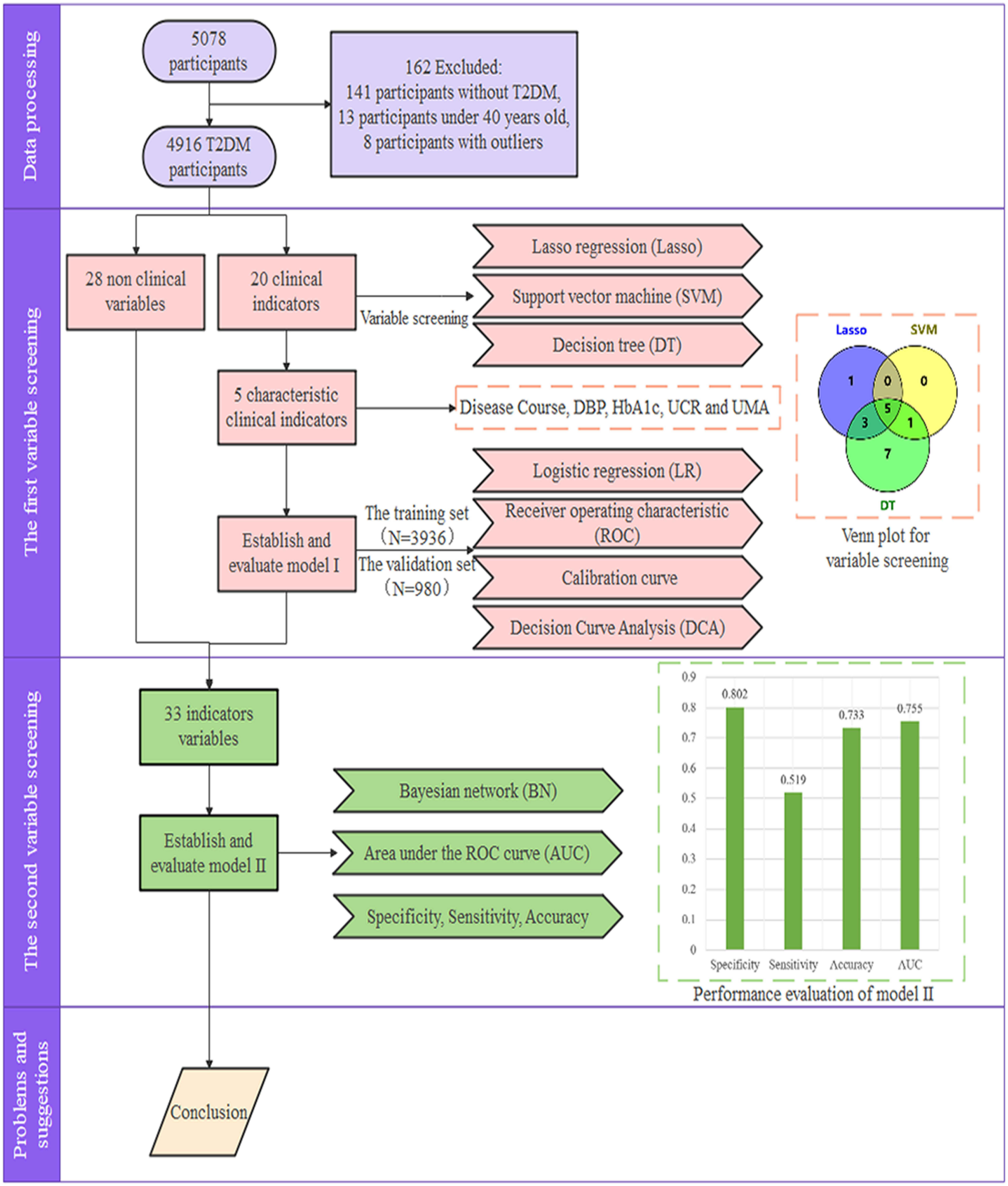

Figure 1 Flow chart of the study design.

community assistance and support, human and material resources, provided to T2DM patients, the relatively lower the risk of DM complications; (3) regular follow-up services by community physicians of T2DM patients were positively associated with the prevalence of DM complications, which may be related to the enhanced disease management approach by the community after the diagnosis of T2DM patients; (4) self-management with diet control had the highest

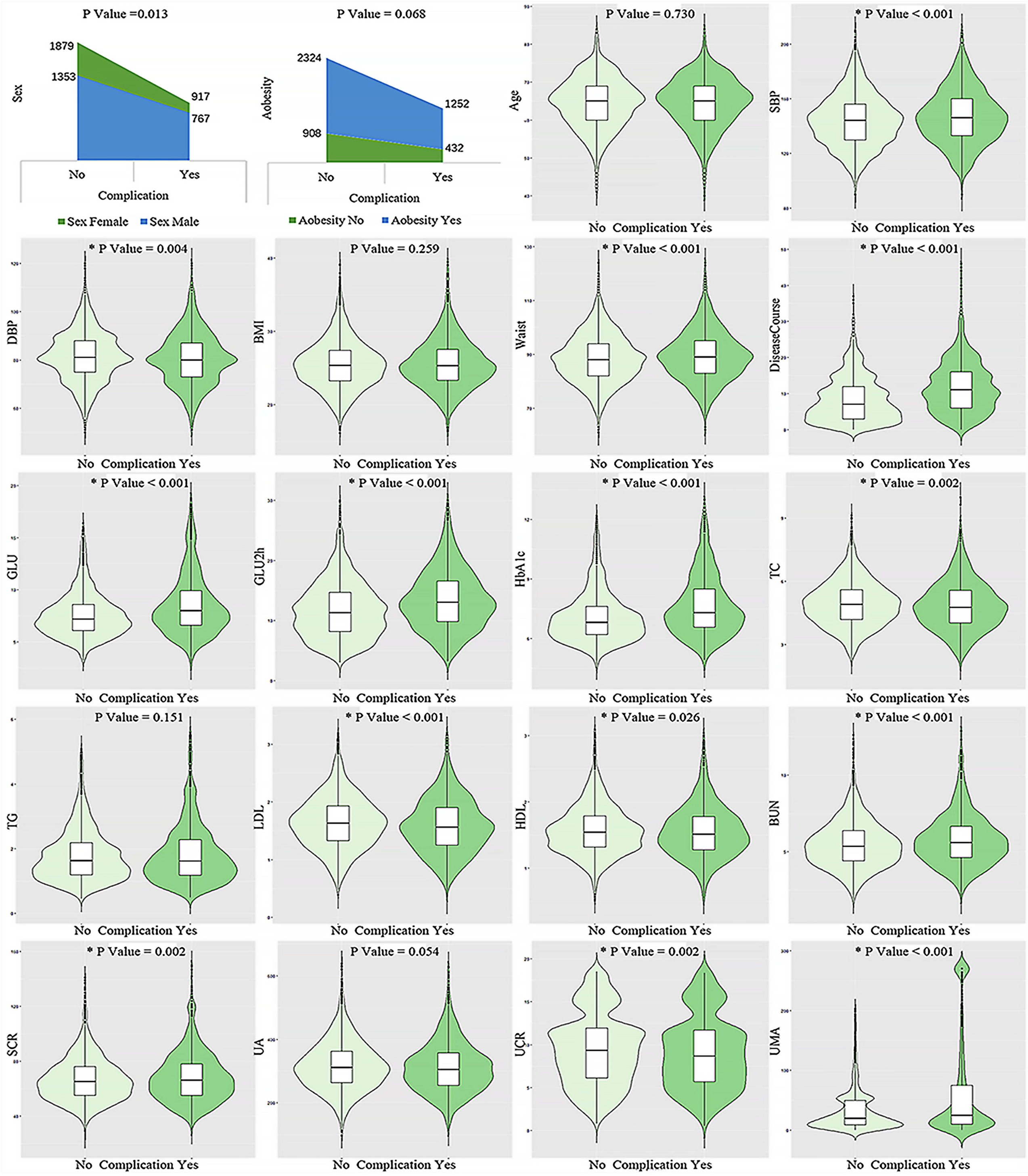

Figure 2 Area and violin plots of the 20 clinical indicators under the condition of whether or not the patient of type 2 diabetes mellitus (T2DM) had diabetes mellitus complications.

Abbreviations: Aobesity, abdominal obesity; SBP, systolic blood pressure; DBP, diastolic blood pressure; Waist, waist circumference; BMI, body mass index; GLU, glucose; GLU2h, 2-hour postprandial plasma glucose; HbA1c, glycosylated hemoglobin; TC, total cholesterol; TG, triglycerides; LDL, low-density lipoprotein; HDL, high-density lipoprotein; BUN, blood urea nitrogen; SCR, serum creatinine; UA, uric acid; UCR, urinary creatinine; UMA, urinary microalbumin.

impact on the occurrence of DM complications, and ranked first in the importance of characteristic variables; (5) patients with T2DM with higher income levels also had a higher risk of DM complications, which is different from the traditional understanding and suggests that the relationship between income level, education level, risk of illness, and other factors still needs further study; (6) patients who had neither consumed alcohol nor smoked had a lower risk of DM

Table 1 Participant Background and Differences Between the No-Complication and Complication Groups [n(%)/Mean ± SD]

Variables	**Complication**		**χ2/Z/T**	**P**
	No	**Yes**		
Education				
Uneducated Low Medium High	166 (5.1) 1910 (59.1) 1040 (32.2) 116 (3.6)	87 (5.2) 1019 (60.5) 513 (30.5) 65 (3.8)	0.822	0.411
Income				
<1500 1500-3500 3500-5000 >5000	275 (8.5) 1656 (51.2) 1026 (31.7) 275 (8.6)	141 (8.4) 897 (53.3) 482 (28.6) 164 (9.7)	0.629	0.529
History of T2DM				
No Yes	1980 (61.3) 1252 (38.7)	915 (54.3) 769 (45.7)	21.946	<0.001*
Hypertension				
No Yes	1247 (38.6) 1985 (61.4)	652 (38.7) 1032 (61.3)	0.008	0.927
Hyperlipidemia				
No Yes	2067 (64.0) 1165 (36.0)	1028 (61.0) 656 (39.0)	4.018	0.045
Coronary heart disease				
No Yes	2823 (87.3) 409 (12.7)	1317 (78.2) 367 (21.8)	69.555	<0.001*
Cerebral apoplexy				
No Yes	3016 (93.3) 216 (6.7)	1459 (86.6) 225 (13.4)	60.460	<0.001*
Malignant tumor				
No Yes	3055 (94.5) 177 (5.5)	1592 (94.5) 92 (5.5)	0.001	0.984
Dietary control				
No Yes	293 (9.1) 2939 (90.9)	172 (10.2) 1512 (89.8)	1.704	0.192
Exercise				
No Yes	2499 (77.3) 733 (22.7)	1302 (77.3) 382 (22.7)	0.001	0.997

(*Continued*)

Table 1 (Continued).

Variables	**Complication**		**χ2/Z/T**	**P**
	No	**Yes**		
Smoking				
Never Quit Smoking Often	2439 (75.5) 193 (6.0) 600 (18.5)	1206 (71.6) 141 (8.4) 337 (20.0)	2.629	0.009*
Drinking				
Never Quit Drinking Often	2468 (76.4) 94 (2.9) 670 (20.7)	1225 (72.7) 72 (4.3) 387 (23.0)	2.604	0.009*
Drinking tea				
Never Occasionally Often	1838 (56.9) 310 (9.3) 1093 (33.8)	915 (54.4) 169 (10.0) 600 (35.6)	1.601	0.109
Diet advice by community doctor				
Never Occasionally Often Every	847 (26.2) 872 (27.0) 1228 (38.0) 285 (8.8)	399 (23.7) 454 (27.0) 637 (37.8) 194 (11.5)	2.568	0.010*
Exercise advice by community doctor				
Never Occasionally Often All the time	879 (27.2) 898 (27.8) 1192 (36.9) 263 (8.1)	422 (25.1) 451 (26.8) 626 (37.2) 185 (11.0)	2.712	0.007*
Health education by community doctor				
Never Occasionally Often All the time	1285 (39.8) 914 (28.3) 874 (27.0) 159 (4.9)	645 (38.3) 453 (26.9) 460 (27.3) 126 (7.5)	2.055	0.040*
Regular outpatient visits				
No Yes	1211 (37.5) 2021 (62.5)	600 (35.6) 1084 (64.4)	1.610	0.204
Regular follow-up services				
No Yes	1683 (52.1) 1549 (47.9)	840 (49.9) 844 (50.1)	2.129	0.145
General practitioner				
No Yes	1952 (60.4) 1280 (39.6)	988 (58.7) 696 (41.3)	1.372	0.241
Know general practitioner's name				
No Yes	2124 (65.7) 1108 (34.3)	1090 (64.7) 594 (35.3)	0.480	0.488

(*Continued*)

Table 1 (Continued).

Variables	Complication		χ2/Z/T	P
	No	Yes		
Know general practitioner's Tel				
No Yes	2307 (71.4) 925 (28.6)	1189 (70.6) 495 (29.4)	0.323	0.570
Family support	10.26±4.112	6.01±2.40	3.854	<0.001*
Community support	2.82±1.28	2.82±1.26	4.779	<0.001*
Staple food intake (g)	243.45±115.10	245.33±122.68	0.531	0.596
Meat intake (g)	73.55±57.79	75.83±63.28	1.273	0.203
Bean product intake (g)	48.77±48.55	53.61±55.29	3.037	0.002*
Vegetable intake (g)	277.97±171.92	285.42±164.03	1.465	0.143
Fruit intake (g)	60.78±62.70	59.49±61.62	0.691	0.490

Note: *p < 0.05, indicating a statistically significant difference.
Abbreviations: SD, standard deviation; T2DM, type 2 diabetes mellitus.

complications compared with those who had recently taken measures to stop drinking and abstaining from alcohol, and the DAG of BN model showed a positive causality between smoking and alcohol consumption; (7) help and support from the family influenced the occurrence of DM complications, and to some extent, the help and support provided to patients in the community; (8) UMA was a risk indicator for early pathological changes in microvascular complications of DM.[46] In this study, we found that increased intake of vegetables reduced the level of UMA, while an extremely high intake of vegetables led to an increased risk of DM complications; therefore, an appropriate intake range and mechanism of action need to be confirmed in further studies. In addition, UMA also directly affected three clinical indicators (DBP, HbA1c, and UCR); and (9) the longer the disease course in patients with T2DM, the higher their HbA1c levels. Figures supporting the results of the causal analysis between DM complications and variables described above are presented in Figure S1.

Discussion

Characteristic Clinical Indicator and DM Complications

The results from model I show that, among the 20 clinical indicator, five indicators were associated with DM complications (disease course, DBP, HbA1c, UCR, and UMA). This study found that the longer the disease course of DM, the higher the risk of DM complications. In previous studies, disease course was also identified as a risk factor for common DM complications such as DR, diabetic foot, diabetic neuropathy, and DN.[47–49] Another study also showed that hypertension is one of the risk factors for microvascular and macrovascular complications in patients with T2DM.[50] However, DBP, an important indicator for the diagnosis of hypertension, was found to be negatively associated with the risk of DM complications in this study. This may be related to the earlier diagnosis and timely care during the natural course of T2DM, as well as the widespread use of statins and blood pressure-lowering therapies worldwide.[51] In addition, a significant reduction in blood pressure may lead to the development of hyperkalemia,[52] which is associated with the development of renal disease and DM.[53] HbA1c, an important indicator of DM, is also strongly associated with DM complications, with higher HbA1c levels leading to an increased risk of DM complications.[54] UMA and UCR are important indicators to evaluate the renal function status of individuals. In some studies on DN, UMA/UCR is often used to assess the effect of certain drugs or interventions on DN, and a lower UMA/UCR indicates a reduced risk of DN or control of the disease in DN patients.[55] In addition, UMA/UCR has also been associated with the incidence and severity of DR indirectly related to kidney disease.[46,56]

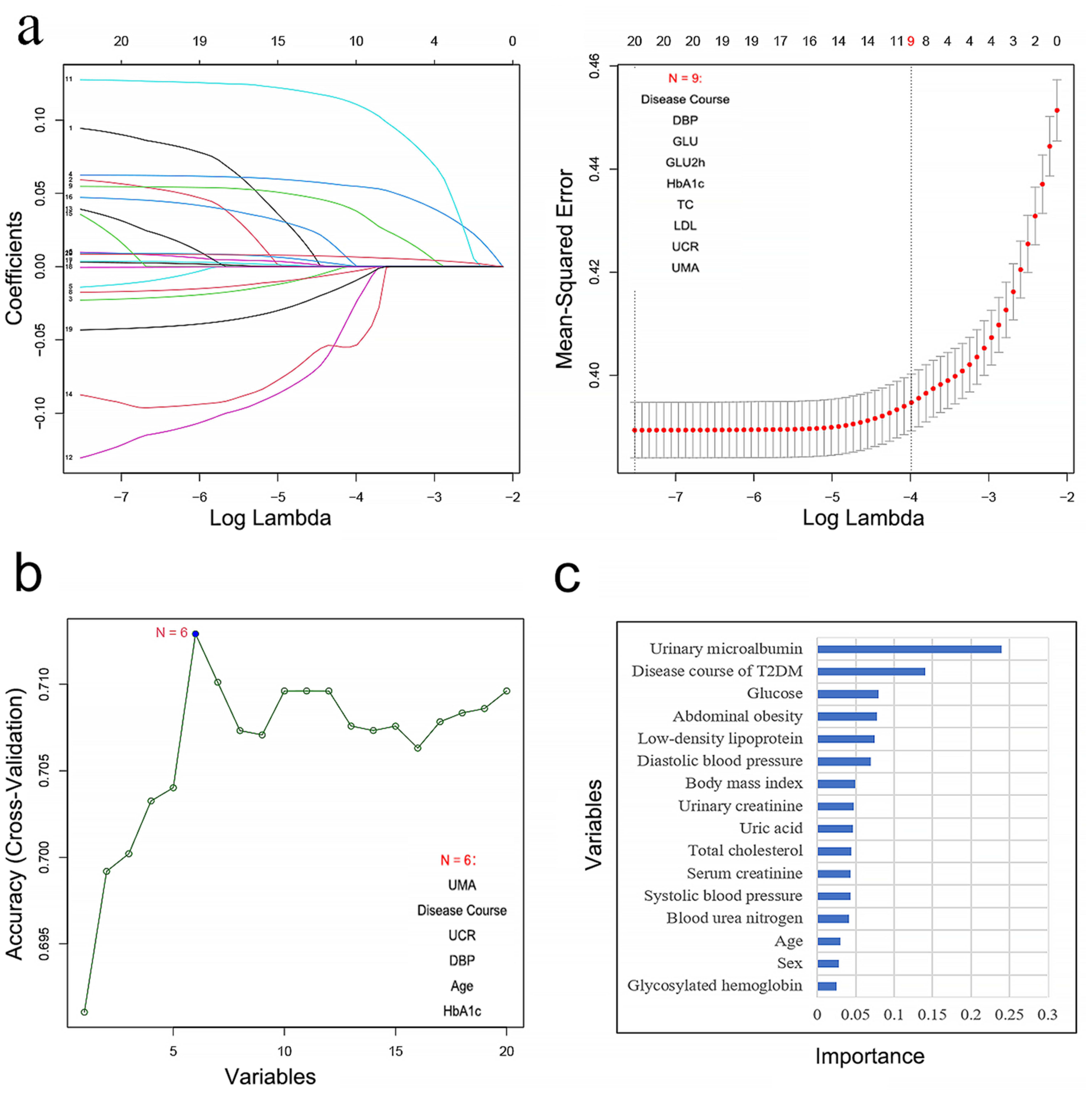

Figure 3 Variable screening results of the three methods. (**a**) Results of Lasso regression analysis. The 9 predictor variables with non-zero coefficients at the best penalty parameter lambda (λ) values are screened out by pooling and normalizing the variables to shrink the coefficients toward zero with the least possible prediction error. (**b**) Results of variable screening by support vector machine (SVM). When the number of variables input into the model is 6, the accuracy of the model is the highest at 0.7129. (**c**) Importance ranking of the 16 characteristic variables screened by the decision tree (DT) model.

Non-Clinical Influencing Factors and Social Determinants of Health

The Bayesian network model constructed in this study revealed a complex probabilistic dependency between clinical indicators and non-clinical factors, including socioeconomic status, social support, lifestyle, and healthcare utilization, which together form an important system influencing the risk of diabetic complications. The model results indicate that the relationship between income level and complication risk is not simply linear. In the complication group, the proportion of individuals with “high” income (9.77%) was slightly higher than in the non-complication group (8.53%), while the proportion of those with “medium” income in the complication group (28.6%) was lower than in the non-complication group (31.7%). This finding contrasts with the conventional understanding that higher

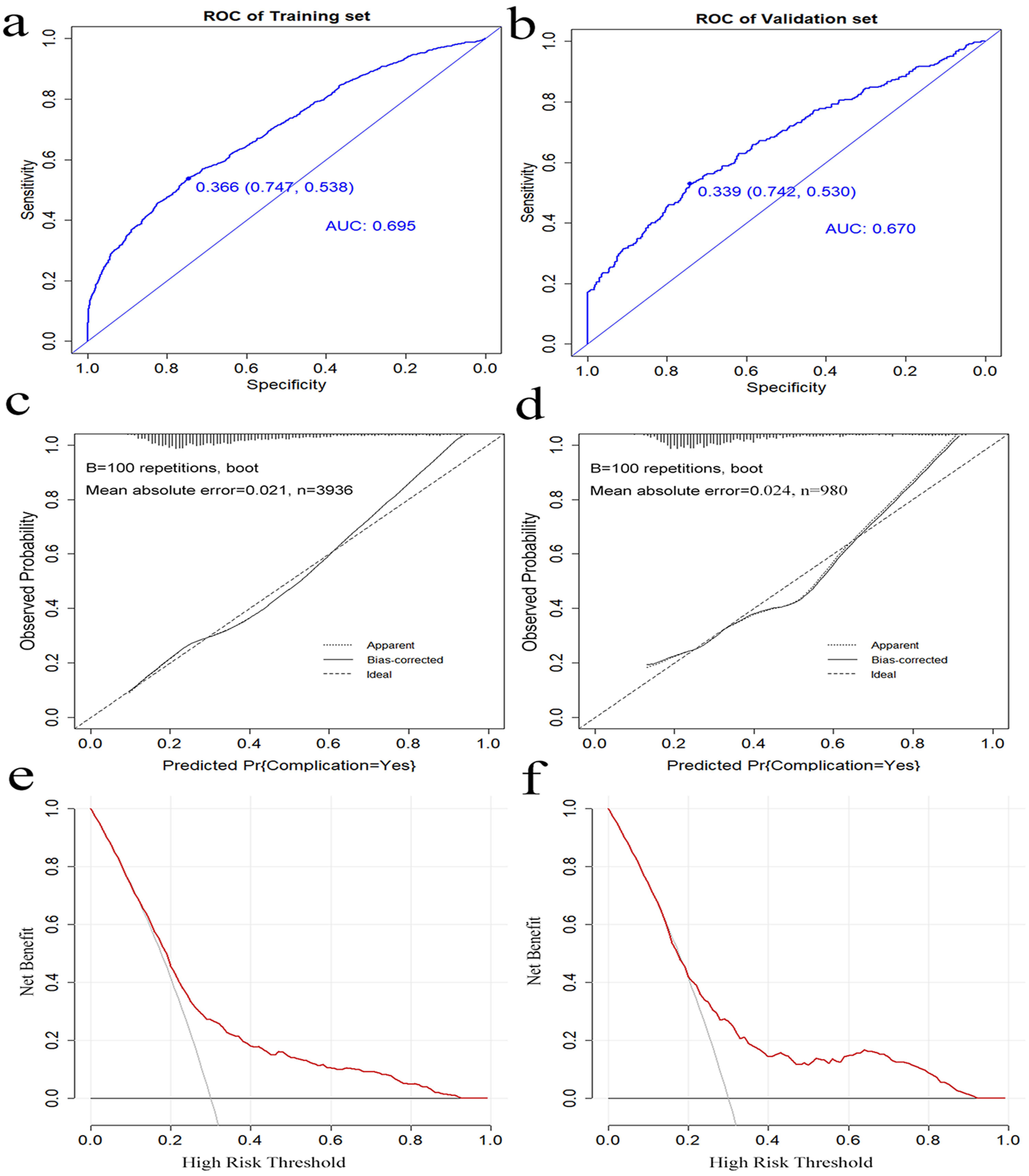

Figure 4 Performance evaluation results of Model I. (**a**) The ROC curve of the model I in the training set. The area under curve (AUC) value was 0.695. The optimal cut-off value for model I is 0.366, with a specificity of 0.747 and a sensitivity value of 0.538. (**b**) The ROC curve of the model I in the validation set. The AUC value was 0.670. The optimal cut-off value for model I is 0.339, with a specificity of 0.742 and a sensitivity value of 0.530. (**c**) The calibration curve of the model I in the training set, which indicates 100 bootstrap replications with a mean absolute error of 0.021. (**d**) The calibration curve of the model I in the validation set, which indicates 100 bootstrap replications with a mean absolute error of 0.024. (**e**) DCA of the Model I in the training set with a risk threshold of 12%-92%. (**f**) DCA of the Model I in the validation set with a risk threshold of 20%-92%.

socioeconomic status leads to better health outcomes,[57] which may be related to the specific study population. Firstly, the study participants were located in urban Shanghai, where higher-income populations may have better healthcare access, leading to earlier and more comprehensive diagnosis and recording of their complications, thus creating detection

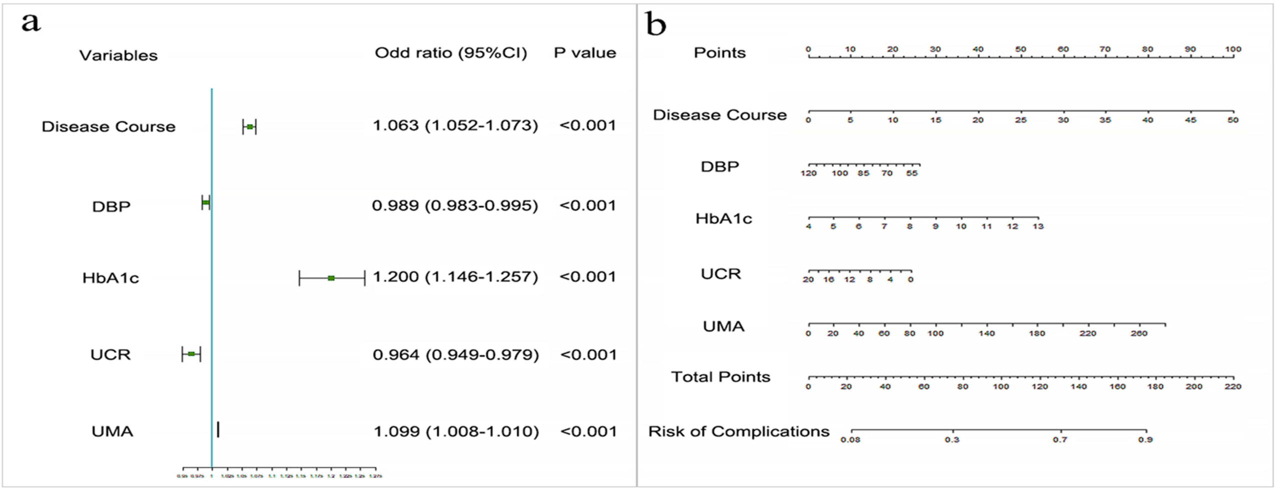

Figure 5 (**a**) Forest plots of the 5 characteristic clinical indicators. (**b**) Nomogram plot of model I. The effect of the 5 indicators (Disease course, DBP, HbA1c, UCR, and UMA) on the occurrence of related complications in patients with T2DM is illustrated. When using the nomogram plot, draw a line perpendicular from the corresponding axis of each predicted variable until it reaches the top line labeled "Points". Sum up the number of points for all predicted variables then draw a line descending from the axis labeled "Total Points" until it intercepts the bottom line labeled "Risk of Complications".

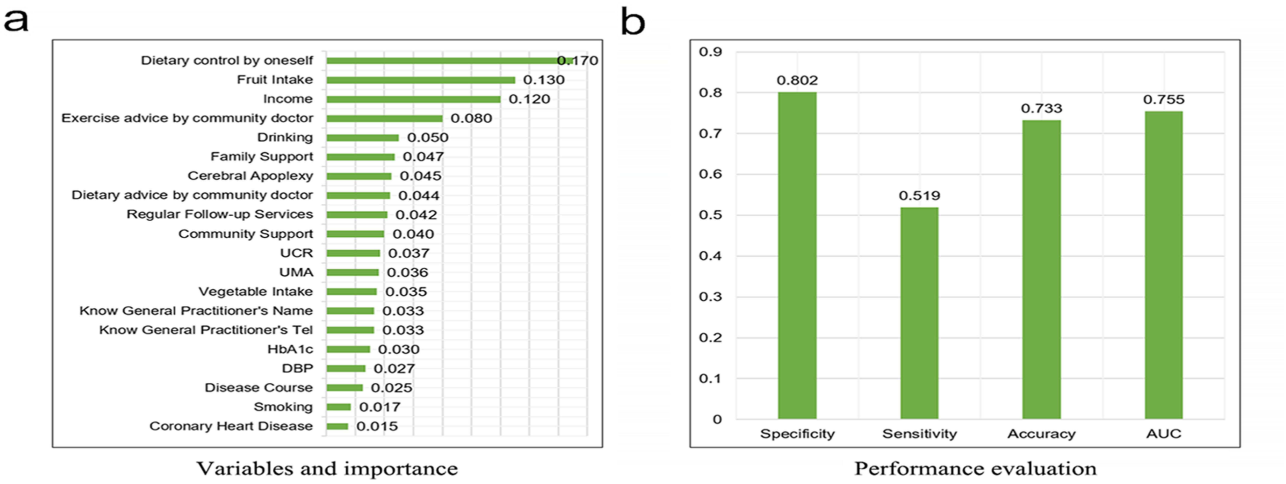

Figure 6 The top 20 variables of importance included in the Bayesian network (BN) model and the performance evaluation results. (**a**) Ranking of the top 20 variables of importance included in the BN model. (**b**) Performance evaluation of BN.

bias. Secondly, higher income may be associated with specific unhealthy lifestyles, such as irregular eating patterns and sedentary behavior resulting from high-stress jobs, factors that may directly or indirectly promote the development of complications.[58] This highlights the complexity of Social Determinants of Health, which can exhibit unique patterns across different cultural and social contexts. Furthermore, social support networks play a significant role in complication management. The Bayesian network showed that a higher level of community support was more prevalent in the non-complication group (45.4%) than in the complication group (40.3%). Strong community support can help patients better manage their condition by providing health information, emotional comfort, and promoting healthy behaviors, thereby delaying the progression of complications.[59] Similarly, family support followed a comparable trend, with the highest level of support being more common in the non-complication group (33.0%) than in the complication group, consistent with previous research identifying family support as a protective factor in chronic disease management.[60] Therefore, health management for community-based T2DM patients must adopt comprehensive strategies.

https://doi.org/10.2147/DMSO.S556130

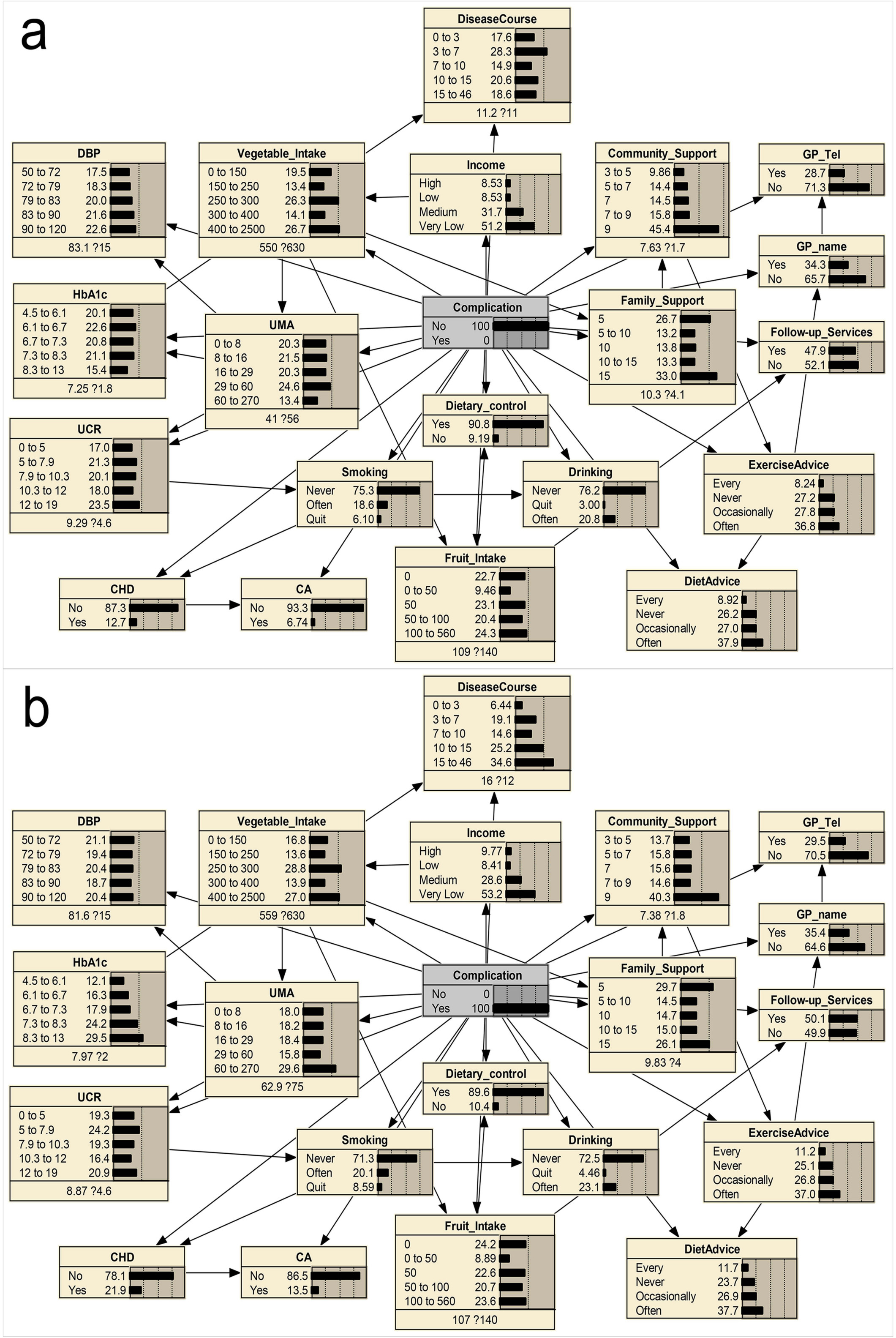

Figure 7 Directed acyclic graphs (DAGs) of the top 20 variables of importance in the BN model. (**a**) DAG of the 20 characteristic variables under the condition of a 0% prevalence of DM complications. (**b**) DAG of the 20 characteristic variables under the condition of a 100% prevalence of DM complications. Coronary heart disease (CHD), cerebral apoplexy (CA), whether the participant had a contracted general practitioner and knew the general practitioner's name (GP name) and telephone number (GP Tel).

Health Management Recommendations for Patients with T2DM in the Community

Some of the studies on community health management have proposed that patient-centered care models,[61] Internet-based management models,[62] appropriate staffing of community health centers,[63] a high level of professionalism among CHWs,[64] community members' ability to actively participate in health management, and application of telemedicine systems can be used effectively in community chronic disease management.[65,66] However, few previous studies have focused on a particular subpopulation with a chronic disease; therefore, this study proposes the following recommendations as a reference for community-based decision making for middle-aged and elderly patients with T2DM.

First, the model based on the five characteristic clinical indicators in this study can be used to screen the population at high risk of DM complications among middle-aged and elderly T2DM patients in the community for early health management. Second, among the screened high-risk groups, we should focus on T2DM patients with longer disease course, higher income level, and those with chronic diseases such as CHD and CA. Long-term and continuous dynamic monitoring of important indicators such as UMA and HbA1c should be adopted. Moreover, in the process of health management, awareness of DM complications should be strengthened among the high-risk groups through health education; particular focus should be given to teaching disease management through diet. Furthermore, interventions on smoking cessation and alcohol abstinence should be carried out simultaneously, and in a timely manner, encouraging long-term abstinence. In addition, health education and training should be provided to patient families, who should be encouraged to actively participate in any health management programs aimed at their loved ones, including support with medication adherence, regular blood glucose monitoring, adherence to a healthy diet, and continued health education. Moreover, the community can provide communication platforms for patients by establishing DM clubs, which could host lectures and other themed events. Finally, this study found that with an increase in prevalence of DM complications, the telephone follow-up rate by community physicians also increased, suggesting that while predicting the risk of DM complications in middle-aged and elderly T2DM patients in the community, it can also be used to estimate the workload of CHWs and the required human resources, which can help community managers to allocate medical resources effectively.

Unlike previous study using the same cohort that focused on models predicting specific combinations of high-risk complications[25], the model developed in this study provides a universal tool for screening overall complication risk in community settings. While previous model emphasized clinical indicators such as BMI and blood lipids[25], our model highlights the importance of disease progression, HbA1c, urinary protein markers, and non-clinical factors through the incorporation of Bayesian networks. Furthermore, unlike most studies relying on hospital specialty data, our model directly serves community health centers—the primary institutions responsible for chronic disease management—thus holding practical application value. The innovation of this study lies in its focus on China's community-based primary healthcare system, developing a complication risk prediction model tailored to this setting. Furthermore, it not only employs machine learning for model development but also integrates Bayesian network algorithms to deeply explore and visually represent the complex probabilistic dependencies between clinical indicators and non-clinical influencing factors, thereby establishing a more multidimensional and comprehensive risk assessment system. Moreover, based on the deep associations revealed by the Bayesian network, we further propose personalized health management strategies for diabetic complications, offering community physicians an interpretable and practical decision-support tool that supports the grassroots implementation of the "Healthy China" initiative. At the same time, this study developed a web-based risk prediction application prototype, aiming to translate the research findings into a tool readily usable by community doctors. Currently, this application exists as a research prototype accessible via an authorized link in standard web browsers. The next steps involve collaborating with community health service centers to conduct formal usability testing and exploring data interface solutions with existing community health information systems, with the goal of integrating it into workflows and ultimately improving the efficiency of grassroots diabetes management.

This study had some limitations. First, the moderate AUC values of Model I (0.695–0.676) and the BN model (0.755) may be partly attributed to the retrospective nature of the study and the use of self-administered questionnaires for non-clinical variables, which are susceptible to recall and social desirability biases. Second, the models were developed and validated on the same dataset, which risks overfitting and limits generalizability; external validation using independent, multi-center cohorts is needed. Third, the inclusion of only middle-aged and elderly diabetic patients restricts the applicability of the findings to younger populations, warranting future studies in broader age groups.

Conclusion

In this study, we developed two models for the prediction of DM-related complications in middle-aged and elderly patients with T2DM, using ML. The prediction model I identified disease course, DBP, HbA1c, UCR, and UMA as influential factors for DM complications, which could be applied to the screening of people at risk of DM complications. According to the BN model, patients with T2DM and chronic diseases (CHD and CA), higher income levels, and longer disease course, and are key populations for the community DM management process. Furthermore, UMA is the key monitoring indicator, while health education, emphasizing self-management using diet, and improving family and community support for patients are also important elements in the process of DM management. This study showcases the significant potential of data-driven approaches to enhance T2DM management at the community level. The models and insights generated not only enable early identification of high-risk patients but also support the design of personalized and resource-efficient intervention strategies, thereby contributing to the prevention and control of diabetic complications in primary care settings. This study has limitations, including the retrospective design and potential biases in self-reported data, which may contribute to the models' moderate predictive performance. Future work should focus on external validation with multi-center data and extending the research to younger diabetic populations.

Data Sharing Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

The study has been ethically approved by the Medical ethics committee of Longhua Hospital, Shanghai University of Traditional Chinese Medicine (Approved No: 2017LCSY069) and in accordance with the Declaration of Helsinki. Written informed consent of all participants were obtained after they were briefed about the study's objectives.

Acknowledgments

We would like to thank Editage (www.editage.cn) for English language editing.

Author Contributions

Conceptualization, A.Z. and J.J.S.; methodology, J.J.S.; formal analysis, X.L.; data curation, J.M.L, H.P., D.Z., H.Y. and Y.Z.; writing—original draft preparation, X.L., L.J.M, H.P., D.Z., H.Y, Y.Z., J.J.S. and A.Z.; writing—review and editing, A.Z. and J.J.S; project administration, A.Z. and J.J.S.; funding acquisition, A.Z. A.Z. and J.J.S. contributed equally to this work and should be considered co-corresponding authors. X.L. and J.M.L. contributed equally to this work. All authors have read and agreed to the published version of the manuscript. All authors took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by grants from the fourth round of three-year action plan for public health - construction of key disciplines of Shanghai Municipal Health Commission (Grant number: 15GWZK1002). Shanghai Municipal Health Commission Health Industry Clinical Research Specialization: A study of precision comorbidity management in community-dwelling elderly patients with type 2 diabetes mellitus (Grant number: 202340189).

Disclosure

The authors declare that they have no conflict of interests.

References

1. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the global burden of disease study 2021. *Lancet*. 2024;403(10440):2100–2132. doi:10.1016/S0140-6736(24)00367-2

2. Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. *Pharmacoeconomics*. 2015;33:811–831. doi:10.1007/s40273-015-0268-9
3. Schmidt AM. Highlighting diabetes: the epidemic continues. *Arter Thromb Vasc Biol*. 2018;38(1):e1–e8. doi:10.1161/ATVBAHA.117.310221
4. Ismail L, Materwala H, Al Kaabi J. Association of risk factors with type 2 diabetes: a systematic review. *Comput Struct Biotechnol J*. 2021;19:1759–1785. doi:10.1016/j.csbj.2021.03.003
5. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet*. 2017;389:2239–2251. doi:10.1016/S0140-6736(17)30058-2
6. Asgari S, Khalili D, Hosseinpanah F, et al. Prediction models for type 2 diabetes risk in the general population: a systematic review of observational studies. *Int J Endocrinol Metab*. 2021;19. doi:10.5812/ijem.109206
7. Ke C, Narayan KMV, Chan JCN, et al. Pathophysiology, phenotypes and management of type 2 diabetes mellitus in Indian and Chinese populations. *Nat Rev Endocrinol*. 2022;18:413–432. doi:10.1038/s41574-022-00669-4
8. Magliano DJ, Boyko EJ, Balkau B, et al. IDF Diabetes Atlas,10th edition. 2021.
9. Papatheodorou K, Banach M, Bekiari E, et al. Complications of diabetes 2017. *J Diabetes Res*. 2018;2018:10–13. doi:10.1155/2018/3086167
10. Hu H, Sawhney M, Shi L, et al. A systematic review of the direct economic burden of type 2 diabetes in China. *Diabetes Ther*. 2015;6:7–16. doi:10.1007/s13300-015-0096-0
11. Wu H, Eggleston KN, Zhong J, et al. How do type 2 diabetes mellitus (T2DM)-related complications and socioeconomic factors impact direct medical costs? A cross-sectional study in rural Southeast China. *BMJ Open*. 2018;8(11):e020647. doi:10.1136/bmjopen-2017-020647
12. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16:377–390. doi:10.1038/s41581-020-0278-5
13. Zhu D, Zhou D, Li N, et al. Predicting diabetes and estimating its economic burden in China using autoregressive integrated moving average model. *Int J Public Health*. 2022;66:1–8. doi:10.3389/ijph.2021.1604449
14. Zhou B, Perel P, Mensah GA, et al. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nat Rev Cardiol*. 2021;18(11):785–802. doi:10.1038/s41569-021-00559-8
15. Fan R, Li S, Wei Y, et al. Integrated metabolomic and transcriptome analyses reveal the regulatory mechanisms of partridge tea phenolics-rich extracts on type 2 diabetes mice. *Food Sci Hum Well*. 2025;14(9):9250219. doi:10.26599/FSHW.2024.9250219
16. Montague TJ, Gogovor A, Krelenbaum M. Time for chronic disease care and management. *Can J Cardiol*. 2007;23:971–975. doi:10.1016/s0828-282x(07)70859-0
17. Zhang K, Ma Y, Luo Y, et al. Metabolic diseases and healthy aging: identifying environmental and behavioral risk factors and promoting public health. *Front Public Health*. 2023;11:1253506. doi:10.3389/fpubh.2023.1253506
18. Lu Y, Qin R, Wang J, et al. The impact of the novel starch-lipid complexes on the glucolipids metabolism, inflammation, and gut dysbiosis of type 2 diabetes mellitus rats. *Food Sci Hum Well*. 2024;13(6):3210–3223. doi:10.26599/FSHW.2023.9250008
19. Wang Z, Li H, Carpenter C, et al. Challenge-enabled machine learning to drug-response prediction. *AAPS J*. 2020;22(5):1–6. doi:10.1208/s12248-020-00494-5
20. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. *BMC Med Res Methodol*. 2021;21:1–18. doi:10.1186/s12874-021-01347-1
21. Silva GFS, Fagundes TP, Teixeira BC, et al. Machine learning for hypertension prediction: a systematic review. *Curr Hypertens Rep*. 2022;24(11):523–533. doi:10.1007/s11906-022-01212-6
22. Zhou Z-R, Wang -W-W, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann Transl Med*. 2019;7:796. doi:10.21037/atm.2019.08.63
23. Arora P, Boyne D, Slater JJ, et al. Bayesian networks for risk prediction using real-world data: a tool for precision medicine. *Value Health*. 2019;22(4):439–445. doi:10.1016/j.jval.2019.01.006
24. Wu Y, Chao J, Bao M, et al. Construction of predictive model for osteoporosis related factors among postmenopausal women on the basis of logistic regression and Bayesian network. *Prev Med Rep*. 2023;35:102378. doi:10.1016/j.pmedr.2023.102378
25. X L, J S, H P, et al. Establishment and health management application of a prediction model for high-risk complication combination of type 2 diabetes mellitus based on data mining. *PLoS One*. 2023;18(8):e289749.
26. Chinese Elderly Type 2 Diabetes Prevention and Treatment of Clinical Guidelines Writing Group; Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Society; Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Health Care Society; Geriatric Professional Committee of Beijing Medical Award Foundation; National Clinical Medical Research Center for Geriatric Diseases (PLA General Hospital). 中国老年2型糖尿病防治临床指南(2022年版)[Clinical guidelines for prevention and treatment of type 2 diabetes mellitus in the elderly in China (2022 edition)]. *Zhonghua Nei Ke Za Zhi*. 2022;61(1):12–50. Chinese. doi:10.3760/cma.j.cn112138-20211027-00751
27. Jia W, Weng J, Zhu D, et al. Standards of medical care for type 2 diabetes in China 2019. *Diabetes Metab Res Rev*. 2019;35(6):e3158. doi:10.1002/dmrr.3158
28. Afkanpour M, Hosseinzadeh E, Tabesh H. Identify the most appropriate imputation method for handling missing values in clinical structured datasets: a systematic review. *BMC Med Res Methodol*. 2024;24(1):188. doi:10.1186/s12874-024-02310-6
29. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
30. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. 2020;51:675–687. doi:10.1016/j.beth.2020.05.002
31. Choi RY, Coyner AS, Kalpathy-Cramer J, et al. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol*. 2020;9:1–12.
32. Uddin S, Khan A, Hossain ME, et al. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019;19(1):1–16. doi:10.1186/s12911-019-1004-8
33. Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018;284:603–619. doi:10.1111/joim.12822
34. Kamkar I, Gupta SK, Phung D, et al. Stable feature selection for clinical prediction: exploiting ICD tree structure using tree-lasso. *J Biomed Inform*. 2015;53:277–290. doi:10.1016/j.jbi.2014.11.013
35. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model via coordinate descent. *J Stat Softw*. 2011;39(5):1–13. doi:10.18637/jss.v039.i05
36. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien_.R package version 1.7-13. 2023.
37. Karatzoglou A, Smola A, Hornik K. kernlab: Kernel-Based Machine Learning Lab_. R package version 0.9-32. 2023.
38. Kuhn M. caret: Classification and Regression Training_. R package version 6.0-93. 2022.
39. Gordon M, Lumley T. _forestplot: Advanced Forest Plot Using “grid” Graphics_. R package version 3.1.1. 2022.

40. Harrell FE. _rms: Regression Modeling Strategies_. R package version 6.5-0. 2023.
41. Jalali A, Roshan D, Alvarez-Iglesias A, Newell J. _DynNom: Visualising Statistical Models using Dynamic Nomograms_. R package version 5.0.2. 2022.
42. Park E, Chang HJ, Nam HS. A Bayesian network model for predicting post-stroke outcomes with available risk factors. *Front Neurol*. 2018;9:1–11. doi:10.3389/fneur.2018.00699
43. Park H, Lee K. Prediction of suicidal ideation in shift workers compared to non-shift workers using machine learning techniques. *J Affect Disord*. 2022;307:125–132. doi:10.1016/j.jad.2022.03.076
44. Ting DSW, Tan KA, Phua V, et al. Biomarkers of diabetic retinopathy. *Curr Diab Rep*. 2016;16. doi:10.1007/s11892-016-0812-9
45. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. doi:10.1186/1471-2105-12-77
46. Zhu CH, Zhang SS, Kong Y, et al. Effects of intensive control of blood glucose and blood pressure on microvascular complications in patients with type II diabetes mellitus. *Int J Ophthalmol*. 2013;6:141–145. doi:10.3980/j.issn.2222-3959.2013.02.06
47. Li M, Yuan T, Li S, et al. Nomogram analysis of the influencing factors of diabetic foot in patients with diabetes mellitus. *Hormones*. 2021;20:333–338. doi:10.1007/s42000-020-00263-1
48. Wu B, Niu Z, Hu F. Study on risk factors of peripheral neuropathy in type 2 diabetes mellitus and establishment of prediction model. *Diabetes Metab J*. 2021;45:526–538. doi:10.4093/dmj.2020.0100
49. Tziomalos K, Athyros VG. Diabetic nephropathy: new risk factors and improvements in diagnosis. *Rev Diabet Stud*. 2015;12:110–118. doi:10.1900/RDS.2015.12.110
50. Pavlou DI, Paschou S, Anagnostis P, et al. Hypertension in patients with type 2 diabetes mellitus: targets and management. *Maturitas*. 2018;112:71–77. doi:10.1016/j.maturitas.2018.03.013
51. Ali MK, Pearson-Stuttard J, Selvin E, et al. Interpreting global trends in type 2 diabetes complications and mortality. *Diabetologia*. 2022;65 (1):3–13. doi:10.1007/s00125-021-05585-2
52. Melmer A, Laimer M. Treatment goals in diabetes. *Endocr Dev*. 2016;31:1–27.
53. Watanabe R. Hyperkalemia in chronic kidney disease. *Rev Assoc Med Bras*. 2020;66:31–36. doi:10.1590/1806-9282.66.s1.31
54. Schnell O, Crocker JB, Weng J. Impact of HbA1c testing at point of care on diabetes management. *J Diabetes Sci Technol*. 2017;11:611–617. doi:10.1177/1932296816678263
55. Sinha N. HbA1c and factors other than diabetes mellitus affecting it. *Singapore Med J*. 2010;51:616–622.
56. Mo R, Shi R, Hu Y, et al. Nomogram-based prediction of the risk of diabetic retinopathy: a retrospective study. *J Diabetes Res*. 2020;2020:1–9. doi:10.1155/2020/7261047
57. Stringhini S, Carmeli C, Jokela M, et al. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. *Lancet*. 2017;389(10075):1229–1237. doi:10.1016/S0140-6736(16)32380-7
58. Li W, Yi G, Chen Z, et al. Is job strain associated with a higher risk of type 2 diabetes mellitus? A systematic review and meta-analysis of prospective cohort studies. *Scand J Work Environ Health*. 2021;47(4):249–257. doi:10.5271/sjweh.3938
59. Kanny S, Hall L, Blackhurst D, et al. Community-based diabetes self-management and support program: addressing quality of life and social vulnerability. *BMC Public Health*. 2025;25(1):1538. doi:10.1186/s12889-025-22627-1
60. Zhu L, Pan Z, Shen F, et al. Effects of family support system on the self-management behaviour of patients with T2DM: a multi-centre cross-sectional study in community settings. *Fam Pract*. 2024;41(2):114–122. doi:10.1093/fampra/cmae010
61. Hudon C, Fortin M, Haggerty J, et al. Patient-centered care in chronic disease management: a thematic analysis of the literature in family medicine. *Patient Educ Couns*. 2012;88:170–176. doi:10.1016/j.pec.2012.01.009
62. Liang G, Jiang H, Huang C, et al. Diabetes health management strategy based on internet plus graded diagnosis and treatment strategy. *Ann Palliat Med*. 2020;9:3915–3922. doi:10.21037/apm-20-1996
63. Thorsen M, McGarvey R, Thorsen A. Diabetes management at community health centers: examining associations with patient and regional characteristics, efficiency, and staffing patterns. *Soc Sci Med*. 2020;255:1–25. doi:10.1016/j.socscimed.2020.113017
64. Hartzler AL, Tuzzio L, Hsu C, et al. Roles and functions of community health workers in primary care. *Ann Fam Med*. 2018;16:240–245. doi:10.1370/afm.2208
65. Plumb J, Weinstein LC, Brawer R, et al. Community-based partnerships for improving chronic disease management. *Prim Care*. 2012;39:433–447. doi:10.1016/j.pop.2012.03.011
66. May CR, Finch TL, Cornford J, et al. Integrating telecare for chronic disease management in the community: what needs to be done? *BMC Health Serv Res*. 2011;11:131. doi:10.1186/1472-6963-11-131