

Development and Validation of a Multidimensional Indicator-Based Risk Prediction Model for Gestational Diabetes Mellitus: A Nested Case-Control Study

Jiajia Chen

The Second Clinical Medical School of Zhengzhou University

Shanshan Yin

Henan Academy of Innovations in Medical Science

Shuling Wang

The Second Clinical Medical School of Zhengzhou University

Shu Li

The Second Clinical Medical School of Zhengzhou University

Ru Feng

The Second Clinical Medical School of Zhengzhou University

Xianqi Wang

The Second Clinical Medical School of Zhengzhou University

Xiao Hao

The Second Affiliated Hospital of Zhengzhou University

Xia Zhang

The Second Affiliated Hospital of Zhengzhou University

Qing Zhang

The Fifth Affiliated Hospital of Zhengzhou University

Guijuan Zhang

The Second Affiliated Hospital of Zhengzhou University

Linlin Hua

hualinlin@zzu.edu.cn

The Second Affiliated Hospital of Zhengzhou University

Research Article

Keywords: GDM, prediction model, LASSO, AUC, DCA, nomogram

Posted Date: July 2nd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6815483/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Gestational diabetes mellitus (GDM) could contribute to significant health risks in both mothers and their offspring. Therefore, this study aims to construct a prediction model to identify women at elevated risk for GDM in early pregnancy.

Methods: This study was a nested case-control study. 346 participants were randomly allocated to the training set (n=242) and the validation set (n=104) at a ratio of 7:3. Sociodemo-graphic characteristics, clinical indicators, and lifestyle behaviors of all participants were obtained at 8–13+6 weeks of gestation. GDM was confirmed through the 75-g oral glucose tolerance test (OGTT). The least absolute shrinkage and selection operator (LASSO) regression was applied to select the most significant factors among candidate variables. We further established a GDM risk prediction model based on the risk factors chosen by the LASSO. The model's calibration, discrimination, and clinical use were assessed using the calibration analysis, area under the receiver operating characteristic (ROC) curve and decision curve analysis (DCA). Finally, we presented the model with a nomogram.

Results: In the study, the prevalence of GDM in the training and validation sets were 24.8% and 26.0%, respectively ($P=0.93$). In the training set, we developed a simple GDM risk prediction model by using family history of diabetes, pre-pregnancy body mass index (BMI), progesterone, aspartate transaminase (AST), activated partial thromboplastin time (APTT), and triglyceride to high-density lipoprotein cholesterol (TG/HDL-C). Among them, family history of diabetes, higher pre-pregnancy BMI, progesterone, AST, and TG/HDL-C levels were associated with increased GDM risk, while higher APTT level was associated with decreased GDM risk. The calibration curve indicated satisfactory accuracy. The ROC curve demonstrated excellent discrimination, with the area under the curve (AUC) of 0.85 (95% confidence interval [CI], 0.80-0.91) and 0.73 (95%CI, 0.62-0.83) for the training and validation set, respectively. The DCA curve demonstrated high net benefit. Furthermore, internal validation with excellent performance demonstrated the generalizability of the model.

Conclusions: The present study developed a model with excellent performance for predicting GDM. Furthermore, a nomogram was constructed to visualize the model. Therefore, this model can serve as an effective GDM prediction tool.

Introduction

Gestational diabetes mellitus (GDM) is a prevalent pregnancy complication characterized by abnormal glucose tolerance with first recognition or onset during pregnancy[1]. In recent years, with increasing maternal age and the increase in rates of obesity in women of reproductive age, the prevalence of GDM has been increasing[2], and it has reached 14.8% in China[3]. Researches have shown that GDM significantly increases risks of preeclampsia, preterm birth, respiratory distress syndrome, abnormal fetal growth, future obesity, cardiovascular diseases, and type 2 diabetes[4-8]. Therefore, the identification of high-risk populations is crucial to prevent the onset of GDM.

The exact pathophysiological mechanisms underlying GDM have not been fully elucidated, which are closely linked to pregnancy-induced dysregulation in glucose homeostasis, lipid metabolism, insulin resistance and chronic low-grade inflammation. In recent years, the development of early prediction models for GDM has garnered increasing global attention, with multiple risk assessment tools emerging. Previous GDM risk prediction models are primarily based on single or multiple maternal characteristics, including pre-pregnancy body mass index (BMI), family history of diabetes, advanced maternal age, hepatic, renal, and coagulation function measures, pregnancy-associated plasma protein A, leptin, lipocalin-2, adiponectin, weight gain, and soft drink intake during pregnancy[9-11]. However, these conventional models demonstrate limited discrimination and calibration ability[9, 10, 12-14]. Therefore, it is urgent to develop a multidimensional, clinically feasible, simple yet accurate GDM prediction model. This model can effectively identify high-risk women in early pregnancy, facilitating timely intervention to improve clinical outcomes.

This study conducted a nested case-control study, aiming to identify the risk factors for GDM and construct a risk prediction model based on sociodemo-graphic characteristics, clinical indicators, and lifestyle behaviors. Finally, a nomogram was developed to visualize the model.

Material and methods

Study design and population

This study was a nested case-control study based on Zhengzhou Birth Cohort. Pregnant women aged 18-45 years and prior to 16 weeks gestation were enrolled to participate in the cohort at their first prenatal care from January 2021 to May 2024. Information on sociodemo-graphic characteristics, detailed medical history, and lifestyle behaviors were recorded through questionnaire by professional doctors at enrollment (**Supplementary file**).

In the present analysis, 1,001 participants with a singleton pregnancy were eligible for inclusion if they were free of preexisting diabetes and cancer. Cases were excluded if they had chronic hypertension, cardiovascular disease (angina, myocardial infarction, coronary revascularization, or stroke), severe hepatic or renal disorders, polycystic ovary syndrome, thrombophilia, antiphospholipid syndrome, or urinary tract infections (n = 33); did not provide information on GDM diagnosis (n = 7); had missing information on covariate (n = 414). Those, with fasting blood glucose (FBG) ≥ 5.1 mmol/L in the first trimester, were excluded (n = 33). Those taking antiplatelet medication (such as aspirin) were also excluded (n = 168). The final analysis consisted of 346 pregnant women after exclusions. The sample size was calculated using the following formula: $n = Z^2 \times P \times (1-P) / e^2$, where n= the required sample size, Z= 1.96 at a 95% confidence interval (CI), P= the prevalence of GDM (14.8%) and e= the margin of error (5%) [3, 15-17]. The overall sample size was calculated to be 194. However, 346 eligible participants were included in this study. PASS (Power Analysis and Sample Size) software was applied to calculate the smallest sample size.

The study was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Zhengzhou University and complied with the tenets of the Declaration of Helsinki for research involving human subjects (No.2021332). All participants provided written informed consent at enrollment.

Blood measurement

In this study, the laboratory data were obtained from the electronic medical record system at 8-13+6 weeks of gestation. Serum thyroid-stimulating hormone (TSH), thyroxine (T4), and progesterone levels were measured using chemiluminescence immunoassay method, serum alanine aminotransferase (ALT), aspartate transaminase (AST), and γ -glutamyl transferase (GGT) levels were measured using continuous monitoring method, serum blood urea nitrogen (BUN) was measured using urease method, serum creatinine (Cr) was measured using creatinase-coupled enzymatic method, serum triglyceride (TG) was measured using oxidase method and serum high-density lipoprotein cholesterol (HDL-C) was measured using direct method on the fully automated analyzers. The TG/HDL-C ratio was calculated as the quotient of TG to HDL-C levels. Plasma blood glucose level was measured using glucose oxidase method, plasma activated partial thromboplastin time (APTT) was determined by optical method and white blood cell count (WBC) was measured using flow cytometry method on the fully automated analyzers. The above measurements were conducted after at least 8-h overnight fast. All biochemical measurements were conducted by trained laboratory personnel adhering to standardized protocols.

Ascertainment of GDM

The primary outcome was the incidence of GDM. Pregnant women were routinely advised to undergo a standardized 75-gram oral glucose tolerance test (OGTT) after at least 8-h overnight fast between the 24th and 28th weeks of gestation. FBG, 1-h postprandial blood glucose (1h-PBG) and 2h-PBG were measured when they received the 75 g oral glucose. According to the report by the International Association of Diabetes and Pregnancy Study Groups (IADPSG), GDM can be diagnosed if any one of the following criteria is met: a fasting plasma glucose level of ≥ 5.1 mmol/L, a 1-hour OGTT plasma glucose level of ≥ 10.0 mmol/L, or a 2-hour OGTT plasma glucose level of ≥ 8.5 mmol/L[18]. In this study, all measurements were conducted by trained laboratory personnel adhering to standardized protocols to ensure the validity of OGTT for GDM diagnosis.

Assessment of covariates

A questionnaire was used to collect participants' lifestyle and clinical risk factors through face-to-face interview, including maternal age, ethnicity, date of last menstrual period, parity, conception method for the current pregnancy, medication use (e.g., aspirin), family history of diabetes, smoking, and alcohol consumption. Medical history details such as diabetes, GDM, chronic hypertension, cardiovascular disease, liver disease, kidney disease, polycystic ovary syndrome, thrombophilia, and antiphospholipid syndrome were recorded. The height and weight (light clothes without shoes) of each subject were measured with standardized protocols. Baseline gestational age was determined by the last menstrual

period. Pre-pregnancy BMI was calculated as self-reported weight (in kilograms) divided by the square of height measured (in meters).

Statistical analysis

Continuous variables were expressed as the mean $\pm$ standard deviation and median (interquartile range), as appropriate. Categorical variables were presented as counts (%). Parametric test (t test) was used for continuous variables with normal distribution, while nonparametric test (Mann-Whitney U test) was applied to compare the non-normally distributed variables. The chi-square test or Fisher's exact test were used to compare categorical variables.

In the training set, the least absolute shrinkage and selection operator (LASSO) regression algorithm was applied to select the most robust and non-redundant features for the prediction model. The iterative selection process was undertaken by conducting 10-fold cross-validation. Only variables with non-zero coefficients were selected in this method. Subsequently, all variables, selected in the LASSO regression, were included in the multivariate logistic regression analysis. Backward stepwise selection method with the Akaike Information Criterion (AIC) was applied to select the optimal independent predictors for constructing the prediction model. The odds ratios (ORs) and 95% CIs for the predictors associated with GDM were estimated using the logistic regression analysis. Finally, a prediction model was created by summing the optimal independent predictors selected by multivariate logistic regression, multiplied with their respective coefficients. Subsequently, the goodness of fit of the developed model was assessed using the Hosmer–Lemeshow test. Calibration curve was applied to evaluate the calibration of the model, visually presenting the result of the Hosmer-Lemeshow test. The discrimination ability of the prediction model was evaluated using the receiver operating characteristic (ROC) curve, and the area under the curve (AUC) was employed to quantify the predictive discrimination. The best threshold in the ROC analysis was determined using the Youden Index. The comparison of the performances in predicting GDM between the combined model and each independent predictive factor was assessed in terms of AUC. Decision curve analysis (DCA) was employed to evaluate the clinical net benefits at different threshold probabilities. Finally, to visualize the model and graphically evaluate variable importance, we developed a nomogram based on the coefficients of the independent predictors.

The model was applied to all patients from the independent validation set to validate the model. The Hosmer–Lemeshow test, calibration curve, ROC curve, AUC value and DCA derived from the validation set were respectively generated to further evaluate the performance of the model.

All statistical analyses were performed using R statistical software (version 4.4.2). All tests were two sided, and *P* values <0.05 were considered statistically significant.

Results

Population characteristics

The detailed distribution of clinical characteristics in the training and validation sets were summarized and presented in Table 1. 346 participants were randomly allocated to the training set (n=242) and the validation set (n=104) at a ratio of 7:3, respectively[19]. The prevalence of GDM in the training and validation sets were 24.8% (n=60) and 26.0% (n=27), respectively ($P=0.93$). All the participants had no smoking and alcohol consumption. No significant differences were observed in other clinical factors between the two datasets ($P>0.05$).

Table 1 Baseline characteristics of the investigated object in the training and validation sets (n =346)

Characteristics	Training set (n = 242)	Validation set (n = 104)	P
Age (mean±SD, years)	32.23±4.22	32.86±4.05	0.20
Ethnicity [n (%)]			0.88
Han	240 (99.2)	104 (100.0)	
Others	2 (0.8)	0 (0.0)	
Parity [n (%)]			1.00
0	148 (61.2)	64 (61.5)	
≥1	94 (38.8)	40 (38.5)	
Assisted reproductive technology [n (%)]			1.00
Yes	18 (7.4)	7 (6.7)	
No	224(92.6)	97(93.3)	
History of GDM [n (%)]			1.00
Yes	1 (0.4)	0 (0.0)	
No	241 (99.6)	104 (100.0)	
Family history of diabetes [n (%)]			0.54
Yes	19 (7.9)	11 (10.6)	
No	223 (92.1)	93 (89.4)	
Smoking [n (%)]			NA
Yes	0 (0.0)	0 (0.0)	
No	242 (100.0)	104 (100.0)	
Alcohol consumption [n (%)]			NA
Yes	0 (0.0)	0 (0.0)	
No	242 (100.0)	104 (100.0)	
Hypothyroidism during pregnancy [n (%)]			0.17
Yes	30 (12.4)	7 (6.7)	
No	212 (87.6)	97 (93.3)	
Pre-pregnancy BMI (mean±SD, kg/m2)	22.99±3.50	23.44±3.75	0.28
Progesterone (mean±SD, ng/ml)	21.12±9.22	22.27±22.19	0.50

WBC (mean±SD, 10 ⁹ /L)	8.35±2.06	8.41±2.85	0.82
FBG (mean±SD, mmol/L)	4.66±0.55	4.67±0.34	0.97
ALT (mean±SD, U/L)	15.62±12.65	18.18±19.34	0.15
AST (mean±SD, U/L)	16.28±6.53	16.96±7.58	0.40
GGT (mean±SD, U/L)	14.94±9.68	16.33±11.15	0.24
BUN (mean±SD, mmol/L)	2.68±0.62	2.78±0.63	0.17
Cr (mean±SD, μmol/L)	44.56±6.74	44.43±5.53	0.86
APTT (mean±SD, s)	34.27±2.58	33.79±2.12	0.10
TG/HDL-C (mean±SD)	0.89±0.32	0.87±0.31	0.60

Abbreviations: GDM, gestational diabetes mellitus; BMI, body mass index; WBC, white blood cell count; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate transaminase; GGT, γ-glutamyl transferase; BUN, blood urea nitrogen; Cr, creatinine; APTT, activated partial thromboplastin time; TG/HDL-C, triglyceride to high-density lipoprotein cholesterol ratio; SD, standard deviation; NA, not applicable for matching variables.

Predictors setting and model establishment

11 candidate variables, including history of GDM, family history of diabetes, pre-pregnancy BMI, progesterone, WBC, FBG, AST, GGT, Cr, APTT, and TG/HDL-C, were significant factors for GDM by the LASSO regression analysis (Fig. 1 a, b). Subsequent multivariate logistic regression analysis indicated that family history of diabetes, pre-pregnancy BMI, progesterone, AST, APTT, and TG/HDL-C were identified as independent risk factors for predicting GDM. Among them, each unit increment of TG/HDL-C was significantly associated with a 5.51-fold (OR, 5.51; 95%CI, 1.87-17.37; $P=0.002$) increased GDM risk. Family history of diabetes was significantly associated with a 3.54-fold (OR, 3.54; 95%CI, 1.09-11.38; $P=0.032$) increased GDM risk. Additionally, each unit increment of pre-pregnancy BMI (OR, 1.20; 95%CI, 1.09-1.34; $P<0.001$), progesterone (OR, 1.10; 95%CI, 1.04-1.17; $P=0.001$), and AST (OR, 1.05; 95%CI, 1.00-1.10; $P=0.047$) were associated with 20%, 10% and 5% increased GDM risk, respectively. However, each unit increment of APTT (OR, 0.76; 95%CI, 0.65-0.88; $P<0.001$) level was significantly associated with 24% decreased GDM risk (Table 2). Thereafter, a model was further constructed by integrating these six predictors and depicted as follows: $\text{Logit}(P) = -0.71 + 1.26 \times \text{Family history of diabetes} + 0.19 \times \text{Pre-pregnancy BMI} + 0.1 \times \text{progesterone} + 0.05 \times \text{AST} - 0.28 \times \text{APTT} + 1.71 \times \text{TG/HDL-C}$.

Table 2 Multivariate logistic regression analysis in the training set.

	Coef	S.E.	Wald Z	OR (95% CI)	Pr(> Z)
Intercept	-0.71	2.79	-0.25	0.49(0.00-138.47)	0.80
Family history of diabetes	1.26	0.59	2.14	3.54(1.09-11.38)	0.032*
Pre-pregnancy BMI	0.19	0.05	3.46	1.20(1.09-1.34)	< 0.001
Progesterone	0.10	0.03	3.19	1.10(1.04-1.17)	0.001*
AST	0.05	0.02	1.98	1.05(1.00-1.10)	0.047*
APTT	-0.28	0.08	-3.62	0.76(0.65-0.88)	< 0.001
TG/HDL-C	1.71	0.56	3.04	5.51(1.87-17.37)	0.002*

Abbreviations: S.E., standard error; OR, odds ratio; CI, confidence interval; BMI, body mass index; AST, aspartate transaminase; APTT, activated partial thromboplastin time; TG/HDL-C, triglyceride to high-density lipoprotein cholesterol ratio; *P <0.05.

Model evaluation and internal validation

The calibration curves of the model for predicting GDM demonstrated excellent concordance between the predicted probabilities and the actual probabilities in the training and validation sets (Fig. 2a, b). Besides, the Hosmer-Lemeshow test yielded nonsignificant *P* values in the training and validation sets (*P*>0.05), respectively, indicating satisfactory calibration power. The ROC curves were observed to present excellent discrimination of the model for differentiating who developed GDM, with AUC values of 0.85 (95%CI, 0.80-0.91) and 0.73 (95%CI, 0.62-0.83) in the training and validation sets, respectively (Fig. 3a, b and Table 3). In addition, as calculated by the Youden Index, the best threshold to differentiate GDM was 0.19 (sensitivity, 0.90; specificity, 0.71; accuracy, 0.76) for the training dataset and 0.19 (sensitivity, 0.78; specificity, 0.58; accuracy, 0.64) for the validation dataset (Fig. 3a, b and Table 3). Compared with each independent predictive factor, the combined model showed greater AUC value in predicting GDM (Fig. 4 and Table 3). Furthermore, the DCA curves demonstrated that using the model to make the decision of whether to treat showed a greater benefit than treating all patients or none across wide range of threshold probabilities in the two sets (Fig. 5a, b). Internal validation with excellent performance demonstrated the reliability and generalizability of the model.

Nomogram construction

A nomogram incorporating these six independent risk factors was constructed to visualize the model (Fig. 6). Each predictor is assigned a corresponding score, with the total score across all variables corresponding to the estimated probability of GDM.

Table 3 ROC curve characteristics

	AUC(95%CI)	Best threshold	Sensitivity	Specificity	Accuracy	Youden Index
Training set	0.85(0.80-0.91)	0.19	0.90	0.71	0.76	0.61
Validation set	0.73(0.62-0.83)	0.19	0.78	0.58	0.64	0.36
Family history of diabetes	0.55(0.50-0.60)	0.50	0.15	0.95	0.75	0.10
Pre-pregnancy BMI	0.68(0.60-0.76)	21.95	0.78	0.52	0.59	0.30
Progesterone	0.68(0.59-0.77)	26.31	0.60	0.91	0.84	0.51
AST	0.62(0.53-0.71)	18.50	0.43	0.86	0.76	0.29
APTT	0.72(0.63-0.80)	34.42	0.80	0.69	0.72	0.49
TG/HDL-C	0.78(0.70-0.86)	0.99	0.75	0.87	0.84	0.62

Abbreviations: AUC, area under the curve; CI, confidence interval; BMI, body mass index; AST, aspartate transaminase; APTT, activated partial thromboplastin time; TG/HDL-C, triglyceride to high-density lipoprotein cholesterol ratio.

Discussion

In the current study, family history of diabetes, higher pre-pregnancy BMI, progesterone, AST, and TG/HDL-C levels were associated with increased GDM risk, while higher APTT level was associated with decreased GDM risk. Based on these six independent predictors, a GDM prediction model with excellent performance was developed. Furthermore, a nomogram was further constructed to assist clinicians in assessing individualized GDM risk and identifying high-risk populations early.

Studies have demonstrated that many risk prediction models for GDM have been constructed. A retrospective study revealed the predictive value of fasting plasma glucose for GDM in early pregnancy[20]. Another retrospective study demonstrated the predictive value of liver function indicator and blood lipid levels[21]. Furthermore, a retrospective cohort study applied univariate and multivariate logistic regression analyses to explore the predictive values of biomarkers and maternal characteristics for GDM[22]. The current prediction model integrates multidimensional clinical parameters, including family history of diabetes, pre-pregnancy BMI, P, AST, APTT, and TG/HDL-C levels, demonstrating greater systematization compared to previous models that include only partial indicators. All incorporated parameters are derived from routine blood tests, characterized by low testing costs and strong clinical accessibility. This model is particularly suitable for identifying high-risk pregnant women in primary

healthcare institutions with relatively limited medical resources. Additionally, compared with previous studies, this study constructs a model with excellent performance utilizing LASSO regression for variable selection to address multicollinearity and complex correlations among traditional biochemical indicators of GDM. Moreover, internal validation demonstrates reliability and generalizability of the model. The nomogram demonstrates excellent clinical applicability due to its intuitive visualization, user-friendly design, and high predictive accuracy.

Researches have examined the association between family history of diabetes and GDM risk. A prospective study demonstrated that a family history of diabetes (OR, 1.68; 95%CI, 1.39-2.04) was a significant risk factor for GDM[23]. Additionally, a meta-analysis also confirmed that a family history of diabetes was a significant risk factor for GDM[24]. Furthermore, clinical prediction models suggested that incorporating a family history of diabetes with parameters such as BMI and age significantly enhanced the early predictive accuracy for GDM[25]. These findings align closely with the results of the study. This study further demonstrates that a family history of diabetes is an independent risk factor for GDM and associated with a 3.54-fold increased GDM risk. Molecular mechanism studies had further elucidated the critical role of genetic factors in the pathogenesis of GDM, particularly the association of specific gene variants (e.g., TCF7L2) with insulin secretion and sensitivity, which may significantly increase susceptibility to GDM[26]. The underlying mechanisms involve multiple pathways, including genetic predisposition, metabolic abnormalities, and epigenetic regulation, highlighting the central role of genetic background in the development of GDM[27].

Studies have shown that pre-pregnancy overweight or obesity is a major risk factor for GDM. A cohort study in southwestern China revealed that women who were overweight or obese before pregnancy were more likely to develop GDM[28]. Meta-analysis further supported this conclusion, demonstrating that each 1 kg/m² increase in pre-pregnancy BMI raised the risk of GDM by approximately 8%[29]. Furthermore, a cross-sectional study also demonstrated a positive correlation between the prevalence of GDM and pre-pregnancy BMI[30]. Consistent with previous studies, the study demonstrates that each unit increment of pre-pregnancy BMI is associated with 20% increased GDM risk and serves as an independent predictive factor for GDM development. Notably, animal studies revealed that miRNA-containing exosomes secreted by adipose tissue macrophages in obese mice, when transplanted into lean mice, could induce glucose intolerance and insulin resistance[31]. This mechanism suggests that specific factors secreted by adipose tissue may increase the risk of GDM by interfering with glucose metabolism and inducing insulin resistance. Based on these findings, enhancing pre-pregnancy weight management holds significant clinical importance for the prevention of GDM.

Numerous studies have demonstrated the correlation between progesterone level in early pregnancy and GDM risk. A previous study reported that elevated progesterone levels during the first trimester were associated with increased risk of GDM[32]. A systematic review and meta-analysis revealed that the use of 17 α -hydroxyprogesterone caproate for the prevention of recurrent preterm delivery had a significantly higher risk of developing GDM[33]. Further supporting this finding, a clinical prediction model revealed that the level of progesterone in early pregnancy was a risk factor for GDM[34]. However, some studies

presented conflicting conclusions, indicating that progesterone medication does not increase GDM risk. It should be noted that these studies primarily investigated vaginal progesterone administration rather than systemic medication[35]. These findings suggest that hormonal regulation during early pregnancy may exert a comprehensive influence on GDM development. Similar to previous findings, this study reveals that each unit increment of progesterone level is associated with 10% increased GDM risk and progesterone is an independent risk factor for GDM. Therefore, pregnant women should exercise caution medication intake that may influence progesterone levels (such as dydrogesterone). Previous animal studies demonstrated that exogenous administration of progesterone in rats can lead to decreased insulin sensitivity and elevated blood glucose levels[36, 37].

The AST, one of the liver enzymes, reflects the status of the liver. The liver, one of the key target organs of insulin, plays a critical role in maintaining systemic metabolic balance by regulating glucose and lipid metabolism[38]. However, research on the association between AST and GDM risk remains limited. A study found that GDM patients often exhibit elevated liver enzymes, including ALT, AST, GGT, and alkaline phosphatase (ALP)[39]. Additionally, a prospective study demonstrated positive correlation between AST and GDM risk[40]. These findings are largely consistent with the study. This study further confirms that each unit increment of AST level is associated with 5% increased GDM risk and AST is an independent risk factor for GDM. However, another study did not find a significant association between AST and GDM risk[41], which diverges from this study. This inconsistency may stem from factors such as smaller sample sizes, differences in liver biomarker detection methods, variations in study design, heterogeneity in study populations, and the adjustment of covariates in the models. Although elevated liver enzyme levels may positively correlate with insulin resistance[42], whether increased liver enzymes directly contribute to GDM risk requires further validation through larger-scale and rigorously designed studies.

Studies have confirmed the association between APTT and GDM risk. A case-control study demonstrated that compared with control group, GDM group showed shorter APTT[43]. Additionally, a case-control study found that shortened APTT was associated with poor glycemic control in the GDM group[44]. Similar to previous studies, this study confirms that each unit increment of APTT level is significantly associated with 24% decreased GDM risk and APTT is identified as an independent risk factor for GDM development. During pregnancy, women undergo a series of physiological changes in coagulation function to meet the demands of gestation, such as maintaining placental blood supply. APTT, a key indicator for assessing the intrinsic coagulation system, typically reflects a hypercoagulable state when shortened. The reduced APTT levels observed in GDM patients during early pregnancy suggest enhanced blood coagulability and an increased risk of thrombosis, which may be linked to the pathophysiological mechanisms of GDM. A hypercoagulable state can potentially induce vascular endothelial damage, triggering inflammatory responses and exacerbating insulin resistance, thereby elevating the risk of GDM. Therefore, monitoring coagulation indicators such as APTT in early pregnancy is valuable for assessing maternal coagulation status and provides critical insights for early identification of high-risk GDM populations and the implementation of preventive interventions.

Numerous studies have demonstrated the association between TG/HDL-C ratio and GDM risk. A prospective study demonstrated that elevated TG level was positively correlated with GDM risk[45]. Furthermore, a systematic review and meta-analysis indicated that HDL-C was inversely related to GDM risk[46]. Since the TG/HDL-C ratio integrates both TG and HDL-C metrics, it is closely linked to GDM development. A prospective study further confirmed that, after adjusting for confounders, the relative risk of GDM in women in the highest tertile of the TG/HDL-C ratio was 3.90 times higher than that in the lowest tertile[47]. Consistent with previous research, this study further confirm that each unit increment of TG/HDL-C level is significantly associated with a 5.51-fold increased GDM risk, indicating a stable relationship between the TG/HDL-C ratio and GDM risk. During pregnancy, elevated estrogen levels and insulin resistance can promote hepatic lipid synthesis[48]. Increased TG levels in early pregnancy lead to elevated free fatty acids in the blood, and high free fatty acid levels may impair insulin sensitivity[49], creating a vicious cycle between high TG levels and insulin resistance, which may be a key mechanism in GDM development[50]. Animal studies also suggest that low HDL-C levels may impair glucose homeostasis by reducing insulin secretion and sensitivity[51, 52].

Several limitations should be considered. First, the relatively small sample size presents certain constraints, but the sample size was determined to be statistically adequate for model development through formal power calculations. Second, the performance of the model has been not confirmed through external validation in independent populations. Notably, internal validation with excellent performance was conducted to demonstrate reliability and generalizability of the model. Third, regarding clinical variable selection, while incorporating additional unmeasured risk factors (e.g., genetic markers and ultrasound data) could enhance model performance, but this study focused on the most clinically prevalent and readily available indicators. Fourth, the incidence of GDM is higher after rigorous exclusion, what could impact the model performance and subsequent clinical application, where GDM prevalence is lower. Notably, the observed GDM incidence of 15.8% in our cohort closely approximates the reported prevalence of 14.8% in China. This prediction model requires external validation in larger, multi-center cohort studies to verify its generalizability. Fifth, the inclusion of established high-risk factors (maternal age and FBG) unexpectedly degraded model performance, potentially due to limited sample size. Future studies with larger cohorts are needed to optimize this model. Lastly, as the study participants were primarily of Han ethnicity, the findings might not be directly generalizable to other populations with different ethnics due to differences in dietary habits and genetic backgrounds. Therefore, it is necessary to validate the results in diverse ethnic populations.

Conclusions

In summary, based on family history of diabetes, pre-pregnancy BMI, P, AST, APTT, and TG/HDL-C levels, the present study developed a model with excellent for predicting GDM. Furthermore, a nomogram was constructed to visualize the model. This model demonstrated reliable performance for identifying high-risk pregnant women and optimizing the clinical prevention of GDM.

Abbreviations

GDM Gestational diabetes mellitus

BMI Body mass index

FBG Fasting blood glucose

CI Confidence interval

PASS Power Analysis and Sample Size

TSH Thyroid-stimulating hormone

T4 Thyroxine

ALT Alanine aminotransferase

AST Aspartate transaminase

GGT γ -glutamyl transferase

BUN Blood urea nitrogen

Cr Creatinine

TG Triglyceride

HDL-C High-density lipoprotein cholesterol

APTT Activated partial thromboplastin time

WBC White blood cell count

OGTT Oral glucose tolerance test

PBG Postprandial blood glucose

IADPSG International Association of Diabetes and Pregnancy Study Groups

LASSO Least absolute shrinkage and selection operator

AIC Akaike Information Criterion

OR Odds ratio

ROC Receiver operating characteristic

AUC Area under the curve

DCA Decision curve analysis

SD Standard deviation

NA Not applicable for matching variables.

S.E. Standard error

ALP Alkaline phosphatase

Declarations

Acknowledgements

We thank all participants of the study for their valuable contributions.

Author contributions

HLL and CJJ conceived and planned the study. CJJ, YSS, WSL, LS, FR, WXQ, HX, ZX, and ZGJ collected the data. CJJ and YSS performed the statistical analysis. CJJ drafted the original manuscript. HLL, ZQ and CJJ reviewed and revised the draft of the manuscript. HLL and ZQ supervised the study. All authors approved the final manuscript for submission.

Funding

The study was supported by National Natural Science Foundation of China (82404282), Henan Province Key Research and Development Project (221111310700), Henan Medical Science and Technology Research and Development Program (LHGJ20220451, LHGJ20240319, LHGJ20220515), and Henan Province Youth Health Science and Technology Innovation Talent Training Project (YXKC2022051). HLL conceived and planned the study. YSS, HX, ZX, and ZGJ collected the data. Other funders had no role in the study design, implementation, analysis, decision to publish, or reparation of the manuscript.

Data availability

The datasets analysed in the study are not publicly available due to privacy but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was performed by the principles of the Declaration of Helsinki. The study was approved by the Ethics Committee of the Second Affiliated Hospital of Zhengzhou University. The patients were well informed, and consent forms were signed.

Consent for publication

Not Applicable.

Competing interests

The authors declare no conflict of interest.

References

1. Kautzky-Willer A, Winhofer Y, Kiss H, Falcone V, Berger A, Lechleitner M, et al. [Gestational diabetes mellitus (Update 2023)] [J]. *Wien Klin Wochenschr.* 2023; 135(Suppl 1): 115-28.
2. Sweeting A, Wong J, Murphy HR, Ross GP. A Clinical Update on Gestational Diabetes Mellitus [J]. *Endocr Rev.* 2022; 43(5): 763-93.
3. Gao C, Sun X, Lu L, Liu F, Yuan J. Prevalence of gestational diabetes mellitus in mainland China: A systematic review and meta-analysis [J]. *J Diabetes Investig.* 2019; 10(1): 154-62.
4. Coustan DR. Pregnancy outcomes in women with and without gestational diabetes mellitus according to the international association of the diabetes and pregnancy study groups criteria [J]. *Obstet Gynecol.* 2013; 121(2 Pt 1): 377.
5. Farrar D, Simmonds M, Bryant M, Sheldon TA, Tuffnell D, Golder S, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis [J]. *Bmj.* 2016; 354: i4694.
6. Li M, Rahman ML, Wu J, Ding M, Chavarro JE, Lin Y, et al. Genetic factors and risk of type 2 diabetes among women with a history of gestational diabetes: findings from two independent populations [J]. *BMJ Open Diabetes Res Care.* 2020; 8(1).
7. Lorenzo-Almorós A, Hang T, Peiró C, Soriano-Guillén L, Egido J, Tuñón J, et al. Predictive and diagnostic biomarkers for gestational diabetes and its associated metabolic and cardiovascular diseases [J]. *Cardiovasc Diabetol.* 2019; 18(1): 140.
8. Bianco ME, Josefson JL. Hyperglycemia During Pregnancy and Long-Term Offspring Outcomes [J]. *Curr Diab Rep.* 2019; 19(12): 143.
9. Zhang D, Zhang S, Li G, Lai Y, Hao YT, Chen WQ, et al. A clinical model and nomogram for early prediction of gestational diabetes based on common maternal demographics and routine clinical parameters [J]. *J Obstet Gynaecol Res.* 2022; 48(11): 2738-47.
10. Zhang X, Zhao X, Huo L, Yuan N, Sun J, Du J, et al. Risk prediction model of gestational diabetes mellitus based on nomogram in a Chinese population cohort study [J]. *Sci Rep.* 2020; 10(1): 21223.
11. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, et al. Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis [J]. *J Med Internet Res.* 2022; 24(3): e26634.
12. An R, Ma S, Zhang N, Lin H, Xiang T, Chen M, et al. AST-to-ALT ratio in the first trimester and the risk of gestational diabetes mellitus [J]. *Front Endocrinol (Lausanne).* 2022; 13: 1017448.

13. Benhalima K, Van Crombrugge P, Moyson C, Verhaeghe J, Vandeginste S, Verlaenen H, et al. Estimating the risk of gestational diabetes mellitus based on the 2013 WHO criteria: a prediction model based on clinical and biochemical variables in early pregnancy [J]. *Acta Diabetol.* 2020; 57(6): 661-71.
14. Kotzaeridi G, Blätter J, Eppel D, Rosicky I, Mittlböck M, Yerlikaya-Schatten G, et al. Performance of early risk assessment tools to predict the later development of gestational diabetes [J]. *Eur J Clin Invest.* 2021; 51(12): e13630.
15. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies [J]. *Gastroenterol Hepatol Bed Bench.* 2013; 6(1): 14-7.
16. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies [J]. *Biochem Med (Zagreb).* 2021; 31(1): 010502.
17. Lu L, Li C, Deng J, Luo J, Huang C. Maternal serum NGAL in the first trimester of pregnancy is a potential biomarker for the prediction of gestational diabetes mellitus [J]. *Front Endocrinol (Lausanne).* 2022; 13: 977254.
18. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy [J]. *Diabetes Care.* 2010; 33(3): 676-82.
19. Kang BS, Lee SU, Hong S, Choi SK, Shin JE, Wie JH, et al. Prediction of gestational diabetes mellitus in Asian women using machine learning algorithms [J]. *Sci Rep.* 2023; 13(1): 13356.
20. Tong JN, Wu LL, Chen YX, Guan XN, Tian FY, Zhang HF, et al. Fasting plasma glucose in the first trimester is related to gestational diabetes mellitus and adverse pregnancy outcomes [J]. *Endocrine.* 2022; 75(1): 70-81.
21. Wang H, Li Q, Wang H, Song W. Construction and validation of a line chart for gestational diabetes mellitus based on clinical indicators [J]. *Lipids Health Dis.* 2024; 23(1): 349.
22. Li R, Yuan K, Yu X, Jiang Y, Liu P, Zhang K. Construction and validation of risk prediction model for gestational diabetes based on a nomogram [J]. *Am J Transl Res.* 2023; 15(2): 1223-30.
23. Solomon CG, Willett WC, Carey VJ, Rich-Edwards J, Hunter DJ, Colditz GA, et al. A prospective study of pregravid determinants of gestational diabetes mellitus [J]. *Jama.* 1997; 278(13): 1078-83.
24. Moosazadeh M, Asemi Z, Lankarani KB, Tabrizi R, Maharlouei N, Naghibzadeh-Tahami A, et al. Family history of diabetes and the risk of gestational diabetes mellitus in Iran: A systematic review and meta-analysis [J]. *Diabetes Metab Syndr.* 2017; 11 Suppl 1: S99-s104.
25. Wu Y, Hamelmann P, Van Der Ven M, Asvadi S, Van Der Hout-Van Der Jagt MB, Oei SG, et al. Early prediction of gestational diabetes mellitus using maternal demographic and clinical risk factors [J]. *BMC Res Notes.* 2024; 17(1): 105.
26. Zhang C, Bao W, Rong Y, Yang H, Bowers K, Yeung E, et al. Genetic variants and the risk of gestational diabetes mellitus: a systematic review [J]. *Hum Reprod Update.* 2013; 19(4): 376-90.

27. Lee KW, Ching SM, Ramachandran V, Yee A, Hoo FK, Chia YC, et al. Prevalence and risk factors of gestational diabetes mellitus in Asia: a systematic review and meta-analysis [J]. *BMC Pregnancy Childbirth*. 2018; 18(1): 494.
28. Liang Y, Li DT, Chen MX, Gong YH, Zhang X, Yang WY, et al. [Associations of Pre-pregnancy Body Mass Index and Gestational Weight Gain with Gestational Diabetes Mellitus: a Cohort Study in Southwest China] [J]. *Sichuan Da Xue Xue Bao Yi Xue Ban*. 2019; 50(1): 83-7.
29. Singh J, Huang CC, Driggers RW, Timofeev J, Amini D, Landy HJ, et al. The impact of pre-pregnancy body mass index on the risk of gestational diabetes [J]. *J Matern Fetal Neonatal Med*. 2012; 25(1): 5-10.
30. Groof Z, Garashi G, Husain H, Owayed S, Albader S, Mouhsen H, et al. Prevalence, Risk Factors, and Fetomaternal Outcomes of Gestational Diabetes Mellitus in Kuwait: A Cross-Sectional Study [J]. *J Diabetes Res*. 2019; 2019: 9136250.
31. Ying W, Riopel M, Bandyopadhyay G, Dong Y, Birmingham A, Seo JB, et al. Adipose Tissue Macrophage-Derived Exosomal miRNAs Can Modulate In Vivo and In Vitro Insulin Sensitivity [J]. *Cell*. 2017; 171(2): 372-84.e12.
32. Alyas S, Roohi N, Ahmed S, Ashraf S, Ilyas S, Ilyas A. Lower vitamin D and sex hormone binding globulin levels and higher progesterone, cortisol and t-PA levels in early second trimester are associated with higher risk of developing gestational diabetes mellitus [J]. *J Biol Regul Homeost Agents*. 2020; 34(1).
33. Eke AC, Sheffield J, Graham EM. 17 α -Hydroxyprogesterone Caproate and the Risk of Glucose Intolerance in Pregnancy: A Systematic Review and Meta-analysis [J]. *Obstet Gynecol*. 2019; 133(3): 468-75.
34. Kang M, Zhang H, Zhang J, Huang K, Zhao J, Hu J, et al. A Novel Nomogram for Predicting Gestational Diabetes Mellitus During Early Pregnancy [J]. *Front Endocrinol (Lausanne)*. 2021; 12: 779210.
35. Rosta K, Al-Bibawy K, Al-Bibawy M, Temsch W, Springer S, Somogyi A, et al. Vaginal Progesterone Has No Diabetogenic Potential in Twin Pregnancies: A Retrospective Case-Control Study on 1686 Pregnancies [J]. *J Clin Med*. 2020; 9(7).
36. Picard F, Wanatabe M, Schoonjans K, Lydon J, O'malley BW, Auwerx J. Progesterone receptor knockout mice have an improved glucose homeostasis secondary to beta -cell proliferation [J]. *Proc Natl Acad Sci U S A*. 2002; 99(24): 15644-8.
37. González C, Alonso A, Alvarez N, Díaz F, Martínez M, Fernández S, et al. Role of 17beta-estradiol and/or progesterone on insulin sensitivity in the rat: implications during pregnancy [J]. *J Endocrinol*. 2000; 166(2): 283-91.
38. Bo T, Gao L, Yao Z, Shao S, Wang X, Proud CG, et al. Hepatic selective insulin resistance at the intersection of insulin signaling and metabolic dysfunction-associated steatotic liver disease [J]. *Cell Metab*. 2024; 36(5): 947-68.

39. Clark JM, Brancati FL, Diehl AM. The prevalence and etiology of elevated aminotransferase levels in the United States [J]. *Am J Gastroenterol*. 2003; 98(5): 960-7.
40. Wu P, Wang Y, Ye Y, Yang X, Huang Y, Ye Y, et al. Liver biomarkers, lipid metabolites, and risk of gestational diabetes mellitus in a prospective study among Chinese pregnant women [J]. *BMC Med*. 2023; 21(1): 150.
41. Sridhar SB, Xu F, Darbinian J, Quesenberry CP, Ferrara A, Hedderson MM. Pregravid liver enzyme levels and risk of gestational diabetes mellitus during a subsequent pregnancy [J]. *Diabetes Care*. 2014; 37(7): 1878-84.
42. Gao F, Pan JM, Hou XH, Fang QC, Lu HJ, Tang JL, et al. Liver enzymes concentrations are closely related to prediabetes: findings of the Shanghai Diabetes Study II (SHDS II) [J]. *Biomed Environ Sci*. 2012; 25(1): 30-7.
43. Dong C, Gu X, Chen F, Long Y, Zhu D, Yang X, et al. The variation degree of coagulation function is not responsible for extra risk of hemorrhage in gestational diabetes mellitus [J]. *J Clin Lab Anal*. 2020; 34(4): e23129.
44. Teliga-Czajkowska J, Sienko J, Zareba-Szczudlik J, Malinowska-Polubiec A, Romejko-Wolniewicz E, Czajkowski K. Influence of Glycemic Control on Coagulation and Lipid Metabolism in Pregnancies Complicated by Pregestational and Gestational Diabetes Mellitus [J]. *Adv Exp Med Biol*. 2019; 1176: 81-8.
45. Enquobahrie DA, Williams MA, Qiu C, Luthy DA. Early pregnancy lipid concentrations and the risk of gestational diabetes mellitus [J]. *Diabetes Res Clin Pract*. 2005; 70(2): 134-42.
46. Hu J, Gillies CL, Lin S, Stewart ZA, Melford SE, Abrams KR, et al. Association of maternal lipid profile and gestational diabetes mellitus: A systematic review and meta-analysis of 292 studies and 97,880 women [J]. *EClinicalMedicine*. 2021; 34: 100830.
47. Pazhohan A, Rezaee Moradali M, Pazhohan N. Association of first-trimester maternal lipid profiles and triglyceride-glucose index with the risk of gestational diabetes mellitus and large for gestational age newborn [J]. *J Matern Fetal Neonatal Med*. 2019; 32(7): 1167-75.
48. Rahnemaei FA, Pakzad R, Amirian A, Pakzad I, Abdi F. Effect of gestational diabetes mellitus on lipid profile: A systematic review and meta-analysis [J]. *Open Med (Wars)*. 2022; 17(1): 70-86.
49. Van De Woestijne AP, Monajemi H, Kalkhoven E, Visseren FL. Adipose tissue dysfunction and hypertriglyceridemia: mechanisms and management [J]. *Obes Rev*. 2011; 12(10): 829-40.
50. Manell H, Kristinsson H, Kullberg J, Ubhayasekera SJK, Mörwald K, Staaf J, et al. Hyperglucagonemia in youth is associated with high plasma free fatty acids, visceral adiposity, and impaired glucose tolerance [J]. *Pediatr Diabetes*. 2019; 20(7): 880-91.
51. Di Bartolo BA, Cartland SP, Genner S, Manuneehi Cholan P, Vellozzi M, Rye KA, et al. HDL Improves Cholesterol and Glucose Homeostasis and Reduces Atherosclerosis in Diabetes-Associated Atherosclerosis [J]. *J Diabetes Res*. 2021; 2021: 6668506.
52. Rütli S, Ehses JA, Sibler RA, Prazak R, Rohrer L, Georgopoulos S, et al. Low- and high-density lipoproteins modulate function, apoptosis, and proliferation of primary human and murine

Figures

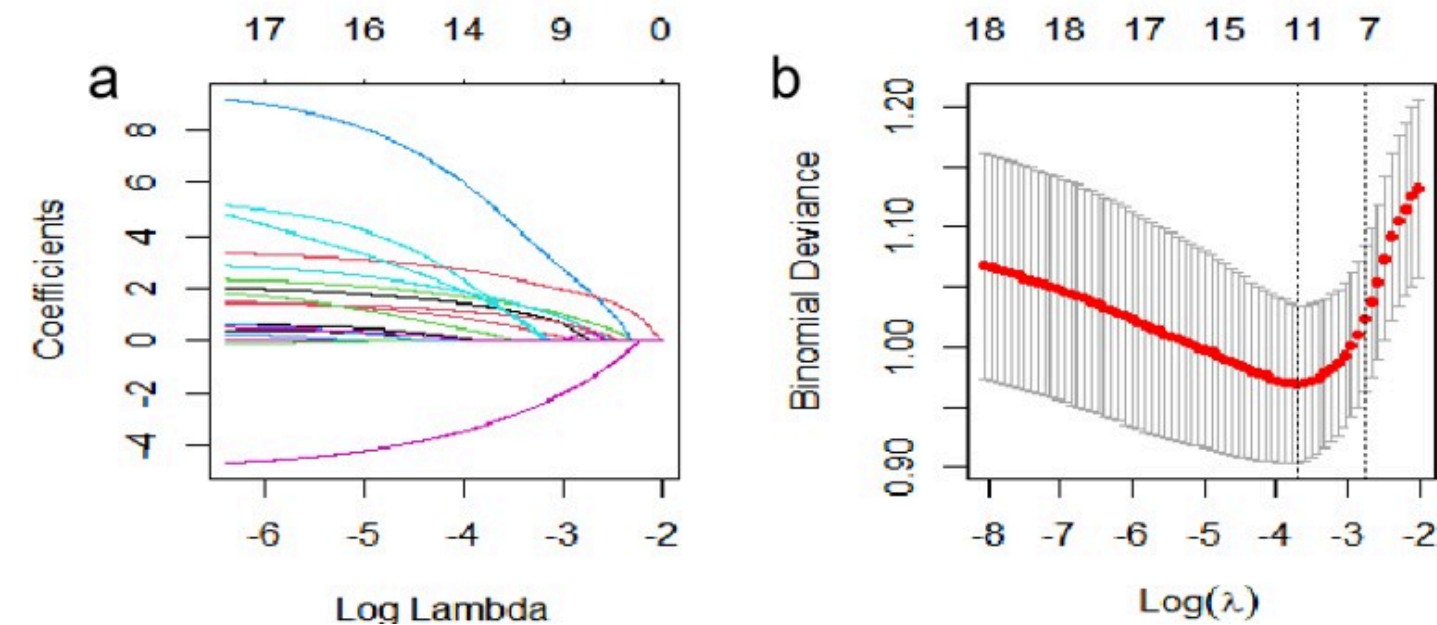


Figure 1

(a) The LASSO coefficient profiles of the 18 indicators. (b) A vertical line was generated at the log (λ) by using 10-fold cross-validation, where the optimal λ value resulted in 11 indicators. The X-axis on the top indicates the number of nonzero coefficient indicators in the model.

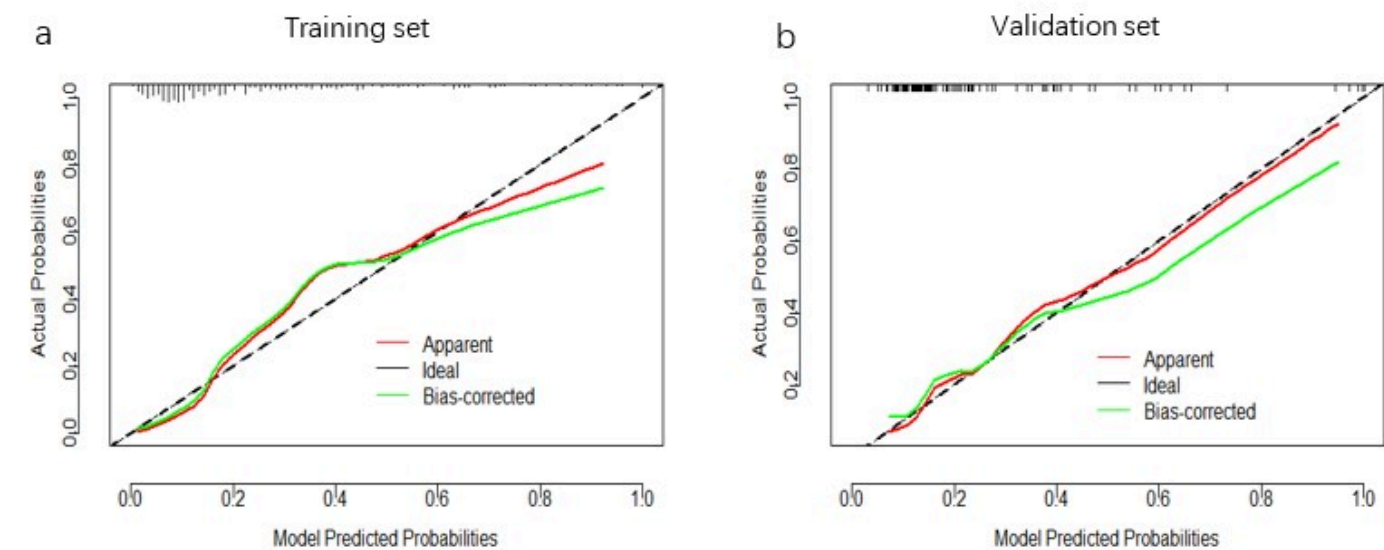


Figure 2

Calibration curve of the model in the training set (a) and validation set (b). For the calibration curve, the x-axis represents the model-predicted probabilities, and the y-axis represents the actual GDM probabilities. The diagonal black dash line represents perfect prediction by an ideal model, and the red solid line represents the predictive performance of the model. The red solid line has a closer fit to the dash line, which indicates a better prediction.

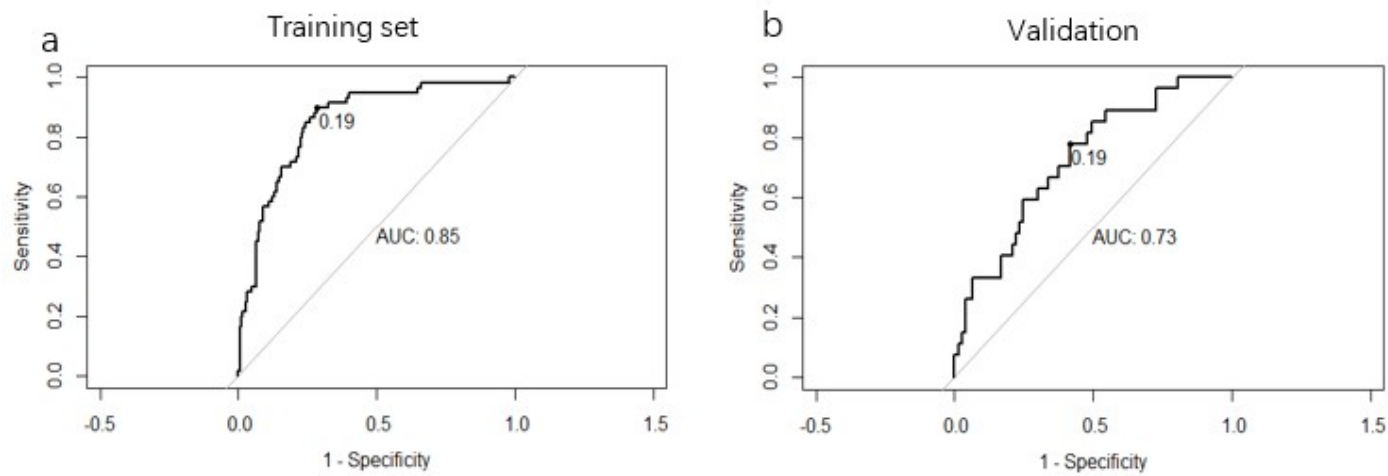


Figure 3

Receiver operator characteristic curve of the model in the training set (a) and validation set (b). AUC: area under the curve.

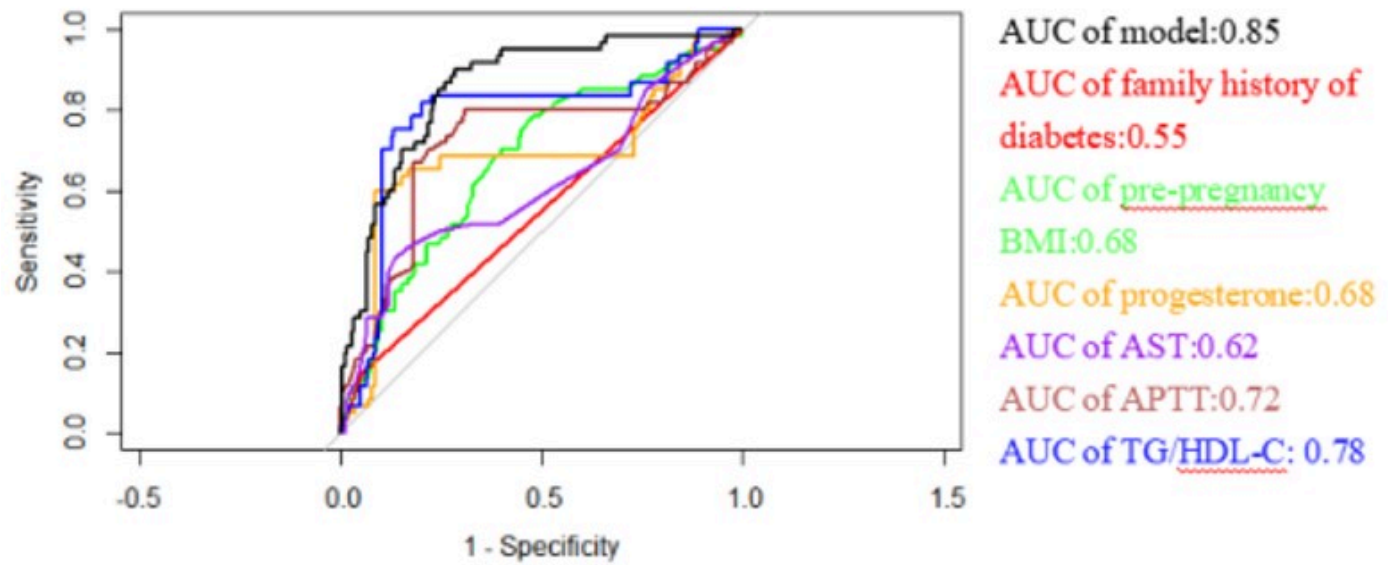


Figure 4

Receiver operator characteristic curves of the models are presented to compare their discriminatory for predicting GDM. AUC, area under the curve; BMI, body mass index; AST, apartate transaminase; APTT, activated partial thromboplastin time; TG/HDL-C, triglyceride to high-density lipoprotein cholesterol ratio.

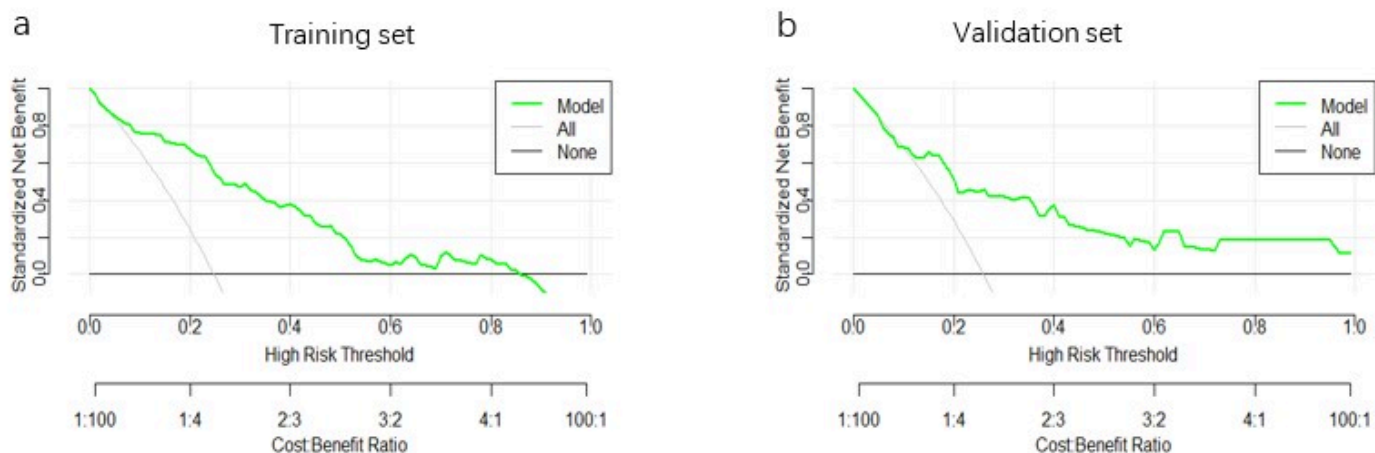


Figure 5

Decision curve analyses for the model in the training set (a) and validation set (b). The x-axis represents the threshold probability. The y-axis represents the net benefit. The horizontal solid black line represents the assumption that no patients with GDM were involved, and the solid gray line represents the assumption that all patients had GDM. The solid green line represents the model.

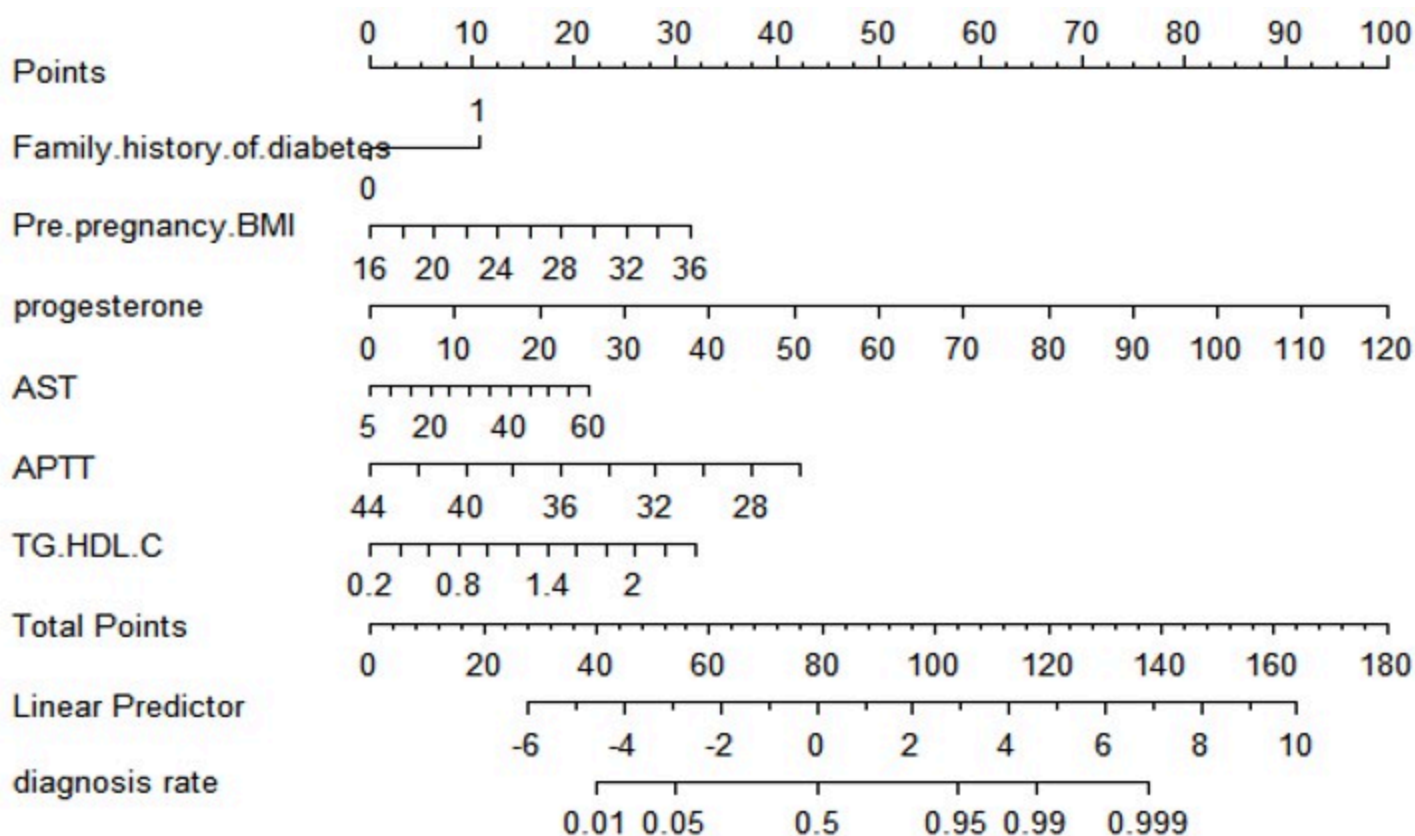


Figure 6

A nomogram for predicting the probability of GDM. Family.history.of.diabetes, family history of diabetes; Pre.pregnancy.BMI, pre-pregnancy body mass index; AST, apartate transaminase; APTT, activated partial

thromboplastin time; TG:HDL.C, triglyceride to high-density lipoprotein cholesterol ratio.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.docx](#)