Received: 17 January 2021 | Revised: 8 April 2021 | Accepted: 5 May 2021

DOI: 10.1002/jcla.23833

RESEARCH ARTICLE

WILEY

Establishment of clinical diagnostic models using glucose, lipid, and urinary polypeptides in gestational diabetes mellitus

Zhiying Hu[1,2] | Man Zhang[1,2]

[1]Clinical Laboratory Medicine, Beijing Shijitan Hospital, Capital Medical University, Beijing, China

[2]Beijing Key Laboratory of Urinary Cellular Molecular Diagnostics, Beijing, China

Correspondence
Man Zhang, Clinical Laboratory Medicine, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China.
Email: zhangman@bjsjth.cn

Funding information
This work was supported by National Key R&D Program of China (No. 2016YFC1000700 2016YFC1000702), the Enhancement Funding of Beijing Key Laboratory of Urinary Cellular Molecular Diagnostics (2020-JS02,) and the Youth Research Funding of Beijing Shijitan Hospital, Capital Medical University (2019-q06)

Abstract

Background: Gestational diabetes mellitus (GDM) has many adverse outcomes that seriously threaten the short-term and long-term health of mothers and infants. This study comprehensively analyzed the clinical diagnostic value of GDM-related clinical indexes and urine polypeptide research results, and established comprehensive index diagnostic models.

Methods: In this study, diagnostic values from the clinical indexes of serum triglyceride (TRIG), high-density lipoprotein cholesterol (HDL-C), fasting plasma glucose (FPG) and glycosylated hemoglobin (HbA1c), and 7 GDM-related urinary polypeptides were analyzed retrospectively. The multiple logistic regression equation, multilayer perceptron neural network model, radial basis function, and discriminant analysis function models of GDM-related indexes were established using machine language.

Results: The results showed that HbA1c had the highest diagnostic value for GDM, with an area under the curve (AUC) of 0.769. When the cut-off value was 4.95, the diagnostic sensitivity and specificity were 70.5% and 70.0%, respectively. Among the seven GDM-related urinary polypeptides, human hemopexin (HEMO) had the highest diagnostic value, with an AUC of 0.690. When the cut-off value was 368.5, the sensitivity and specificity were 79.5% and 43.3%, respectively. The AUC of the multilayer perceptron neural network model was 0.942, followed by binary logistic regression (0.938), radial basis function model (0.909), and the discriminant analysis function model (0.908).

Conclusion: The establishment of a GDM diagnostic model combining blood glucose, blood lipid, and urine polypeptide indexes can lay a foundation for exploring machine language and artificial intelligence in diagnostic systems.

KEYWORDS
clinical diagnostic model, gestational diabetes mellitus, urinary polypeptide

1 | INTRODUCTION

In the normal population, carbohydrates, lipids, and proteins can be transformed into each other, and dynamically balanced to meet the normal physiological needs of the human body. Acetyl CoA, pyruvic acid, α-ketoglutarate, and oxaloacetic acid are four important transfer stations in the process of nutrient transformation. When the plasma glucose level increases, the amounts of acetyl coenzyme A, pyruvic acid, α-ketoglutarate, and oxaloacetic acid also increase, which definitely affects lipid metabolism and protein metabolism. The pathogenesis of gestational diabetes mellitus (GDM) is related to insulin resistance and decreased insulin sensitivity in the second and third trimesters of pregnancy.[1] During pregnancy, the increase in estrogen, progesterone, and cortisol further disrupts glucose insulin balance.[2] Hyperglycemia in GDM patients also affects lipid and protein metabolism.

Studies have shown that dyslipidemia and glucose metabolism disorders often coexist in GDM patients and are closely related to insulin resistance.[3] Pregnancy-induced insulin resistance ensures adequate nutrition for the fetus and placenta. However, in mothers with obesity, metabolic syndrome, or GDM, the excess nutrition places the fetuses at risk of metabolic disease over their lifetime.[4] GDM patients with dyslipidemia have a higher incidence of adverse pregnancy outcomes, especially preterm labor and preeclampsia.[5]

In terms of GDM-related protein metabolism, a prospective case-control study indicated that higher dietary and circulating concentrations of isoleucine, leucine, valine, and total branched-chain amino acids (BCAAs) are related to a greater risk of progression from GDM to type 2 diabetes (T2D) later in life.[6] It is suggested that protein metabolic systems, such as dietary and/or plasma BCAAs, also play an important role in the pathophysiology of GDM and T2D.

Based on the screening of GDM-related clinical indexes and urine polypeptides,[7–9] the clinical diagnostic value of each index was compared horizontally, and the diagnostic models of different statistical classification methods were established in combination with each index. We hope that through the establishment of the diagnosis model, clinical, and research indexes can be applied to the clinical prevention, diagnosis, and dynamic monitoring of GDM. This study preliminarily discussed the clinical application prospects of GDM-related markers before the establishment of an artificial intelligence GDM diagnostic system.

2 | MATERIALS AND METHODS

2.1 | Study population

The subjects were women aged 24–42 years. Based on a previous study,[9] 78 GDM patients (GDM group) were randomly selected between February 2018 and August 2019, and 30 normal pregnant women (N group) were selected in the same period. We had the complete clinical data of all subjects from early pregnancy (8–12 weeks) to 42 days postpartum. A 75 g glucose (one-step method) oral glucose tolerance test (OGTT) was performed at 24–28 weeks of gestation. The diagnosis of GDM met the diabetes diagnosis standard of the American Diabetes Association (ADA) from 2011. In the GDM group, subjects with previous impaired glucose tolerance, acute and chronic infectious diseases, tumors, cardiovascular diseases, and severe liver and kidney dysfunction were excluded. Subjects with other complications of GDM, such as thyroid dysfunction, hypertension, preeclampsia, and anemia, were also excluded.

2.2 | Data collection

The serum triglyceride (TRIG) and high-density lipoprotein cholesterol (HDL-C) values in the first trimester (8–12 weeks), and values of fasting plasma glucose (FPG) and glycosylated hemoglobin (HbA1c) of OGTT in the second trimester (24–28 weeks) were collected by the laboratory information system (LIS) at our hospital. Seven GDM-related urinary peptide markers were successfully identified in previous studies, including coagulation factor IX (F IX), TBC1 family member 5 isoform a (Homosapiens) (TBC1D5a), human immunoglobulin kappa constant (human C_k gene), urine albumin (ALBU), alpha-2-macroglobulin (A2MG), human hemopexin (HEMO), and alpha-1-Microglobulin (AMBP).[8,9] The peak values of the seven urine polypeptides were retrospectively analyzed.

2.3 | Receiver operating characteristic (ROC) analysis

The ROC curves of the 11 indexes were established by SPSS 22.0 software. The area under the curve (AUC) value was calculated to evaluate the diagnostic value of each index. The binary logistic regression equation of the comprehensive indexes was established, and the ROC curve and AUC analysis of the regression equation were carried out to explore the diagnostic value of multiple indexes.

2.4 | Establishment of the clinical diagnosis model

The N and GDM groups were the stratified factors, and the samples in each group were randomly regrouped. The diagnostic models were established using the statistical methods of the multilayer perceptron neural network model, radial basis function, and discriminant analysis function. The values of TRIG, HDL-C, FPG, HbA1c, F IX, TBCID5a, human C_k gene, ALBU, A2MG, HEMO, and AMBP were the input layers, and the N and GDM groups were the output layers. Seventy percent of the samples were randomly used to establish the model, and 30% were used to verify the diagnostic efficiency of the model.

All statistical analyses were performed using SPSS 22.0. Statistical methods included the multilayer perceptron neural network model, radial basis function, and the discriminant analysis function. After establishing the model, the accuracy, sensitivity, and

specificity of the model for disease diagnosis were calculated, and the performance of the model was comprehensively evaluated. The Delong test was performed using SAS 9.4 software to compare the statistical differences between the traditional method and the deep learning approach, based on the ROC curve.

3 | RESULTS

3.1 | Predictive value of lipid metabolism markers in the first trimester of GDM

The clinical characteristics of all the subjects and the levels of lipid metabolism markers were compared (Table 1). There were no significant differences in age, pre-pregnancy body mass index, average gestational age, and average number of pregnancies and births, between the two groups ($p > 0.05$).[8] The serum TRIG and HDL-C levels in the first trimester were significantly different ($p < 0.05$) between the GDM and N groups. The mean value, standard deviation, and P value of TRIG and HDL-C were shown, and the difference in TRIG between the two groups was more significant ($p = 0.02$).

3.2 | ROC analysis

The results of TRIG, HDL-C, FPG, and HbA1c, and the peptide peak values of F IX, TBC1D5a, human C_k gene, ALBU, A2MG, HEMO, and AMBP were collected from 78 GDM patients and 30 normal pregnant women. The ROC curve was established using the above data, and the AUC was calculated (Table 2, Figure 1). The AUC of the 11 indexes were 0.704, 0.565, 0.698, 0.769, 0.612, 0.621, 0.670, 0.641, 0.612, 0.690, and 0.600, respectively. ROC analyses of ALBU, A2MG, HEMO, and AMBP have been reported.[8]

Among the 11 indexes, HbA1c had the highest diagnostic value for GDM, with an AUC of 0.769. When the cut-off value was 4.95, the diagnostic sensitivity and specificity were 70.5% and 70.0%, respectively. Among the seven GDM-related urinary polypeptides, HEMO had the highest diagnostic value for GDM, with an AUC of 0.690. When the cut-off value was 368.5, the sensitivity and specificity were 79.5% and 43.3%, respectively. The other six urinary polypeptides were human C_k gene, ALBU, TBCID5a, F IX, A2MG, and AMBP in descending order of the AUC.

3.3 | Multivariate logistic regression analysis

Multivariate logistic regression analysis was used to analyze the diagnostic value of the 11 GDM-related indexes. The diagnostic formula is as follows:

$$\begin{aligned} Y = \operatorname{logit}(P) = &-19.886 + 1.046X_{TRIG} - 0.187X_{HDL-C} \\ &+1.394X_{FPG} + 2.989X_{HbA1c} - 0.003X_{FIX} - 0.001X_{TBC1D5a} \\ &+0.044X_{humanc\_kgene} + 0.009X_{ABLU} + 0.001X_{A2MG} \\ &-0.009X_{HEMO} - 0.022X_{AMBP} \end{aligned}$$

As shown in Figure 2, the AUC of the multiple logistic regression equation of 11 indexes was 0.938 (95% CI: 0.890–0.985). When the cut-off value was 0.475, the sensitivity and specificity were 94.9% and 76.7%, respectively.

3.4 | Establishment of the multilayer perceptron neural network model

Using SPSS software, a two-layer perceptron neural network model was constructed. The model settings were as follows: the input layer had 11 indexes, the hidden layer had two nodes, and the hidden layer activation function was a hyperbolic tangent function; the output layer had two nodes (N and GDM groups), the output layer activation function was a softmax function, and the cross-entropy error function was used. Among the 108 samples, 74 (68.5%) were in the training group and 34 (31.5%) were in the validation group.

For GDM diagnosis, the accuracy rate of the training group sample classification was 96.1%, and that of the validation group was 96.3% (Table 3). The accuracy of model classification for the GDM group was higher. The classification of GDM samples showed that 3.9% of the training group samples and 3.7% of the validation group samples were incorrectly diagnosed as normal samples (N group)

TABLE 1 Analysis of the clinical characteristics , lipid metabolism markers, TRIG and HDL-C, in the GDM and N groups ($\bar{X} \pm S$).

Parameters	*N* Group (*n* = 30)	GDM group (n = 78)	*p* value
Age(year)*	31.83 ± 3.71	32.88 ± 4.21	0.436
Pre-pregnancy BMI*	21.26 ± 2.52	23.35 ± 3.45	0.054
Average gestational age*	39.54 ± 1.08	38.97 ± 1.95	0.346
Average number of pregnancies*	1.90 ± 0.99	2.13 ± 1.21	0.206
Average number of births*	1.40 ± 0.50	1.41 ± 0.55	0.510
TRIG (mmol/L)	1.13 ± 0.44	1.76 ± 1.23	0.020
HDL-C (mmol/L)	1.55 ± 0.24	1.48 ± 0.32	0.038

*It has been reported.

TABLE 2 ROC analysis

Index	Cut-off value	Sensitivity (%)	Specificity (%)	AUC	95%CI
TRIG	1.14	73.1	63.3	0.704	0.600–0.808
HDL-C	1.42	59.0	53.3	0.565	0.456–0.675
FPG	4.36	69.2	56.7	0.698	0.597–0.798
HbA1c	4.95	70.5	70.0	0.769	0.679–0.859
F IX	2598.5	66.7	53.3	0.612	0.486–0.739
TBC1D5a	853.0	67.9	56.7	0.621	0.502–0.739
Human C_k gene	200.5	69.2	63.3	0.670	0.556–0.785
ALBU*	258.5	75.6	40.0	0.641	0.532–0.750
A2MG*	254.0	78.2	33.3	0.612	0.497–0.726
HEMO*	368.5	79.5	43.3	0.690	0.583–0.796
AMBP*	73.5	74.4	36.7	0.600	0.476–0.724

*It has been reported.

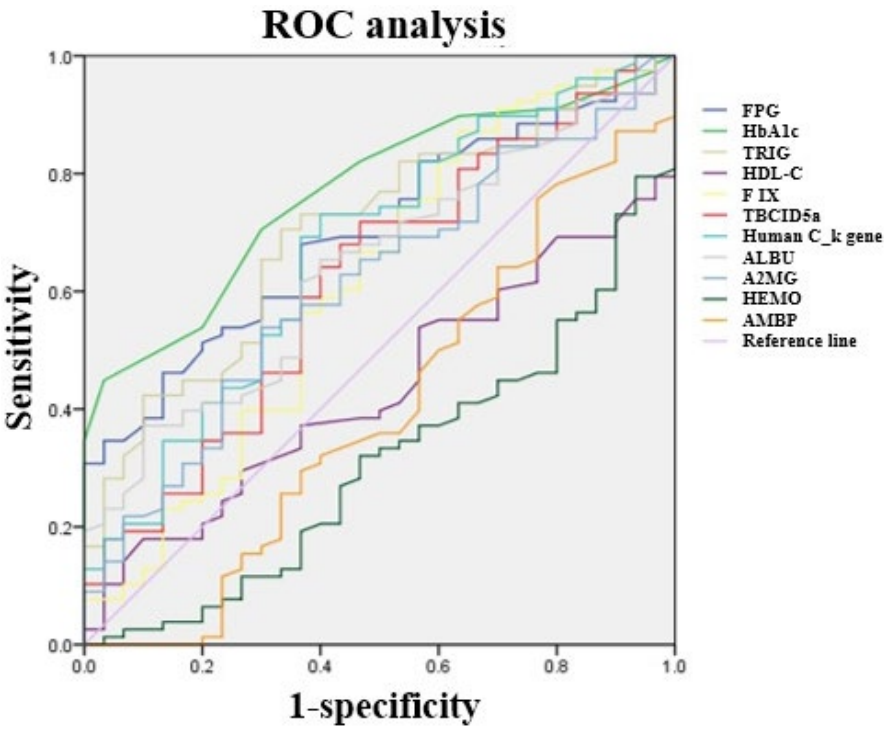

FIGURE 1 ROC analysis of the 11 GDM-related indexes

and were missed diagnoses; the classification of the N group samples showed that 30.4% of the training group samples and 28.6% of the validation group samples were incorrectly diagnosed as GDM samples, and misdiagnosis occurred.

The prediction probability of the GDM group samples was higher than 0.6, and that of the N group samples was higher than 0.4, which indicated that the probability of GDM group samples being correctly predicted by this model was higher than that of the N group samples (Figure 3A). The AUC of the diagnostic value model was 0.942 (Figure 3B). A model scoring system was used to screen cases. About 70% of GDM patients were screened out if 50% of high-score patients were included; about 90% of GDM patients could be screened out if 70% of high-score patients were included (Figure 3C, 3D).

The importance of the 11 independent variables was analyzed (Table 4). The importance of the independent variables was ranked from high to low. Among all the independent variables, HbA1c contributed the most to the model, with an importance coefficient of 0.140 (100%).

3.5 | Establishment of the radial basis function model

A two-layer radial basis function model was constructed using the radial basis function module of SPSS to classify normal pregnant women and GDM patients. There were 11 indexes in the input layer, six nodes in the hidden layer, and the activation function was

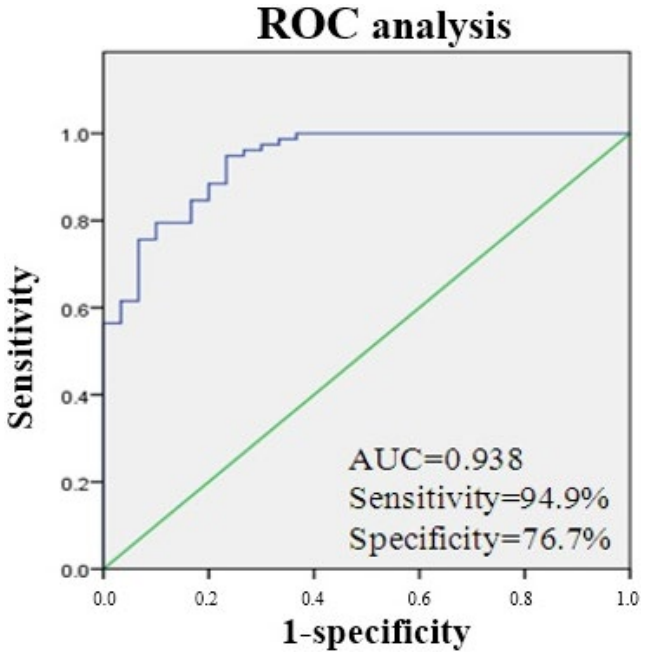

FIGURE 2 Multivariate logistic regression ROC analysis

a softmax function. The output layer had two nodes (N group and GDM group), and the activation function was the identity function; the sum of squares error function was used. Among the 108 samples, 74 (68.5%) were in the training group and 34 (31.5%) were in the validation group.

The accuracy rates of the GDM training and validation groups were 86.3% and 92.6%, respectively (Table 5). According to the classification of GDM group, 13.7% of the training group samples and 7.4% of the GDM validation group samples were misdiagnosed as the N group. A total of 34.8% of the training group samples and 28.6% of the verification group samples were misdiagnosed as GDM.

The prediction probability of the GDM group was higher than 0.6 and that of the N group was higher than 0.4, which indicated that the prediction probability of GDM patients was better than that of normal pregnant women (Figure 4A). This model is more suitable for screening diagnosis. The diagnostic value of the model showed that the AUC was 0.909 (Figure 4B). A model scoring system was used to screen cases. Nearly 60% of GDM patients were screened out if 50% of high-score patients were included; about 80% of GDM patients could be screened out if 70% of high-score patients were included (Figure 4C, 4D).

The importance of the 11 independent variables was analyzed (Table 6) and ranked from high to low. Among all the independent variables, AMBP contributed the most to the model, with an importance coefficient of 0.159 (100%).

3.6 | Establishment of the discriminant analysis function model

The discriminant functions of the 11 independent variables in the N and GDM groups were established using the SPSS discriminant analysis model. The Bayes discriminant function analysis method was used. All samples were preliminarily classified and cross-validated to clarify the diagnostic efficiency of the discriminant analysis function model.

Bayes discriminant function:

$$\begin{aligned} Y_{Ngroup} = & -68.225 - 6.988X_{FPG} + 23.351X_{HbA1c} + 0.330X_{TRIG} \\ & + 14.487X_{HDL-C} + 0.015X_{FIX} + 0.005X_{TBC1D5a} \\ & - 0.131X_{humanc\_k\ gene} + 0.015X_{ABLU} + 0.018X_{HEMO} + 0.002X_{AMBP} \end{aligned}$$

$$\begin{aligned} Y_{GDMgroup} = & -72.992 - 7.261X_{FPG} + 24.808X_{HbA1c} + 0.661X_{TRIG} \\ & + 13.766X_{HDL-C} + 0.013X_{FIX} + 0.005X_{TBC1D5a} - 0.104X_{human\ c\_k\ gene} \\ & + 0.019X_{A2MG} + 0.011X_{HEMO} - 7.86 \times 10^{-5}X_{AMBP} \end{aligned}$$

The accuracies of the GDM preliminary classification and cross-validation groups were 82.1% and 80.8%, respectively (Table 7). In the GDM group, 17.9% of the preliminary classification samples and 19.2% of the cross-validation samples were incorrectly classified as the N group, and were missed diagnoses; for the N group samples, 20% of the initial samples and 30% of the cross-validation samples were incorrectly classified as GDM.

After establishing the Bayes discriminant function of the 11 independent variables, the probability values (P values) of two new variables used to distinguish the N group and GDM group were saved, namely Dis-N and Dis-GDM. The ROC curve was

TABLE 3 The prediction classification results of multilayer perceptron neural network model

Known samples	Observation value	Predictive value		
		N group	GDM group	Accuracy
Training group *n* = 74	N group	16	7	69.6%
	GDM group	2	49	96.1%
	Overall percentage	24.3%	75.7%	87.8%
Validation group *n* = 34	N group	5	2	71.4%
	GDM group	1	26	96.3%
	Overall percentage	17.6%	82.4%	91.2%

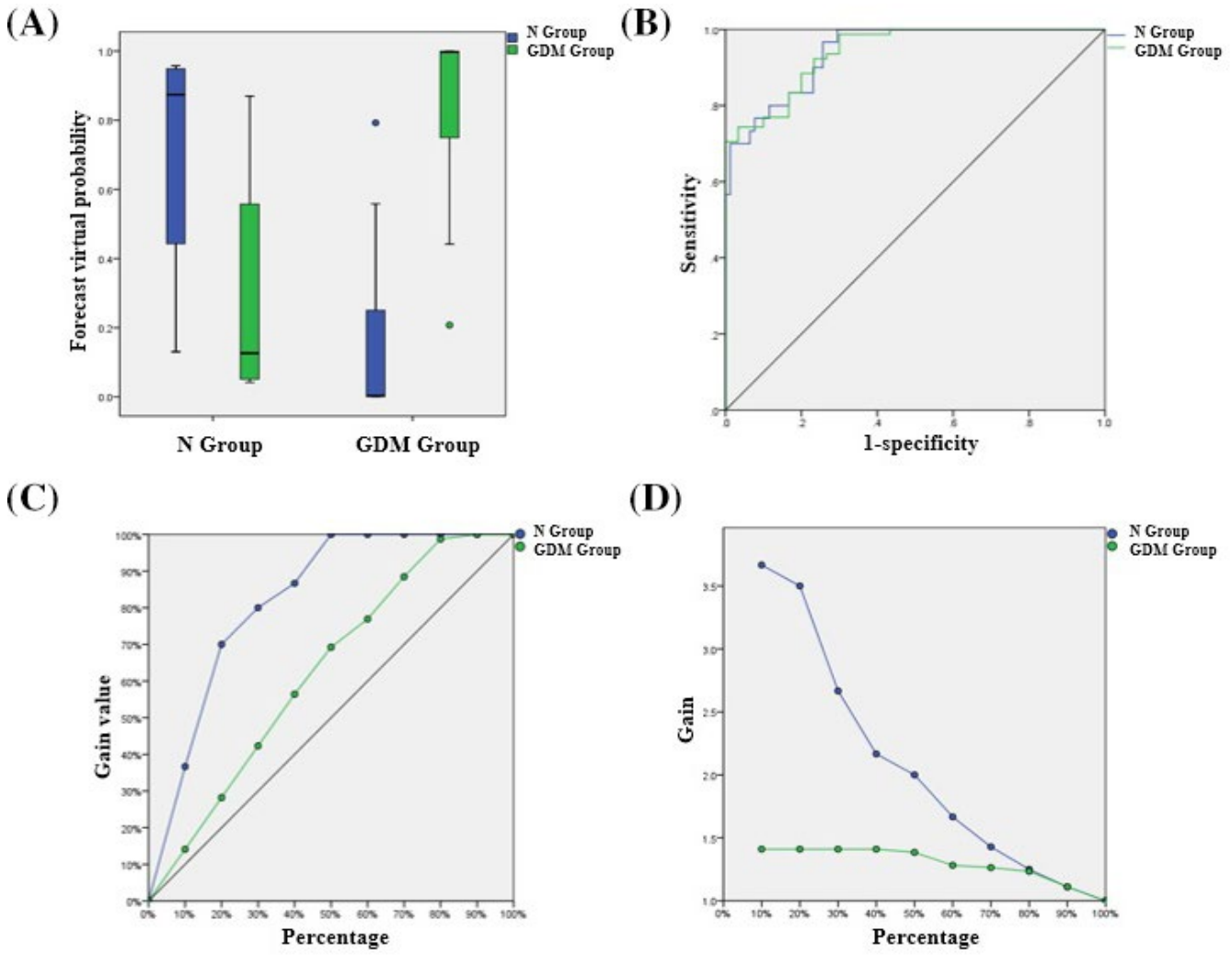

FIGURE 3 Classification statistics of the multilayer perceptron neural network model

TABLE 4 Importance of independent variables in the multilayer perceptron neural network model

Independent variable	Importance	Standardization importance
HbA1c	0.140	100%
ALBU	0.137	97.7%
FPG	0.126	89.5%
human C_k gene	0.118	83.9%
AMBP	0.113	80.6%
F IX	0.092	65.3%
HEMO	0.090	64.4%
TRIG	0.056	40.2%
A2MG	0.048	34.1%
HDL-C	0.048	33.9%
TBCID5a	0.033	23.3%

analyzed using Dis-N and Dis-GDM, and the AUC was calculated. Through the images and result analysis, it was found that the ROC curves of Dis-N and Dis-GDM used the reference line as the symmetry lines (Figure 5). The AUC of the Dis-GDM curve was 0.908 (95% CI: 0.853–0.963). When the cut-off value was 0.59, the diagnostic sensitivity and specificity were 90.0% and 78.2%, respectively.

3.7 | Delong test

The Delong test was performed using SAS 9.4 software to compare the statistical differences between different machine learning methods based on the ROC curve (Table 8). The results showed that there was no significant difference ($p < 0.05$) in the ROC curve between different machine learning methods.

TABLE 5 The prediction classification results of the radial basis function model

Known samples	Observation value	Predictive classification		
		N group	GDM group	Accuracy
Training group *n* = 74	N group	15	8	65.2%
	GDM group	7	44	86.3%
	Overall percentage	29.7%	70.3%	79.7%
Validation group *n* = 34	N group	5	2	71.4%
	GDM group	2	25	92.6%
	Overall percentage	20.6%	79.4%	88.2%

TABLE 6 Importance of independent variables in the radial basis function model

Independent variable	Importance	Standardization importance
AMBP	0.159	100%
HEMO	0.154	96.8%
ALBU	0.123	77.4%
HbA1c	0.103	64.5%
FPG	0.097	61.1%
human C_k gene	0.074	46.3%
F IX	0.066	41.3%
TRIG	0.062	38.9%
TBCID5a	0.057	35.9%
HDL-C	0.057	35.8%
A2MG	0.048	30.3%

4 | DISCUSSION

4.1 | Glucose, lipid, and protein metabolism in GDM

With the extension of pregnancy, insulin sensitivity in normal pregnant women decreases significantly (40%–80%).[10–12] Compared with women with normal glucose tolerance, GDM patients have lower insulin sensitivity, lower insulin response, and lower insulin secretion.[12] They may have decreased glucose-insulin sensitivity because of a receptor defect in peripheral tissue, most probably in skeletal muscle.[13] At the same time, GDM patients also have chronic β-cell dysfunction,[14] which manifests by insufficient compensation of the β cells for the severe acquired insulin resistance in late pregnancy. [13] β-cell defects may occur prior to pregnancy and persist after delivery. Compared with normal pregnant women, those with a history of GDM have a >7-fold risk of developing T2D.[15] GDM patients are at high risk of developing T2D in the order of 20%–60% in the first 5 years following an index pregnancy,[16,17] and the degree of decline in β-cell function varies in relation to the severity of gestational dysglycemia.[3]

GDM patients are often accompanied by lipid metabolism disorders, which are also closely related to insulin resistance.[3] In patients with poor GDM control, TRIG levels are higher in the second and third trimesters.[18] In patients with mild GDM whose FPG is lower than 105 mg/dL, the increase in TRIG levels and the decrease in TRIG metabolism may be the cause of fetal macrosomia. Dyslipidemia in GDM patients increases the risk of vascular injury, which may lead to endothelial dysfunction, and is also a pathogenic factor for preeclampsia in GDM patients.[19] Together with plasma and cell membranes, lipids are the main components of the arterial wall. Therefore, abnormal lipids may lead to vascular injury, which plays an extremely important role in the pathogenesis of GDM.[20] This study retrospectively analyzed the predictive value of TRIG and HDL-C levels in the first trimester of GDM. In agreement with the same conclusion as previous studies,[7] two indexes showed

TABLE 7 Prediction classification results of the discriminant analysis model.

Known samples		Model prediction classification		
		N group	GDM group	Accuracy
Preliminary classification	N group	24	6	80.0%
	GDM group	14	64	82.1%
	Overall percentage	35.2%	64.8%	81.5%
Cross-validation*	N group	21	9	70.0%
	GDM group	15	63	80.8%
	Overall percentage	33.3%	66.7%	77.8%

*Cross-validation was performed only for the individual case (observation) in the analysis. In cross-validation, each individual case is classified by functions derived from all cases except the individual case.

TABLE 8 Delong test

Method	ROC Area	SD	95%CI	*p* value
Multivariate logistic regression analysis	0.938	0.024	0.890–0.985	0.744
multilayer perceptron neural network model	0.942	0.021	0.900–0.984	
Multivariate logistic regression analysis	0.938	0.024	0.890–0.985	0.338
radial basis function model	0.910	0.028	0.854–0.965	
Multivariate logistic regression analysis	0.938	0.024	0.890–0.985	0.070
discriminant analysis function model	0.910	0.028	0.853–0.963	
multilayer perceptron neural network model	0.942	0.021	0.900–0.984	0.271
radial basis function model	0.910	0.028	0.854–0.965	
multilayer perceptron neural network model	0.942	0.021	0.900–0.984	0.091
discriminant analysis function model	0.910	0.028	0.853–0.963	
radial basis function model	0.910	0.028	0.854–0.965	0.966
discriminant analysis function model	0.910	0.028	0.853–0.963	

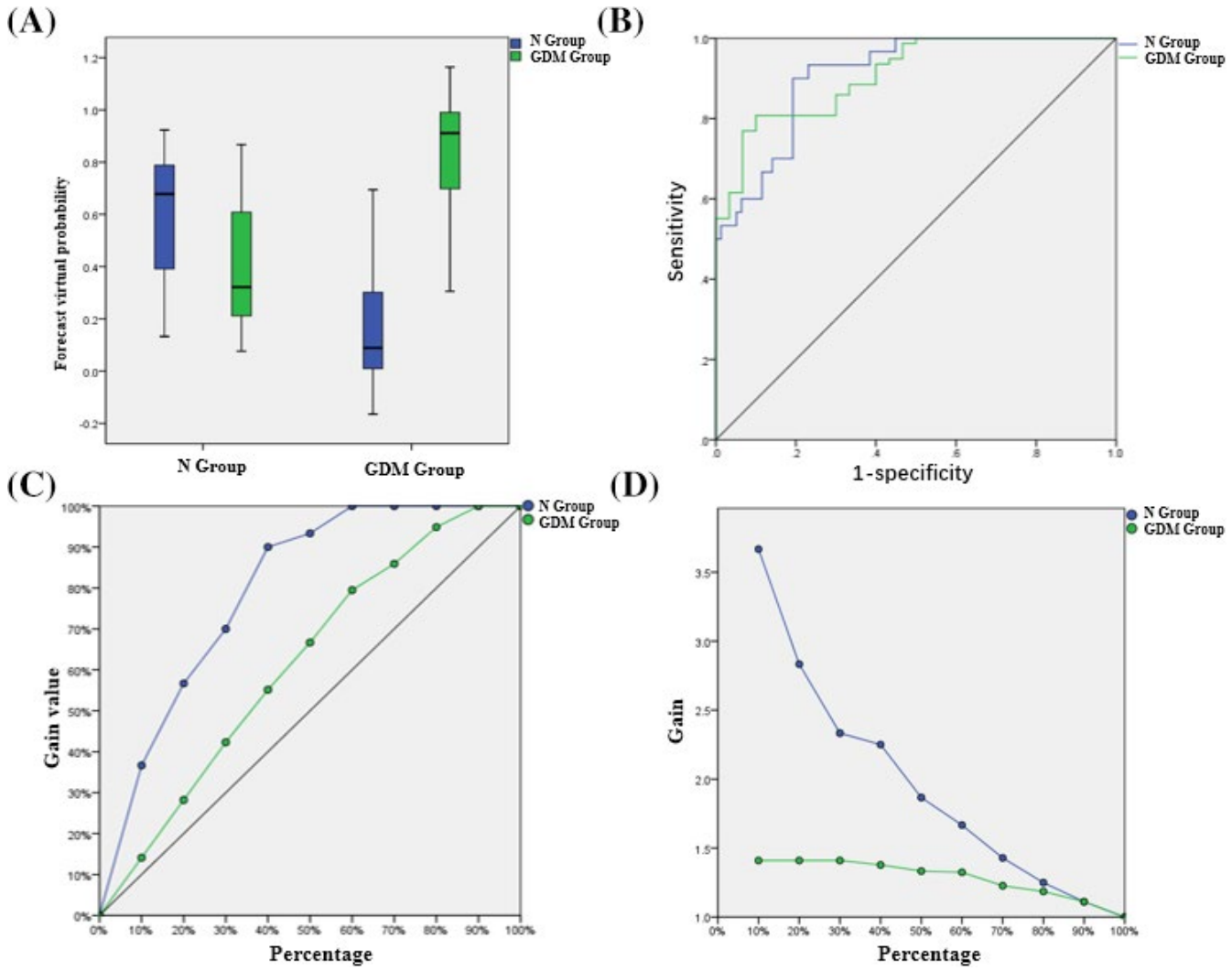

FIGURE 4 Classification statistics of the radial basis function model

significant differences between the N group and GDM group, and the TRIG level in the GDM group was significantly higher than that in the N group. Moreover, the HDL-C level in the GDM group was significantly lower than that in the N group.

Studies on GDM and protein metabolism have indicated that GDM causes elevated plasma amino acids, especially BCAAs, which cannot be synthesized by the human body and must be obtained mainly from the diet. [21,22] The increased circulating concentrations

of BCAAs directly affect the development of insulin resistance and T2D, possibly by increasing the presence of toxic BCAA intermediate metabolites, which in turn interferes with β-cell mitochondrial function.[23] Therefore, BCCAs may be potential biomarkers of GDM, and plasma BCAAs can serve as reliable biomarkers for the prevention, early diagnosis, and treatment of diabetes.[24] Similarly, urine-excreted metabolites, such as those related to the serotonin system, non-polar amino acids, and ketone bodies, may complete a predictive or early diagnostic panel of biomarkers for GDM.[25] Studies have shown that urinary tryptophan metabolism-related metabolites were significantly increased in the third trimester in women with GDM.[26] The upregulation of these pathways could trigger insulin resistance and may respond to oxidative stress and inflammation during GDM.[25] Our previous study on urine peptides in GDM patients[8,9] also indicated that the changes in glucose metabolism had an impact on protein and amino acid metabolism, which can be reflected in urine. The relationship and mechanism between urinary polypeptides and GDM have been detailed in a previous paper.

Therefore, better metabolic control, including blood lipid regulation, and blood glucose and protein intake control, may play a preventive role in the occurrence of adverse prognoses in mothers and infants.[27–30]

4.2 | ROC analysis of GDM-related indexes

The ROC curve was used to analyze the clinical diagnostic value of 11 GDM-related indexes. The indexes included TRIG, HDL-C, FPG, HbA1c, F IX, TBCID5a, human C_k gene, ALBU, A2MG, HEMO, and AMBP. Among all the indexes, HbA1c had the highest diagnostic value, with an AUC of 0.769. When the cut-off value was 4.95, the sensitivity and specificity of diagnosis were 70.5% and 70.0%, respectively. Among the seven GDM-related urinary polypeptides, HEMO had the highest diagnostic value for GDM, with an AUC of 0.690. When the cut-off value was 368.5, the sensitivity and specificity were 79.5% and 43.3%, respectively. The other six urinary polypeptides were human C_k gene, ALBU, TBCID5a, F IX, A2MG, and AMBP in descending order.

4.3 | Clinical value of the diagnostic model

Binary logistic regression analysis, multilayer perceptron neural network model, radial basis function model, and discriminant analysis function model were established to comprehensively evaluate the diagnostic value of the 11 indexes in distinguishing the N and GDM groups. Among the four multi-parameter diagnostic models, the AUC of the multilayer perceptron neural network model was 0.942, followed by binary logistic regression (0.938), radial basis function model (0.909), and discriminant analysis function model (0.908). The results showed that the multilayer perceptron neural network model had the greatest diagnostic value for GDM. In the process of establishing the models, all training samples (or initial samples) and validation samples were comprehensively analyzed. The diagnostic accuracy of the multilayer perceptron neural network model for the N and GDM groups were 70% and 96.2%, respectively, which was higher than that of the other two diagnostic models (the difference was not statistically significant). The diagnostic accuracy of the three models for GDM patients was higher than that of normal pregnant women. Moreover, the sensitivity of disease diagnosis was high, which was suitable for preliminary clinical screening of all pregnant women. However, even with the deep learning approach,

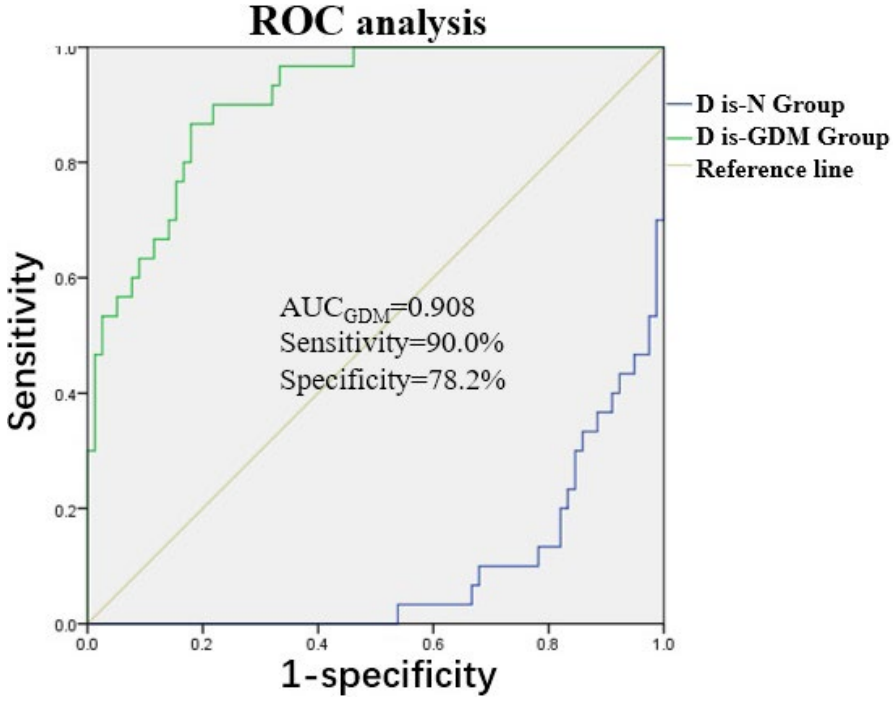

FIGURE 5 ROC analysis of the discriminant analysis model

the models still had low specificity. On the follow-up study, we will continue to improve the model by expanding the number of samples, introducing new parameters closely related to the disease, and refining the influencing factors.

Compared with single index analysis, the multi-index comprehensive analysis method can use the diagnostic advantages of the many parameters, and increase the sensitivity and specificity of disease diagnosis. The establishment of diagnostic models will be helpful for early screening, prevention, and judging the prognosis of GDM, as well as facilitate clinical application and improve the prognosis of patients. In the absence of acknowledged GDM guidelines, research on intelligent diagnostic models is expected to provide a new diagnosis and treatment model for GDM prevention and judgment of prognosis.

Compared with traditional statistical methods, machine learning methods, especially deep learning methods, have a strong ability for feature extraction and disease prediction. The performance of deep learning methods in some complex tasks has reached or even exceeded the level of human decision-making, and it is increasingly used in medical diagnosis, emotion analysis, and other medical fields.[31,32] However, the shortcomings of machine learning methods are also very clear. For example, the logistic regression method requires strict assumptions, which are easy to under-fit, and the classification accuracy is not high; the discriminant classification method tends to over-fit easily, and its effect is not good if the sample attributes are related. The deep learning method requires a lot of data for training, and the model is in a black box state, making it difficult to understand the internal mechanism. Therefore, it is necessary to clarify the advantages and disadvantages of different machine learning methods and run them correctly in order to achieve the expected goal of the research.

5 | CONCLUSIONS

Through this study, we explored the diagnostic value of GDM-related indexes, established the GDM diagnosis model, and comprehensively analyzed the relationship between glucose metabolism, lipid metabolism, and urine polypeptides in GDM patients. The GDM-related urine molecular peptide results, combined with FPG and other clinical indexes, are expected to serve as the basis for the development of urine glucose metabolism detection kits for GDM patients. The establishment of diagnostic models can lay the foundation for exploration of artificial intelligence and machine language in diagnostic systems. The study also provided the possibility of early prediction and monitoring of glucose metabolism in patients with GDM. Comprehensive analysis of the relationship between glucose metabolism, lipid metabolism, and urinary polypeptides in GDM patients can also provide support for disease-related pathophysiological research.

ACKNOWLEDGMENTS

We gratefully acknowledge all the volunteers for their generous donation of urine and serum specimens, as well as the staff at the Clinical Laboratory Medicine of Beijing Shijitan Hospital, Capital Medical University, for their enthusiastic assistance.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Z Hu analyzed the data and wrote the manuscript. M Zhang designed the study and drafted the manuscript. All authors reviewed, edited, and approved the submitted manuscript.

DATA AVAILABILITY STATEMENT

Available at https://doi.org/10.6084/m9.figshare.14214380.v1

ORCID

Zhiying Hu https://orcid.org/0000-0002-6186-051X
Man Zhang https://orcid.org/0000-0001-5166-6804

REFERENCES

1. Drobny J. Metabolic syndrome and the risk of pre-eclampsia. *Bratisl Lek Listy*. 2009;110(7):401-403.
2. Sun Y, Yang H, Sun WJ. Risk factors for pre-eclampsia in pregnant Chinese women with abnormal glucose metabolism. *Int Gynaecol Obstet*. 2008;101(1):74-76.
3. Retnakaran R, Qi Y, Sermer M, et al. Beta-cell function declines within the first year postpartum in women with recent glucose intolerance in pregnancy. *Diabetes Care*. 2010;33(8):1798-1804.
4. Barbour LA. Metabolic Culprits in obese pregnancies and gestational diabetes mellitus: big babies, big twists, big picture: the 2018 norbert freinkel award lecture. *Diabetes Care*. 2019;42(5):718-726.
5. Nordin NM, Wei JWH, Naing NN, et al. Comparison of maternal-fetal outcomes in gestational diabetes and lesser degrees of glucose intolerance. *J Obstet Gynaecol Res*. 2006;32(1):107-114.
6. Tobias DK, Clish C, Mora S, et al. Dietary intakes and circulating concentrations of branched-chain amino acids in relation to incident type 2 diabetes risk among high-risk women with a history of gestational diabetes mellitus. *Clin Chem*. 2018;64(8):1203-1210.
7. Hu ZY, Hu M, Han J, et al. Clinical value of blood lipid metabolism in patients with gestational diabetes mellitus in early pregnancy. *Labeled Immunoassays & Clin Med*. 2019;26(9):1476-1479. http://en.cnki.com.cn/Article_en/CJFDTotal-BJMY201909007.htm
8. Hu ZY, Tian YP, Li J, et al. Urinary peptides associated closely with gestational diabetes mellitus. *Dis Markers*. 2020;2020:1-11. eCollection.
9. Hu ZY, Zhang M, Tian YP. Screening and analysis of small molecular peptides in urine of gestational diabetes mellitus. *Clin Chim Acta*. 2020;2020(502):174-182.
10. Sivan E, Chen X, Homko CJ, et al. Longitudinal study of carbohydrate metabolism in healthy obese pregnant women. *Diabetes Care*. 1997;20(9):1470-1475.
11. Buchanan TA, Metzger BE, Freinkel N, et al. Insulin sensitivity and b-cell responsiveness to glucose during late pregnancy in lean and moderately obese women with normal glucose tolerance or mild gestational diabetes. *Am J Obstet Gynecol*. 1990;162(4):1008-1014.
12. Ryan EA, O'Sullivan MJ, Skyler JS. Insulin action during pregnancy: studies with the euglycemic clamp technique. *Diabetes*. 1985;34(4):380-389.
13. Catalano PM, Huston L, Amini SB, et al. Longitudinal changes in glucose metabolism during pregnancy in obese women with normal

glucose tolerance and gestational diabetes mellitus. *Am J Obstet Gynecol*. 1999;180(4):903-916.
14. Buchanan TA, Xiang AH. Gestational diabetes mellitus. *J Clin Invest*. 2005;115(3):485-491.
15. Bellamy L, Casas JP, Hingorani AD, et al. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet*. 2009;373(9677):1773-1779.
16. Feig DS, Zinman B, Wang X, et al. Risk of development of diabetes mellitus after diagnosis of gestational diabetes. *CMAJ*. 2008;179(3):229-234.
17. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care*. 2002;25(10):1862-1868.
18. Teliga-Czajkowska J, Sienko J, Zareba-Szczudlik J, et al. Influence of glycemic control on coagulation and lipid metabolism in pregnancies complicated by pregestational and gestational diabetes mellitus. *Adv Exp Med Biol*. 2019;1176:81-88.
19. Gaillard R, Durmuş B, Hofman A, et al. Risk factors and outcomes of maternal obesity and excessive weight gain during pregnancy. *Obesity (Silver Spring)*. 2013;21(5):1046-1055.
20. Guariguata L, Hambleton WI, Beagley J, et al. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract*. 2014;103(2):137-149.
21. Park S, Park JY, Lee JH, et al. Plasma levels of lysine, tyrosine, and valine during pregnancy are independent risk factors of insulin resistance and gestational diabetes. *Metab Syndr Relat Disord*. 2015;13(2):64-70.
22. Rahimi N, Razi F, Nasli-Esfahani E, et al. Amino acid profiling in the gestational diabetes mellitus. *J Diabetes Metab Disord*. 2017;16:13.
23. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat Rev Endocrinol*. 2014;10(12):723-736.
24. Zhao L, Wang M, Li J, et al. Association of circulating branched-chain amino acids with gestational diabetes mellitus: a meta-analysis. *Int J Endocrinol Metab*. 2019;17(3):e85413.
25. Lorenzo-Almorós A, Hang T, Peiró C, et al. Predictive and diagnostic biomarkers for gestational diabetes and its associated metabolic and cardiovascular diseases. *Cardiovasc Diabetol*. 2019;18(1):140.
26. López-Hernández Y, Herrera-Van Oostdam AS, Toro-Ortiz JC, et al. Urinary metabolites altered during the third trimester in pregnancies complicated by gestational diabetes mellitus: relationship with potential upcoming metabolic disorders. *Int J Mol Sci*. 2019;20(5):1186.
27. Ersanli ZO, Damci T, Sen C, et al. Lipid metabolism alterations in patients with gestational diabetes mellitus associated fetal macrosomia. *Annlst Super Sanita*. 1997;33(3):411-415.
28. Crowther CA, Hiller JE, Moss JR, et al. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med*. 2005;352(24):2477-2486.
29. Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med*. 2009;361(14):1339-1348.
30. Cosson E, Vicaut E, Sandre-Banon D, et al. Initially untreated fasting hyperglycaemia in early pregnancy: prognosis according to occurrence of gestational diabetes mellitus after 22 weeks' gestation: a case-control study. *Diabetic Med*. 2020;37(1):123-130.
31. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci Biobehav Rev*. 2017;74(Pt A):58-75. https://doi.org/10.1016/j.neubiorev.2017.01.002. Epub 2017 Jan 10.
32. Zhang Y, Wang C, Guo WY, et al. Multi-source news comment sentiment prediction based on two-way hierarchical semantic model (in Chinese). *J Computer Res Development*. 2018;55(5):933-944.

How to cite this article: Hu Z, Zhang M. Establishment of clinical diagnostic models using glucose, lipid, and urinary polypeptides in gestational diabetes mellitus. *J Clin Lab Anal*. 2021;35:e23833. https://doi.org/10.1002/jcla.23833