The Journal of Maternal-Fetal & Neonatal Medicine

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ijmf20

Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques

Yan Xiong , Lu Lin , Yu Chen , Stephen Salerno , Yi Li , Xiaoxi Zeng & Huafeng Li

To cite this article: Yan Xiong , Lu Lin , Yu Chen , Stephen Salerno , Yi Li , Xiaoxi Zeng & Huafeng Li (2020): Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques, The Journal of Maternal-Fetal & Neonatal Medicine, DOI: 10.1080/14767058.2020.1786517

To link to this article: https://doi.org/10.1080/14767058.2020.1786517

Published online: 06 Aug 2020.

THE JOURNAL OF MATERNAL-FETAL & NEONATAL MEDICINE
https://doi.org/10.1080/14767058.2020.1786517

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques

Yan Xiong[a]*, Lu Lin[a]*, Yu Chen[b,c], Stephen Salerno[d], Yi Li[d], Xiaoxi Zeng[c] and Huafeng Li[e]

[a]College of Mechanical Engineering, Sichuan University, Chengdu, China; [b]Department of Applied Mechanics, Sichuan University, Chengdu, China; [c]Medical Big Data Center, Sichuan University, Chengdu, China; [d]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; [e]West China Second University Hospital, Sichuan University, Chengdu, China

ABSTRACT

Aim: Our objective was to develop a first 19 weeks risk prediction model with several potential gestational diabetes mellitus (GDM) predictors including hepatic and renal and coagulation function measures.

Methods: A total of 490 pregnant women, 215 with GDM and 275 controls, participated in this case-control study. Forty-three blood examination indexes including blood routine, hepatic and renal function, and coagulation function were obtained. Support vector machine (SVM) and light gradient boosting machine (lightGBM) were applied to estimate possible associations with GDM and build the predict model. Cutoff points were estimated using receiver operating characteristic curve analysis.

Results: It was observed that a cutoff of Prothrombin time (PAT-PT) and Activated partial thromboplastin time (PAT-APTT) could reliably predict GDM with sensitivity of 88.3% and specificity of 99.47% (AUC of 94.2%). If we only use hepatic and renal function examination, a cutoff of DBIL and FPG with sensitivity of 82.6% and specificity of 90.0% (AUC of 91.0%) was obvious and a negative correlation with PAT-PT (r=–0.430549) and patient activated partial thromboplastin time (PAT-APTT) (r=–0.725638). A negative correlation with direct bilirubin (DBIL) (r=–0.379882) and positive correlation with fasting plasma glucose (FPG) (r = 0.458332) neglect coagulation function examination.

Conclusion: The results of this study point out the possible roles of PAT-PT and PAT-APTT as potential novel biomarkers for the prediction and earlier diagnosis of GDM. A first 19 weeks risk prediction model, which incorporates novel biomarkers, accurately identifies women at high risk of GDM, and relevant measures can be applied early to achieve the prevention and control effects.

ARTICLE HISTORY
Received 22 December 2019
Revised 17 April 2020
Accepted 19 June 2020

KEYWORDS
First trimester; gestational diabetes mellitus; machine learning; screening; prediction

1. Introduction

Gestational diabetes mellitus (GDM) is one of the most common complications observed in pregnant women. GDM can lead to preterm labor, pre-eclampsia, nephropathy, birth trauma, cesarean section and subsequent impaired wound healing, among others [1]. Additionally, GDM imparts an increased health risk to both the mother and fetus. There is a higher prevalence of obesity and metabolic syndrome in children born to GDM mothers and a higher maternal risk for developing type 2 diabetes mellitus in the future [2]. Currently, screening for GDM is done at 24–28 weeks gestation, however, this often does not provide sufficient time for adequate intervention [3]. Thus, earlier identification of women at risk for GDM may allow for earlier diagnosis and consequent reduction in associated comorbidities [3].

Recently, there has been a growing interest in the discovery and validation of GDM-associated biomarkers. For example, one study found sex hormone binding globulin (SHBG) levels to be predictive of type 2 diabetes mellitus in women. It was shown that a threshold of 50 nmol/L of SHBG had a 90% sensitivity and a 96% specificity rate in diagnosing GDM [4]. Gözükara et al. found age, total testosterone, and body mass index (BMI) to be independent predictors of GDM, with a sensitivity of 63.6% and a specificity of 62.7% for a total testosterone level of 0.45 ng/mL [3]. Mierzyński et al. further found nesfatin-1 and vaspin

CONTACT Yu Chen yu_chen@scu.edu.cn Department of Applied Mechanics, Sichuan University, Sichuan, Chengdu, China; Huafeng Li lihuafeng@scu.edu.cn West China Second University Hospital, Sichuan University, Chengdu, China
*Both authors contributed equally to this work.

to be potential novel biomarkers of GDM [5]. Riskin-Mashiah et al. observed fasting glucose levels of 80–85 mg/dL yielded sensitivities of 75–55% and specificities of 52–75% for GDM prediction [6]. Other studies suggest that first trimester serum uric acid levels, hyperinsulinemia, serum secreted frizzled-related protein-5 levels, and the 75 g oral glucose tolerance test are predictive of GDM and impaired glucose tolerance [7–10]. During the second trimester, unconjugated estriol in the maternal serum is a useful predictor of GDM development [11].

A common theme in the above studies is that they examined one or two biomarkers in isolation, rather than studying the predictive capacity of several potential biomarkers simultaneously. In contrast, multivariable models may have greater predictive power than these univariate indicators. For example, one study found that both elevated tissue plasminogen activator and low HDL cholesterol levels were significantly associated with GDM development [12]. Isik et al. studied the prediction of GDM during the first trimester using the Homeostatic Model Assessment for Insulin Resistance (HOMA), BMI, and the waist-to-hip ratio [13]. Another study found the HOMA, SHBG, triglycerides, and LDL cholesterol levels to be predictors of GDM in low-risk pregnancies [14]. Rasanen et al. found that first-trimester serum concentrations of glycosylated fibronectin, adiponectin, high-sensitivity C-reactive protein, and placental lactogen were significantly associated with GDM [15]. Lastly, Arianne et al. found that family history of diabetes, previous GDM, South/East Asian ethnicity, parity, BMI, pregnancy-associated plasma protein-A, triglycerides, and lipocalin-2 levels were also predictive of GDM [16]. The aim of this study is the early detection of GDM using routine laboratory examination measures. We illustrate a new screening approach for GDM by utilizing measures from the recommended maternal health examination in China at10–19 weeks gestation. We then discuss the relationship between GDM and these hematologic/biochemical parameters to assess their potential role in the prediction of GDM.

2. Materials and methods

2.1. Study design and population

This is a case-control study of pregnant women presenting to the West China Second University Hospital between December 2016 and December 2018. Participants who provided blood samples at 10–19 weeks gestation, completed routine prenatal care, and delivered a viable, full-term infant at our institution were included in the study. A total of 44 hematologic and biochemical parameters from routine blood tests, hepatic and renal function, and coagulation function examinations were collected. These predictors, in addition to demographic characteristics, were taken from maternal medical records (see Table 1).

Our outcome variable was an indicator of GDM, whereby each record was labeled as (1) for GDM cases and (0) for controls. Cases were considered eligible for inclusion if GDMwas diagnosed before or during index pregnancy and if the routine Chinese maternal health exam was completed at the West China Second University Hospital. GDM cases were diagnosed according to the World Health Organization (WHO) [17] and Adults Treatment Panel III guidelines [18] ($\geq$7.0 mmol/L for fasting glucose and/or $\geq$7.8 mmol/L for 2-h post-glucose) during the second trimester (26–28 weeks) gestation.

The medical records were also screened for healthy pregnant women (controls). The exclusion criteria were as follows: (1) preexisting diabetes; (2) a twin pregnancy; (3) a previous history of hypertension or heart disease; (4) a history of other severe metabolic disease such as an autoimmune disease. All women had phenotypically normal neonates. This study was approved by the corresponding Hospital Ethics Committee (Table 1).

2.2. Statistical analysis

All analyses were performed using SPSS (the Statistical Package for the Social Sciences) for Windows, version 16. Descriptive statistics for the data were presented as means and standard deviations, stratified by GDM status. Bivariate tests of unadjusted associations between each predictor and GDM status were carried out via Student's unpaired *t*-test or the Mann–Whitney *U*-test. Correlations between predictor variables were estimated via Pearson's correlation. Fisher's exact test was performed when appropriate. A Bonferroni correction was used to adjust the significance level for multiple comparisons. The test was considered significant with a p value less than 0.001. Results from these bivariate tests, clinical diagnoses, and previous literature were used to preliminarily screen the complete set of variables for potential important predictors of GDM.

2.3. Predictive model

Predictive modeling was carried out via (1) support vector machines (SVMs) and (2) generalized boosted

Table 1. List of variables collected from routine blood, hepatic and renal function, and coagulation function examinations.

Routine Blood Examination	Hepatic and renal function examination
Measured variables (abbreviation, SI)	*Measured variables (abbreviation, SI)*
White blood cell count (WBC, $\times 10^9$/L)	Alanine aminotransferase (ALT, U/L)
Percent neutrophils (NEUT, %)	Aspartate aminotransferase (AST, U/L)
Percent monocytes (MONO, %)	Total bilirubin (TB, μmol/L)
Percent lymphocytes (LYMPH, %)	Direct bilirubin (DBIL, μmol/L)
Percent eosinophils (EOS, %)	Total protein (TP, g/L)
Percent basophils (BASO, %)	Albumin (ALB, g/L)
Red blood cell count (RBC, $\times 10^{12}$/L)	Glutamyl transpeptidase (r-GT, U/L)
Hemoglobin (HGB, g/L)	Lactate dehydrogenase (LDH, U/L)
Hematocrit (HCT, %)	Urea (UN, mmol/L)
Platelets (PLT, $\times 10^9$/L)	Creatinine (Cr, μmol/L)
Mean platelet volume (MPV, fL)	Fasting plasma glucose (FPG, mmol/L)
Mean corpuscular volume (MCV, fL)	*Calculated variables (abbreviation, SI)*
Mean corpuscular hemoglobin (MCH, pg)	Indirect bilirubin (IDIL, μmol/L)
Mean corpuscular hemoglobin concentration (MCHC, g/L)	Globulin (GLB, g/L)
Red cell distribution width-CV (RDW-CV, %)	**Coagulation function examination**
Red cell distribution width-SD (RDW-SD, fL)	*Measured variables (abbreviation, SI)*
Platelet distribution width (PDW, fL)	*Prothrombin time*
Platelet large cell ratio (P-LCR, %)	Patient-PT (PAT-PT, s)
Plateletocrit (PCT, %)	Reference-PT (REF-PT, s)
Calculated variables (abbreviation, SI)	*Activated partial thromboplastin time*
Neutrophil count (NEUT, $\times 10^9$/L)	Patient-APTT (PAT-APTT, s)
Monocyte count (MONO, $\times 10^9$/L)	Reference-APTT (REF-APTT, s)
Lymphocyte count (LYMPH, $\times 10^9$/L)	
Eosinophil count (EOS, $\times 10^9$/L)	Fibrinogen (Fg, mg/dL)
Basophil count (BASO, $\times 10^9$/L)	Thrombin time (TT, s)

modeling, which have faster speed and lower cost compared to experimental bioassays. One of its most important features is the calculation of the variable importance, which measures the association between a given variable and the accuracy of the prediction, based on the percentage of increase in the mean square-error [19].

SVM maps the feature data which is not separable to a higher dimensional space, makes these features becomes linear separable in this space and then uses a maximum boundary interval hyperplane to classify them.

Light gradient boosting machine (lightGBM) is a framework to implement GBDT (Gradient Boosting Decision Tree), and GBDT is an algorithm which is based on iterative accumulation of many decision trees. It constructs a group of weak learners (decision trees) to form a strong learner, and accumulates the results of multiple decision trees as the final prediction output. Based on GBDT, lightGBM has many improvements on algorithm and has more advantages, such as faster training speed, lower memory consumption, better accuracy and fast processing of massive data.

For the purposes of this analysis, the full study population was divided into two groups. Specifically, the original dataset was divided 70% into a training set, in order to generate a predictive model for GDM, and 30% into a testing set validate the model. After splitting the data, it was ensured that the proportion of GDM patients to healthy persons in the testing and the training sets were the same as the original data set.

The predictive models were first developed on the training datasets and then to predict the GDM status of patients in the testing dataset. The fitted models were then applied to the testing datasets, which cross-validated by *k*-fold cross validation. Predictive accuracy was evaluated via the area under the receiver operating characteristic curves (AUCs) for these methods.

3. Results

3.1. Statistic results

13,329 medical records were obtained on pregnant women presenting to the West China Second University Hospital from December 2016 to December 2018. These records were first filtered and preprocessed. After removing patients who did not meet the inclusion criteria or did not have complete medical records, 215 cases and 275 healthy controls were carried forwardas the study sample.

Table 2 reports descriptive statistics for the porential predictors, stratified by case status. As shown, GDM cases had significantly higher

percent monocytes (MONO), percent basophils (BASO), platelets (PLT), plateletocrit (PCT), and monocyte count (MONO), but significantly lower mean corpuscular hemoglobin concentration (MCHC) ($p < .001$ for all unadjusted comparisons). For all parameters

Table 2. Comparison of routine blood work examination, hepatic and renal function examination and coagulation function examination between 215 GDM cases and 275 healthy controls.

Variables	GDM Cases ($n = 215$)	Controls ($n = 275$)	p Value
Measured variables (abbreviation), SI	**Mean (±s)**	**Mean (±s)**	
Routine Blood Examination			
White blood cell count (WBC), $\times 10^9$/L	8.58 ± 1.78	8.27 ± 1.68	.044
Percent neutrophils (NEUT), %	74.74 ± 6.16	74.80 ± 4.84	.906
Percent lymphocytes (LYMPH), %	19.06 ± 7.02	19.55 ± 4.37	.367
Percent monocytes (MONO), %	5.17 ± 1.33	4.44 ± 1.17	<.001
Percent eosinophils (EOS), %	1.08 ± 0.98	1.02 ± 0.85	.476
Percent basophils (BASO), %	0.22 ± 0.13	0.18 ± 0.12	.001
Red blood cell count (RBC), $\times 10^{12}$/L	4.17 ± 0.37	4.14 ± 0.33	.455
Hemoglobin (HGB), g/L	125.45 ± 10.53	126.60 ± 8.95	.188
Hematocrit (HCT), %	37.05 ± 2.63	36.70 ± 2.52	.138
Mean corpuscular volume (MCV), fL	89.21 ± 5.51	88.70 ± 3.10	.224
Mean corpuscular hemoglobin (MCH), pg	30.44 ± 4.59	30.60 ± 1.15	.615
Mean corpuscular hemoglobin concentration (MCHC), g/L	338.02 ± 9.75	344.96 ± 7.22	<.001
Red cell distribution width-CV (RDW-CV), %	14.06 ± 9.01	13.47 ± 0.75	.336
Red cell distribution width-SD (RDW-SD), fL	43.56 ± 3.42	43.49 ± 2.89	.804
Platelets (PLT), $\times 10^9$/L	215.92 ± 62.85	198.55 ± 46.36	.001
Platelet distribution width (PDW), fL	14.29 ± 3.44	14.03 ± 2.81	.379
Platelet large cell ratio (P-LCR), %	35.53 ± 11.50	34.41 ± 9.27	.244
Mean platelet volume (MPV), fL	11.29 ± 1.40	11.19 ± 1.17	.379
Plateletocrit (PCT), %	0.24 ± 0.06	0.22 ± 0.04	<.001
Calculation variables (abbreviation), SI	Mean (±s)	Mean (±s)	
Neutrophil count (NEUT), $\times 10^9$/L	6.45 ± 1.55	6.22 ± 1.46	.093
Monocyte count (MONO), $\times 10^9$/L	0.44 ± 0.14	0.37 ± 0.12	<.001
Lymphocyte count (LYMPH), $\times 10^9$/L	1.59 ± 0.50	1.59 ± 0.38	.956
Eosinophil count (EOS), $\times 10^9$/L	0.09 ± 0.09	0.08 ± 0.07	.134
Basophil count (BASO), $\times 10^9$/L	0.02 ± 0.01	0.02 ± 0.01	.001
Hepatic and Renal function examination			
Alanine aminotransferase (ALT), U/L	25.40 ± 20.48	21.10 ± 15.85	.011
Aspartate aminotransferase (AST), U/L	22.94 ± 10.86	21.88 ± 8.18	.232
Total bilirubin (TB), μmol/L	9.51 ± 3.41	10.79 ± 3.35	<.001
Direct bilirubin (DBIL), μmol/L	2.45 ± 0.97	3.37 ± 1.19	<.001
Total protein (TP), g/L	71.23 ± 4.30	72.42 ± 3.90	.001
Albumin (ALB), g/L	42.47 ± 8.20	43.30 ± 3.52	.163
Glutamyl transpeptidase (r-GT), U/L	18.60 ± 14.19	14.65 ± 8.80	<.001
Lactate dehydrogenase (LDH), U/L	154.56 ± 22.46	162.97 ± 25.32	<.001
Urea (UN), mmol/L	3.13 ± 0.68	3.17 ± 0.73	.600
Creatinine (Cr), μmol/L	41.62 ± 5.20	43.17 ± 5.43	.002
Fasting plasma glucose (FPG), mmol/L	4.77 ± 0.51	4.34 ± 0.02	<.001
Calculation variables (abbreviation), SI	Mean (±s)	Mean (±s)	
Indirect bilirubin (IDIL), μmol/L	7.07 ± 2.59	7.42 ± 2.42	.123
Globulin (GLB), g/L	29.71 ± 2.87	29.02 ± 2.67	.006
Coagulation Function Examination			
Fibrinogen (Fg), mg/dL	401.55 ± 66.34	377.38 ± 70.30	<.001
Thrombin time (TT), s	16.40 ± 0.62	16.46 ± 0.53	.274
Prothrombin time			
PAT-PT, s (PT)	11.01 ± 0.56	11.61 ± 0.67	<.001
REF-PT, s	12.14 ± 0.34	11.96 ± 0.34	<.001
Activated partial thromboplastin time			
PAT-APTT, s	25.80 ± 1.93	32.01 ± 3.49	<.001
REF-APTT, s	27.54 ± 0.23	31.45 ± 1.07	<.001

higher counts/concentrations suggest a greater risk of developing GDM.

In examining the hepatic and renal function variables, it was shown that GDM cases had higher glutamyl transpeptidase (r-GT), fasting plasma glucose (FPG), and fibrinogen (Fg) levels ($p < .001$). The opposite phenomenon was shown for total bilirubin (TB), direct bilirubin (DBIL), total protein (TP), and lactate dehydrogenase (LDH) levels ($p < .001$). Higher values of these parameters are also indicative of a greater the likelihood of developing GDM.

Despite significant differences in FPG for two groups, mean FPG levels were within the normal range in both the GDM and control groups. Therefore, early prediction based on other indicators is very important. It means early screening based on FPG levels is not sufficient. Combined with coagulation function examination indices, the early prediction would be relatively easy. And for coagulation function examination indices only the thrombin time (TT) level did not differ between the two groups, the other coagulation indices have significant difference in coagulation function examination. First 19 weeks coagulation function examination variables including PAT-PT, patient activated partial thromboplastin time (PAT-APTT) and reference activated partial thromboplastin time

(REF-APTT) level were significantly lower among GDM cases as compared to healthy controls. .

3.2. ROC curve

The ROC curves of GDM with/without coagulation are shown in Figure 1. An ROC was used to determine whether PAT-PT and REF-APTT level alone may be used to predict the development of GDM with sensitivity of 92.5% and a specificity of 99.2%. The AUC was found to be 0.998; a DBIL and FPG value of hepatic and renal function examination was found to predict development of GDM with sensitivity of 77.6% and a specificity of 89.3% without coagulation function examination. The AUC was found to be 0.878.

We determined cutoff values based on clinical information for the early screening. We stratified training set ($n = 490$) into two groups including observation group ($n = 215$, GDM patients) and control group ($n = 275$, healthy pregnant women). The AUC was used as the main score to compare the two methods. Predictive accuracy was assessed by calculating AUC.

3.3. Correlation analysis

A positive and significant correlation of GDM with coagulation function examination indices PAT-PT ($r = -0.396715$) and REF-APTT ($r = -0.911669$), with hepatic and renal function indices DBIL ($r = -0.367481$) and FPG ($r = 0.467626$). Three items PAT-PT, REF-APTT, and DBIL negative correlation, FPG positive correlation of hepatic and renal and coagulation function examination could be predictors of GDM accurately. We think hepatic and renal and coagulation is reasonable combination.

4. Discussion

Early screening and detection of pregnant women is important for the identification and treatment of GDM. The establishment of a reliable GDM screening procedure would allow for informed clinical decision making and therapeutic interventions that reduce the long-term complications of GDM. Existing research has identified several potential biomarkers such as HOMA and SHBG. However, limitations such as complex detection methods, low sensitivity, and low accuracy, hinder the use of these markers in clinical practice. The aim of this study was to determine whether parameters found in the routine Chinese maternal examination at 10-19 weeks gestation are predictive of GDM. We found that differences in direct

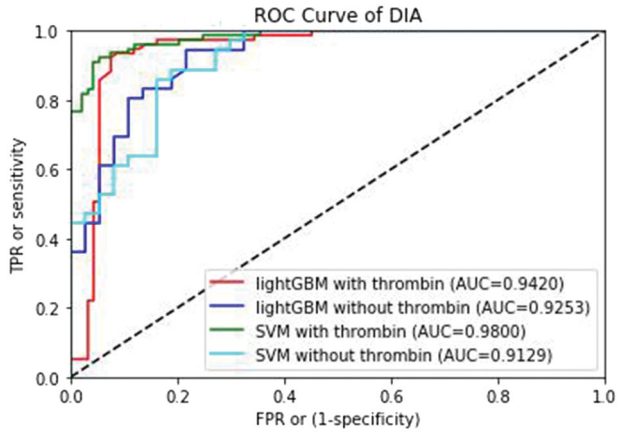

Figure 1. ROC curve of GDM.

bilirubin and fasting plasma glucose levels, prothrombin time, and activated partial thromboplastin time were predictive of GDM development using two machine learning approaches. We further characterized the correlation between GDM and these parameters, to assess their potential role in predictingof GDM. Our study reported that first 19 weeks DBIL and FPG values were independent predictors for the development of GDM in light GBM (AUC of 87.81%) and SVM (AUC of 84.79%). Our study also demonstrated that the potential novel biomarkers PAT-PT and REF-APTT are used for the prediction and earlier diagnosis. To our best knowledge, this is the first study to put forward that coagulation function indexes during first 19 weeks of pregnancy are useful predictors of GDM development. In the study, PAT-PT and REF-APTT values were found to be a better marker in two different models: light GBM (AUC of 99.83%) and SVM (AUC of 99.74%) respectively than the other factors.

Oxygen intake during pregnancy is required to maintain oxygen consumption between the pregnant women and the fetus. If the cardiopulmonary function is insufficient to maintain the above balance, the pregnant women will suffer from relative hypoxia and may activate the body's compensatory response. In the study, blood work examination indexes-MCHC value in the GDM group were significantly lower than that in the control one, while the levels of MONO and PLT were significantly higher than those in the control one. PLT reflects the platelet maturation and peripheral blood marrow hematopoiesis directly. It may be caused by mild activity in hematopoietic function because of platelet depletion.

The energy metabolism of tissue cells is mainly sugar metabolism. The body's compensatory response is activated when the tissue is relatively hypoxic and

energy and substance production are insufficient. Possible adaptive mechanisms include raising blood glucose levels and slowing blood flow to prolong oxygen uptake in the tissue.

GDM is defined as a form of hyperglycemia which normally develops in the second or third trimester during pregnancy among women without pre-pregnancy diabetes mellitus [20]. Although women with GDM typically have normal blood glucose levels in the first trimester; some indicators associated with transient insulin resistance may present earlier in gestation. Studies have shown that the maternal cardiovascular system during pregnancy undergoes complicated hemodynamic changes, such as an increase in cardiac output, a reduction in blood pressure and a lower peripheral vascular resistance [21]. In this study, FPG levels at 10-19 weeks were significantly higher in women who eventually developed GDM. It has previously been shown that hyperglycemia is associated with vasoconstriction, decreased blood flow, and tissue hypoxia [22], further resulting in endothelial dysfunction, inflammation [23], and compensatory vasodilation.Thus, despite FPG levels observed within the normal range in our study,the differences in these levels between eventual GDM cases and controls suggests its utility as a predictive measure. Pregnant is a pro- inflammatory state.There is a balancebetween pro-inflammatory cytokines and anti-inflammatory cytokines, which are necessary for the normal development of the fetus. In the setting of increased inflammation, this vital balance is compromised, placing risk on the fetus. Janus et al. [24] hypothesizethat endothelial dysfunction is an underlying factor in the pathophysiologic progression from GDM to longer-term consequences. Injury to the vascular endothelium changes blood flow and blood composition, leading to abnormal coagulation and a series of concurrent symptoms [25]. During pregnancy, the body is usually in a hypercoagulable state, which manifests as increased plasma thrombin, coagulation factors and fibrinogen. This is associated with a decrease in anticoagulation factors and fibrinolytic system activity, as well as mild coagulation in the blood vessels. The decrease in blood flow velocity increases the risk of undesirable coagulation. There are antagonistic coagulation and anticoagulation pathways in the circulatory system, and under normal circumstances, there is a dynamic balance between them. If the dynamic balance between these pathways is disrupted, as seen during injury of vascular endothelium, changes in blood flow and composition beget imbalance in favor of coagulation.

Thus, PAT-PT, PAT-APTT and REF-APTT may reflect these endogenous and exogenous coagulation pathways in the context of ourstudy. To the best of our knowledge, we are the first to posit the indexes of coagulation function examination as independent predictors of GDM at 10-19 weeks gestation. Although FPG level does not seem to be an ideal marker with a little lower sensitivity and specificity than that of the coagulation levels PAT-PT and REF-APTT, it still has been associated with the development of GDM and definitely have a role in combination with other markers, such as coagulation function levels PAT-PT and REF-APTT.

In the study, machine learning algorithms such as lightGBM and SVM are used in the prediction models. Nowadays, machine learning algorithms are increasingly used to explore the risk factors of disease occurrence and development, establish disease classification and prediction models, and provide decision supports for clinical diagnosis and treatment.

This study has a few potential limitations. First, the patient records were taken from the medical record of the West China Second University Hospital. It may have induced selection bias into our study. However, this bias is not expected to be differential in nature. Future studies should include more patients from multiple centers. .

Second, this study focuses on retrospectively selecting patients based on indicators measured in the laboratory, which may carry measurement error. A prospective study may mitigate technical variation in measured laboratory values in the future.

Lastly, we initially screened for many indexes when designing this experiment. This led to a large number of patients being excluded due to incomplete data. Future efforts should focus on those significant predictors idntified here.

5. Conclusion

In this paper, first 19 weeks coagulation levels of PAT-PT and REF-APTT were found to be significantly higher in patients who subsequently developed GDM, and the level of these biomarkers have high sensitivity (92.5%) and specificity (99.2%) in prediction of GDM development. It provides a new idea for early prediction of GDM. Our study's evidence supports that early pregnancy blood routine examination is conducive to GDM screening. Identifying the prediction model which contains 2–3 kinds of markers at greatest risk for GDM has implications for prevention, early detection and intervention.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Key R&D Program of China [2018YFC2001800].

ORCID

Yu Chen http://orcid.org/0000-0002-8931-8079

References

[1] Yogev Y, Visser GHA. Obesity, gestational diabetes and pregnancy outcome. Semin Fetal Neonatal Med. 2009;14(2):77–84.
[2] Kralisch S, Klein J, Lossner U, et al. Hormonal regulation of the novel adipocytokine visfatin in 3T3-L1 adipocytes. J Endocrinol. 2005;185(3):R1–R8.
[3] Gözükara YM, Aytan H, Ertunc D, et al. Role of first trimester total testosterone in prediction of subsequent gestational diabetes mellitus. J Obstet Gynaecol Res. 2015;41(2):193–198.
[4] Tawfeek MA, Alfadhli EM, Alayoubi AM, et al. Sex hormone binding globulin as a valuable biochemical marker in predicting gestational diabetes mellitus. BMC Womens Health. 2017;17(1):18.
[5] Mierzyński R, Poniedziałek-Czajkowska E, Dłuski D, et al. Nesfatin-1 and vaspin as potential novel biomarkers for the prediction and early diagnosis of gestational diabetes mellitus. Int J Mol Sci. 2019;20(1):159.
[6] Riskin-Mashiah S, Damti A, Younes G, et al. First trimester fasting hyperglycemia as a predictor for the development of gestational diabetes mellitus. Eur J Obstet Gynecol Reprod Biol. 2010;152(2):163–167.
[7] Aker SŞ, Yüce T, Kalafat E, et al. Association of first trimester serum uric acid levels gestational diabetes mellitus development. Turk J Obstet Gynecol. 2016;13(2):71–74.
[8] Grewal E, Kansara S, Kachhawa G, et al. Prediction of gestational diabetes mellitus at 24 to 28 weeks of gestation by using first-trimester insulin sensitivity indices in Asian Indian subjects. Metab Clin Exp. 2012;61(5):715–720.
[9] Oztas E, Ozler S, Ersoy E, et al. Prediction of gestational diabetes mellitus by first trimester serum secreted frizzle-related protein-5 levels. J Matern Fetal Neonat Med. 2015;29(9):1–5.
[10] Yeral MI, Erdinc ASO, Uygur D, et al. 239: prediction of gestational diabetes mellitus in the first trimester; comparison of fasting plasma glucose, two-step and one-step methods. Am J Obstet Gynecol. 2013;208(1):S110–S111.
[11] Junguk H, Eun-Hee C, Kwang-Hyun B, et al. Prediction of gestational diabetes mellitus by unconjugated estriol levels in maternal serum. Int J Med Sci. 2017;14(2):123–127.
[12] Savvidou M, Nelson SM, Makgoba M, et al. First-trimester prediction of gestational diabetes mellitus: examining the potential of combining maternal characteristics and laboratory measures. Diabetes. 2010;59(12):3017–3022.
[13] Isik H. Predicting gestational diabetes mellitus during the first trimester using anthropometric measurements and HOMA-IR. J Endocrinol Invest. 2015;39(5):577–583.
[14] Kumru P, Arisoy R, Erdogdu E, et al. Prediction of gestational diabetes mellitus at first trimester in low-risk pregnancies. Taiwan J Obstet Gynecol. 2016;55(6):815–820.
[15] Rasanen JP, Snyder CK, Rao PV, et al. Glycosylated fibronectin as a first-trimester biomarker for prediction of gestational diabetes. Obstet Gynecol. 2013;122(3):586–594.
[16] Arianne NS, Jencia W, Heidi A, et al. A novel early pregnancy risk prediction model for gestational diabetes mellitus. Fetal Diagn Ther. 2018;45:1–9.
[17] Kuzuya T, Nakagawa S, Satoh J, et al. Report of the Committee on the Classification and Diagnostic Criteria Of Diabetes Mellitus. Diabetes Res Clin Pract. 2002;55(1):65–85.
[18] Li L-J, Tan KH, Aris IM, et al. Retinal vasculature and 5-year metabolic syndrome among women with gestational diabetes mellitus. Metabolism. 2018;83:216–224.
[19] Fang X, Liu W, Ai J, et al. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. BMC Infect Dis. 2020;20(1):222–230.
[20] American Diabetes Association. (2) Classification and diagnosis of diabetes. Diabetes Care. 2017;40(Suppl. 1):S11.
[21] Li LJ, Cheung CY, Ikram MK, et al. Blood pressure and retinal microvascular characteristics during pregnancy: Growing Up in Singapore Towards Healthy Outcomes (GUSTO) Study. Hypertension. 2012;60(1):223–230.
[22] Benitez-Aguirre PZ, Sasongko MB, Craig ME, et al. Retinal vascular geometry predicts incident renal dysfunction in young people with type 1 diabetes. Diabetes Care. 2012;35(3):599–604.
[23] Festa A, D'Agostino R Jr., Tracy RP, et al. Elevated levels of acute-phase proteins and plasminogen activator inhibitor-1 predict the development of type 2 diabetes: the insulin resistance atherosclerosis study. Diabetes. 2002;51(4):1131–1137.
[24] Janus A, Szahidewicz-Krupska E, Mazur G, et al. Insulin resistance and endothelial dysfunction constitute a common therapeutic target in cardiometabolic disorders. Mediators Inflamm. 2016;2016:3634948.
[25] Yang Y. Effect of gynecological laparoscopy on the blood coagulation and it's mechanism. Int J Pathol Clin Med. 2011;31(2):112–116.