CLINICAL and MOLECULAR HEPATOLOGY

The forum for latest knowledge of hepatobiliary diseases

Nonalcoholic fatty liver disease in pregnancy

Cerebral atherosclerosis in cirrhosis

NAFLD and prediction of gestational diabetes

Low urea level and hepatic decompensation

Cost-effectiveness of HCV screening in Korea

KASL

CLINICAL and MOLECULAR HEPATOLOGY

Original Article (Special Issue)

Nonalcoholic fatty liver disease and early prediction of gestational diabetes mellitus using machine learning methods

Seung Mi Lee[1,2,*], Suhyun Hwangbo[3,*], Errol R. Norwitz[4], Ja Nam Koo[5], Ig Hwan Oh[5], Eun Saem Choi[2], Young Mi Jung[1,2], Sun Min Kim[1,6], Byoung Jae Kim[1,6], Sang Youn Kim[7], Gyoung Min Kim[8], Won Kim[9,10], Sae Kyung Joo[9,10], Sue Shin[11,12], Chan-Wook Park[1,2], Taesung Park[3,13], and Joong Shin Park[1,2]

[1]Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul; [2]Department of Obstetrics and Gynecology, Seoul National University Hospital, Seoul; [3]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Korea; [4]Department of Obstetrics and Gynecology, Tufts University School of Medicine, Boston, MA, USA; [5]Seoul Women's Hospital, Incheon; [6]Department of Obstetrics and Gynecology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul; [7]Department of Radiology, Seoul National University College of Medicine, Seoul; [8]Department of Radiology, Yeonsei University College of Medicine, Seoul; [9]Department of Internal Medicine, Seoul National University College of Medicine, Seoul; [10]Department of Internal Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul; [11]Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul; [12]Department of Laboratory Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul; [13]Department of Statistics, Seoul National University, Seoul, Korea

Graphical Abstract

Machine Learning Model
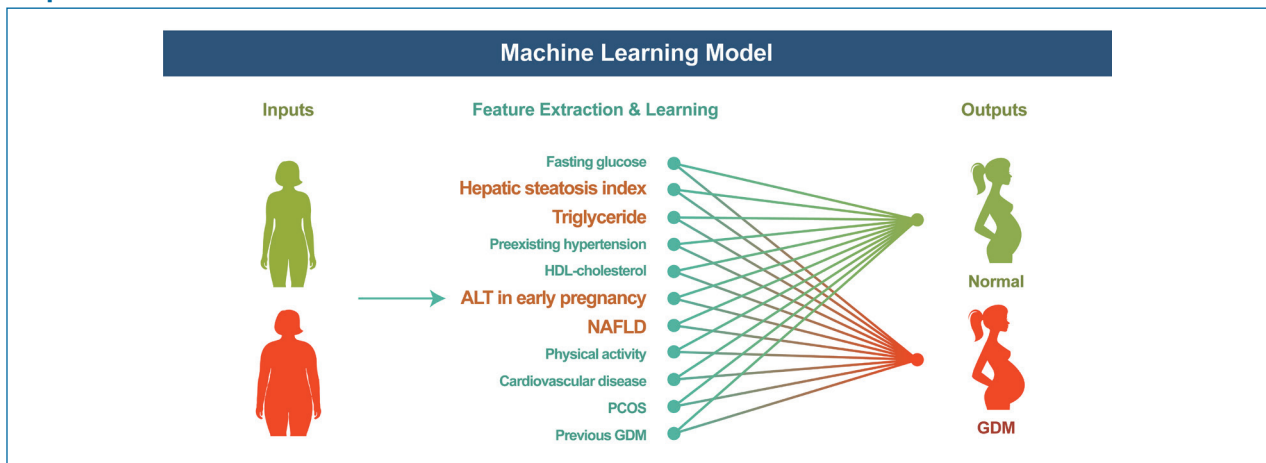
Abbreviations:

ACOG, American College of Obstetricians and Gynecologists; ADA, American Diabetes Association; ALT, alanine aminotransferase; AUC, area under the receiver operating characteristic curve; BMI, body mass index; DNN, deep neural network; GDM, gestational diabetes mellitus; HDL, high-density lipoprotein; HSI, hepatic steatosis index; LR, logistic regression; NAFLD, nonalcoholic fatty liver disease; OGTT, oral glucose tolerance test; RF, random forest; SVM, support vector machine; TG, triglycerides

*Both authors contributed equally to this work as co-first authors.

Editor: Sung Won Lee, The Catholic University of Korea, Korea

Corresponding author : Joong Shin Park

Department of Obstetrics & Gynecology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
Tel: +82-2-2072-3199, Fax: +82-2-762-3599
E-mail: jsparkmd@snu.ac.kr
https://orcid.org/0000-0002-5246-0477

Taesung Park

Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
Tel: +82-2-880-8924, Fax: +82-2-883-6144
E-mail: tspark@stats.snu.ac.kr
https://orcid.org/0000-0002-8294-590X

Background/Aims: To develop an early prediction model for gestational diabetes mellitus (GDM) using machine learning and to evaluate whether the inclusion of nonalcoholic fatty liver disease (NAFLD)-associated variables increases the performance of model.

Methods: This prospective cohort study evaluated pregnant women for NAFLD using ultrasound at 10–14 weeks and screened them for GDM at 24–28 weeks of gestation. The clinical variables before 14 weeks were used to develop prediction models for GDM (setting 1, conventional risk factors; setting 2, addition of new risk factors in recent guidelines; setting 3, addition of routine clinical variables; setting 4, addition of NALFD-associated variables, including the presence of NAFLD and laboratory results; and setting 5, top 11 variables identified from a stepwise variable selection method). The predictive models were constructed using machine learning methods, including logistic regression, random forest, support vector machine, and deep neural networks.

Results: Among 1,443 women, 86 (6.0%) were diagnosed with GDM. The highest performing prediction model among settings 1–4 was setting 4, which included both clinical and NAFLD-associated variables (area under the receiver operating characteristic curve [AUC] 0.563–0.697 in settings 1–3 vs. 0.740–0.781 in setting 4). Setting 5, with top 11 variables (which included NAFLD and hepatic steatosis index), showed similar predictive power to setting 4 (AUC 0.719–0.819 in setting 5, *P*=not significant between settings 4 and 5).

Conclusions: We developed an early prediction model for GDM using machine learning. The inclusion of NAFLD-associated variables significantly improved the performance of GDM prediction. (ClinicalTrials.gov Identifier: NCT02276144) **(Clin Mol Hepatol 2022;28:105-116)**

Keywords: Nonalcoholic fatty liver disease; Diabetes, Gestational; Machine learning; Prediction; Pregnancy, High-risk

Study Highlights
- Using machine learning, a high-performance prediction model was developed.
- Among the variables, the addition of NAFLD and NAFLD-associated variables into the prediction model improved its performance.
- Incorporation of NAFLD and NAFLD-associated variables into the prediction model can improve prediction of GDM. However, further evaluation through large prospective studies is needed before these factors can be incorporated in routine practice.

INTRODUCTION

Gestational diabetes mellitus (GDM) complicates 5–10% of all pregnancies, and is associated with increased maternal morbidity and fetal/neonatal complications. In accordance with the rising incidence of obesity and metabolic complications worldwide, the incidence of GDM is also increasing.[1] Therefore, early and accurate prediction of GDM is critical to ensure preventive strategies are effective.

The American College of Obstetricians and Gynecologists (ACOG) has long recommended identifying women at high risk of developing GDM by screening them in early pregnancy using a series of clinical and demographic risk factors.[2] According to these criteria, women with one or more risk factors (including a personal history of GDM or impaired glucose tolerance, a family history of diabetes, obesity, or glucosuria) should be identified and tested for GDM in early pregnancy. In 2017, based on the recommendations of the American Diabetes Association (ADA), the ACOG re-

vised its guidelines to incorporate these new criteria.[3,4] However, these guidelines have not been universally adopted and they have low accuracy,[5] which has severely limited our ability to prevent pregnancy-related complications.

Nonalcoholic fatty liver disease (NAFLD) refers to hepatic fat accumulation in the absence of excessive alcoholic consumption. It is a common cause of chronic liver dysfunction,[6-11] and recent data suggest that NAFLD is the early hepatic manifestation of metabolic syndrome.[12-14] NAFLD has also been identified as a risk factor for pregnancy complications such as GDM, preeclampsia, and fetal growth abnormalities.[15-18] However, there is a paucity of information regarding whether testing for NAFLD in early pregnancy can inform the prediction model for GDM.

Machine learning algorithms derived from computational learning methodologies are being increasingly used in medical informatics. A few recent studies have included machine learning methods in the development of prediction models for GDM.[19-21] However, these studies had the some limitations, including the

following: 1) they used clinical/demographic variables collected in the second trimester, although prediction of GDM in the first trimester is more important in clinical practice; and 2) they mostly developed prediction models using already established clinical variables and did not include NAFLD-associated variables. In addition, the development of a prediction model with a relatively small number of variables is also an important task, since the use of many variables in the prediction model of machine learning may keep the clinician away from a wide acceptance of the developed prediction model.

In the current study, we developed prediction models for GDM with three main goals: 1) to predict the risk of GDM in the first trimester; 2) to use machine learning methodology to select essential variables to be included in the model and to develop the best predictive model; and 3) to evaluate whether the inclusion of NAFLD-associated variables improved the performance of the predictive model.

MATERIALS AND METHODS

Study population

This was a secondary analysis of a prospective cohort study of "Fatty Liver in Pregnancy" (NCT02276144).[5,16,17,22] In this cohort, women with singleton pregnancies who visited either Incheon Seoul Women's Hospital or Seoul Metropolitan Government Seoul National University Boramae Medical Center in Seoul, Korea for routine antenatal care in the first trimester were invited to enroll in the study. The enrolled subjects were routinely evaluated for fatty liver by ultrasound, taken for fasting blood at 10–14 weeks, and then followed at delivery. The cohort included all of the enrolled women, including women with fatty liver and those without. The current analysis included enrolled women who delivered between June 2015 and April 2020. The study was approved by the Institutional Review Board of Seoul Metropolitan Government Seoul National University Boramae Medical Center and the Public Institutional Review Board designated by the Ministry of Health and Welfare of Korea (No. 1308-116-518). Each participant provided informed written consent, and the study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki. All authors had access to the study data, and they reviewed and approved the final version of the manuscript before submission.

Data collection

At the time of enrollment, patients with chronic liver diseases, such as hepatitis, primary biliary or sclerosing cholangitis, hemochromatosis, and Wilson disease, were not invited to enroll in the study cohort. After enrollment, basic clinical and demographic factors, including medical and family history, were retrieved using a questionnaire. Alcohol consumption was self-reported using the validated cut-annoyed-guilty-eye questionnaire[23] to exclude alcoholic fatty liver. Patients with pre-GDM, incomplete records for classical risk factors for GDM, previable birth before 24 weeks, or incomplete follow-up were also excluded from the current study. The laboratory results that were routinely measured in early pregnancy during antenatal care, such as complete blood count, serology for syphilis or hepatitis, and the presence or absence of glycosuria, were also retrieved by reviewing the medical records of the patients.

During the routine antenatal visit at 10–14 weeks of gestation, fasting blood samples were collected and stored for subsequent analysis. At the same visit, liver ultrasound was performed to detect NAFLD. Hepatic steatosis was assessed using a semi-quantitative grading system (grades 0–3), and NAFLD was defined as hepatic steatosis grades 1–3.[24,25] The hepatic steatosis index (HSI) was calculated using the following equation: HSI = 8 × alanine aminotransferase (ALT) / aspartate aminotransferase + body mass index (BMI) + 2 (if type 2 diabetes) + 2 (if female).[26] The remaining information regarding antenatal care and pregnancy outcomes were extracted from the patient's medical charts by trained researchers.

Diagnosis of GDM

It is a routine practice in the participating hospitals to diagnose GDM using a two-stage approach. First, all patients were screened at 24–28 weeks of gestation using a 50-g oral glucose tolerance test (OGTT).[27] In screen-positive cases (defined as ≥140 mg/dL), patients underwent a 100-g 3-hour OGTT (Supplementary Fig. 1). A diagnosis of GDM was made in patients with two or more blood glucose levels higher than the established cut-off values (≥95 mg/dL fasting, ≥180 mg/dL at 1 hour, ≥155 mg/dL at 2 hours, and/or ≥140 mg/dL at 3 hours).[28]

Definition of high-risk women using the old and new ACOG criteria

GDM risk factors used in the original ACOG guidelines were derived from the 4th International Workshop Conference on GDM in 1998, which defined high-risk women as those with at least one of the following risk factors: strong family history of diabetes, obesity, previous GDM history, impaired glucose tolerance, or glucosuria (Supplementary Fig. 2).[2] GDM risk factors in the most recent 2018 ACOG guidelines[4] were derived from the recommendations of the ADA, which defined high-risk women as those who are overweight or obese with one of the following risk factors: physical inactivity, family history of type 2 diabetes, high-risk race or ethnicity, previous delivery of a macrosomic infant, previous GDM history, preexisting hypertension, low high-density lipoprotein (HDL) or high triglyceride, personal history of polycystic ovarian syndrome or cardiovascular disease, and/or other conditions such as severe obesity.[3,4] According to the World Health Organization, the categories of overweight, obesity, and severe obesity for Asian population are defined as BMI of 23–25 kg/m^2, 25–30 kg/m^2, and >30 kg/m^2, respectively.[29,30]

Statistical analysis

Clinical/demographic variables collected before 14 weeks were used to develop prediction models for GDM in four settings (settings 1, 2, 3, and 4). To limit the number of variables and identify those that are most important, a stepwise variable selection methodology was used to define one additional model [Setting 5]. The variables used in each setting are summarized in Supplementary Table 1. In brief: setting 1, conventional risk factors from the 4th International ADA Workshop; setting 2, setting 1 + revised risk factors from the ADA; setting 3, setting 2 + additional clinical variables; setting 4, setting 3 + additional variables associated with NAFLD; setting 5, using only 11 selected important variables.

To avoid overfitting, the study population was randomly divided into a model development and a test dataset with a 2:1 ratio (Fig. 1) in a stratified manner, taking into account the ratio of the GDM group and the non-GDM group. Prediction models were developed using the model development dataset. To identify important predictors for GDM, the area under the receiver operating characteristic curve (AUC)-based stepwise selection was performed via 5-fold cross validation on the model development set.[31] Variables with the highest mean validation AUCs were selected at each step of the stepwise selection process. To construct predictive models, we considered four machine learning models: logistic regression (LR), random forest (RF),[32] support vector machine (SVM),[33] and deep neural network (DNN). For each model based on RF, SVM, and DNN, we tuned the hyperparameters to select the optimal combination with the highest mean AUC using 5-fold cross vali-
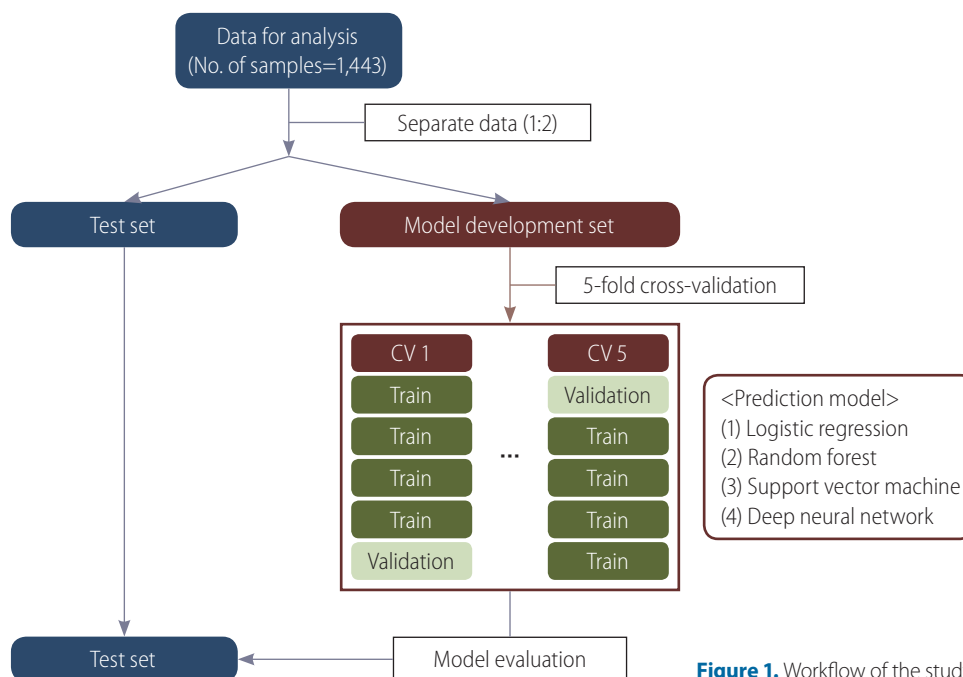

Figure 1. Workflow of the study.

dation. Then, the final prediction model was evaluated using the test dataset.

To evaluate the predictive power for each setting more systematically and to avoid data split-dependent results, we repeated the data split process into a model development set and a test set a total of 10 times, and then compared the mean AUC of each model. All analyses were performed using R (version 3.6.1).

RESULTS

Subject population

A total of 1,709 women were enrolled in the original cohort. After excluding women with pre-GDM (n=27), incomplete records for classical risk factors for GDM (n=147), previable birth before 24 weeks (n=11), or incomplete follow-up (n=81), 1,443 women were included in the current analysis (Supplementary Fig. 3). Among them, 86 women (6.0%) were subsequently diagnosed with GDM. Table 1 presents the baseline clinical/demographic

Table 1. Baseline features and pregnancy outcomes of the study population

Characteristic	No GDM (n=1,357)	GDM (n=86)	*P*-value
Baseline characteristic			
Age (years)	32.3±4.0	32.4±4.6	0.758
Nulliparity	716 (52.8)	51 (59.3)	0.286
BMI before pregnancy (kg/m^2)	22.1±3.6	25.1±4.9	<0.001
WC before pregnancy (cm) (n=1,418)	70.9±5.7	74.5±7.3	<0.001
Laboratory result in early pregnancy			
Gestational age at measurement	7.8±1.4	7.8±1.5	0.952
Hemoglobin (g/dL)	12.7±1.0	13.0±1.0	0.006
Platelet counts (×10^3/uL)	250.7±53.6	273.6±57.1	<0.001
AST (U/L)	16.0±5.2	18.8±15.5	0.098
ALT (U/L)	14.3±10.1	20.9±20.7	0.005
Laboratory and ultrasound result at 10–14 weeks			
Gestational age at measurement	12.4±0.5	12.3±0.6	0.209
AST (U/L)	16.6±10.7	17.6±9.1	0.819
ALT (U/L)	12.8±14.4	16.5±14.1	0.001
Cholesterol (mg/dL)	172.1±30.3	179.2±29.5	0.028
HDL cholesterol (mg/dL)	68.6±14.2	63.6±15.3	0.012
LDL cholesterol (mg/dL)	81.4±22.3	84.3±25.4	0.150
Triglycerides (mg/dL)	111.0±43.1	151.7±77.6	<0.001
γ-GT (U/L)	13.7±8.4	16.1±10.1	0.001
Fasting glucose (mg/dL)	79.6±8.9	88.7±13.0	<0.001
HSI	30.3±5.0	34.5±5.6	<0.001
NAFLD by liver ultrasound	158 (11.8)	32 (37.6)	<0.001
Pregnancy outcome	1,327	85	
Gestational age at delivery (weeks)	38.9±1.4	38.5±1.7	0.033
Birthweight (kg)	3.2±0.4	3.2±0.5	0.998
Large-for-gestational age neonates	137 (10.3)	15 (17.6)	0.053

Values are presented as mean±standard deviation or number (%).
GDM, gestational diabetes mellitus; BMI, body mass index; WC, waist circumference; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; γ-GT, gamma-glutamyl transferase; HSI, hepatic steatosis index; NAFLD, nonalcoholic fatty liver disease.

characteristics and pregnancy outcomes according to GDM status. Women who subsequently developed GDM were more likely to have a higher pre-pregnancy BMI and waist circumference compared to those who did not. With regard to pregnancy outcomes, women with GDM delivered at an earlier gestational age (P=0.033) and, despite an earlier gestational age, their neonates were more likely to be large-for-gestational age, although the difference was not statistically significant (P=0.053).

Risk stratification according to conventional guidelines

Table 2 summarizes the categorization of women at high risk for developing GDM according to the 1998[2] and 2018 criteria.[4]

Among those who developed GDM (n=86), the old criteria identified 59.3% (51/86) of women to be at high risk, and the frequency of all variables included was indeed higher in women who developed GDM. The new criteria identified only 41.9% (36/86) of women as high-risk, and the current analysis showed that some of the risk factors included in this ACOG-recommended model (physical inactivity, previous delivery of a macrosomic infant, low HDL, and a personal history of polycystic ovarian syndrome or cardiovascular disease) did not differ between the two groups in the cohort.

Selection of important variables

Among the variables retrieved, we selected important variables

Table 2. Comparison of risk factors in the study population

Characteristic	No GDM (n=1,357)	GDM (n=86)	P-value
Risk factors in old criteria, 1998[2]			
Classified as high-risk women by old criteria	387 (28.5)	51 (59.3)	<0.001
Severe obesity, BMI ≥30 kg/m^2	51 (3.8)	13 (15.1)	<0.001
Family history of type 2 diabetes	290 (21.4)	31 (36.0)	0.002
Previous GDM	24 (1.8)	7 (8.1)	<0.001
Impaired fasting glucose	20 (1.5)	18 (20.9)	<0.001
Glucosuria	35 (2.6)	8 (9.3)	0.001
Risk factors in new ACOG criteria, 2018[4]			
Classified as high-risk women by new criteria	194 (14.3)	36 (41.9)	<0.001
Overweight or obese, BMI ≥23 kg/m^2	418 (30.8)	47 (54.7)	<0.001
Physical inactivity	161 (11.9)	10 (11.6)	1.000
Family history of type 2 diabetes	290 (21.4)	31 (36.0)	0.002
High-risk race or ethnicity	0 (0.0)	0 (0.0)	-
Previous macrosomia	15 (1.1)	1 (1.2)	1.000
Previous GDM	24 (1.8)	7 (8.1)	<0.001
Preexisting hypertension	11 (0.8)	3 (3.5)	0.059
Low HDL, <35 mg/dL	13/1,350 (1.0)	1/84 (1.2)	1.000
High TG, >250 mg/dL	14/1,350 (1.0)	6/84 (7.1)	<0.001
PCOS	23 (1.7)	2 (2.3)	0.993
Impaired fasting glucose	20 (1.5)	18 (20.9)	<0.001
History of cardiovascular disease	8 (0.6)	1 (1.2)	1.000
Severe obesity, BMI ≥30 kg/m^2	51 (3.8)	13 (15.1)	<0.001

Values are presented as number (%).
The risk factors in the old criteria were from the 4th International Workshop Conference on GDM in 1998;[2] the risk factors in the new criteria were based on the recommendation of the American Diabetes Association, which defined high-risk women as overweight or obese women with one of the risk factors.[3]
GDM, gestational diabetes mellitus; BMI, body mass index; ACOG, American College of Obstetricians and Gynecologists; HDL, high-density lipoprotein; TG, triglycerides; PCOS, polycystic ovarian syndrome.

for the early prediction of GDM using AUC-based stepwise selection with 5-fold cross validation. The top 11 variables were fasting glucose, HSI, triglyceride level, HDL level, ALT in early pregnancy, preexisting hypertension, cardiovascular disease, polycystic ovarian syndrome, NAFLD, previous GDM history, and physical inactivity.

Evaluation of the predictive modeling

Supplementary Table 2 compares the baseline clinical characteristics of the model development and test dataset. There were no differences in the characteristics between the model development and test dataset.

The prediction model for GDM was developed using LR, RF,

SVM, and DNN for settings 1–5. The AUC was used as an evaluative measure. The results are presented in Table 3. Among settings 1–4, the test AUC was the highest in setting 4 (highest AUC in setting 1–3, 0.617 by LR, 0.643 by RF, 0.697 by SVM, and 0.609 by DNN; AUC in setting 4, 0.740 by LR, 0.781 by RF, 0.756 by SVM, and 0.745 by DNN; Table 3), indicating that the addition of NAFLD-associated variables significantly improved performance of the prediction model. Then, we compared the test AUC of the models between setting 5 and settings 1–4. The test AUC of the developed prediction model using the top 11 variables [Setting 5] was similar to that of the model using setting 4, with the highest test AUC from the model developed by SVM (0.719 by LR, 0.763 by RF, 0.819 by SVM, and 0.777 by DNN; Table 3). Using SVM,

Table 3. Results of predictive modeling

Setting	Variables used	Prediction model	Model development set				Test set				P-value
			AUC	Sen	Spe	P-value	AUC	Sen	Spe	P-value	
Setting 1	(1) Conventional ACOG risk factors	LR	0.728	0.649	0.723	<0.001	0.609‖	0.483	0.698	0.041	0.194*
		RF	0.667	0.368	0.961	<0.001	0.565	0.172	0.962	0.082	0.003†
		SVM	0.713	0.649	0.723	<0.001	0.600	0.483	0.698	0.053	0.003‡
		DNN	0.683	0.525	0.817	<0.001	0.585	0.359	0.796	0.042	0.023§
Setting 2	(1) + (2) New ACOG risk factors form 2017	LR	0.777	0.719	0.734	<0.001	0.563	0.481	0.728	0.364	0.105*
		RF	0.702	0.456	0.945	<0.001	0.578	0.222	0.951	0.069	0.009†
		SVM	0.729	0.737	0.667	<0.001	0.697‖	0.704	0.666	<0.001	0.084‡
		DNN	0.686	0.631	0.672	<0.001	0.609	0.548	0.616	0.135	0.054§
Setting 3	(1) + (2) + (3) Routine clinical variables	LR	0.842	0.809	0.761	<0.001	0.617	0.520	0.758	0.104	0.297*
		RF	0.983	0.915	0.955	<0.001	0.643‖	0.440	0.859	0.033	0.167†
		SVM	0.810	0.638	0.870	<0.001	0.605	0.520	0.725	0.095	0.008‡
		DNN	0.615	0.545	0.599	0.035	0.597	0.480	0.628	0.250	0.014§
Setting 4	(1) + (2) + (3) + (4) Variables associated with NAFLD	LR	0.881	0.800	0.868	<0.001	0.740	0.500	0.929	<0.001	0.652*
		RF	1.000	1.000	1.000	<0.001	0.781‖	0.750	0.670	<0.001	0.647†
		SVM	1.000	1.000	1.000	<0.001	0.756	0.708	0.747	<0.001	0.246‡
		DNN	0.800	0.572	0.807	<0.001	0.745	0.517	0.836	<0.001	0.457§
Setting 5	Top 11 important variables selected	LR	0.840	0.778	0.779	<0.001	0.719	0.542	0.872	0.001	1
		RF	1.000	1.000	0.996	<0.001	0.763	0.708	0.755	<0.001	1
		SVM	0.800	0.733	0.775	<0.001	0.819‖	0.708	0.866	<0.001	1
		DNN	0.806	0.759	0.678	<0.001	0.777	0.750	0.654	<0.001	1

Sen (i.e., sensitivity) and Spe (i.e., specificity) are represented as the values at the threshold with the maximum balanced accuracy.
AUC, area under the receiver operating characteristic curve; Sen, sensitivity; Spe, specificity; ACOG, American College of Obstetricians and Gynecologists; LR, logistic regression; RF, random forest; SVM, support vector machine; DNN, deep neural network; NAFLD, nonalcoholic fatty liver disease.
*P-value when compared with the LR model in setting 5 in the test dataset.
†P-value when compared with the RF model in setting 5 in the test dataset.
‡P-value when compared with the SVM model in setting 5 in the test dataset.
§P-value when compared with the DNN model in setting 5 in the test dataset.
‖The maximum test AUC for each setting.

the sensitivity and specificity in the test set were 70.8% and 86.6%, respectively, in setting 5 (Supplementary Fig. 4).

Figure 2 shows the receiver operating characteristic curves of the best prediction model in each setting. Prediction models from settings 4 and 5 had higher AUCs compared to those from settings 1–3. The old criteria of ACOG had a sensitivity of 59.3% and specificity of 71.5%, and the new criteria of ACOG had a sensitivity of 41.9% and specificity of 85.9% for the prediction of GDM, as previously shown in Table 2.

We further identified which variable contributed the most to the prediction result, focusing on the SVM model that had the highest test AUC for setting 5. To this end, the 11 predictors used in setting 5 were systematically excluded, and the effect on the prediction result was thus evaluated. The more important the variable, the more the AUC value was expected to decrease once the variable was excluded. Fasting glucose had the highest effect, followed by HSI (Fig. 3).

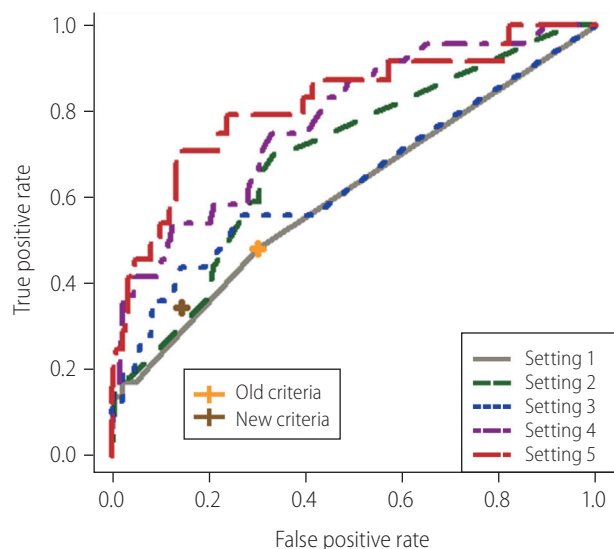To evaluate the predictive power for each setting more systematically by avoiding data split-dependent results, we repeated the data split process into a model development set and a test set 10 times and compared the mean AUC of each model (Supplementary Fig. 5). These data confirmed that the predictive model with the top 11 variables had the highest predictive performance, regardless of the data split. We compared the predictive power of the models in setting 5 with that in settings 1–4 using the Wilcoxon rank-sum test (Supplementary Table 3). This analysis confirmed that the predictive performance of setting 5 was significantly higher than that of settings 1–3 and similar to that of setting 4.

DISCUSSION

This study demonstrated that the addition of NAFLD-associated variables significantly improved the prediction model performance for GDM in early pregnancy. In addition, the model with selected important variables [Setting 5] showed similar predictive power as the model derived from all clinical variables, including NAFLD-associated variables [Setting 4].

Overall, we suggest a final model with 11 important variables [Setting 5]. The model showed the highest predictive power for the AUC among the five settings. The predictive performance of the final model was much higher than that of settings 1–3 with a small number of variables. Setting 5 also showed similar predictive performance as setting 4, which used all available clinical
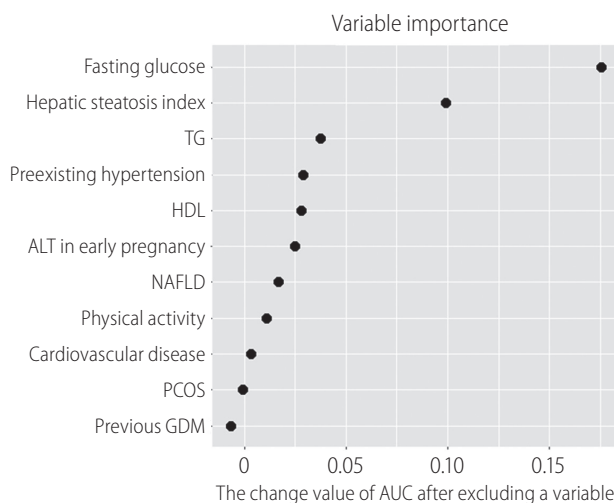

Figure 2. Receiver operating characteristic curves of the best prediction model for gestational diabetes in settings 1–5. Setting 1, conventional risk factors using older ACOG criteria. Setting 2, addition of new ACOG risk factors to setting 1. Setting 3, addition of routine clinical variables to setting 2. Setting 4, addition of variables associated with NAFLD to setting 3. Setting 5, top 11 variables. High risk 1, old criteria (from the 4th international workshop) had a sensitivity of 59.3% and specificity of 71.5% for GDM. High risk 2, new criteria (from the ADA) had a sensitivity of 41.9% and specificity of 85.9% for GDM. ACOG, American College of Obstetricians and Gynecologists; NAFLD, nonalcoholic fatty liver disease; GDM, gestational diabetes mellitus; ADA, American Diabetes Association.

Figure 3. Variable importance of the top 11 selected variables in support vector machine model. TG, triglycerides; HDL, high-density lipoprotein; ALT, alanine aminotransferase; NAFLD, nonalcoholic fatty liver disease; PCOS, polycystic ovarian syndrome; GDM, gestational diabetes; AUC, area under the receiver operating characteristic curve.

variables and NAFLD-associated variables. We confirmed that these trends were not data-split-dependent results through statistical analysis. Based on the ultimate goal of developing a parsimonious model with high predictive power, we suggest a prediction model in setting 5 as the final model.

NAFLD is strongly associated with the development of type 2 diabetes, hypertension, metabolic syndrome, and other cardiovascular complications.[12-14,34] Several recent studies have reported that NAFLD is also a risk factor for GDM, which is consistent with the observation that pregnancy can unmask subclinical metabolic disorders in patients at risk of metabolic diseases later in life.[15,16,35,36] The molecular mechanisms underlying the relationship between NAFLD and metabolic complications in later life appear to be related to hepatic insulin resistance and lipotoxicity in the setting of excessive free fatty acids, hepatokines, or cytokines, and peripheral adiposity, which leads to oxidative stress, activation of proinflammatory cytokines, and fibrosis.[16,37-39]

In previous studies, NAFLD identified in early pregnancy was shown to increase the risk of GDM, with the odds ratios ranging from 2.2 to 6.5.[40] Moreover, several biomarkers related to NAFLD, such as ALT, triglycerides (TG), and gamma-glutamyl transferase, have been independently reported as risk factors for GDM.[19,41,42] However, whether we should evaluate NAFLD-associated factors in early pregnancy for GDM prediction has not been evaluated to date. In this study, we showed that the addition of NAFLD-associated variables [Setting 4] significantly improved the prediction model performance for GDM by both traditional machine learning algorithms and DNN. In addition, NAFLD, HSI, TG, and ALT in the first trimester were identified among the top 11 important selected variables for GDM prediction [Setting 5].

Several recent studies have used machine learning algorithms to develop prediction models for GDM, although most of the included clinical variables were retrieved in the second trimester. Ye et al.[19] failed to show better performance of machine learning algorithms for GDM prediction compared to traditional LR analysis, although other studies have reported improved performance. For example, Xiong et al.[20] reported higher accuracy with gradient boosting and SVM with clinical variables up to 19 weeks; however, they used a case-control study design, which did not accurately represent the real-world context in which we see GDM. Artzi et al.[21] used real-world data from retrospective electronic health records retrieved up to 20 weeks of gestation and developed a prediction model for GDM with a gradient boosting model, with variable success. Theoretically, machine learning should improve the performance of predictive models due to its ability to learn from non-linear and complex relationships among risk factors in real-world datasets.[43] In addition, machine learning can also identify important variables that may not have been identified in other analyses. In the current study, we showed that the machine learning algorithm performed better than the conventional LR analysis, and identified important variables in the prediction model.

Interventions such as lifestyle modification, weight gain optimization, and regular exercise starting early in pregnancy may prevent some cases of GDM.[44-46] As such, there is an increasing demand for early prediction of GDM. Both the ACOG and ADA have developed guidelines to identify high-risk women in early pregnancy.[2-4] However, predictive modeling using both original and revised criteria has poor accuracy in identifying women at risk, and there is a high demand for a more accurate model in early pregnancy.[5] In the current study, we developed a prediction model using clinical/demographic variables collected prior to 14 weeks, enabling accurate prediction of GDM in the first trimester. After acute prediction of GDM in early pregnancy, we can modify our screening strategies for GDM or intervention in high-risk women with lifestyle modification or regular exercise.

To establish a prediction model that could be implemented more easily in routine clinical practice, we tried to limit the number of variables in this study.[47] The model with the 11 most important variables [Setting 5] showed similar or better performances than settings 1–4 (Supplementary Fig. 1), making this probably the most useful of the five models. However, whether the suggested model can be used in clinical practice requires further larger randomized studies.

The strengths of this study are that it used data that were collected prospectively, and included data on both NAFLD and HSI. In the current study, NAFLD was defined by liver ultrasound and not by histologic examination, although sonographic evaluation of the fatty liver is subjective and may not be able to detect small amounts of fat accumulation.[48] However, histological confirmation of the liver was not possible in asymptomatic pregnant women. Instead, we evaluated HSI, a metric derived from laboratory results, which is a more objective marker of hepatic dysfunction than ultrasound alone.[49] One of the limitations of the current study is that all study subjects were of Korean ethnicity. Whether these findings can be generalized to other racial/ethnic groups is not clear. In addition, we enrolled pregnant women only when they denied a history of chronic liver disease and agreed to enroll in a prospective cohort study. Given that there could be differences between women who agreed to participate in the study and those who refused, there is a possibility of selection bias at enroll-

ment. Moreover, we did not include a cost-effective analysis. For this prediction model, participants should undergo further examinations, such as liver ultrasound and laboratory tests, at 10–14 weeks in addition to routine laboratory tests in early pregnancy. Whether the suggested model with additional evaluation can be used in practice also requires a cost-effective analysis. In addition, we did not use an external validation dataset to evaluate the prediction model. However, the evaluation of liver ultrasound and sampling of fasting blood in early pregnancy has not been a routine practice in obstetrics, and we failed to find another pregnancy population dataset for external validation that had similar data regarding NAFLD as ours. Instead, 1) we split the study population into a development and test dataset before performing any analysis, and showed the performance of the final prediction model in a test dataset; and 2) we repeated this data split process 10 times to avoid split-dependent results, and showed that the predictive model with the top 11 variables had the highest predictive performance regardless of the data split. Nevertheless, the study population was based on a single cohort. Further studies with external datasets are needed to confirm the usefulness of the proposed prediction model. Lastly, we could not evaluate the influence of mild or significant fibrosis in the liver, although fibrosis itself might be more problematic in terms of metabolic outcomes. Moreover, fibrosis has been reported to be associated with adverse outcomes in non-pregnant patients with NAFLD.[50]

In conclusion, we developed early prediction models for GDM using machine learning, which performed better than the models using only clinical/demographic variables recognized by the ACOG and ADA. The inclusion of NAFLD-associated variables significantly improved the performance of early GDM prediction. However, further evaluation in large prospective studies is needed before these models can be incorporated into routine practice.

Authors' contribution

Drs. SM Lee, S Hwangbo, T Park, and JS Park had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: SM Lee, S Hwangbo, T Park, and JS Park; Acquisition, analysis, or interpretation of data: All authors; Drafting of the manuscript: SM Lee, S Hwangbo, T Park, and JS Park; Critical revision of the manuscript for important intellectual content: All authors; Statistical analysis: SM Lee, S Hwangbo, T Park; Study supervision: T Park, and JS Park.

Conflicts of Interest

The authors have no conflicts to disclose.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Clinical and Molecular Hepatology website (http://www.e-cmh.org).

REFERENCES

1. Hunt KJ, Schuller KL. The increasing prevalence of diabetes in pregnancy. Obstet Gynecol Clin North Am 2007;34:173-199, vii.

2. Metzger BE, Coustan DR. Summary and recommendations of the Fourth International Workshop-Conference on gestational diabetes mellitus. The Organizing Committee. Diabetes care 1998;21 Suppl 2:B161-B167.

3. American Diabetes Association. Standards of medical care in diabetes--2012. Diabetes Care 2012;35 Suppl 1(Suppl 1):S11-S63.

4. ACOG practice bulletin No. 190: gestational diabetes mellitus. Obstet Gynecol 2018;131:e49-e64.

5. Hong S, Lee SM, Kwak SH, Kim BJ, Koo JN, Oh IH, et al. A comparison of predictive performances between old versus new criteria in a risk-based screening strategy for gestational diabetes mellitus. Diabetes Metab J 2020;44:726-736.

6. Cho HC. Prevalence and factors associated with nonalcoholic fatty liver disease in a nonobese Korean population. Gut Liver 2016;10:117-125.

7. Choi SY, Kim D, Kim HJ, Kang JH, Chung SJ, Park MJ, et al. The relation between non-alcoholic fatty liver disease and the risk of coronary heart disease in Koreans. Am J Gastroenterol 2009;104:1953-1960.

8. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther 2011;34:274-285.

9. Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013;10:686-690.

10. Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 2011;140:124-131.

11. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73-84.

12. Ramesh S, Sanyal AJ. Evaluation and management of non-alcoholic steatohepatitis. J Hepatol 2005;42 Suppl(1):S2-S12.

13. Targher G, Bertolini L, Padovani R, Rodella S, Tessari R, Zenari L, et al. Prevalence of nonalcoholic fatty liver disease and its association with cardiovascular disease among type 2 diabetic patients. Diabetes Care 2007;30:1212-1218.

14. Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. Diabetes 2001;50:1844-1850.

15. Hagström H, Höijer J, Ludvigsson JF, Bottai M, Ekbom A, Hultcrantz R, et al. Adverse outcomes of pregnancy in women with non-alcoholic fatty liver disease. Liver Int 2016;36:268-274.

16. Lee SM, Kwak SH, Koo JN, Oh IH, Kwon JE, Kim BJ, et al. Non-alcoholic fatty liver disease in the first trimester and subsequent development of gestational diabetes mellitus. Diabetologia 2019;62:238-248.

17. Lee SM, Kim BJ, Koo JN, Norwitz ER, Oh IH, Kim SM, et al. Nonalcoholic fatty liver disease is a risk factor for large-for-gestational-age birthweight. PLoS One 2019;14:e0221400.

18. Sarkar M, Grab J, Dodge JL, Gunderson EP, Rubin J, Irani RA, et al. Non-alcoholic fatty liver disease in pregnancy is associated with adverse maternal and perinatal outcomes. J Hepatol 2020;73:516-522.

19. Ye Y, Xiong Y, Zhou Q, Wu J, Li X, Xiao X. Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: a retrospective cohort study. J Diabetes Res 2020;2020:4168340.

20. Xiong Y, Lin L, Chen Y, Salerno S, Li Y, Zeng X, et al. Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. J Matern Fetal Neonatal Med. 2020 Aug 6. doi: 10.1080/14767058.2020.1786517.

21. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nat Med 2020;26:71-76.

22. Jung YM, Lee SM, Hong S, Koo JN, Oh IH, Kim BJ, et al. The risk of pregnancy-associated hypertension in women with nonalcoholic fatty liver disease. Liver Int 2020;40:2417-2426.

23. Ewing JA. Detecting alcoholism. The CAGE questionnaire. JAMA 1984;252:1905-1907.

24. Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, Hurley M, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. Gastroenterology 2002;123:745-750.

25. Taylor KJ, Riely CA, Hammers L, Flax S, Weltin G, Garcia-Tsao G, et al. Quantitative US attenuation in normal liver and in patients with diffuse liver disease: importance of fat. Radiology 1986;160:65-71.

26. Lee JH, Kim D, Kim HJ, Lee CH, Yang JI, Kim W, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis 2010;42:503-508.

27. Practice bulletin No. 137: gestational diabetes mellitus. Obstet Gynecol 2013;122(2 Pt 1):406-416.

28. Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. Am J Obstet Gynecol 1982;144:768-773.

29. International Diabetes Institute. The Asia-Pacific perspective: redefining obesity and its treatment. Geneva: World Health Organization, 2000:1-55.

30. Kim Y, Suh YK, Choi H. BMI and metabolic disorders in South Korean adults: 1998 Korea National Health and Nutrition Survey. Obes Res 2004;12:445-453.

31. Kim SI, Song M, Hwangbo S, Lee S, Cho U, Kim JH, et al. Development of web-based nomograms to predict treatment response and prognosis of epithelial ovarian cancer. Cancer Res Treat 2019;51:1144-1155.

32. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18-22.

33. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Adv Neural Inf Process Syst 1997;9:155-161.

34. Armstrong MJ, Adams LA, Canbay A, Syn WK. Extrahepatic complications of nonalcoholic fatty liver disease. Hepatology 2014;59:1174-1197.

35. De Souza LR, Berger H, Retnakaran R, Vlachou PA, Maguire JL, Nathens AB, et al. Non-alcoholic fatty liver disease in early pregnancy predicts dysglycemia in mid-pregnancy: prospective study. Am J Gastroenterol 2016;111:665-670.

36. Mousa N, Abdel-Razik A, Shams M, Sheta T, Zakaria S, Shabana W, et al. Impact of non-alcoholic fatty liver disease on pregnancy. Br J Biomed Sci 2018;75:197-199.

37. Sharma DL, Lakhani HV, Klug RL, Snoad B, El-Hamdani R, Shapiro JI, et al. Investigating molecular connections of non-alcoholic fatty liver disease with associated pathological conditions in West Virginia for biomarker analysis. J Clin Cell Immunol 2017;8:523.

38. Misu H, Takamura T, Takayama H, Hayashi H, Matsuzawa-Nagata N, Kurita S, et al. A liver-derived secretory protein, selenoprotein P,

causes insulin resistance. Cell Metab 2010;12:483-495.

39. Peverill W, Powell LW, Skoien R. Evolving concepts in the pathogenesis of NASH: beyond steatosis and inflammation. Int J Mol Sci 2014;15:8591-8638.

40. Hershman M, Mei R, Kushner T. Implications of nonalcoholic fatty liver disease on pregnancy and maternal and child outcomes. Gastroenterol Hepatol (N Y) 2019;15:221-228.

41. Liu H, Li J, Leng J, Wang H, Liu J, Li W, et al. Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China. Diabetes Metab Res Rev 2021;37:e3397.

42. Zheng T, Ye W, Wang X, Li X, Zhang J, Little J, et al. A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women. BMC Pregnancy Childbirth 2019;19:252.

43. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018;131:129-133.

44. Koivusalo SB, Rönö K, Klemetti MM, Roine RP, Lindström J, Erkkola M, et al. Gestational diabetes mellitus can be prevented by lifestyle intervention: the finnish gestational diabetes prevention study (RADIEL): a randomized controlled trial. Diabetes Care 2016;39:24-30.

45. Wang C, Wei Y, Zhang X, Zhang Y, Xu Q, Sun Y, et al. A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women. Am J Obstet Gynecol 2017;216:340-351.

46. Song C, Li J, Leng J, Ma RC, Yang X. Lifestyle intervention can reduce the risk of gestational diabetes: a meta-analysis of randomized controlled trials. Obes Rev 2016;17:960-969.

47. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis 2019;11(Suppl 4):S574-S584.

48. Bril F, Ortiz-Lopez C, Lomonaco R, Orsak B, Freckleton M, Chintapalli K, et al. Clinical value of liver ultrasound for the diagnosis of nonalcoholic fatty liver disease in overweight and obese patients. Liver Int 2015;35:2139-2146.

49. Khov N, Sharma A, Riley TR. Bedside ultrasound in the diagnosis of nonalcoholic fatty liver disease. World J Gastroenterol 2014;20:6821-6825.

50. Kabarra K, Golabi P, Younossi ZM. Nonalcoholic steatohepatitis: global impact and clinical consequences. Endocr Connect 2021;10:R240-R247.