

Predictive models using machine learning to identify fetal growth restriction in pre-eclampsia patients: Development and Evaluation Study

Qing Hua, Fengchun Yang, Yadan Zhou, Fenglian Shi, Xiaoyan You, Jing Guo, Li
Li

Submitted to: Journal of Medical Internet Research
on: December 29, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 24

 Figures 25

 Figure 1..... 26

 Figure 2..... 27

 Figure 3..... 28

 Figure 4..... 29

 Figure 5..... 30

 Figure 6..... 31

 Figure 7..... 32

 Figure 8..... 33

 Multimedia Appendixes 34

 Multimedia Appendix 1..... 35

Predictive models using machine learning to identify fetal growth restriction in pre-eclampsia patients: Development and Evaluation Study

Qing Hua^{1*}; Fengchun Yang^{2, 3*}; Yadan Zhou¹; Fenglian Shi¹; Xiaoyan You¹; Jing Guo¹; Li Li^{1*} A

¹Department of Obstetrics and Gynecology Zhengzhou Central Hospital Affiliated to Zhengzhou University Zhengzhou CN

²Institute of Medical Information Chinese Academy of Medical Sciences & Peking Union Medical College Beijing CN

³Department of Infectious Diseases, Tongji Hospital Tongji Medical College and State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Disease, Huazhong University of Science and Technology Wuhan CN

*these authors contributed equally

Corresponding Author:

Li Li A

Department of Obstetrics and Gynecology

Zhengzhou Central Hospital Affiliated to Zhengzhou University

No.16 Tongbai North Road, Zhongyuan District

Zhengzhou

CN

Abstract

Background: Fetal growth restriction (FGR) is a common complication of preeclampsia (PE). FGR in patients with PE increases the risk of neonatal perinatal mortality and morbidity. However, early prediction methods for FGR are biased or clinically unexplainable, which is difficult to apply to clinical practice, leading to a relative delay in intervention and a lack of effective treatments.

Objective: The study aims to develop an auxiliary diagnostic model based on machine learning (ML) to predict the occurrence of FGR in PE patients.

Methods: This study utilized a retrospective case-control approach to analyze 38 features, including the basic medical history and peripheral blood laboratory test results of pregnant patients with PE, either or not complicated by FGR. ML models were constructed to evaluate the predictive value of maternal parameter changes on the PE combined with FGR. Multiple algorithms were tested, including logistic regression, light gradient boosting (LGB), random forest (RF), extreme gradient boosting (XGB), multilayer perceptron (MLP), naive Bayes (NB) and support vector machine (SVM). The model performance was identified by the area under the receiver operating characteristic curve (AUC). The SHapley Additive exPlanation method (SHAP) was adopted to rank the feature importance and explain the final model for clinical application.

Results: The random forest (RF) model performed best in discriminative ability among the 7 ML models. After reducing features according to importance rank, an explainable final RF model was established with 9 features, including Urinary protein quantification (UPQ), Gestational week of delivery (GWD), Umbilical artery Systolic-to-Diastolic Ratio (S/D), Amniotic fluid Index (AFI), Triglyceride, D-dimer, Weight, Height and Maximum systolic pressure (MSP). The model could accurately predict FGR for 513 PE patients in the training and testing dataset (AUC = 0.83 ± 0.03 , $n=149$ FGR and 364 without FGR) using 5-fold cross-validation, which was closely validated for 103 PE patients in an external dataset (AUC = 0.82 ± 0.048 , $n=45$ FGR and 58 without FGR). On the whole, UPQ, Umbilical artery S/D and GWD exhibited the highest contributions to the model performance ($c=0.45, 0.34$ and 0.33) based on SHAP analysis. For specific individual patient, SHAP analysis can reveal the protective and risk factors to develop RGR for interpreting the model's clinical significance. Finally, the model has been translated into a convenient webpage tool to facilitate its utility in clinical settings.

Conclusions: We successfully developed a model that accurately predicts FGR development in patients with PE. The SHAP method captures highly relevant risk factors for model interpretation, alleviating concerns about the "black box" problem of indirect interpretation of ML techniques.

(JMIR Preprints 29/12/2024:70068)

DOI: <https://doi.org/10.2196/preprints.70068>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org/](#)

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Predictive models using machine learning to identify fetal growth restriction in pre-eclampsia patients: Development and Evaluation Study

Qing Hua^{1*}, Fengchun Yang^{2,3*}, Yadan Zhou¹, Fenglian Shi¹, Xiaoyan You, Jing Guo, Li Li¹

¹Department of Obstetrics and Gynecology, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, Henan, China

²Institute of Medical Information, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

³Department of Infectious Diseases, Tongji Hospital, Tongji Medical College and State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Disease, Huazhong University of Science and Technology, Wuhan, China

* These authors contributed equally

Corresponding author,

LiLi, Department of Obstetrics and Gynecology, Zhengzhou Central Hospital Affiliated to Zhengzhou University, No.16 Tongbai North Road, Zhongyuan District, Zhengzhou, 450007, Henan, China. Phone: 86-13683816225; E-mail: zzsylili@zzu.edu.cn;

Abstract

Background: Fetal growth restriction (FGR) is a common complication of preeclampsia (PE). FGR in patients with PE increases the risk of neonatal perinatal mortality and morbidity. However, previous prediction methods for FGR are class biased or clinically unexplainable, which is difficult to apply to clinical practice, leading to a relative delay in intervention and a lack of effective treatments.

Objective: The study aims to develop an auxiliary diagnostic model based on machine learning (ML) to predict the occurrence of FGR in PE patients.

Methods: This study utilized a retrospective case-control approach to analyze 38 features, including the basic medical history and peripheral blood laboratory test results of pregnant patients with PE, either or not complicated by FGR. ML models were constructed to evaluate the predictive value of maternal parameter changes on the PE combined with FGR. Multiple algorithms were tested, including logistic regression, light gradient boosting (LGB), random forest (RF), extreme gradient boosting (XGB), multilayer perceptron (MLP), naive Bayes (NB) and support vector machine (SVM). The model performance was identified by the area under the receiver operating characteristic curve (AUC) and other evaluation indexes. The SHapley Additive exPlanation method (SHAP) was adopted to rank the feature importance and explain the final model for clinical application.

Results: The RF model performed best in discriminative ability among the 7 ML models. After

reducing features according to importance rank, an explainable final RF model was established with 9 features, including Urinary protein quantification (UPQ), Gestational week of delivery (GWD), Umbilical artery Systolic-to-Diastolic Ratio (S/D), Amniotic fluid Index (AFI), Triglyceride, D-dimer, Weight, Height and Maximum systolic pressure (MSP). The model could accurately predict FGR for 513 PE patients (149 with FGR and 364 without) in the training and testing dataset ($AUC = 0.83 \pm 0.03$) using 5-fold cross-validation, which was closely validated for 103 PE patients (45 with FGR and 58 without) in an external dataset ($AUC = 0.82 \pm 0.048$). On the whole, UPQ, Umbilical artery S/D and GWD exhibited the highest contributions to the model performance ($c=0.45, 0.34$ and 0.33) based on SHAP analysis. For specific individual patient, SHAP results reveal the protective and risk factors to develop FGR for interpreting the model's clinical significance. Finally, the model has been translated into a convenient webpage tool to facilitate its utility in clinical settings.

Conclusions: We successfully developed a model that accurately predicts FGR development in PE patients. The SHAP method captures highly relevant risk factors for model interpretation, alleviating concerns about the “black box” problem of ML techniques.

Key words

machine learning; fetal growth restriction; preeclampsia; random forest; SHAP

Introduction

Pre-eclampsia (PE) is a hypertensive disorder of pregnancy that originates at the maternal-fetal interface and affects multiple organ systems with an incidence rate ranging from 2% to 8%[1, 2]. Severe forms of preeclampsia can lead to complications such as maternal death, fetal growth restriction and stillbirth[3]. A research data from China revealed that the incidence rate of fetal growth restriction (FGR) in early-onset preeclampsia is as high as 59.1%[4].

FGR is defined as the failure of a fetus to achieve its genetic growth potential [5]. They have an increased risk of perinatal or long-term complications, such as fetal asphyxia, neurodevelopmental disabilities, cardiovascular disease and type 2 diabetes, compared with normally grown fetuses[6, 7]. It is often accompanied by pre-eclampsia, which may result in increased perinatal mortality. Thus the early identification and intervention of FGR are crucial for improving its perinatal outcomes. However, early prediction methods are limited, and even in developed countries, more than 50% of FGR cases are not detected before birth[8], leading to a relative delay in intervention and a lack of effective treatments.

Ultrasound measurements have been widely validated for identifying FGR, yet there is still a certain rate of missed diagnosis[9]. This is because ultrasound measurements serve as bone markers, ignoring soft tissues like fetal fat and muscle, which can lead to estimation errors[10]. Besides, errors

can also arise from the selection of section surfaces and the influence of fetal position[11]. If a feasible prediction method can be identified, it would be highly valuable in improving outcomes for PE complicated by FGR.

An appropriate prediction model is needed for PE patients accompanied by FGR. Several studies have attempted to develop models for the prediction of FGR. For example, Feng et al. established a combined first-and second-trimester prediction model for screening late-onset FGR in fetuses using multivariate logistic regression[12]. Machine learning (ML) algorithms are increasingly favored for such tasks due to their superior ability to handle complex, nonlinear relationships[13]. For instance, Huang et al. developed an artificial neural network model to predicting the occurrence of PE complicated by FGR based on maternal peripheral blood parameters and clinical indicators[14]. However, most existing models always were developed from datasets with highly imbalanced classes[15]. Imbalanced data causes model bias, overfitting, feature importance distortion, and unreliable performance metrics due to the dominance of the majority class. This issue is particularly pronounced in PE and FGR prediction, where class imbalance can lead to overly optimistic model performance estimates[16]. In addition, previous studies tend to focus narrowly on specific subgroups, such as early- or late-onset FGR or PE[17], limiting their generalizability. Furthermore, since the inherent “black box” of ML models prevents clear interpretation of predictive features, which poses a barrier to clinical adoption. In summary, existing prediction models suffer from narrow population scope, class imbalance, and lack of clinical interpretability.

Considering that FGR often coexists with PE, it is of great significance to predict the occurrence of FGR in patients with PE. Thus, the study aims to develop an auxiliary diagnostic model based on ML to predict the occurrence of FGR in PE patients. Specifically, we constructed and optimized the ML model for predicting the occurrence of FGR in PE patients. Meanwhile, in order to facilitate clinical understanding and application, we adopted SHapley Additive exPlanation method (SHAP) method to capture highly relevant risk factors for model interpretation.

Methods

Data collect

We selected preeclampsia (PE) patients who gave birth in the Obstetrics Department of Zhengzhou Central Hospital affiliated to Zhengzhou University, from January 2021 to April 2024. The gestational age was determined based on the fetus size indicated by the last menstrual period and early pregnancy color ultrasound. If there was inconsistency between the two methods, the determination was made based on the results of the early pregnancy color ultrasound.

The diagnosis of PE is based on the ACOG criteria[2]: 1) systolic blood pressure ≥ 140 mmHg and/or diastolic blood pressure ≥ 90 mmHg after 20 weeks of gestation, accompanied by urine protein quantification of 300 mg/24 h, or random urine protein (+); 2) in the absence of proteinuria, preeclampsia can also be diagnosed if newly diagnosed hypertension occurs along with thrombocytopenia (platelet count $< 100 \times 10^9/L$), impaired liver function (transaminase two times higher than the normal upper limit), new development of renal insufficiency (serum creatinine $> 97 \mu\text{mol/L}$ or higher than two times the normal upper limit), pulmonary edema, or new onset cerebral or visual disturbances. FGR diagnostic criteria (Chinese population's birth weight chart) [7] is as follows: the estimated fetal weight below the 10th percentile compared with the population norms on growth charts. Specifically, fetuses whose birth weight is less than 10% of the standard growth value for the corresponding gestational week are defined as FGR, and the others are non-FGR.

After excluding patients with multiple pregnancies, test-tube baby, fetal chromosomal anomaly, miscarriage in the second trimester, stillbirth, and incomplete medical records, we included a total of 513 cases for analysis among 570 PE patients, consisting of 149 patients with FGR and 364 patients without FGR. The selection of research objects is shown in Figure 1. Following the same criteria as above, we also collected a dataset for external validation of the model. This dataset includes records from 103 pregnant women, among whom 45 delivered infants with FGR and 58 without FGR.

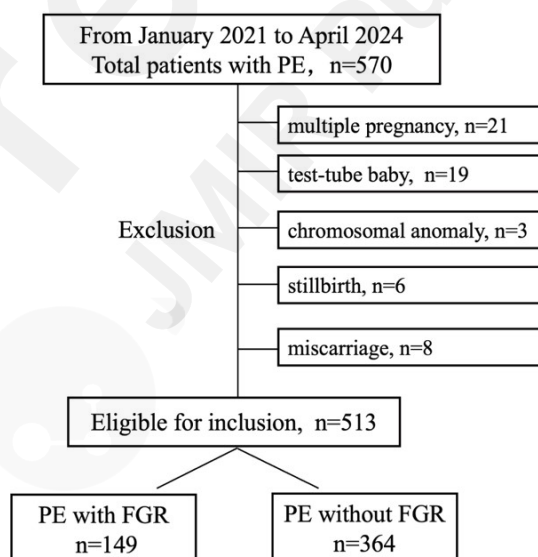


Figure 1 Flow chart of patient selection.

We initially identified 38 features potentially related to FGR through a literature review. Recorded patient characteristics included baseline characteristics, laboratory results in late pregnancy (before delivery) and the weight of neonates. Baseline characteristics included maternal age, gravidity, parity, pre-pregnancy body mass index (BMI), blood pressures, gestational ages, personal

history (PE, repeated abortions, missed abortion history), family history (hypertensive disorders), with or without chronic hypertension, co-existence of gestational diabetes mellitus during pregnancy, S/D value, amniotic fluid index, gestational age at birth. Relevant laboratory tests included urine protein quantitation, routine blood tests, hepatic and renal function, lipid profiles, coagulation function and serum haptoglobin. In past studies, these features of preeclampsia were strongly associated with FGR. Subsequently, we used feature importance to select more critical features for the final model construction.

Ethics Approval

The protocol was approved by the Ethics Committee of Zhengzhou Central Hospital Affiliated to Zhengzhou University. Considering its retrospective nature, the requirement for informed consent was waived. All relevant clinical data and outcome information were collected from the hospital's electronic medical record system.

Model Training and validation

Before model training, we applied identical preprocessing steps to the training, validation, and test sets. First, we removed variables with missing rates exceeding 10%, then imputed the remaining missing values using a random forest-based method-missForest. For continuous variables, we conducted standardization through min-max scaling. We divided the collected maternal data into a training set and a test set with 70% utilized for training and 30% for validation (internal validation), in order to avoid problems with overfitting. The SHapley Additive exPlanation method (SHAP) was adopted to rank the feature importance. For features selection, we added features to the model training process step by step according to the order of feature importance shown by SHAP.

For training models, we applied 7 widely used ML algorithms in assisted diagnosis prediction, including logistic regression[18], tree-based model (light gradient boosting machine [LightGBM]) [19], random forest (RF)[20], and extreme gradient boosting (XGBoost)[21], multilayer perceptron[22], naive Bayes[23], and support vector machine-based methods[24]. The logistic regression model represents the linear model, which is a state-of-the-art classification model for a baseline construction. RF, LightGBM, and XGBoost were representative ensemble learning models, which do well in dealing with multiple types of features. RF is a bagging ensemble algorithm containing multiple decision trees, which uses the voting method to classify samples and integrate the final voting results produced by multiple decision trees. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements ML algorithms under the gradient boosting framework. LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It uses a gradient-based one-sided sampling algorithm to reduce

the sample dimension and a mutually exclusive feature bundling algorithm to reduce the feature dimension.

The total of 38 features mentioned above were used to develop the prediction models. In order to optimize the prediction model, grid search combined with manual fine tuning was applied to obtain the final hyperparameters (see in the Supplementary material). Commonly used evaluation indexes, such as the area under the receiver-operating-characteristic (ROC) curve (AUC), sensitivity (recall), specificity, accuracy, precision, recall, and F1-score were used to evaluate the reliability of these models. During the training of our prediction models, we employed the 5-fold cross-validation method for model training. The SHAP method was able to select clinically important features that contribute significantly to model predictions. It can improve the robustness of the model to a certain extent by removing irrelevant or noisy features and avoiding the feature collinearity problem. Finally, the prediction model was further evaluated in an external dataset.

Model Interpretation

To enhance the clinical applicability, we combined the diagnostic prediction results with clinical feature importance explanations to identify a PE patient with FGR. In this study, except for interpreting feature contributions, SHAP values were used to assess the clinical significance of predictive models[25].

For the ML model interpretation, we used both globally explainable and locally explainable methods[26]. The global explanation determines the importance of features by comparing the magnitude of the model prediction error change before and after replacing a feature. If the prediction error changes more, it indicates that the feature is more important. Meanwhile, we used SHAP values to explore individual-based decision-making processes in the view of local explanation.

Model Deployment

Webpage deployment tool based on streamlit framework to facilitate the utility of the model in clinical settings, the final prediction model was implemented into a web application established based on the Streamlit Python-based framework. When the values of corresponding features from the final model are provided, the application can return the probability of FGR and the force plot for the individual.

Results

Patient Characteristics

We collected information on 513 parturients and newborns, of which 149 were diagnosed with FGR. Among the information collected before delivery, 29 variables were continuous and 9 were discrete.

We statistically analyzed information on the group differences between these variables between newborns with FGR and those without. Demographics, pregnancy characteristics and neonatal outcomes for eligible preeclampsia patients are described in **Table 1**. Maternal demographics in regard to age, height, pre-pregnancy BMI have no difference between the 2 groups ($P \geq 0.05$) but differed in gestational age at diagnosis of preeclampsia, weight, weight gain during pregnancy, maximum systolic or diastolic pressure (mmHg), Umbilical artery Systolic-to-Diastolic Ratio (S/D) and amniotic fluid Index. Previous preeclampsia, onset period of hypertension, chronic hypertension, combined with gestational diabetes, history of FGR and family history of hypertension were strong risk factors for FGR in preeclampsia patients in our cohort. The following laboratory parameters with statistically significant differences were screened by intergroup difference analysis ($P < 0.05$): hemoglobin, total protein, albumin, uric acid, fibrinogen, total cholesterol and serum haptoglobin (HP) (**Table1**). Besides, there was also a difference in the sex of the newborns between the two groups.

Table 1 Comparison of demographic and clinical characteristics and outcomes between non-FGR and FGR in the cohort

	No-FGR group	FGR group	<i>P</i>
Age (years)	31.24(4.43)	31.81(4.55)	0.2085
Onset period of hypertension (weeks)	32.3(10.68)	30.91(9.73)	0.1883
Gestational age at diagnosis of preeclampsia (weeks)	36.73(3.71)	34.48(4.49)	0
Gestational week of delivery (weeks)	37.81(2.32)	35.45(3.6)	0
Maximum systolic pressure(mmHg)	152.95(15.92)	173.56(146.12)	0.0105
Maximum diastolic pressure(mmHg)	100.44(10.64)	111.79(87.52)	0.019
Height (cm)	161.28(4.75)	160.41(4.65)	0.0681
Weight (kg)	78.8(11.55)	74.98(10.91)	0.0009
Pre-pregnancy BMI (kg/m ²)	24.54(4.17)	23.89(3.86)	0.1159
Weight gain during pregnancy (kg)	15.05(7.59)	13.28(4.64)	0.0108
Urinary protein quantification (mg/24h)	1362.1(2310.7)	3032.6(3678.96)	0
Platelets (*10 ⁹ /L)	197.05(65.93)	188.67(60.65)	0.1976
Hematocrit (%)	36.12(20.46)	36.4(4.18)	0.8743
Hemoglobin (g/L)	116.33(15.84)	121.63(14.91)	0.0008
Total protein (g/L)	58.75(6.11)	57.15(7.19)	0.0139
Albumin (g/L)	31.79(4.38)	30.9(3.75)	0.0361
Globulin (g/L)	26.97(3.83)	26.45(3.97)	0.1814
ALT(U/L)	22.64(47.85)	26.44(35.88)	0.3984
AST(U/L)	31.5(24.74)	34.21(32.11)	0.3222
ALP(U/L)	163.68(85.42)	150.33(63.63)	0.0974
Total bilirubin (umol/L)	11.83(6.73)	13.44(24.19)	0.2576
Uric acid (umol/L)	369.88(101.3)	405.6(107.17)	0.0006
D-dimer (mg/L)	2.1(1.71)	2.12(3.71)	0.9469
Fibrinogen (mg/L)	4.33(1.07)	4.07(1.1)	0.0157

Total cholesterol (mmol/L)	6.48(1.7)	8.32(15.34)	0.0294
Triglyceride (mmol/L)	4.74(2.57)	4.62(2.79)	0.6393
serum haptoglobin (HP) (g/L)	0.7(0.35)	0.57(0.35)	0.0002
Amniotic fluid Index (mm)	115.5(36.75)	103.45(32.6)	0.0008
Umbilical artery S/D (free segment)	2.33(0.74)	2.73(0.85)	0
Chronic hypertension (hypertension diagnosed before 20 weeks of pregnancy)	428	85	0.764
History of FGR	500(97.4)	13(2.5)	0.000
History of cesarean section	384(74.8)	129(25.2)	0.8282
History of adverse pregnancy	416(81.1)	97(18.9)	0.0671
Primipara	219(42.6)	294(57.3)	0.990
Combined with gestational diabetes	365(71.2)	148(28.8)	0.5935
Family history of hypertension	424(82.6)	89(17.3)	0.9283
Fetal sex	243(47.3)	270(52.6)	0.0009

Continuous values were presented as median (Standard deviation). Categorical values were presented as number (Percentage).

Model development and performance comparison

The data collected were used to generate 7 ML models to predict FGR developed during the prenatal testing in pregnant woman. The discriminative performances (including AUC, Accuracy, Sensitivity, Specificity (Recall), Precision and F1-score) of these models are showed in **Table 2** and **Figure 2**. Among the 7 models, RF model (AUC = 0.812) had the best predictive effect for FGR, followed by LightGBM model (AUC = 0.806) and GBM model (AUC = 0.803). For the RF model, the SHAP summary plots of the top 20 features (**Figure 3**) for the RF model are presented.

Table 2 The discriminative performances of these 7 models

Model	AUC	Accuracy	Specificity (Recall)	Sensitivity	Precision	F1-score
RF	0.812 (0.751-0.872)	0.595 (0.449-0.76)	0.497 (0.225-0.852)	0.83 (0.511-0.997)	0.565 (0.415-0.795)	0.583 (0.441-0.787)
LGB	0.806 (0.744-0.861)	0.739 (0.709-0.784)	0.756 (0.726-0.802)	0.697 (0.588-0.883)	0.541 (0.481-0.643)	0.61 (0.563-0.745)
XGB	0.803 (0.727-0.862)	0.689 (0.595-0.801)	0.652 (0.471-0.874)	0.777 (0.602-0.96)	0.483 (0.467-0.997)	0.593 (0.545-0.627)
LR	0.733 (0.669-0.793)	0.653 (0.513-0.716)	0.676 (0.378-0.899)	0.595 (0.24-0.857)	0.431 (0.324-0.496)	0.504 (0.472-0.581)
MLP	0.701 (0.647-0.744)	0.645 (0.571-0.725)	0.662 (0.478-0.819)	0.603 (0.395-0.793)	0.421 (0.375-0.997)	0.502 (0.438-0.652)
SVM	0.699 (0.643-0.794)	0.345 (0.285-0.536)	0.104 (0.0-0.468)	0.933 (0.7-1.0)	0.303 (0.231-0.426)	0.452 (0.348-0.524)
NB	0.697 (0.651-0.746)	0.721 (0.706-0.737)	0.89 (0.864-0.916)	0.309 (0.208-0.397)	0.531 (0.475-0.623)	0.394 (0.298-0.473)

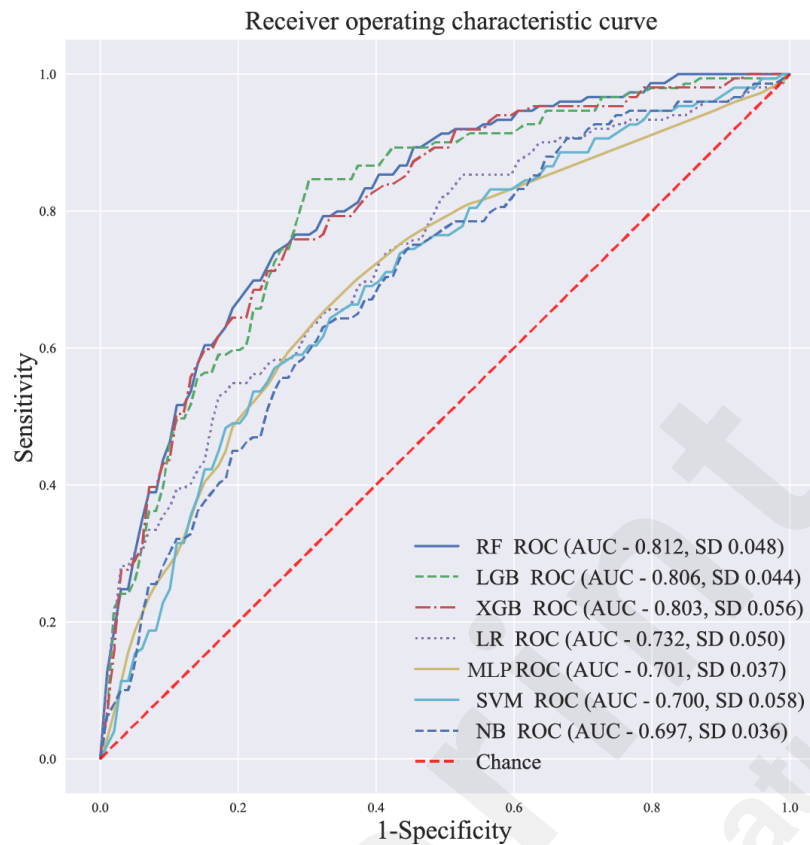


Figure 2 ROC curves of the 7 ML models. ROC: receiver operating curve; AUC: area under characteristic curve; SD: Standard deviation LGB: light gradient boosting machine; LR: logistic regression; MLP: multilayer perceptron; NB: naive Bayes; RF: random forest; SVM: support vector machine; XGB: extreme gradient boosting.

Figure 3 The SHAP summary plots of the top 20 features for the RF model.

Identification of the final model

In order to study the influence of the number of features selected during model construction on model performance, we added features to the model training process step by step according to the order of feature importance shown by SHAP. The final model was identified during the feature reduction of the RF model. As displayed in **Figure 4**, the 38-feature model was significantly better than 5-feature model ($\Delta\text{AUC} = 0.036$, $P = 0.004$), but not significantly better than the 10-feature model ($\Delta\text{AUC} = 0.011$, $P = 0.161$) in predicting of FGR prediction among the prenatal testing in pregnant woman.

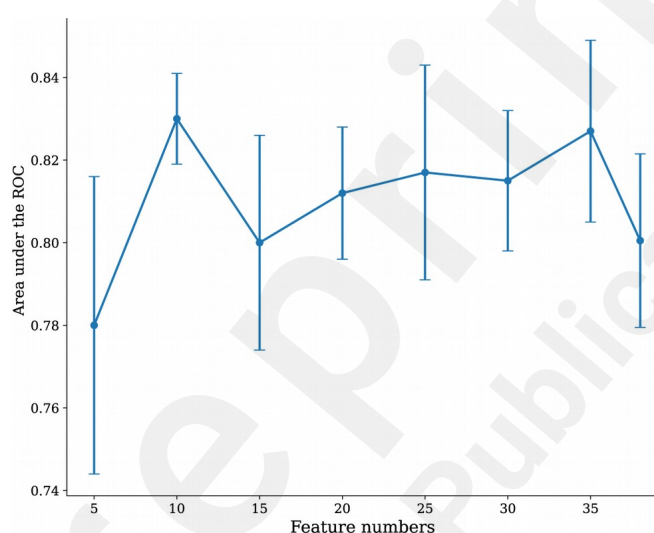


Figure 4 Random Forest model AUC value changes with different numbers of features.

The 10-feature model had a good net benefit and a high threshold probability, comparable to the 38-feature model. Hence, we focused on the 10-feature RF model. Considering the adverse effect of FGR predicted by fetal sex on sociodemographic sex ratio, we retained the other 9 features (including Urinary protein quantification(mg/24h), Gestational week of delivery(GWD), Umbilical artery S/D (free segment), Amniotic fluid Index (mm,AFI), Triglyceride (mmol/L), D-dimer (mg/L), Weight(kg), Height (cm), and Maximum systolic pressure (mmHg, MSP) as the final model for further analysis. For the model building process referred to 513 patients (**Figure 5A**), after 5-fold cross-validation, the final RF model achieved an AUC of 0.830 with a sensitivity of 0.886, a specificity of 0.866, an accuracy of 0.872 for predicting FGR the prenatal testing in pregnant woman. The external validation dataset was collected referred to 103 patients using the same standardized protocols and inclusion/exclusion criteria as the internal dataset. Statistical tests confirmed that no significant differences (all $p > 0.05$) existed in key variables between the two cohorts (Table S1). For

the external validation (**Figure 5B**), the final model gave an AUC of 0.820, which was similar to that in the internal validation ($\Delta\text{AUC} = 0.05$, $P = 0.477$), indicating that the final model showed great performance both in internal and external validations.

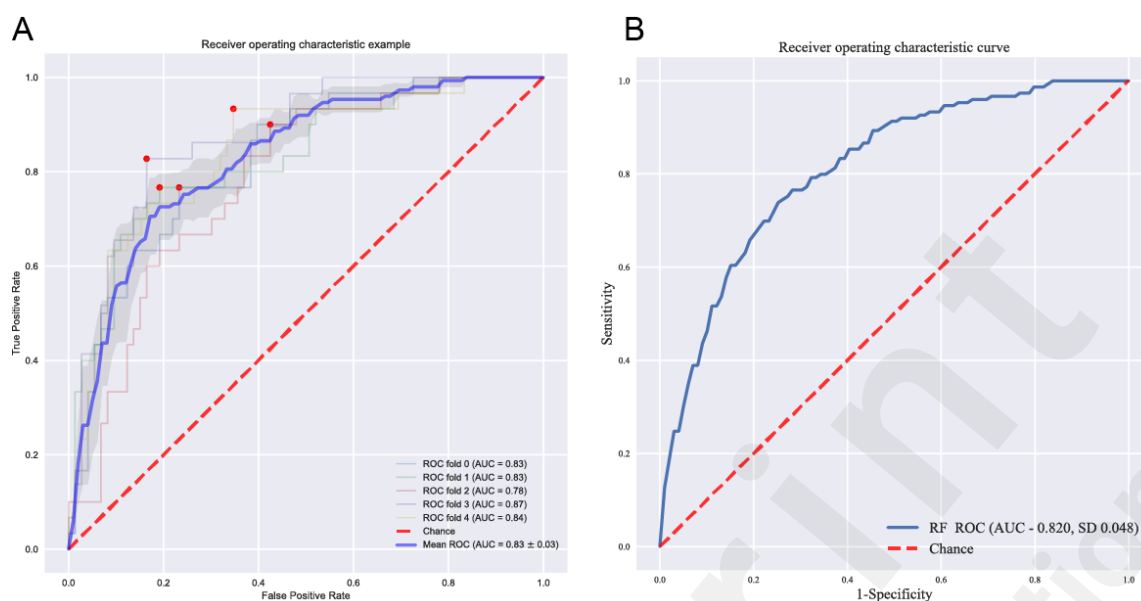


Figure 5 Performance of the 10-feature RF final model. (A) The 5-fold cross-validation AUC curve of the final model; (B) The external validation AUC curve.

Model explanation

Since it is difficult for clinicians to accept a prediction model that is not directly explainable and interpretable, the SHAP method is utilized to interpret the output of the final model by calculating the contribution of each variable to the prediction. This method could provide two types of explanations: global explanation of the model at the feature level and local explanation at the individual level. Global explanation described the overall functionality of the model. As shown in SHAP summary plots (**Figure 6**), the contributions of the feature to the model were evaluated using the average SHAP values and exhibited in descending order. Urinary protein quantification (UPQ), Umbilical artery S/D and Gestational week of delivery (GWD) showed the top three contributions ($c=0.45, 0.34, 0.33$) to the final model.

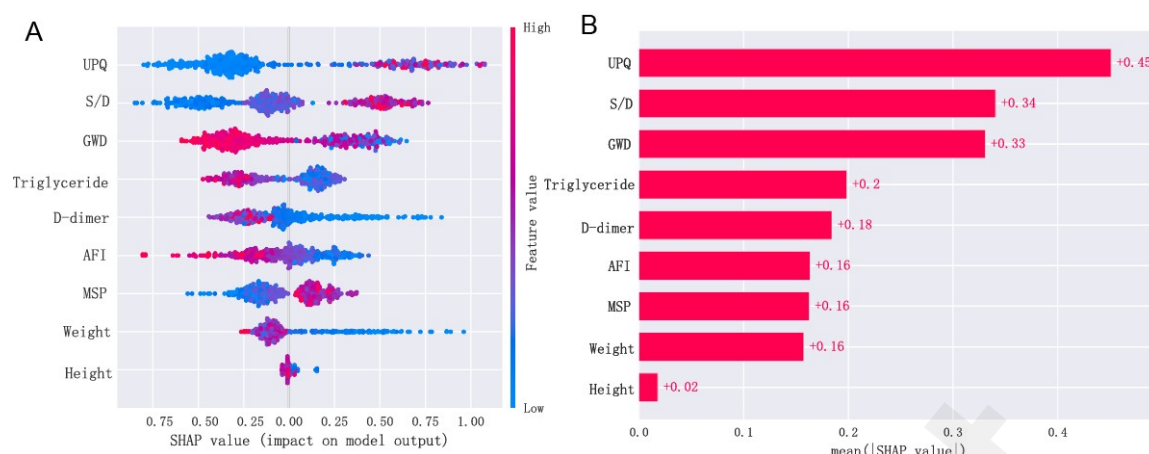


Figure 6 Global model explanation by the SHAP method. (A) SHAP summary dot plot. (B) SHAP summary bar plot.

Additionally, the SHAP dependence plot can facilitate understanding how a single feature affects the output of the prediction model. The real values versus the SHAP values of these 9 features are shown in **Figure 7**, and SHAP values that are higher than zero correspond to a positive class prediction in the model, in other words, a higher risk of FGR. For instance, pregnant woman with an Urinary protein quantification (UPQ) > 2500ml or a Maximum systolic pressure (mmHg) > 150 mmHg had SHAP values higher than zero, which pushed the decision towards the “FGR” class. In addition, Gestational week of delivery >37 Week or a Amniotic fluid Index <100 mm pushed the decision towards the “non-FGR” class, as well as Triglyceride (mmol/L) ≤5 or a high actual value ≥3 of Umbilical artery S/D (free segment).

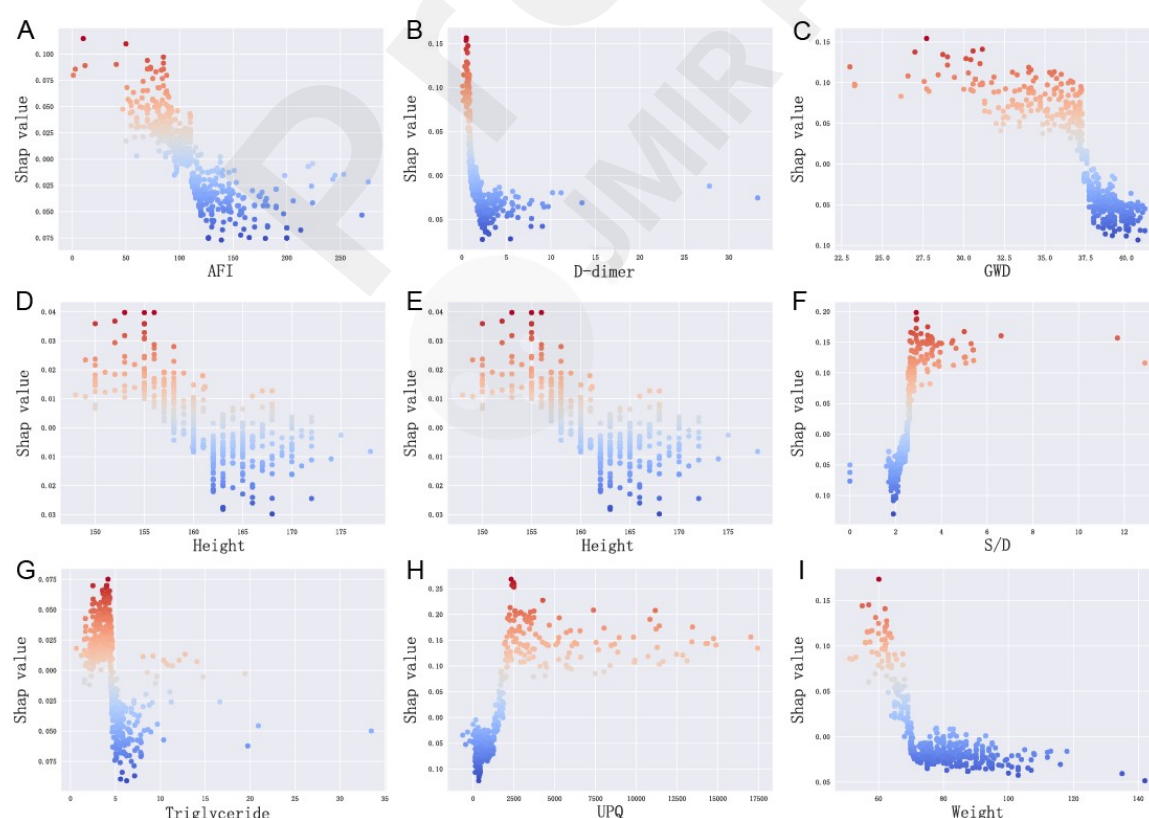


Figure 7 SHAP dependence plot. Each dependence plot shows how a single feature affects the output of the prediction model, and each dot represents a single patient. The darker the red color of the dot, the higher the risk of FGR, while the darker the blue color of the dot, the lower the FGR risk.

Furthermore, local explanation analyzed how a certain prediction was made for a specific individual by incorporating the individualized input data. The actual measured values of features were also displayed in the waterfall plot (**Figure 8**). **Figure 8A** showed a patient whose child did not develop FGR during the pregnant stay. As observed, the values of GWD (weeks), Umbilical artery S/D (free segment), Maximum systolic pressure (mmHg), Urinary protein quantification (mg/24h), D-dimer (mg/L), and body weight pushed the decision towards the “non-FGR” class, but Triglyceride did not. **Figure 8B** shows another patient whose child developed FGR during pregnancy. At present, multiple indicators such as abnormal urinary protein quantification (mg/24h) and Umbilical artery S/D (free segment) can increase the risk of FGR, pushed the decision towards the “FGR” class.

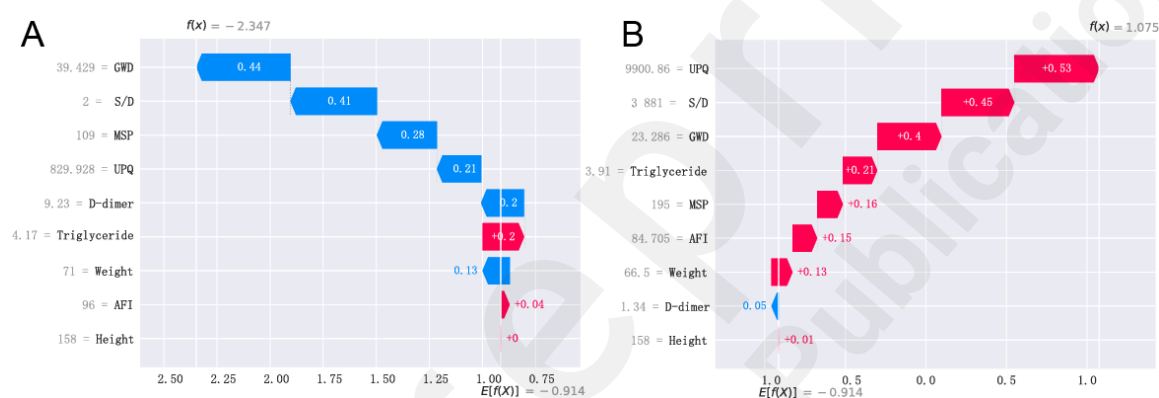


Figure 8 Local model explanation by the SHAP method. (A) PE without FGR; (B) PE with FGR.

Convenient application for clinical utility

The final prediction model was implemented into the web application to facilitate its utility in clinical scenarios. When the actual values of the 9 features required for the model are entered, this application will automatically predict the risk of FGR for an individual child. The web application is accessible online at <https://predictionmodel-for-FGR.streamlit.app>.

Discussion

Principal Findings

Among the 7 ML models used to predict FGR from PE patients, the RF model had the best performance with a good net benefit and a high threshold probability in feature reduction. Thus we employed the RF algorithm to develop the final model with 10 features. These features can be easily obtained or evaluated during pregnancy, so the model is promising as an early identification tool for the occurrence of FGR in pregnant women with PE.

In our study, training and testing data were collected from real clinical environments, ensuring that the model is exposed to a diverse and representative dataset. The random sampling of data helps to minimize bias and ensures that the model is robust to variations in input data. And the model's performance on real-world data validates its ability to generalize well and maintain high predictive accuracy under different input assumptions.

The ML technique has been described as a “black box” with little explanation about how predictions are derived. This may result in clinicians refusing to use it because they are hesitant to make medical decisions based on opaque information. This brought up another advantage of this study: we utilized the SHAP approach to explain the “black-box” of ML models. The SHAP method could provide a global explanation that describes the overall functionality of a model and a local explanation that details how a certain prediction is made for an individual by inputting the individualized data. Moreover, with a convenient tool based on the Streamlit framework, this prediction model can be used on the webpage and shared with more clinicians.

Our final model performed well both in internal and external validations, with an AUC of 0.83 and 0.82, respectively. Except for Fetal sex, there are 9 main characteristics for predicting FGR in our model, including Urinary protein quantification, Maximum systolic pressure, Triglyceride, Gestational week, D-dimer, Weight and Height of pregnant woman, Umbilical artery S/D, and Amniotic fluid Index. Urinary protein quantity and blood pressure are important indicators to reflect the severity of PE. Our prediction model found a positive correlation between urinary protein quantity, maximum systolic pressure and FGR, especially when the 24-hour urinary protein quantity was greater than 2,500 mg and the maximum systolic pressure was greater than 160mmHg.

A Mendelian randomization study has revealed a strong correlation between triglyceride levels in pregnant women's blood and the occurrence of preeclampsia[27]. Previous research has found that patients with PE have higher serum triglyceride levels than normal pregnant women[28, 29], and those with FGR have significantly higher levels than those without FGR[30]. Our study further discovered that when triglyceride levels in patients with PE exceed 5 mmol/L, the occurrence of FGR increases. The mechanism may be due to the presence of dyslipidemia may result in shallow chorionic villus implantation, ultimately leading to reduced blood flow to the placenta and uterus. This, in turn, can hinder nutrient supply and fetal development, thereby increasing the likelihood of FGR occurrence in patients with PE[14, 31, 32]. M Bozkurt et.al discovered that compared with normal pregnant women, preeclampsia patients had higher D-dimer levels[33]; other studies have also found a correlation between D-dimer levels and the severity of preeclampsia[34]. These findings provide strong support for our study that the risk of fetal growth restriction (FGR) increases when D-

dimer levels in preeclampsia patients reach 2.5mg/L. Among ultrasound indicators of fetal, AFI and S/D value can reflect the occurrence of FGR. The AFI was negatively correlated with the occurrence of FGR, especially when it was less than 90mm;The probability of FGR is significantly increased when S/D value is greater than 3.

Class imbalance can indeed affect the performance of ML Models in multiple aspects[35]. Sampling methods such as SMOTE and Resampling (including undersampling and oversampling) can provide solutions to this problem[17]. In this study, the number of positive (PE complicated with FGR) and negative (PE without FGR) sample are 149 and 364 in model building process, as well as 45 and 58 in external dataset. The ratios between the classes are 1:2.5 and 1:1.3, respectively, those are much lower than the commonly recognized class imbalance threshold of 1:10. Therefore, we believe that the class imbalance in this study has an extremely negligible impact on the construction of the model, and thus no special treatment is carried out.

For future FGR prediction, integrating current models based on maternal physiological, ultrasound, and genetic data into real-time clinical decision support systems is vital. These systems, connected to hospital Electronic Health Records (EHRs), can provide instant risk assessments during prenatal check-ups, guiding pregnancy management. Additionally, exploring new features like pregnant women's metabolomic profiles, maternal-fetal interface microbiomes, and longitudinal placental data can enhance model accuracy. Incorporating these into existing models will likely lead to more precise FGR prediction and better maternal-fetal outcomes.

We propose integrating the predictive model into the hospital information system (HIS). During routine prenatal check-ups, healthcare professionals can collect data through the information system, after which the model will automatically generate predictions. Healthcare providers can then perform interventions or follow-ups based on these results. Furthermore, using model interpretation techniques to explain the predictions will provide crucial decision-making support for clinicians. For high-risk pregnancies, doctors can develop individualized management plans in advance, such as enhanced monitoring, nutritional guidance, and medication, to improve neonatal growth and prognosis.

Limitations

The study has several limitations. First, clinical data of PE patients was retrospectively collected from electronic records, it's necessary to collect some prospective cases and observe the follow-up results to verify the accuracy of the model. Second, due to the lack of guidelines or consensus for selecting features for the prediction model, how many features should be included in the model remains elusive. Although more features may provide more information for the prediction model,

including a large number of features may limit the clinical use of the model, and including non-causal features may reduce the accuracy of the prediction. Nevertheless, further studies are needed to simplify and validate the application of the model in clinical conditions.

Conclusion

The prediction of fetal growth restriction in patients with preeclampsia offers a valuable guide for improving perinatal outcomes. Combined with basic medical history and peripheral blood laboratory test results, this machine prediction model exhibits a reliable predictive value in predicting the occurrence of preeclampsia complicated by fetal growth restriction. The model could be employed as a decision-making tool to support clinicians.

Acknowledgments

This work was supported by the medical science and technology research project of Henan Province (LHGJ20240966), the key scientific research projects of colleges and universities of Henan Province (24A320037) and the scientific and technological research projects of Henan province (242102310008).

Author contributions

Li Li contributed to the conception and design of the study. Qing Qua and Fengchun Yang equally contributed to analyze, interpret the data and write the manuscript; Yadan Zhou, Fenglian Shi, Xiaoyan You and Jing Guo contributed to data collection and manuscript revision. All authors read and approved the final manuscript.

Conflict of interest

The authors declare that there are no conflicts of interest in this research.

Data Availability Statement

The datasets generated and/or analyzed during this study are not publicly available due to patient privacy concerns. Requests for access to the data may be sent to email address of corresponding author. The use of the data is subject to relevant privacy regulations and is restricted to non-commercial research purposes. Applicants are required to provide a valid reason and usage plan for our review.

Reference

1. Ives, C.W., et al., *Preeclampsia-Pathophysiology and Clinical Presentations: JACC State-of-the-Art Review*. J Am Coll Cardiol, 2020. **76**(14): p. 1690-1702.
2. ACOG Practice Bulletin No. 202: *Gestational Hypertension and Preeclampsia*. Obstet Gynecol, 2019. **133**(1): p. 1.
3. Jim, B. and S.A. Karumanchi, *Preeclampsia: Pathogenesis, Prevention, and Long-Term Complications*. Semin Nephrol, 2017. **37**(4): p. 386-397.
4. Shen, H., et al., *Severe early-onset PE with or without FGR in Chinese women*. Placenta,

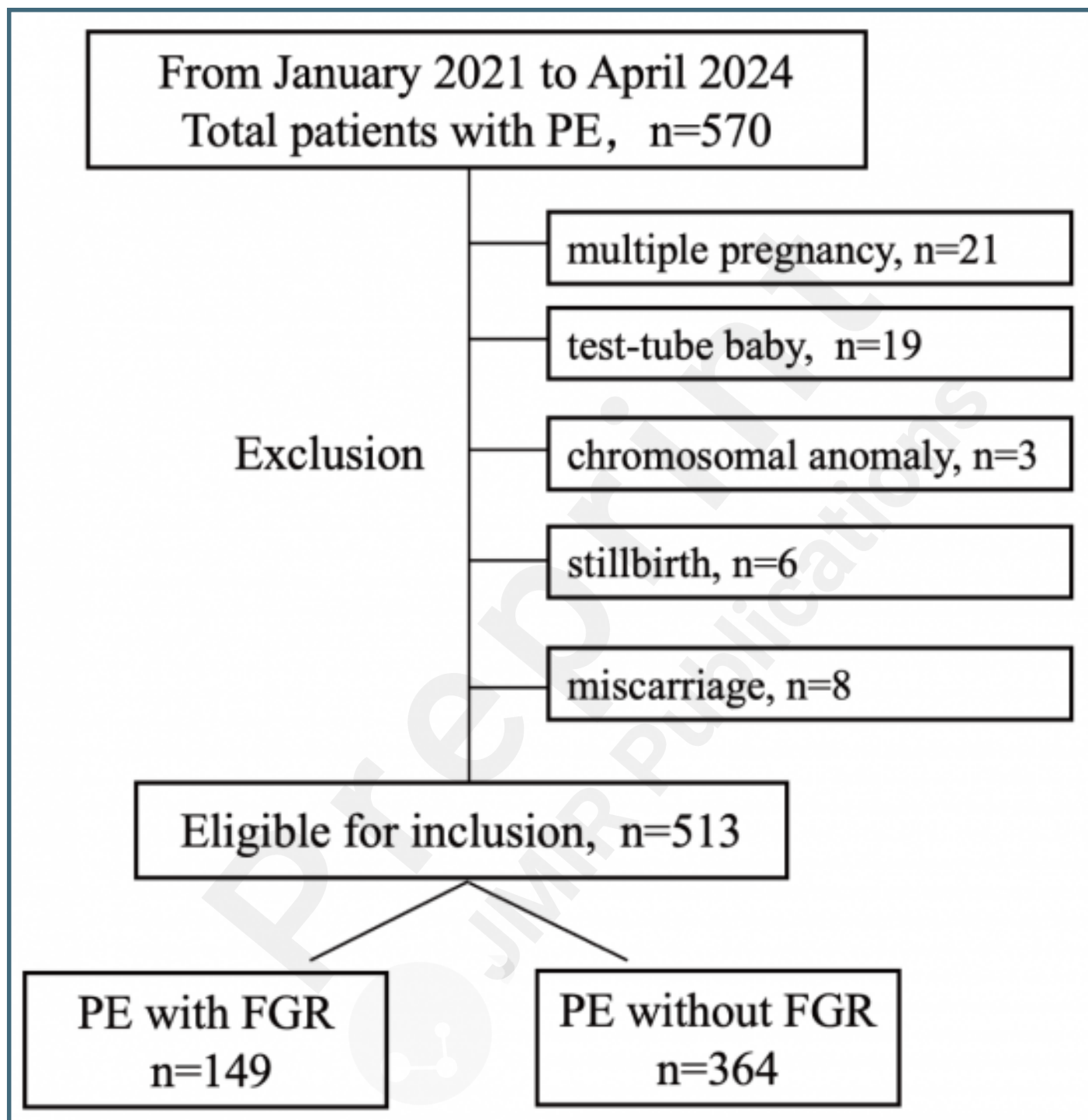
2020. **101**: p. 108-114.
5. Sharma, D., et al., *Intrauterine growth restriction - part 1*. J Matern Fetal Neonatal Med, 2016. **29**(24): p. 3977-87.
6. McIntire, D.D., et al., *Birth weight in relation to morbidity and mortality among newborn infants*. N Engl J Med, 1999. **340**(16): p. 1234-8.
7. *Fetal Growth Restriction: ACOG Practice Bulletin, Number 227*. Obstet Gynecol, 2021. **137**(2): p. e16-e28.
8. Malhotra, A., et al., *Neonatal Morbidities of Fetal Growth Restriction: Pathophysiology and Impact*. Front Endocrinol (Lausanne), 2019. **10**: p. 55.
9. Caradeux, J., et al., *Diagnostic performance of third-trimester ultrasound for the prediction of late-onset fetal growth restriction: a systematic review and meta-analysis*. Am J Obstet Gynecol, 2019. **220**(5): p. 449-459 e19.
10. Mustafa, H.J., et al., *Diagnostic performance of 32 vs 36 weeks ultrasound in predicting late-onset fetal growth restriction and small-for-gestational-age neonates: a systematic review and meta-analysis*. Am J Obstet Gynecol MFM, 2024. **6**(1): p. 101246.
11. Sehgal, A., et al., *Clinician performed ultrasound in fetal growth restriction: fetal, neonatal and pediatric aspects*. J Perinatol, 2017. **37**(12): p. 1251-1258.
12. Feng, Y., et al., *Prediction of late-onset fetal growth restriction using a combined first- and second-trimester screening model*. J Gynecol Obstet Hum Reprod, 2022. **51**(2): p. 102273.
13. Agarwal, V., et al., *Learning statistical models of phenotypes using noisy labeled training data*. J Am Med Inform Assoc, 2016. **23**(6): p. 1166-1173.
14. Huang, K.H., et al., *Prediction of pre-eclampsia complicated by fetal growth restriction and its perinatal outcome based on an artificial neural network model*. Front Physiol, 2022. **13**: p. 992040.
15. Mula, R., et al., *Angiogenic factors assessment in pre-eclampsia high-risk population for the prediction of small-for-gestational age neonates: A prospective longitudinal study*. Int J Gynaecol Obstet, 2023. **161**(2): p. 439-446.
16. Chicco, D., *Ten quick tips for machine learning in computational biology*. BioData Min, 2017. **10**: p. 35.
17. Ren, Y., et al., *Issue of Data Imbalance on Low Birthweight Baby Outcomes Prediction and Associated Risk Factors Identification: Establishment of Benchmarking Key Machine Learning Models With Data Rebalancing Strategies*. J Med Internet Res, 2023. **25**: p. e44081.
18. Stoltzfus, J.C., *Logistic regression: a brief primer*. Acad Emerg Med, 2011. **18**(10): p. 1099-104.
19. Yan, J., et al., *LightGBM: accelerated genomically designed crop breeding through ensemble learning*. Genome Biol, 2021. **22**(1): p. 271.
20. Chen, X. and H. Ishwaran, *Random forests for genomic data analysis*. Genomics, 2012. **99**(6): p. 323-9.
21. Zou, Y., et al., *Extreme gradient boosting model to assess risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: Individual prediction using SHapley Additive exPlanations*. Comput Methods Programs Biomed, 2022. **225**: p. 107038.
22. Seidel, P., A. Seidel, and O. Herbarth, *Multilayer perceptron tumour diagnosis based on chromatography analysis of urinary nucleosides*. Neural Netw, 2007. **20**(5): p. 646-51.
23. Sugahara, S. and M. Ueno, *Exact Learning Augmented Naive Bayes Classifier*. Entropy (Basel), 2021. **23**(12).
24. Pal, S., et al., *A support vector machine-based cure rate model for interval censored data*. Stat Methods Med Res, 2023. **32**(12): p. 2405-2422.
25. Loh, H.W., et al., *Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022)*. Comput Methods Programs Biomed, 2022. **226**: p. 107161.
26. Champendal, M., et al., *A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging*. Eur J Radiol, 2023. **169**: p. 111159.
27. Hosier, H., et al., *Dyslipidemia and Risk of Preeclampsia: A Multiancestry Mendelian Randomization Study*. Hypertension, 2023. **80**(5): p. 1067-1076.
28. Ray, J.G., et al., *Brief overview of maternal triglycerides as a risk factor for pre-eclampsia*. BJOG, 2006. **113**(4): p. 379-86.

29. Liu, Q., et al., *Elevated mid-trimester 4-h postprandial triglycerides for predicting late-onset preeclampsia: a prospective screening study*. J Transl Med, 2022. **20**(1): p. 81.
30. Chassen, S.S., et al., *Altered Cord Blood Lipid Concentrations Correlate with Birth Weight and Doppler Velocimetry of Fetal Vessels in Human Fetal Growth Restriction Pregnancies*. Cells, 2022. **11**(19).
31. Opichka, M.A., et al., *Vascular Dysfunction in Preeclampsia*. Cells, 2021. **10**(11).
32. Jung, E., et al., *The etiology of preeclampsia*. Am J Obstet Gynecol, 2022. **226**(2S): p. S844-S866.
33. Bozkurt, M., et al., *Troponin I and D-Dimer levels in preeclampsia and eclampsia: prospective study*. Clin Exp Obstet Gynecol, 2015. **42**(1): p. 26-31.
34. Shao, H., et al., *The association of antenatal D-dimer and fibrinogen with postpartum hemorrhage and intrauterine growth restriction in preeclampsia*. BMC Pregnancy Childbirth, 2021. **21**(1): p. 605.
35. Lanera, C., et al., *Screening PubMed abstracts: is class imbalance always a challenge to machine learning?* Syst Rev, 2019. **8**(1): p. 317.

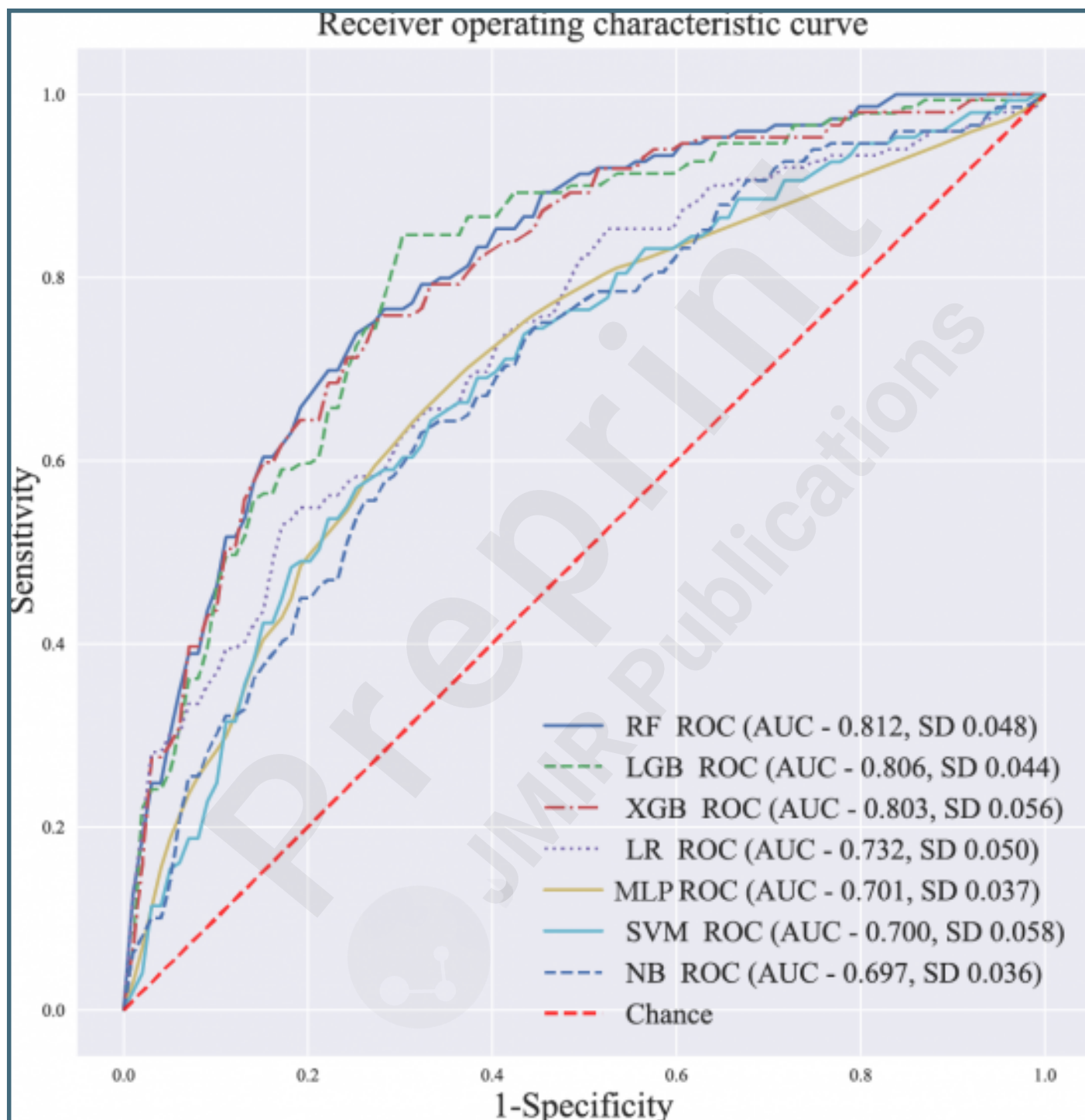
Supplementary Files

Figures

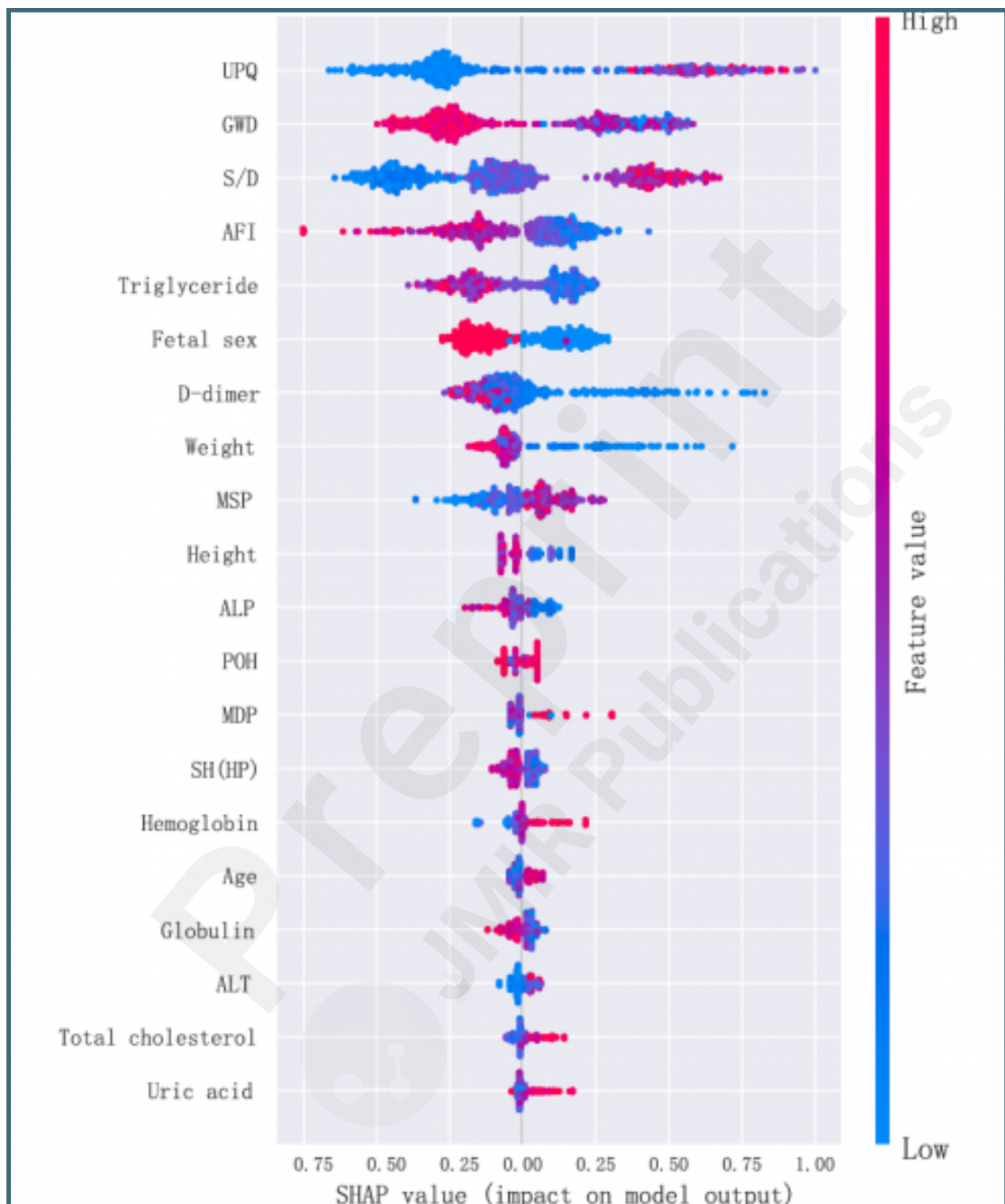
Flow chart of patient selection.



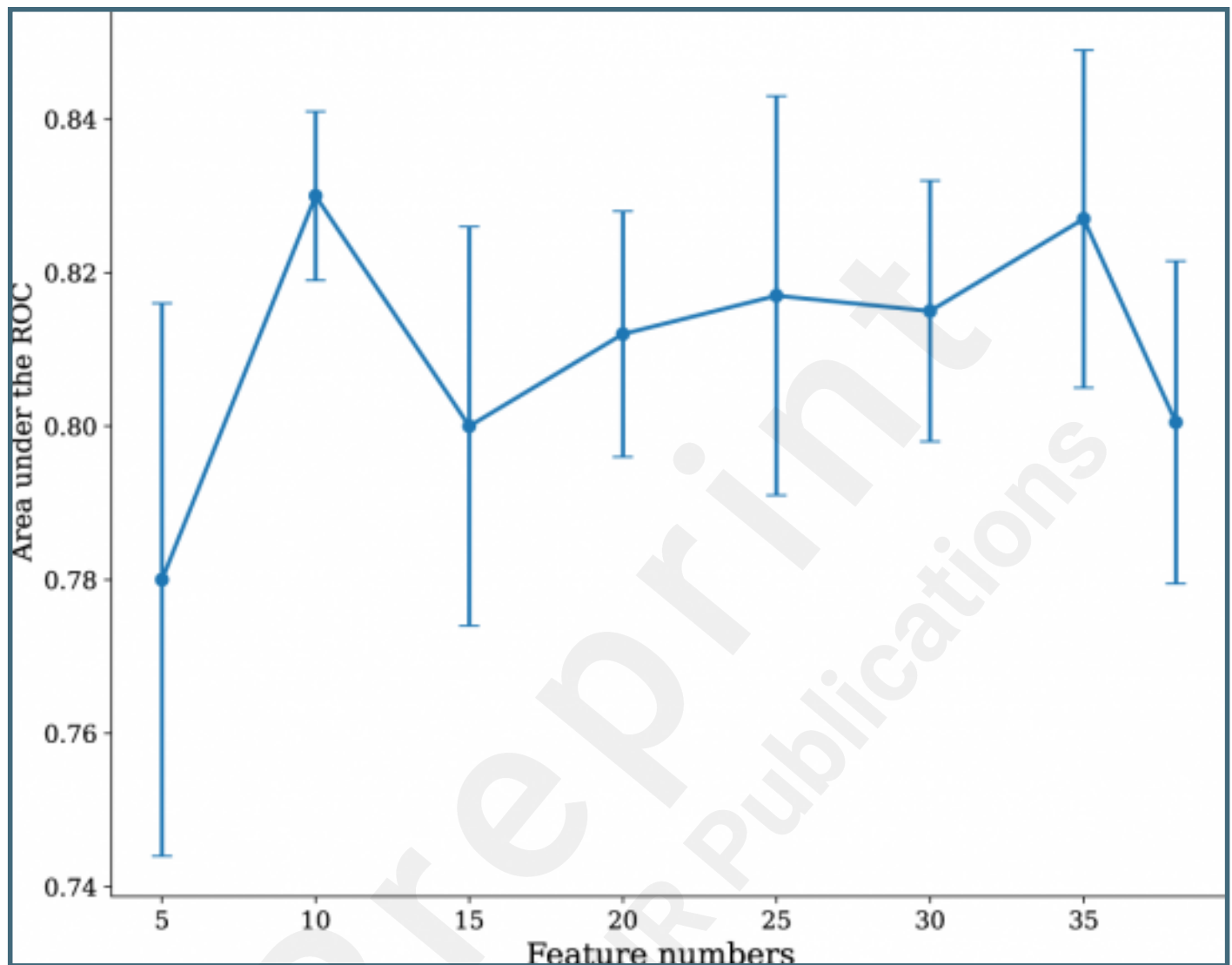
ROC curves of the 7 ML models. ROC: receiver operating curve; AUC: area under characteristic curve; SD: Standard deviation
LGB: light gradient boosting machine; LR: logistic regression; MLP: multilayer perceptron; NB: naive Bayes; RF: random forest; SVM: support vector machine; XGB: extreme gradient boosting.



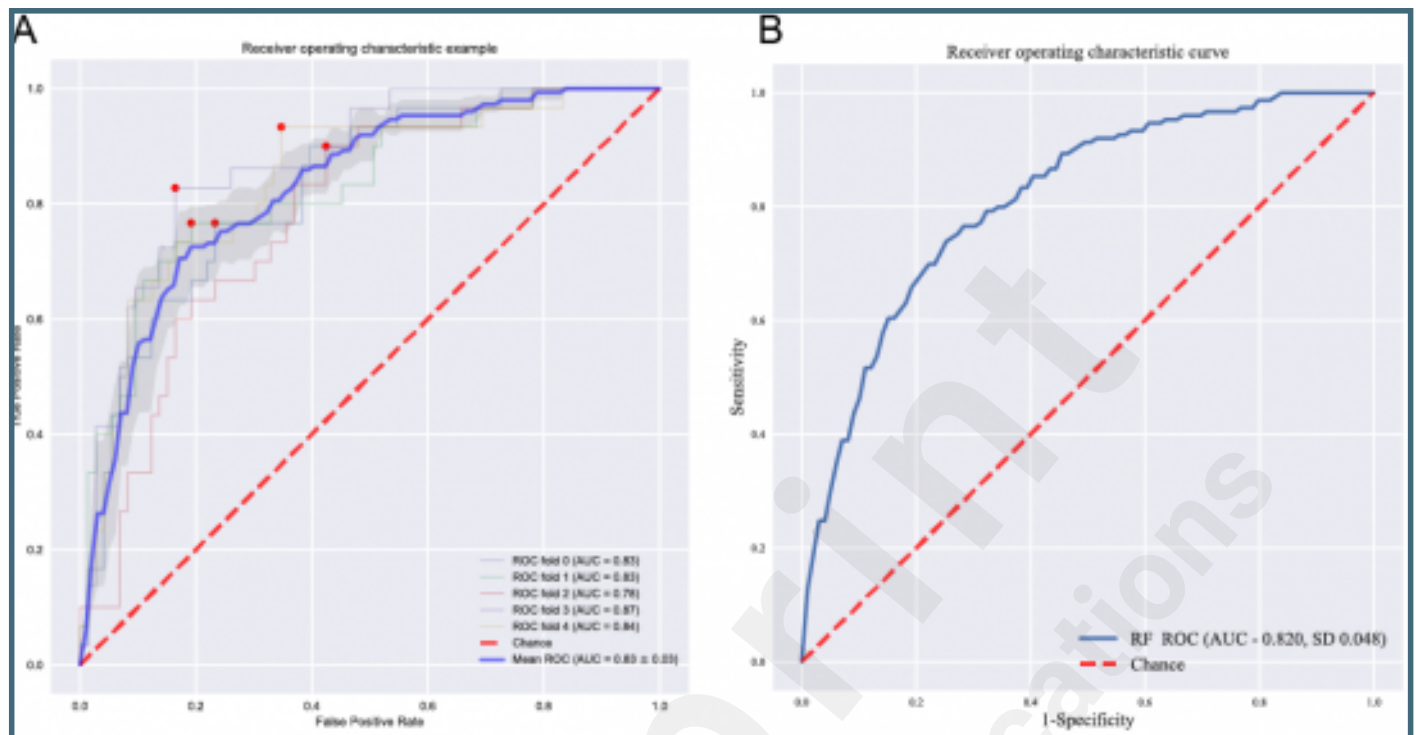
The SHAP summary plots of the top 20 features for the RF model.



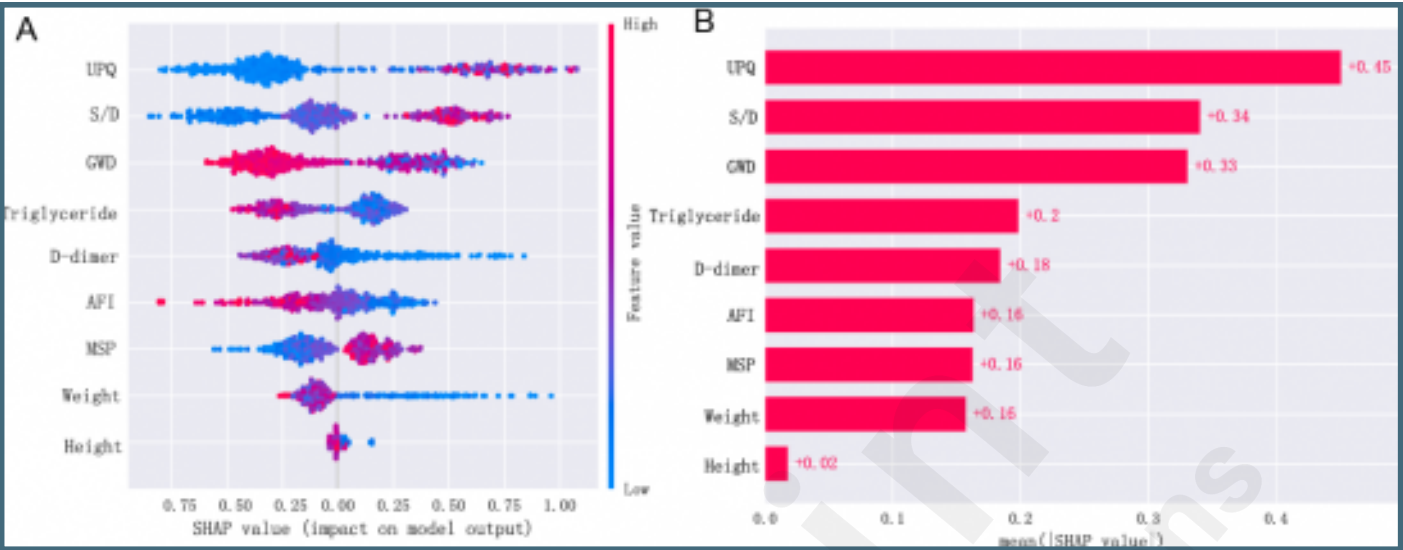
Random Forest model AUC value changes with different numbers of features.



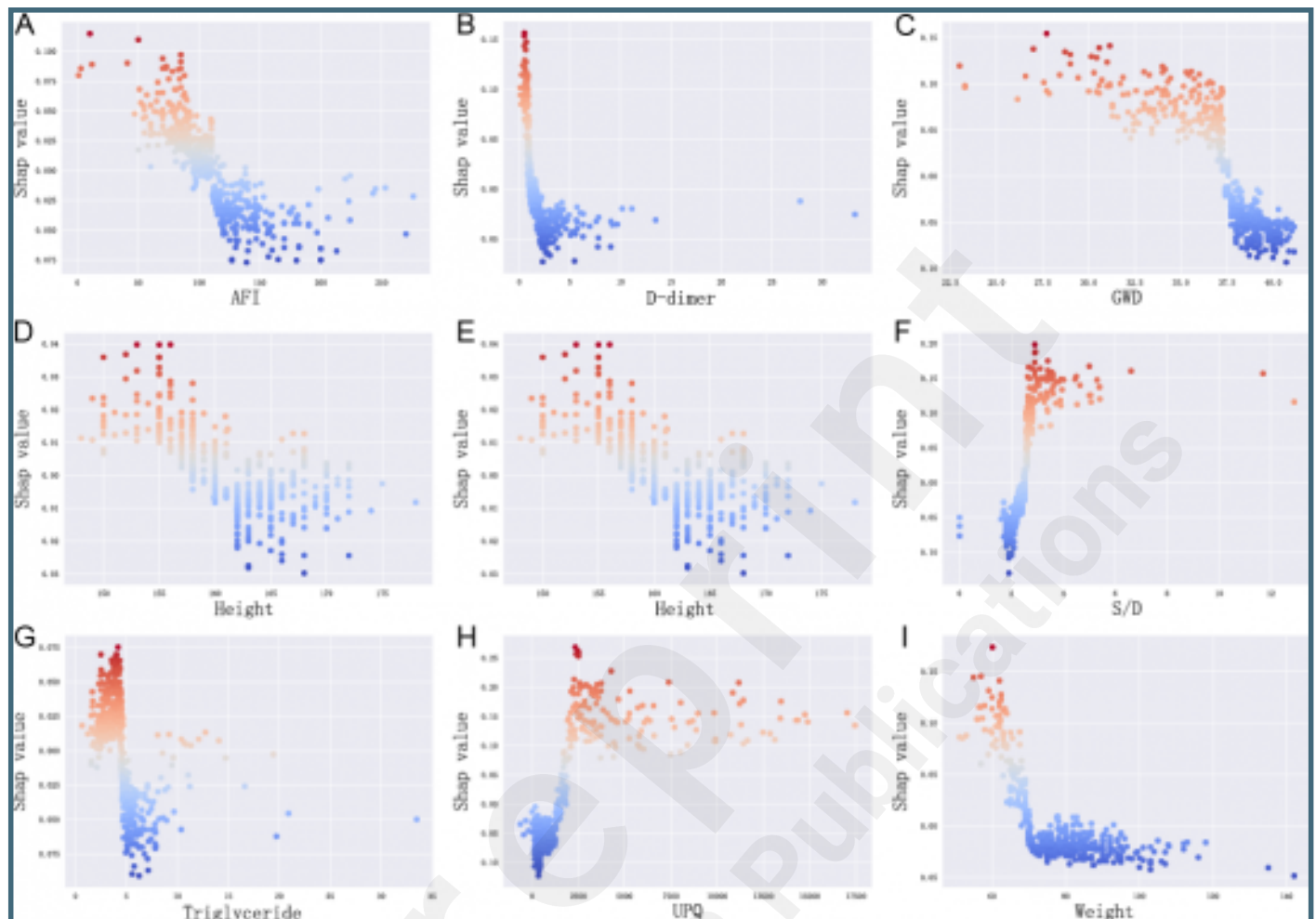
Performance of the 10-feature RF final model. (A) The 5-fold cross-validation AUC curve of the final model; (B) The external validation AUC curve.



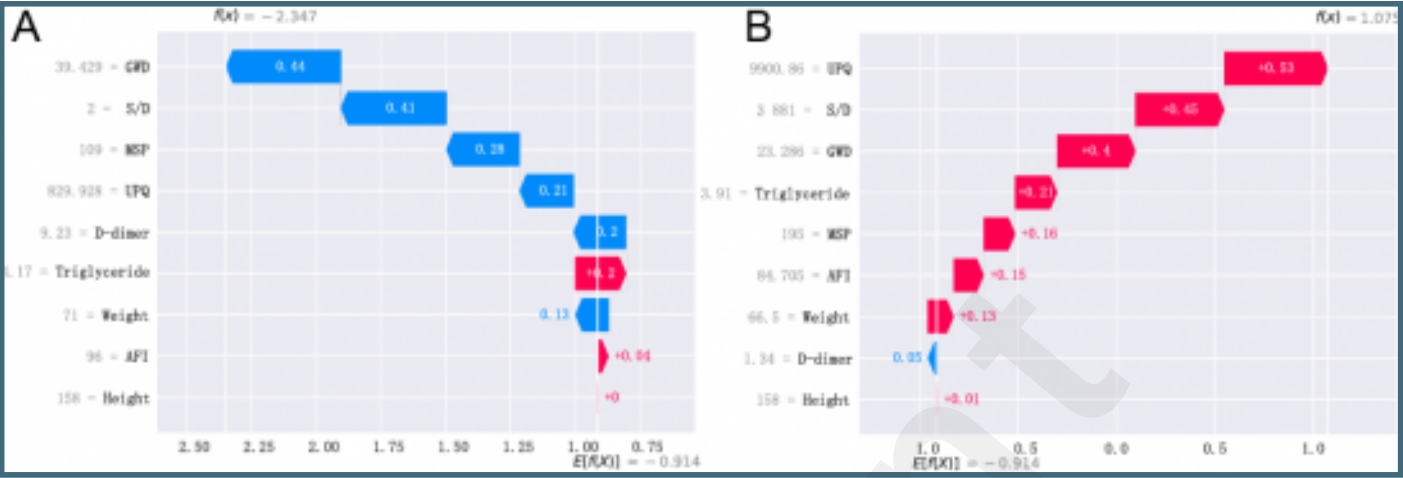
Global model explanation by the SHAP method. (A) SHAP summary dot plot. (B) SHAP summary bar plot.



SHAP dependence plot. Each dependence plot shows how a single feature affects the output of the prediction model, and each dot represents a single patient.



Local model explanation by the SHAP method. (A) PE without FGR; (B) PE with FGR.



Multimedia Appendixes

Parameter configurations and supplementary table.

URL: <http://asset.jmir.pub/assets/3772a44649a06063132c43f198205192.docx>

