BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Predicting preeclampsia in early pregnancy using clinical and laboratory data via machine learning model

Songchang Chen[1†], Jia Li[2,3†], Xiao Zhang[2,3†], Wenqiu Xu[2,3†], Zhixu Qiu[2,3†], Siyao Yan[4†], Wenrui Zhao[2,3], Zhiguang Zhao[2,3], Peirun Tian[3], Qiang Zhao[5,6], Qun Zhang[5], Weiping Chen[5], Huahua Li[5], Xiaohong Ruan[5], Gefei Xiao[7], Sufen Zhang[7], Liqing Hu[7], Jie Qin[7], Wuyan Huang[7], Zhongzhe Li[8], Shunyao Wang[3,9], Rui Zhang[10], Shang Huang[11,12], Xin Wang[11,13], Yao Yao[11], Jian Ran[11], Danling Cheng[11], Qi Luo[14], Teng Pan[11], Ruyun Gao[15], Jing Zheng[15], Yuxuan Wang[15], Cong Liu[1], Xianling Cao[1], Xuanyou Zhou[1], Naixin Xu[1], Lanlan Zhang[16], Xu Han[16], Haolin Wang[17], Suihua Feng[5*], Shuyuan Li[16*], Jianguo Zhang[2,3,4*], Lijian Zhao[2,3,4*] and Fengxiang Wei[11*]

Abstract

Background This study was performed to characterize the relationship of various laboratory test indicators with clinical information and Preeclampsia (PE) development. Then, prediction models for early-onset preeclampsia (EOPE), late-onset preeclampsia (LOPE), and preterm preeclampsia (Preterm PE) were developed using maternal characteristics and laboratory data.

Methods Between January 2019 and December 2021, we retrospectively recruited 144 EOPE, 363 LOPE, 231 Preterm PE, and 1458 healthy participants from six hospitals. We utilized all available clinical and laboratory data obtained during routine prenatal visits in early pregnancy. The models for EOPE, LOPE, and Preterm PE were created using ensemble machine learning models with patient clinical and laboratory data. Results: By comparing laboratory variables between PE patients and healthy controls, we identified 7, 18, 8, 15, 7,29 laboratory markers for EOPE, LOPE, and Preterm PE, severe PE, superimposed PE, first-time PE respectively. The ensemble EOPE and LOPE models incorporating clinical and laboratory predictors outperformed the clinical factor models respectively. The ensemble

Background

Preeclampsia is a severe and multifaceted condition that typically arises after the 20th week of pregnancy, characterized by a sudden increase in blood pressure. This condition can lead to a range of complications, including proteinuria, dysfunction of the mother's organs, and uteroplacental dysfunction. Globally, preeclampsia affects approximately 4.6% of pregnancies, resulting in an estimated 4 million cases each year[1]. Tragically, this condition is a leading cause of maternal and perinatal morbidity and mortality, contributing to the deaths of over 70,000 women and 500,000 babies worldwide[2, 3].

†Songchang Chen, Jia Li, Xiao Zhang, Wenqiu Xu, Zhixu Qiu and Siyao Yan contribute equally to this study and share first authorship.

*Correspondence:
Suihua Feng
13702285687@139.com
Shuyuan Li
shuyuanli816@163.com
Jianguo Zhang
zhangjg@bgi.com
Lijian Zhao
zhaolijian@bgi.com
Fengxiang Wei
haowei727499@163.com

Full list of author information is available at the end of the article

EOPE model demonstrated good sensitivity (72.22%,95% confidence interval [CI]: 57.59%-86.85%) and specificity (85.25%,95% CI: 80.54%-89.97%) in distinguishing EOPE from controls in early pregnancy. Similarly, the ensemble LOPE model showed good accuracy in differentiating LOPE from healthy participants (sensitivity: 69.57%, 95% CI: 56.27%-82.86%; specificity: 85.25%, 95% CI: 80.54%-89.97%). The prediction scores demonstrated notable positive correlations with blood pressure at admission, while they showed inverse correlations with 24-hour urine protein levels and fetal growth restriction among PE patients. In conclusion, our study identified key laboratory indicators for forecasting PE. The developed models exhibited good predictive capability for assessing preeclampsia risk and severity based on clinical and laboratory data.

Clinical trial number Not applicable.

Keywords Preeclampsia, Laboratory data, Machine learning model

Preeclampsia results from placental dysfunction causing syncytiotrophoblast stress, which releases pro-inflammatory cytokines, extracellular vesicles, reactive oxygen species, anti-angiogenic agents into the maternal circulation [4]. These factors lead to the maternal clinical manifestation of pre-eclampsia. Although the precise cause of preeclampsia is not yet fully understood, recognized risk factors for this condition include maternal age 35 years or older, body mass index (BMI) > 30 kg/m^2, assisted reproductive technologies, primiparity, diabetes mellitus, preexisting hypertension, renal disease, family history of PE, and occurrence of PE in a previous pregnancy [3].

Daily administration of low-dose aspirin (150 mg) starting from < 16 to 36 weeks of gestation could significantly reduce the occurrence of preeclampsia, particularly of EOPE [5], suggesting the importance of early screening and intervention for pregnant women at high risk of developing PE. Predicting preeclampsia in early pregnancy (less than 16 weeks) is highly challenging due to its poorly understood causes, various risk factors, and likely multiple pathogenic phenotypes. The current screening models recommended by The International Federation of Gynecology and Obstetrics (FIGO) incorporates maternal characteristics, uterine artery Doppler measurements and two protein biomarkers, including placental growth factor (PlGF) and pregnancy-associated plasma protein A (PAPPA) (6, 7. 8) to early predict preeclampsia. Recent studies have validated the performance of these models in Asian population but reported lower sensitivity and positive predictive values in screening PE as compared to western population [9–12]. Therefore, identifying new biomarkers with good predictive power is still greatly needed in the development of more precise prediction models.

Two recent studies have investigated the predictive capability of laboratory data for PE, Maric et al. developed a prediction model using the elastic net algorithm by incorporating clinical and laboratory variables. However, their findings indicated that laboratory results did not significantly enhance the predictive capability of the clinical factor models [13]. In a separate study by Li et al., predictors for PE at 12 weeks of gestation were identified, including maternal characteristics and certain laboratory variables from routine blood tests, however, their studies didn't conduct classification analysis on PE patients, their gradient boosting model exhibited relatively poor performance in screening for PE in early pregnancy [14]. Machine learning methods are well-equipped to deal with many variables, such as clinical, laboratory data from patients, and automatically select the most informative features. They have been increasingly used in the prediction of PE in recent years[15]. In this study, we first applied multiple statistical methods to pinpoint the clinical factors and laboratory test variables linked to distinct PE subtypes, and compared these findings across the various PE subtypes. Next, we conducted a correlation analysis between clinical data and laboratory test variables to uncover key laboratory indicators associated with PE. Furthermore, we strived to identify novel predictive markers and establish comprehensive predictive models for EOPE, LOPE, and Preterm PE by integrating clinical features and laboratory markers from early pregnancy. We then independently validated these models and analyzed their relationship with severity of PE. Ultimately, our objective is to leverage these discoveries for early PE screening and to inform therapeutic strategies.

Methods and materials

Participants and study design

During early pregnancy, Chinese pregnant women underwent routine prenatal laboratory tests, including routine blood test, hepatic and renal function tests, routine urine test, thyroid hormones, hepatitis B antigens and antibodies. This study is a retrospective study. After the pregnancy outcomes were finalized, we retrospectively collected clinical and laboratory data of participants and recruited 1965 participants with singleton pregnancy (144 EOPE, 363 LOPE, 1458 healthy controls) between January 2019 and December 2021 from six hospitals, including Zhuhai Center for Maternal and Child Health Care, Jiangmen Central Hospital, Shenzhen Bao'an District Maternal and Child Health Hospital,

Longgang District Maternity & Child Healthcare Hospital of Shenzhen City, Obstetrics & Gynecology Hospital of Fudan University, International Peace Maternity and Child Health Hospital (Fig. 1). The patients diagnosed with PE were included in this study based on the guidelines provided by the International Society for the Study of Hypertension in Pregnancy[2]. These patients experienced high blood pressure after 20 weeks of pregnancy, with systolic blood pressure exceeding 140 mm Hg and/or diastolic blood pressure exceeding 90 mm Hg on at least two occasions, with a minimum time gap of 4 hours. Additionally, proteinuria of 300 mg or more in a 24-hour urine collection was required for PE diagnosis. In cases where 24-hour urine protein quantitation was not available, a diagnosis of PE required at least one reading of + on dipstick analysis of urine specimens. PE patients were further classified as early-onset pre-eclampsia (EOPE) or late-onset pre-eclampsia (LOPE) based on the time of development of preeclampsia before or after 34 weeks of gestation, respectively. As preterm PE is associated with a relatively poor maternal and fetal prognosis, we also investigated the laboratory biomarkers related to preterm PE, which helps us to establish prediction models to screen for preterm PE and finally improve maternal and fetal prognosis. Preterm PE was defined as preeclampsia in patients who delivered before 37 weeks of gestation. We followed the guidelines provided by the Chinese Society of Obstetrics and Gynecology, Chinese Medical Association to define mild and severe PE[16]. PE patients with severe features were diagnosed, if they met any of the following conditions [1]: The continuous increase in blood pressure is uncontrollable: systolic blood pressure ≥ 160 mmHg and/or diastolic blood pressure ≥ 110 mmHg [2]; Persistent headache, visual impairment, or other central nervous system abnormalities [3]; Persistent upper abdominal pain and subcapsular hematoma or liver rupture [4]; Abnormal transaminase levels: elevated levels of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) in the blood [5]; Renal function impairment: Urinary

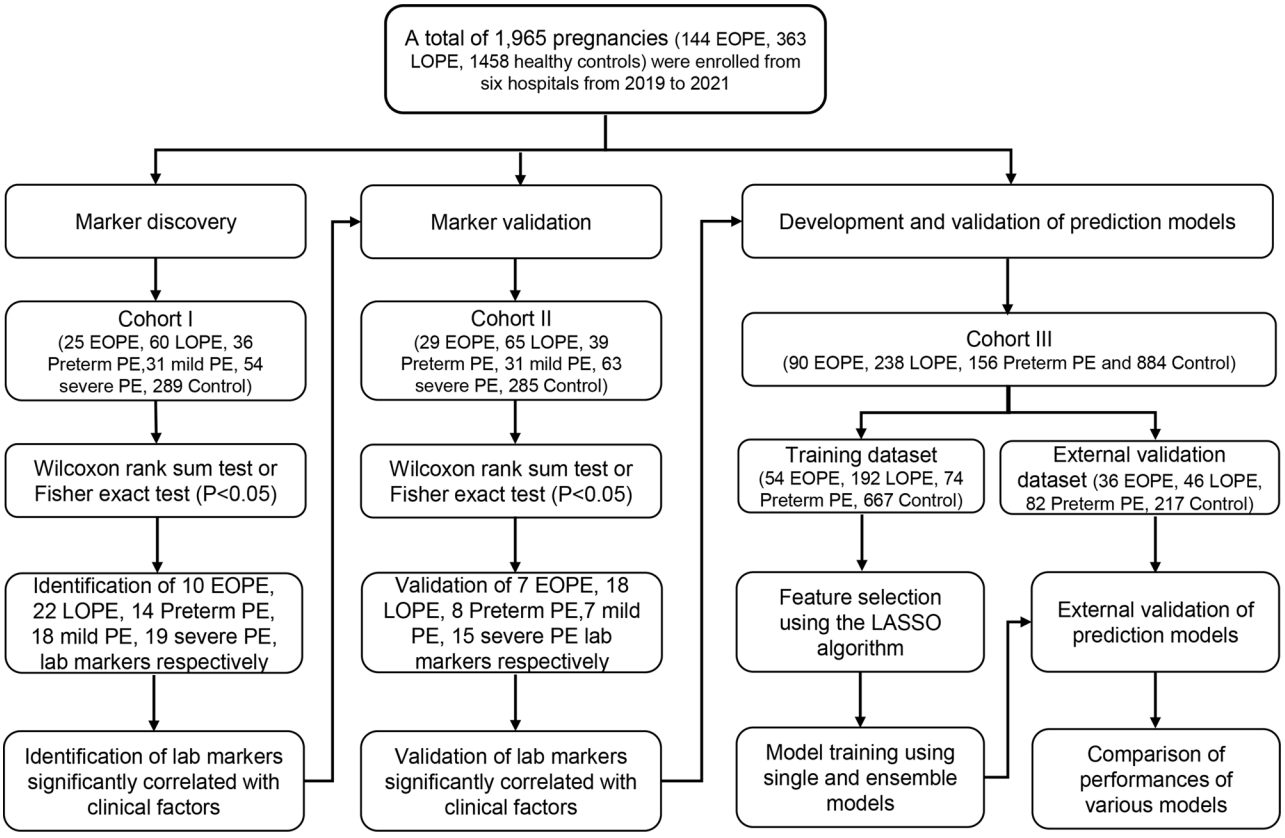


Fig. 1 The schematic workflow of this study. The study involved 1,965 pregnancies consisting of 144 EOPE, 363 LOPE, 1458 healthy controls enrolled between 2019 and 2021 from six hospitals. The participants were divided into three cohorts. Cohort I, comprising 25 EOPE, 60 LOPE, 36 Preterm PE, 31 mild PE, 54 severe PE, and 289 healthy controls, was used to identify laboratory markers associated with different subtypes of PE and clinical characteristics. Cohort II, consisting of 29 EOPE, 65 LOPE, 39 Preterm PE, 31 mild PE, 63 severe PE and 285 healthy controls, was used to validate the results. Cohort III included 90 EOPE, 238 LOPE, 156 Preterm PE and 884 healthy participants and was further divided into a training dataset and an EV dataset. The training dataset was used to select optimal features for predicting PE and develop single and ensemble machine learning models. The EV dataset was used to validate and assess the performance of the established models

protein quantification > 2.0 g/24 h; Oligouria (24-hour urine output < 400 ml, or hourly urine output < 17 ml), or blood creatinine level > 10^6 μmol/L [6]; Hypoalbuminemia with ascites, pleural effusion, or pericardial effusion [7]; Hematology abnormalities: Platelet count shows a continuous decrease below 100×10^9/L; Intravascular hemolysis, manifested as anemia, elevated levels of lactate dehydrogenase (LDH) in the blood, or jaundice [8]; Heart failure [9]; Pulmonary edema [10]; Fetal growth restriction (FGR) or oligohydramnios, fetal death in the uterus, placental abruption, etc. The PE patients without the above severe features were considered as mild PE. The definition of FGR is a birth weight that is less than the 10th percentile for gestational age [17]. Healthy controls were defined as pregnant individuals without any obstetric, medical, or surgical complications during pregnancy, and who delivered at full-term. All participants who took aspirin treatment were eliminated from this study.

The participants from Zhuhai Center for Maternal and Child Health Care and Jiangmen Central Hospital were combined and randomly split into two cohorts at a ratio of 1:1, including the cohort I (25 EOPE, 60 LOPE, 36 Preterm PE, 289 healthy controls) and cohort II (29 EOPE, 65 LOPE, 39 Preterm PE, 285 healthy controls). The former was used to identify predictive clinical and laboratory test markers associated with different subtypes of PE and analyze correlations between clinical information and laboratory test variables. Then, the results were validated in the cohort II dataset. The participants from the other four hospitals were assigned to the cohort III, which includes a training dataset and external validation (EV) dataset. The training dataset comprises 54 EOPE, 192 LOPE, 74 Preterm PE, 667 healthy controls from Shenzhen Bao'an District Maternal and Child Health Hospital, Longgang District Maternity & Child Healthcare Hospital of Shenzhen City, The Obstetrics & Gynecology Hospital of Fudan University. The training dataset was utilized to validate the identified laboratory markers, perform feature selection and train the prediction models. The participants from International Peace Maternity and Child Health Hospital were used as an external validation for the prediction models (Fig. 1). All experiments and methods were performed in accordance with relevant guidelines and regulations. This study was approved by the Ethics Committees of Beijing Genomics Institute (BGI-IRB 22026) and the Ethics Committee of each participating hospital.

Analysis of clinical and laboratory data

During routine prenatal visits in early pregnancy, all necessary clinical and laboratory data were retrospectively collected. The study analyzed various clinical factors including the participant's age, BMI, diastolic blood pressure (DBP), systolic blood pressure (SBP), birth times, history of recurrent pregnancy loss (RPL, > 2 times of miscarriages or induced abortions), in vitro fertilization (IVF), and past medical history (PMH). The PMH included complications such as diabetes mellitus, PE, family history of PE, chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome. Additionally, blood pressure at admission, urine protein, FGR, birth weight and gestational weeks at birth were recorded. Mean arterial pressure (MAP) was calculated using a specific equation[18]:

$$\mathrm{MAP} = \mathrm{DBP} + \mathrm{PP} \times \frac{(27.07 + 0.181 \times \mathrm{DBP} + 2.303)}{100}.$$

in which, PP is the difference between systolic and diastolic blood pressure.

All available laboratory data were obtained during routine prenatal visits in early pregnancy with a median gestational week of 12.57 (25th - 75th percentile:11.4 ~ 15.43 weeks of gestation). The laboratory data include 45 routine prenatal laboratory test results from five main types of laboratory tests [1]: routine blood test: white blood cells (WBC), red blood cells (RBC), hemoglobin (Hb), hematocrit (Hct), monocytes(MO), lymphocytes(LY), eosinophils (EO), neutrophils (NE), basophils (BA), platelet count (PLT), platelet distribution width (PDW), plateletcrit (PCT), mean platelet volume (MPV), mean corpuscular hemoglobin (MCH), mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), etc [2]; hepatic and renal function tests: ALT, AST, Total Bilirubin Test (T-BIL), total protein (TP), fasting glucose (GLU), creatinine(CRE), uric acid (UA), urea, etc [3]; routine urine test: urine blood (BLD), protein (PRO), glucose, urinary white blood cells (UWBC), urinary red blood cells (URBC) [4]; thyroid hormones: thyroid-stimulating hormone (TSH), free thyroxine (FT4) and thyroid peroxidase antibody (TPOAb) [5]; hepatitis B antigens and antibodies. Wilcoxon sum rank test and Fisher exact test were utilized to analyze continuous and categorical variables between PE patients and healthy controls respectively. Correlations between clinical information and laboratory test variables were analyzed using cor.test and visualized using the R package pheatmap. $P < 0.05$ was considered statistically significant.

Feature selection with the LASSO algorithm

The LASSO algorithm, known as Least Absolute Shrinkage and Selection Operator, is widely used for feature selection, especially in high-dimensional data analysis [19]. In our study, we utilized the LASSO model with an alpha value of 1 and performed a search for the optimal regularization parameter, Lambda, which controls

the strength of the penalty. This search was conducted using the cv.glmnet() function, which implements 10-fold cross-validation for glmnet. The function provided two sets of variable combinations at two specific Lambda values: lambda.min and lambda.1se. Lambda.min represents the Lambda value that maximizes the area under the curve (AUC), while lambda.1se represents the Lambda value that yields a more regularized model with a cross-validated AUC within one standard error of the minimum. Variables with coefficients shrunk to zero or close to zero were considered unimportant and were excluded from further model development. The selected variables at the Lambda.min value were considered the optimal combination of predictors.

The establishment and validation of the ensemble machine learning models

The model development process involved two main steps: predictor selection and model development. To mitigate the impact of differences in predictor scale on prediction models, patient BMI, MAP, and selected laboratory test variables were normalized by dividing their raw values by the corresponding median values of healthy participants. Missing values for each laboratory test were replaced with the median values of all participants. The LASSO algorithm was then used to determine the best predictors for EOPE, LOPE and Preterm PE separately. Seven single machine learning models including Generalized Linear Model (GLM), Neural Network(nnet), random forest(rf), Gradient Boosting Model(gbm), Support Vector Machines with Radial Basis Function Kernel (svmRadial), Multivariate Adaptive Regression Splines (earth) and glmnet and 120 ensemble models were developed in the training dataset using the R package caret. The 120 ensemble models included any combination of 7 single machine learning models and were developed using the R package caretEnsemble. The package has 3 primary functions: caretList, caretEnsemble and caretStack. caretList is used to build lists of caret models on the same training data, with the same re-sampling parameters. caretEnsemble and caretStack are used to create ensemble models from such lists of caret models. caretEnsemble uses a glm to create a simple linear blend of models and caretStack uses a caret model to combine the outputs from several component caret models (https://cran.r-project.org/web/packages/caretEnsemble/index.html).The performances of these models were independently validated in the EV set. Receiver operating characteristic (ROC) curves were plotted and AUC values were computed using probabilities predicted by the prediction models. DeLong's test was utilized to compare the difference between two ROC curves. Correlations between predictors, gestational weeks at delivery, birth weight, blood pressure at admission, 24-hour urine protein and PE risk prediction were analyzed using Pearson correlation. Differences in PE risk prediction were analyzed between FGR and controls as well as subgroups of urine protein using wilcoxon rank sum test.

Table 1 Comparison of maternal obstetric characteristics and pregnancy outcome of the women who did and did not develop PE in the discovery cohort

Clinical feature		EPE (n = 25)	LPE (n = 60)	Preterm PE (n = 36)	Mild PE (n = 31)	Severe PE (n = 54)	Control (n = 289)
The onset weeks of gestation		28.94 ± 4.05	37.29 ± 1.56	32.01 ± 4.27	35.94 ± 3.79	36.45 ± 4.82	
Age		33.36 ± 5.43***	32.07 ± 4.83***	32.69 ± 5.05***	32.65 ± 4.6***	32.33 ± 5.28***	29.35 ± 4.40
BMI		23.91 ± 4.17***	23.85 ± 5.24***	23.83 ± 4.26***	24.58 ± 4.2***	23.22 ± 5.22**	21.18 ± 2.87
MAP		93.64 ± 9.15***	91.28 ± 10.14***	91.76 ± 9.99***	93.14 ± 10.89***	91.15 ± 9.22***	83.42 ± 8.24
Primiparous woman	No	14	26	18	12	28	131
	Yes	11	34	18	19	26	157
RPL	No	19**	54	31	27	46	269
	Yes	6	6	5	4	8	19
Past medical history	No	21***	59	33**	30	50***	289
	Yes	4	1	3	1	4	0
IVF	No	20***	49***	28***	23***	46***	284
	Yes	5	11	8	8	8	5
Gestational age at delivery (weeks)		32.42 ± 4.37***	37.76 ± 1.53***	33.22 ± 3.72***	36.96 ± 2.99***	35.75 ± 3.89***	39.08 ± 0.93
Birthweight (g)		1872.22 ± 678.89***	2785.09 ± 601.70***	1982.14 ± 601.91***	2966.43 ± 544.83**	2336.98 ± 727.83***	3194.69 ± 312.72

*, **, *** denote *P* value < 0.05, < 0.01 and < 0.001 respectively. RPL is defined as a woman who has more than 2 miscarriages or induced abortions. Past medical history denotes a woman has at least one of the following complications: medical history of pregnancy diabetes, pre-eclampsia, family history of pre-eclampsia, chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome

Results

Identification of clinical risk factors and associated laboratory test variables for PE

Maternal characteristics, demographics, gestational ages at delivery and birth weight are shown in Table 1. The EOPE, LOPE, Preterm PE, mild PE, severe PE, Superimposed PE, First-time PE patients presented a significantly older age, higher BMI and MAP values, higher prevalence of IVF, smaller gestational age at delivery, lower birthweight than healthy controls in the cohort I (Table 1 and Table S1, $P<0.05$ for all cases, Wilcoxon rank sum test or Fisher exact test). Moreover, the participants who experienced RPL were more frequent to develop EOPE. Participants with past medical history involving diabetes, pre-eclampsia, family history of pre-eclampsia, chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome had increased risk of EOPE and Preterm PE (Table 1, $P<0.05$ for all cases, Fisher exact test). Furthermore, the majority of the results have been validated in the cohort II (Table S2, $P<0.05$ for all cases, Wilcoxon rank sum test or Fisher exact test). After further analysis, significant associations were observed between clinical features and laboratory test indicators in both cohort I and II datasets. BMI displayed a notably negative correlation with AST/ALT and positive correlations with GLU, RBC, and UA. Meanwhile, MAP showed a positive correlation with PDW and RBC, but an inverse correlation with MCH. Disease severity positively correlated with GLU, RBC, and UA, while negatively correlating with MCH, MCHC, and MCV. Additionally, MCH exhibited a correlation with gestational age at delivery in cohort I, with an absolute value of correlation coefficients exceeding 0.2 and a statistical significance of $p<0.05$ in all cases (Figure S1A). These associations were consistently validated in cohort II, with statistical significance ($P<0.05$) observed in all cases (Figure S1B).

Identification of laboratory test biomarkers for different subtypes of PE

First, we aimed to identify the laboratory test results associated with PE and performed differential expression analysis among EOPE, LOPE, Preterm PE and healthy controls in the cohort I. The EOPE patients exhibited significantly elevated ALT, CRE, Hct, MO, RBC, UA,

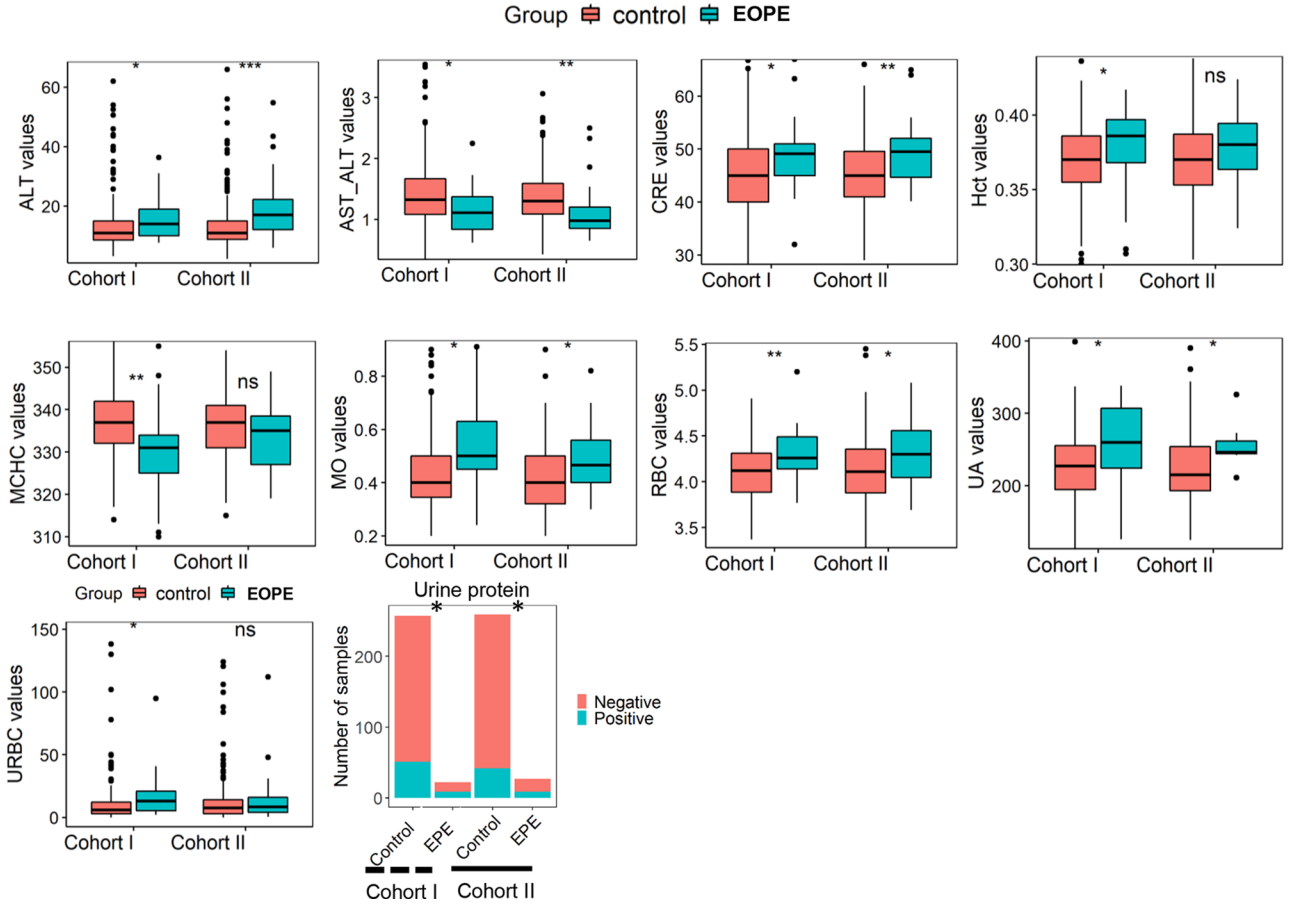


Fig. 2 Identification and validation of laboratory test markers for EOPE. The study compared the differences in expression or prevalence of 10 laboratory test markers between EOPE and healthy participants in cohorts I and II. As a result, seven laboratory markers were confirmed to have predictive value for EOPE (ns: not significant, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$)

URBC levels, decreased AST/ALT, MCHC levels and higher prevalence of urinary protein than healthy controls (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Fig. 2). A total of 22 laboratory test variables were significantly different between LOPE and healthy controls, including 13 up-regulated (ALT, LY, NE, EO, PCT, PDW, UA, URBC, GLU, MO, WBC, RBC, PLT), 7 down-regulated laboratory parameters (AST/ALT, MPV, MCH, MCHC, FT4, MCV,TP) and two laboratory parameters (PRO and BLD) with more occurrences in diseased participants (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Figure S2 and Figure S3). With respect to the laboratory parameters associated with Preterm PE, ALT, LY, PDW, PLT, CRE, RBC, MO, UA, GLU, NE, WBC were significantly increased, while, AST/ALT, MCHC, FT4 were markedly decreased in Preterm PE patients as compared to healthy controls (P value < 0.05 for all cases, Wilcoxon rank-sum test, Figure S4). Further validation of the results in the cohort II resulted in 7 EOPE, 18 LOPE, 8 Preterm PE laboratory markers (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Figure2, Figure S2, 3 and 4).

Severe preeclampsia has grave consequences for both maternal and neonatal health, contributing to 50,000–100,000 annual deaths globally, as well as serious fetal and neonatal morbidity and mortality [20]. We also investigated the laboratory test variables associated with severe PE. There are 18 laboratory test variables that have significantly different concentrations between severe PE patients and healthy controls in the cohort I, with MCHC, PDW and FT4 ranking the top three (Fig. 3A). Further analysis revealed severe PE were significantly associated with eleven up-regulated (ALT, AST, GLU, LY, MO, NE, PLT, RBC, URBC, WBC and UA), three down-regulated (AST/ALT, MCH, MCHC) laboratory markers and higher frequency of UPRO in the validation cohort (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Fig. 3B-P). While, mild PE exhibited significantly increased ALT, PLT, RBC, URBC, UA and decreased MPV values as well as higher fraction of BLD-positive cases as compared to healthy controls (Figure S5A–H). Furthermore, ALT, PLT, RBC, URBC, UA were common biomarkers for severe PE and mild PE, while AST, AST/ALT, GLU, LY, MCH, MCHC, MO, NE, WBC were specific for severe PE (Figure S5I).

As risk disorders, such as chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome, have a direct impact on lab data, abnormal kidney or blood indices (PCV, Platelets, etc). We also separately analyzed the laboratory test variables related to PE patients with underlying disorders (superimposed pre-eclampsia) and those without such underlying disorders (first-time pre-eclampsia). The results demonstrated superimposed PE patients (n = 22) exhibited significantly higher MO, UA, WBC, Hct, NE, more frequent UPRO and Hepatitis B surface antigen (HBsAg) as compared to healthy controls (n = 1458, P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Figure S6A–B). In addition to the above 7 laboratory test variables, we also identified 22 laboratory test variables showing significant difference between first-time PE patients (n = 485) and healthy controls, such as ALT, AST, AST/ALT, CRE and EO, suggesting the two subtypes of PE might be largely different regarding the pathogenesis (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Figure S6C–E). Lastly, we conducted a comparison of the laboratory variables associated with various subtypes of PE. UA was a common variable observed in seven types of PE, ALT, RBC, MO, AST/ALT, WBC and UPRO were associated with more than five subtypes of PE. On the other hand, PEO, Hb, MCV, TP, hepatitis B core antibody (HBcAb), and hepatitis B e antibody (HBeAb) were specific markers for first-time PE occurrences (Fig. 4).

Feature selection with the LASSO algorithm

We analyzed the identified clinical and laboratory test markers between EOPE and healthy controls in the training dataset and confirmed higher BMI, MAP, RPL and IVF were risk clinical factors for EOPE development, moreover, increases in ALT, UA and MO levels, as well as decrease in AST/ALT levels were related to increased risk of EOPE (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Table S3 and Figure S7). Then, the LASSO algorithm was used to determine the optimal feature combination for predicting EOPE. The model identified 7 features (BMI, MAP, RPL, IVF, UA, MO and AST/ALT) as the optimal predictors for EOPE (mean AUC: 0.879, log(λ): – 6.03, Fig. 5A). MAP, UA, and RPL were the top three positively correlated predictors, while AST/ALT was a negative factor for EOPE prediction (Fig. 5B). Following the same strategy, 2 clinical factors (BMI and MAP) and 15 laboratory markers were identified as having predictive value for LOPE (Figure S8). The feature selection process further narrowed down these variables to the top 15 with the highest predictive capabilities for LOPE, achieving a mean AUC of 0.81 and a log(λ) value of – 7.26 (Fig. 6A and B). Finally, we found that 2 clinical markers (BMI and MAP) and five laboratory markers (ALT, UA, MO, WBC, and AST/ALT) were significantly associated with Preterm PE (P value < 0.05 for all cases, Wilcoxon rank-sum test or Fisher exact test, Table S3 and Figure S9). These features showed the best performance for predicting Preterm PE, with a mean AUC of 0.764 and a log(λ) value of – 7.38 (Figure S10A and B).

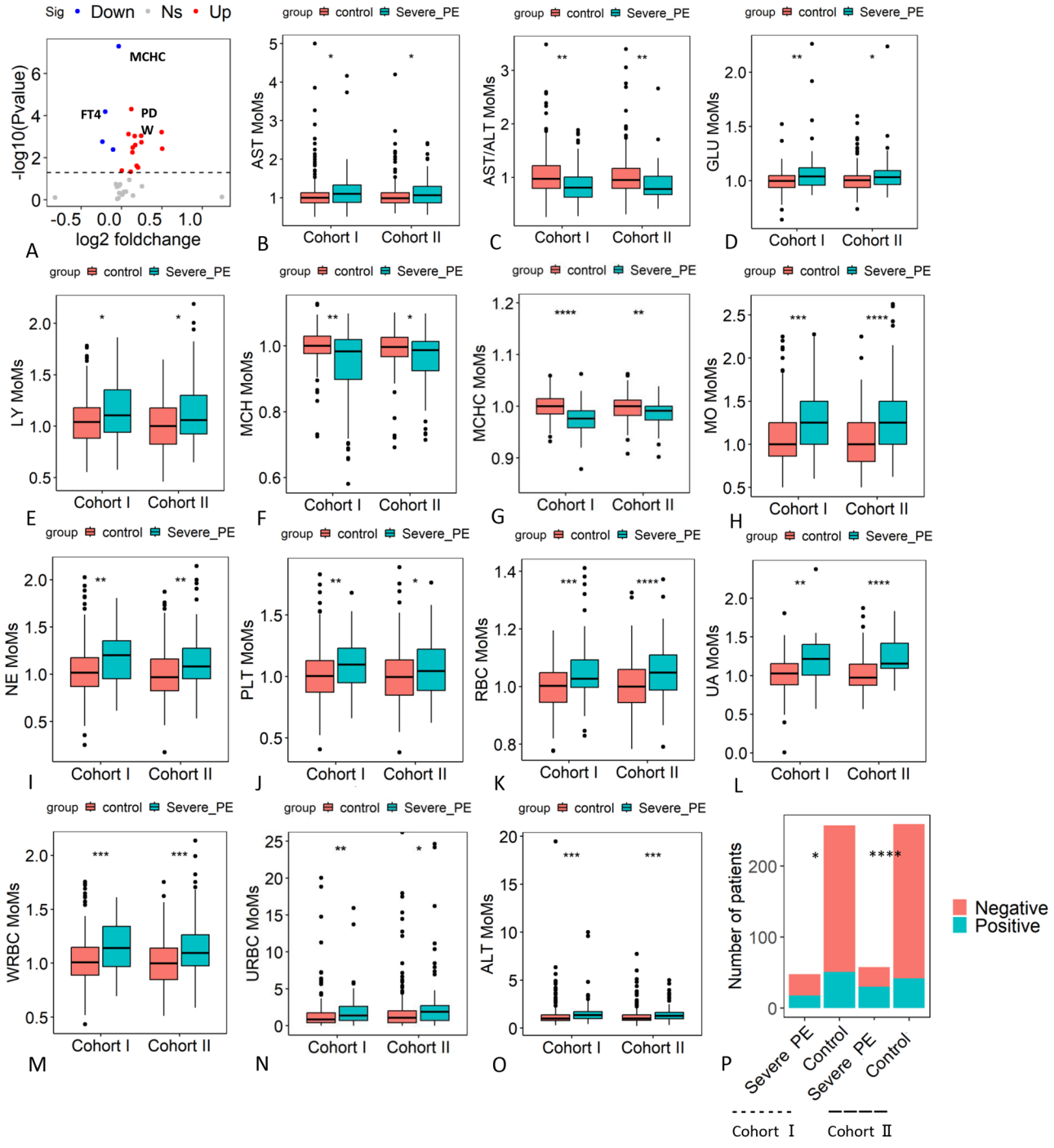


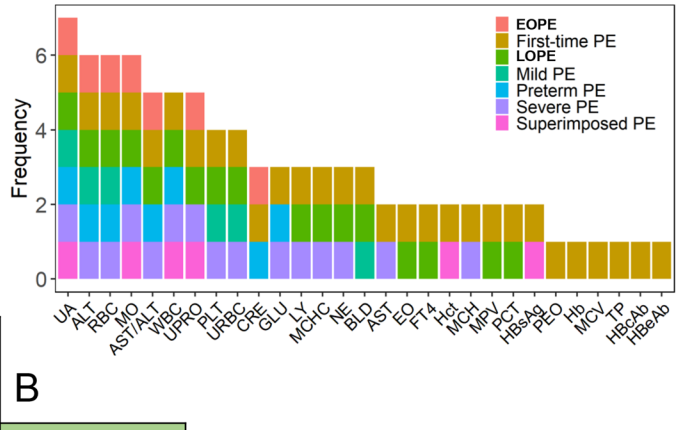
Fig. 3 Analysis of laboratory test variables associated with severe PE. **A**. A Volcano plot illustrates the 18 laboratory test variables that show significant expression difference between severe PE and healthy controls, with red, blue, grey representing significantly up-regulated, down-regulated, not significantly differentially expressed test variables respectively. The dashed line indicates a significance level of $P < 0.05$. Ns: not significant. **B–O**. The comparison of concentration differences of 14 laboratory test variables between severe PE and healthy controls in the cohort I and II datasets. **P**. A barplot compares the distribution differences of UPRO positivity for severe PE patients and healthy controls

Prediction of preeclampsia in early pregnancy

In the study, different single and ensemble machine learning models were established using clinical markers alone or in combination with laboratory predictors in the training set and confirmed in the EV datasets. The ensemble EOPE model comprising gbm and svmRadial showed the highest mean AUC value as compared to seven single models and other ensemble models (Figure S11A). The ensemble EOPE model incorporated 4 clinical factors and 3 laboratory biomarkers and presented

Laboratory variables	EOPE	LOPE	Preterm PE	Severe PE	Mild PE	Superimposed PE	First-time PE
EOPE	ALT, UPRO AST/ALT,CRE,MO,RBC, UA						
LOPE	ALT,AST/ALT,MO,RBC,UA, UPRO	ALT,AST/ALT, EO,FT4, LY, MCHC, RBC, MO, MPV, NE, PCT,PDW,PLT,UA,BLD,URBC, WBC, UPRO					
Preterm PE	ALT, UA, AST/ALT,CRE,MO,RBC,	ALT, AST/ALT, MO, RBC, UA, WBC	AST/ALT,MO, ALT, CRE, GLU,RBC, UA,WBC				
Severe PE	ALT,MO, RBC,UA, UPRO, AST/ALT	AST/ALT,ALT, LY, MCHC, MO,NE,PLT,RBC,URBC,WBC,UA,UPRO	AST/ALT, ALT,GLU, MO,RBC, UA,WBC	ALT, AST, GLU, LY, MO, NE, PLT, RBC,URBC,WBC, UA,AST/ALT,MCH,MCHC,UPRO			
Mild PE	ALT, RBC, UA,	ALT, PLT,UA, RBC,URBC, MPV,BLD	ALT,RBC,UA	ALT, PLT, RBC, URBC, UA	ALT, PLT, RBC,UA,MPV,BLD,URBC,		
Superimposed PE	MO, UA, UPRO,	MO, UA, WBC, NE, UPRO	UA,MO, WBC	MO, UA,WBC, NE, UPRO	UA	MO, UA, WBC, Hct, NE, UPRO, HBsAg	
First-time PE	ALT,AST/ALT,CRE,MO,RBC,UA, UPRO	ALT,EO,FT4, AST/ALT,LY, MCHC, MO, MPV, NE,UA, PCT,PLT, RBC,URBC, WBC,BLD,UPRO	AST/ALT, ALT, CRE, GLU,MO,RBC,UA,WBC	ALT, AST, GLU, LY, MO, NE, PLT, RBC, URBC,WBC, UA,AST/ALT,MCH,MCHC,UPRO	ALT, PLT, RBC, URBC, UA,MPV, BLD	MO, UA, WBC, Hct,NE,UPRO, HBsAg	ALT,AST,AST/ALT, CRE,EO,PEO,FT4,GLU,Hb,Hct, LY, MCH,MCHC,NE, MCV, MO, MPV, PCT, PLT, RBC, TP, URBC, WBC, UA, BLD, UPRO, HBsAg,HBcAb,HBeAb

A



B

Fig. 4 Analysis of laboratory test variables associated with different types of PE. **A**. The identification of laboratory test variables specifically associated with each subtype of PE, along with the common laboratory variables shared among various subtypes of PE. **B**. The frequencies of PE-associated laboratory test variables, categorized according to the different subtypes of PE

higher AUC value than the clinical factors model in the EV dataset ($P < 0.05$, DeLong's test, Fig. 5C). This indicates that the addition of laboratory predictors improved the performance of the model. The ensemble EOPE model demonstrated good sensitivity (72.22%, 95% confidence interval [CI]: 57.59%-86.85%) and specificity (85.25%, 95% CI: 80.54%-89.97%, Table 2) in distinguishing EOPE from healthy controls in early pregnancy. The correlations between each predictor and prediction model scores were also analyzed, revealing positive correlations with MAP, IVF, MO, UA, BMI and RPL ($r \geq 0.15$, $p < 0.001$ for all cases, Pearson correlation, Fig. 5D), while a negative correlation with AST/ALT ($r = -0.14$, $p < 0.001$, Fig. 5D). These findings suggest that the model captures both clinical and laboratory expression differences. The prediction scores exhibited significantly negative correlations with gestational weeks at delivery and birth weight ($r = -0.25$ and -0.24 respectively, $p < 0.001$ for all cases, Pearson correlation, Fig. 5E-F). Furthermore, we also investigated the relationships between PE risk prediction and blood pressure at admission, urine protein levels and FGR in EOPE patients. The results demonstrated significant negative correlations between PE risk prediction and urine protein values ($p < 0.05$, Pearson correlation or wilcoxon rank sum test, Fig. 5G and H).

Using the same approach, the ensemble LOPE model comprising six machine learning models (nnet, glmnet, glm, rf, gbm, svmRadial) exhibited the best performance and slightly better performance than the clinical factor model in the EV dataset (Figure S11B and Fig. 6C).

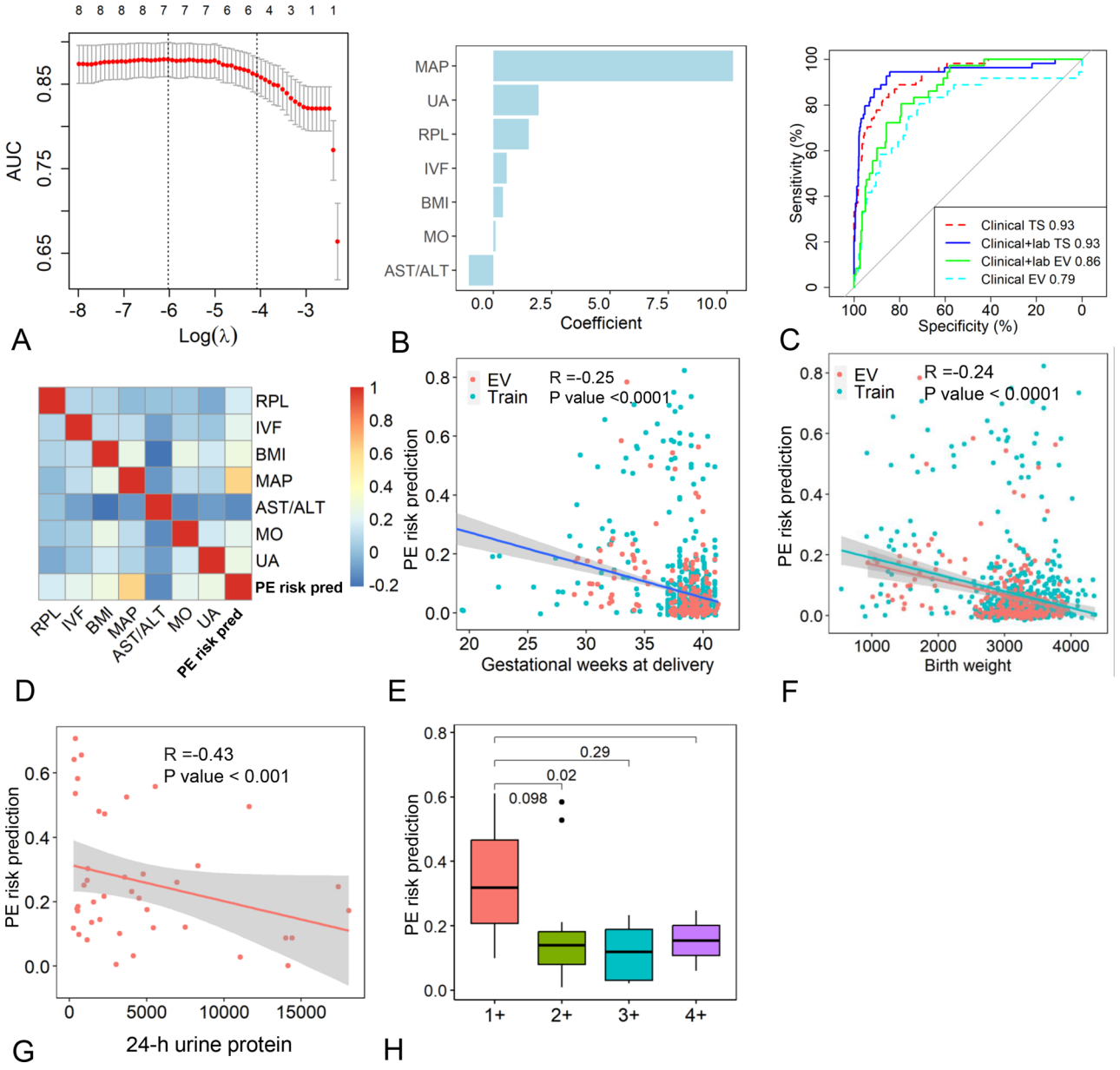


Fig. 5 The establishment and validation of the prediction models for EOPE. **A**. The ten-fold cross-validation for glmnet examined the relationship between the number of predictors and AUC values at different log (lambda) values in the training set. The left dashed line represents the Lambda.min value that maximizes the AUC value with the optimal combination of predictors, while the right dashed line indicates the lambda.1se value that yields a more regularized model with a cross-validated AUC within one standard error of the minimum. **B**. The coefficients of seven clinical and laboratory markers for predicting EOPE are presented, with MAP and AST/ALT being most positively and negatively correlated with EOPE, respectively. **C**. ROC curves are shown for the clinical and laboratory marker models as well as the clinical factor models in different datasets, including the training set (TS) and external validation set (EV). **D**. A pheatmap visualizes the correlations between seven predictors and the PE risk predicted by the ensemble EOPE model in the training and EV set. The vertical bar represents correlation coefficients, with red and blue showing high and low correlation respectively. **E–F**. The prediction scores presented significantly negative correlations with gestational weeks at delivery (E) and birth weight (F). **G**. The prediction scores showed significantly negative correlations with twenty-four hour urine protein levels in EOPE patients. **H**. Comparison of PE risk prediction among EOPE samples with different readings of + on dipstick analysis of urine specimens

The model exhibited good sensitivity (69.57%, 95% CI: 56.27%-82.86%) and specificity (85.25%, 95% CI: 80.54%-89.97%) in distinguishing LOPE from healthy controls in early pregnancy (Table2). The prediction scores showed significantly positive correlations with MAP, BMI, PLT, WBC, UA, PCT ALT, RBC, MO, LY and ALT ($r > 0.25$, $p < 0.001$ for all cases, Fig. 6D), negative correlations with AST/ALT, gestational weeks at delivery and birth weight ($r = -0.29, -0.16$ and -0.18 respectively, $p < 0.001$ for all cases, Fig. 6D-F).Furthermore, the PE risk prediction

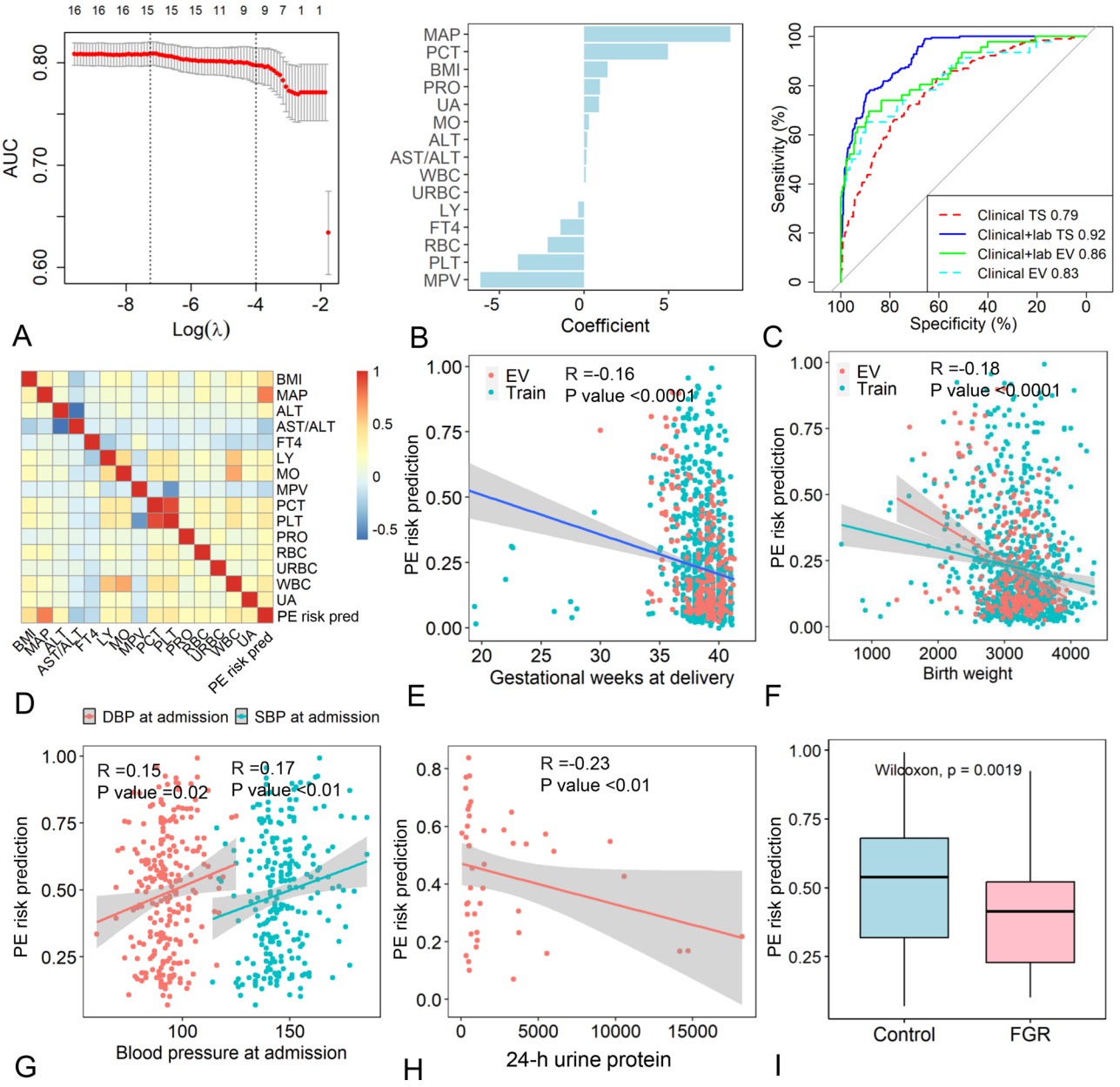


Fig. 6 The establishment and validation of the prediction models for LOPE. **A**. The glmnet method utilized ten-fold cross-validation to analyze the relationship between the number of predictors and AUC values across various log(lambda) settings in the training dataset. **B**. A bar plot displays the coefficients of 15 clinical and laboratory indicators associated with LOPE, with MAP, MPV, and PCT being the three most positively correlated with LOPE, respectively. **C**. ROC curves are presented for both the clinical and laboratory marker models, as well as the clinical factor models, across diverse datasets. **D**. A pheatmap visualizes the correlations between 15 predictors and the PE risk prediction of the LOPE model in the training and EV set. **E–F**. The prediction scores exhibit significant negative correlations with gestational weeks at delivery (E) and birth weight (F). **G**. The prediction scores show significant positive correlations with SBP and DBP at admission among LOPE samples. **H**. The prediction scores demonstrate significant negative correlations with 24-hour urine protein levels in LOPE patients. **I**. A comparison of PE risk prediction of LOPE models is provided between FGR and control groups within LOPE samples

demonstrated significant positive correlations with blood pressure at admission and negative correlation with urine protein levels in LOPE patients ($p < 0.05$ for all cases, Pearson correlation, Fig. 6G and H). LOPE samples with higher PE risk prediction were less likely to develop FGR as compared to those with lower PE risk prediction ($p < 0.01$, wilcoxon rank sum test, Fig. 6I).

Finally, we demonstrated the ensemble Preterm PE model consisting of glm, glmnet and nnet outperformed the clinical factors model in the EV dataset (Figure S11C, Figure S10C). The ensemble Preterm PE model exhibited

Table 2 The performances of the PE models in the validation set of 36 EPE, 46 LPE, 82 Preterm PE and 217 healthy participants

	Group	True positive	False positive	True negative	False negative	Sensitivity	Specificity	PPV	NPV	Cut-off
Clinical factors	EPE	21	32	185	15	58.33% (95% CI, 42.23%-74.44%)	85.25% (95% CI,80.54%-90%)	39.62% (95% CI,26.45%-52.79%)	92.5% (95% CI,88.85%-96.15%)	0.1
	Preterm PE	50	32	185	32	60.98% (95% CI, 50.42%-71.53%)	85.25% (95% CI,80.54%-90%)	60.98% (95% CI,50.42%-71.53%)	85.25% (95% CI,80.54%-90%)	0.17
	LPE	30	32	185	16	65.22% (95% CI, 51.45%-78.98%)	85.25% (95% CI, 80.54%-89.97%)	48.39% (95% CI,35.95%-60.83%)	92.04% (95% CI,88.3%-95.78%)	0.35
Clinical factors + lab variables	EPE	26	32	185	10	72.22% (95% CI, 57.59%-86.85%)	85.25% (95% CI, 80.54%-89.97%)	44.83% (95% CI,32.03%-57.63%)	94.87% (95% CI,91.78%-97.97%)	0.09
	Preterm PE	52	32	185	30	63.41% (95% CI, 52.99%-73.84%)	85.25% (95% CI, 80.54%-89.97%)	61.9% (95% CI,51.52%-72.29%)	86.05% (95% CI, 81.41%-90.68%)	0.18
	LPE	32	32	185	14	69.57% (95% CI, 56.27%-82.86%)	85.25% (95% CI, 80.54%-89.97%)	50% (95% CI,37.75%-62.25%)	92.96% (95% CI,89.41%-96.52%)	0.36

good accuracy in separating Preterm PE from healthy participants in early pregnancy (sensitivity: 63.41%, 95% CI, 52.99%-73.84%; specificity: 85.25%, 95% CI: 80.54%-89.97%, Table2) in the EV set. The predictor MAP had highest correlation with PE risk prediction, followed by BMI, UA, MO, WBC, ALT, AST/ALT ($p < 0.001$ for all cases, Pearson correlation, Figure S10D), indicating the model captures differences in both clinical and laboratory markers. The prediction scores obtained from the ensemble Preterm model exhibited significantly negative correlations with gestational weeks at delivery and birth weight ($r = -0.24$ and -0.23 respectively, $p < 0.001$ for all cases, Pearson correlation, Figure S10E-F). The PE risk prediction exhibited a notable positive association with blood pressure at admission and a negative correlation with urine protein levels in patients with preterm preeclampsia, with statistical significance observed in all cases ($p < 0.05$, using Pearson correlation; see Figures S10G-I).

Discussion

In the present study, we systematically investigated the predictive values of various clinical characteristics and 45 routine prenatal laboratory test parameters for different subtypes of PE in early pregnancy using all available clinical and laboratory data from six hospitals. We confirmed that pregnant women with higher MAP and BMI, IVF, all known PE risk factors, had a significantly higher risk for PE than those with lower MAP, BMI and without IVF. Our study also found that participants with RPL are more likely to develop EOPE. While the association between RPL and pre-eclampsia is not yet fully understood, several studies have investigated the relationship between RPL and PE. These studies revealed that RPL is strongly associated with preterm PE [21, 22]. Trogstad et al reported a significantly elevated risk of PE in cases of RPL, only when there was a history of assisted reproduction [23]. Two other studies have failed to find a higher risk for PE in RPL women, which could be attributed to the small size of women with RPL [24] and the lack of a strict definition of consecutive miscarriages [25]. Our study showed the association with RPL was statistically significant for EOPE rather than LOPE. Although an increased risk of preterm PE was also observed in women with recurrent miscarriages, the association was not statistically significant, possibly because of the smaller size of our study as compared with previous publications [21, 22].

Based on the analysis of laboratory test variables, several laboratory test results were found to be significantly associated with different subtypes of PE, such as UA, ALT, RBC, MO, AST/ALT, WBC and UPRO. In line with our findings, higher plasma levels of ALT and lower AST/ALT are significantly associated with elevated risk of PE in early pregnancy [26, 27]. We found that increased serum UA in early pregnancy was an independent risk factor of PE, our results are consistent with previous studies [27, 28]. Many experimental studies report that UA might play a pivotal role in the development of hypertension. Hyperuricemia-induced hypertension can be prevented by UA-lowering levels [29]. Hyperuricemia also affects the renin-angiotensin system, induces endothelial dysfunction and inhibits neuronal nitric oxide synthase[29, 30]. Additionally, UA can induce trophoblastic production of pro-inflammatory interleukin-1β through activation of inflammatory pathways [31], which may contribute to the progression of high blood pressure and the development of PE during pregnancy. In our study, we have identified PE biomarkers from routine blood tests that have been previously reported in

the literature. For instance, biomarkers such as WBC, LY, RBC, PLT, and MO were found to be up-regulated in LOPE patients, which is consistent with previous studies[13, 14, 32]. While, some biomarkers such as MPV and FT4, were found to be negatively associated with LOPE, which contradicts previous findings[33, 34]. Monteith' et al reported MPV was significantly increased in the third trimester but exhibited no significant difference during the first trimester in EOPE patients as compared to heathy participants [33]. However, Oğlak' et al study has uncovered a significant increase in first trimester MPV values in patients who developed preeclampsia in later pregnancy [34]. Shan-Shan Lin' et al has identified MPV was down-regulated at 8–12 gestational weeks but significantly up-regulated after 16 weeks of gestation in PE patients in comparison with healthy controls [35], which supports the finding in our study. The expression of FT4 in preeclampsia (PE) compared to healthy controls has shown inconsistent results [36, 37]. In our study, involving a significant number of PE and control subjects from multiple clinical centers, we confirmed that FT4 expression is decreased in PE cases and can serve as an informative biomarker for PE prediction. Additionally, we identified a new biomarker called PCT, which is increased in early pregnancy and can be useful in evaluating the risk of preeclampsia.

Maric et al. developed a prediction model using the elastic net algorithm by incorporating clinical and laboratory variables. However, their findings indicated that laboratory results did not significantly enhance the predictive capability of the clinical factor models [13]. In a separate study by Li et al., predictors for PE at 12 weeks of gestation were identified, including maternal characteristics and certain laboratory variables from routine blood tests. Their gradient boosting model exhibited relatively poor performance in screening for PE in early pregnancy [14]. In comparison to these previous studies, our study offers several advantages. Firstly, the clinical and laboratory data of the above two studies come from one or two medical centers, the results of their studies are lack of indEOPEndent validation. We conducted a multicenter case-control study, collecting clinical and laboratory data from six hospitals, providing independent and external validation for our results. This strengthens the reliability and credibility of our findings. Secondly, while Maric et al focused on predicting preeclampsia and EOPE separately, and Li et al evaluated all PE patients, we conducted a comprehensive analysis by investigating predictive markers for different subtypes of PE and reported the common and specific biomarkers for seven PE types. Lastly, we established three distinct models and thoroughly assessed their performances, providing a more comprehensive understanding of PE prediction. Furthermore, not only did the models provide risk assessment for different subtypes of PE but also evaluate the severity of PE, such as blood pressure at admission, 24-hour urine protein levels and FGR. These predictions will help clinicians take preventive measures to reduce the incidence of PE and improve clinical outcomes.

Limitations of this study

The established integrated models present superior performance to clinical factor models in predicting EOPE and LOPE. These laboratory variables are easily accessible through routine prenatal laboratory tests, enhancing the practicality of our integrated models for early pregnancy PE prediction, but it is important to acknowledge its limitations. The models have been established based on a multi-center case-control study. However, we are uncertain about their effectiveness in screening for PE in the general population. To address this, we plan to conduct a large prospective cohort study involving more tertiary centers in the future. The screening models recommended by FIGO incorporates maternal characteristics, uterine artery Doppler measurements and PLGF and PAPPA [6–8] to early predict preeclampsia. Since our study was retrospective, we were unable to evaluate the predictive performance of our model when combined with biochemical markers such as PLGF and PAPPA, as well as Doppler ultrasound imaging. Future studies should be conducted to create prediction models that incorporate these established predictive biomarkers and laboratory parameters, we believe that the performance of the combined model could be further enhanced.

Abbreviations

PE	Preeclampsia
EOPE	Early-onset preeclampsia
LOPE	Late-onset preeclampsia
CI	Confidence interval
LC-MS-MS	Liquid Chromatography with tandem mass spectrometry
SD	Standard deviation
BMI	Body mass index
DBP	Diastolic blood pressure
SBP	Systolic blood pressure
IVF	In vitro fertilization
PMH	Past medical history
MAP	Mean arterial pressure
AUC	Area under the curves
ROC	The receiver operating characteristic curves
MCHC	Mean corpuscular hemoglobin concentration
FT4	Free Thyroxine
PMO	Proportion of monocytes
EO	Eosinophils count
CRE	Creatinine
URBC	Urine red blood cell
AST	Aspartate aminotransferase
LY	Lymphocytes count
ALT	Alanine Transaminase
T-BIL	Total Bilirubin Test
SG	Urine specific gravity
PEO	Proportion of Eosinophils
RBC	Red blood cell count
MPV	Mean platelet volume
HBsAg	Hepatitis B surface antigen
MO	Monocytes count

HCT	Hematocrit
Hb	Hemoglobin
BA	Basophils count
UWBC	Urine white blood cell
WBC	White blood cell count
TSH	Thyroid stimulating hormone
X-TAL	Crystals
BLD	Urine blood
EV	External validation
NE	Neutrophils
PLT	Platelet count
PDW	Platelet distribution width
PCT	Plateletcrit
MCH	Mean corpuscular hemoglobin
MCV	Mean corpuscular volume
T-BIL	Total Bilirubin Test
TP	Total protein
GLU	Fasting glucose
CRE	Creatinine
UA	Uric acid
PRO	Protein
TSH	Thyroid-stimulating hormone
FT4	Free thyroxine
TPOAb	Thyroid peroxidase antibody
PlGF	Placental growth factor
PAPPA	Pregnancy-associated plasma protein A
FGR	Fetal growth restriction

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02999-5.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
None.

Author contributions
Suihua FENG, Shuyuan LI,Jianguo ZHANG, Lijian ZHAO, Fengxiang WEI: Conceptualization. Songchang CHEN, Jia LI, Xiao ZHANG, Wenqiu XU, Zhixu QIU, Siyao YAN: Methodology, Software. Wenrui ZHAO, Zhiguang ZHAO, Peirun TIAN, Qiang ZHAO, Qun ZHANG, Weiping CHEN, Huahua LI, Xiaohong RUAN, Gefei XIAO, Sufen ZHANG, Liqing HU, Jie QIN, Wuyan HUANG, Zhongzhe LI, Shunyao WANG, Rui ZHANG, Shang HUANG, Xin WANG, Yao YAO, Jian RAN, Danling CHENG, Qi LUO, Teng PAN, Ruyun GAO, Jing ZHENG, Yuxuan WANG, Cong LIU, Xianling CAO, Xuanyou ZHOU, Naixin XU, Lanlan ZHANG, Xu HAN, Haolin WANG: Data curation. Jia LI, Xiao ZHANG, Wenqiu XU, Zhixu QIU: Writing- Original draft preparation. Qi LUO, Teng PAN,Ruyun GAO, Jing ZHENG, Yuxuan WANG: Visualization, Investigation. Wenrui ZHAO, Zhiguang ZHAO, Peirun TIAN: Software, Validation. Suihua FENG, Shuyuan LI,Jianguo ZHANG, Lijian ZHAO, Fengxiang WEI: Writing- Reviewing and Editing.

Funding
The study was supported by the research grant from the sub-project of the National Key R&D Program (No: 2023YFC2705600), the Shanghai Municipal Commission of Science and Technology Program (23ZR1408000), Shanghai Municipal Commission of Health and family planning (2024ZZ2017), Zhuhai Social Development Science and Technology Plan Project (No:2220004000295), Shenzhen Key Medical Discipline Construction Fund (No: SZXK028), Shenzhen Science and Technology Innovation Committee (No: KJYY20180703173402020), S&T Program of Shijiazhuang (No: 235790429H) and Key Research and Development Projects of Hebei Province (No:21377720D). Shenzhen Science, Technology, and Innovation Committee (No: JCYJ20220530162412029); Science and Technology Development Special Fund of Shenzhen Longgang District (No: LGKCYLWS2022008).

Data availability
The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0004774.

Declarations

Ethics approval and consent to participate
All experiments and methods were performed in accordance with relevant guidelines and regulations. This study was approved by the Ethics Committees of Beijing Genomics Institute (BGI-IRB 22026) and the Ethics Committee of each participating hospital. Clinical trial number is not applicable.

Consent for publication
Informed consent was waived by the institutional review boards (IRBs) to allow us to use the clinical data for this study. All data underwent deidentification processes.

Competing interests
The authors report no conflict of interest.

Author details
[1]Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai 200011, China
[2]Hebei Industrial Technology Research Institute of Genomics in Maternal & Child Health, Clin Lab, BGI Genomics, Shijiazhuang, Hebei Province 050011, China
[3]BGI Genomics, Shenzhen, Guangdong Province 518083, China
[4]Hebei Medical University, Shijiazhuang, Hebei Province 050011, China
[5]Department of Obstetrics and Gynecology, Jiangmen Central Hospital, Jiangmen, Guangdong Province 529030, China
[6]Clinical Transformation and Application Key Lab for Obstetrics and Gynecology, Pediatrics, and Reproductive Medicine of Jiangmen, Jiangmen, Guangdong Province 529030, China
[7]Department of Medical Genetics and Prenatal Diagnosis, Zhuhai Center for Maternal and Child Health Care, Zhuhai, Guangdong Province 519000, China
[8]Department of Prevention and Health Care, Zhuhai Center for Maternal and Child Health Care, Zhuhai, Guangdong Province 519000, China
[9]Hunan Provincial Key Laboratory of Regional Hereditary Birth Defects Prevention and Control, Changsha Hospital for Maternal & Child Health Care Affiliated to Hunan Normal University, Changsha, Hunan Province 410007, China
[10]Division of Maternal-Fetal Medicine, Shenzhen Bao'an Women's and Children's Hospital, Shenzhen, Guangdong Province 518102, China
[11]The Genetics Laboratory, Longgang District Maternity and Child Healthcare Hospital of Shenzhen City (Longgang Maternity and Child Institute of Shantou University Medical College), Shenzhen, Longgang District, Guangdong Province 518712, China
[12]Longgang District Maternity and Child Healthcare Hospital of Shenzhen City (Longgang Maternity and Child Institute of Shantou University Medical College), Guangdong Province 518712 Shenzhen, China
[13]Department of Blood Transfusion, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School, Nanjing Medical University, 80 Tianedang Road, Suzhou 518712 Jiangsu, China
[14]School of Basic Medical Sciences, Jiamusi University, Jiamusi, Heilongjiang Province 154007, China
[15]School of Public Health, Hebei Medical University, Hebei Province Key Laboratory of Environment and Human Health, Shijiazhuang, Hebei Province 050017, China
[16]International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200030, China
[17]School of Computer Science, Guangzhou College of Technology and Business, Guangdong Province 510850, China

Received: 6 January 2025 / Accepted: 8 April 2025
Published online: 01 May 2025

References

1. Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2013 Sep;170(1):1–7.
2. Poon LC, Shennan A, Hyett JA, Kapur A, Hadar E, Divakar H, et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: a pragmatic guide for first-trimester screening and prevention. Int J Gynecol Obstet. Internet]. 2019 May 1;145(S1):1–33. Available from: https://doi.org/10.1002/ijgo.12802.
3. Magee LA, Nicolaides KH, von Dadelszen PP. N Engl J Med. 2022; May;386(19):1817–32
4. Dimitriadis E, Rolnik DL, Zhou W, Estrada-Gutierrez G, Koga K, Francisco RPV, et al. Pre-eclampsia. Nat Rev Dis Primers. 2023; Feb;9(1):8.
5. Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, de Paco Matallana C, et al. Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. N Engl J Med. 2017; Aug;377(7):613–22.
6. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol. 2015 Jul;213(1):62.e1–62.e10.
7. Odibo AO, Zhong Y, Goetzinger KR, Odibo L, Bick JL, Bower CR, et al. First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. Placenta. 2011; Aug;32(8):598–602.
8. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. Am J Obstet Gynecol. 2020 Jul;223(1):12–23.e7.
9. Cheng Y, Leung TY, Law LW, Ting YH, Law KM, Sahota DS. First trimester screening for pre-eclampsia in Chinese pregnancies: case-control study. BJOG. 2018 Mar;125(4):442–49.
10. Hu J, Gao J, Liu J, Meng H, Hao N, Song Y, et al. Prospective evaluation of first-trimester screening strategy for preterm pre-eclampsia and its clinical applicability in China. Ultrasound Obs Gynecol. 2021; Oct;58(4):529–39.
11. Chaemsaithong P, Pooh RK, Zheng M, Ma R, Chaiyasit N, Tokunaka M, et al. Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population. Am J Obstet Gynecol. 2019; Dec;221(6):650.e1–650.e16.
12. Chaemsaithong P, Sahota DS, Poon LC. First trimester preeclampsia screening and prediction. Am J Obstet Gynecol Internet]. 2022;226(2):S1071–S1097.e2. Available from: https://doi.org/10.1016/j.ajog.2020.07.020.
13. Marić I, Tsur A, Aghaeepour N, Montanari A, Stevenson DK, Shaw GM, et al. Early prediction of preeclampsia via machine learning. Am J Obstet Gynecol MFM 2020;2:1–17.
14. Li S, Wang Z, Vieira LA, Zheutlin AB, Ru B, Schadt E, et al. Improving pre-eclampsia risk prediction by modeling pregnancy trajectories from routinely collected electronic medical record data. Npj Digit Med 2022;5:1–16.
15. Ranjbar A, Taeidi E, Mehrnoush V, Roozbeh N, Darsareh F. Machine learning models for predicting pre-eclampsia: a systematic review protocol. BMJ Open. 2023 Sep;13(9):e074705.
16. Yang Z, Zhang W. Diagnosis and treatment of hypertension and pre-eclampsia in pregnancy: a clinical practice guideline in China (2020). Chin J Obs Gynecol. 2020 Apr;55(4):227–38.
17. Mikolajczyk RT, Zhang J, Betran AP, Souza JP, Mori R, Gülmezoglu AM, et al. A Global Reference for Fetal-weight and Birthweight Percentiles. Lancet (London, England:2011. May;377(9780). 1855–61.
18. Grillo A, Salvi P, Furlanis G, Baldi C, Rovina M, Salvi L, et al. Mean arterial pressure estimated by brachial pulse wave analysis and comparison with currently used algorithms. J Hypertens 2020;38:2161–68.
19. Tay JK, Narasimhan B, Hastie T. Elastic Net Regularization Paths for All Generalized Linear Models. J Stat Softw. 2023;106.
20. Oyston CJ, Stanley JL, Baker PN.Potential targets for the treatment of pre-eclampsia. Expert Opin Ther Targets. 2015;19(11):1517–30.
21. Gunnarsdottir J, Stephansson O, Cnattingius S, Åkerud H, Wikström AK. Risk of placental dysfunction disorders after prior miscarriages: a population-based study. Am J Obstet Gynecol Internet]. 2014;211(1):34.e1–34.e8. Available from: https://doi.org/10.1016/j.ajog.2014.01.041.
22. Roepke ER, Christiansen OB, Källén K, Hansson SR.Women with a history of recurrent pregnancy loss are a high-risk population for adverse obstetrical outcome: a retrospective cohort study. J Clin Med. 2021;10(2):1–12.
23. Trogstad L, Magnus P, Moffett A, Stoltenberg C.The effect of recurrent miscarriage and infertility on the risk of pre-eclampsia. BJOG An Int J Obstet Gynaecol. 2009;116(1):108–13.
24. Dempsey MA, Flood K, Burke N, Fletcher P, Kirkham C, Geary MP, et al. Perinatal outcomes of women with a prior history of unexplained recurrent miscarriage. J Matern Neonatal Med 2015;28:522–25.
25. Field K, Murphy DJ.Perinatal outcomes in a subsequent pregnancy among women who have experienced recurrent miscarriage: a retrospective cohort study. Hum Reprod. 2015;30(5):1239–45.
26. Zhang Y, Sheng C, Wang D, Chen X, Jiang Y, Dou Y, et al. High-normal liver enzyme levels in early pregnancy predispose the risk of gestational hypertension and preeclampsia: a prospective cohort study. Front Cardiovasc Med. 2022;9.
27. Chen Y, Ou W, Lin D, Lin M, Huang X, Ni S, et al. Increased Uric Acid, Gamma-Glutamyl Transpeptidase and Alkaline Phosphatase in Early-Pregnancy Associated With the Development of Gestational Hypertension and Preeclampsia. Front Cardiovasc Med 2021;8:1–11.
28. Wolak T, Sergienko R, Wiznitzer A, Paran E, Sheiner E.High uric acid level during the first 20 weeks of pregnancy is associated with higher risk for gestational diabetes mellitus and mild preeclampsia. Hypertens Pregnancy. 2012;31(3):307–15.
29. Liu C-W, Ke S-R, Tseng G-S, Wu Y-W, Hwang -J-J. Elevated serum uric acid is associated with incident hypertension in the health according to various contemporary blood pressure guidelines. Nutr, Metab Cardiovasc Dis. 2021 Apr;31(4):1209–18.
30. Ponticelli C, Podestà MA, Moroni G. Hyperuricemia as a trigger of immune response in hypertension and chronic kidney disease. Kidney Int. 2020 Nov;98(5):1149–59.
31. Mulla MJ, Myrtolli K, Potter J, Boeras C, Kavathas PB, Sfakianaki AK, et al. Uric acid induces trophoblast IL-1β production via the inflammasome: implications for the pathogenesis of preeclampsia. Am J Reprod Immunol. 2011; Jun;65(6):542–48.
32. Castleman JS, Lip GYH, Shantsila E. Monocytes are increased in pregnancy after gestational hypertensive disease. Sci Rep Internet]. 2022;12(1):1–8. Available from: https://doi.org/10.1038/s41598-022-13606-2.
33. Monteith C, Egan K, O'Connor H, Maguire P, Kevane B, Szklanna PB, et al. Early onset preeclampsia is associated with an elevated mean platelet volume (MPV) and a greater rise in MPV from time of booking compared with pregnant controls: results of the CAPE study. J Perinat Med 2018;46:1010–15.
34. Oğlak SC, Tunç Ş, Ölmez F.First Trimester Mean Platelet Volume, Neutrophil to Lymphocyte Ratio, and Platelet to Lymphocyte Ratio Values Are Useful Markers for Predicting Preeclampsia. Ochsner J. 2021;21(4):364–70.
35. Lin SS, Wang CR, Wei DM, Lu JH, Chen XJ, Chen QZ, et al. Incremental predictive value of platelet parameters for preeclampsia: results from a large prospective cohort study. BMC Pregnancy Childbirth. Internet]. 2023;23(1):1–11. Available from: https://doi.org/10.1186/s12884-023-05661-y.
36. Maduka I, Obi N, Egwu M.Evaluation of Thyroid Hormones and Thyroid Antibodies in Nigerian Pregnant Women with Pre-eclampsia. J Med Lab. 2020;30(4):52–61.
37. Misra M, Singh L, Thakur A, Singh S, Kumar J, Juyal D. Comparative Analysis of Serum Calcium Level and Thyroid Profile in Pregnant Normotensive and Preeclamptic Women: a Case-control Study. J Clin Diagn Res. 2020;5–9.

Publisher's Note