


RESEARCH

Open Access



# Prediction of postpartum depression in women: development and validation of multiple machine learning models

Weijing Qi<sup>1†</sup>, Yongjian Wang<sup>1,2†</sup>, Yipeng Wang<sup>1</sup>, Sha Huang<sup>1</sup>, Cong Li<sup>1</sup>, Haoyu Jin<sup>1</sup>, Jinfan Zuo<sup>1</sup>, Xuefei Cui<sup>1</sup>, Ziqi Wei<sup>1</sup>, Qing Guo<sup>3\*</sup> and Jie Hu<sup>1\*</sup> 

## Abstract

**Background** Postpartum depression (PPD) is a significant public health issue. This study aimed to develop and validate machine learning (ML) models using biopsychosocial predictors to predict the risk of PPD for perinatal women and to provide several risk assessment tools for the early detection of PPD.

**Methods** Candidate predictors, including history of mental illness and demographic, psychosocial, and physiological factors, were obtained from 1138 perinatal women between August 2021 and August 2022. The primary outcome of PPD was measured with the Edinburgh Postnatal Depression Scale at 6 weeks postpartum. Seven feature selection methods and six ML algorithms were employed to develop models, and their prediction performances were compared.

**Results** A total of 11 potential predictive factors associated with PPD were identified and subsequently used to construct prenatal and postpartum predictive models for PPD. The cross-validation results showed that the models built on logistic regression (LR) [area under the curve (AUC): 0.801, 0.858] and artificial neural network (ANN) (AUC: 0.787, 0.844) algorithms exhibited the best prediction performance. In contrast to the prenatal models, the addition of postpartum predictors (primary caregiver and mother-in-law's care) remarkably improved the predictive performance of the postpartum models. The risk-stratification score, the nomogram, and the Shapley additive explanation were used to visualize and interpret the risk prediction model for predicting PPD in the early stage.

**Conclusions** The LR and ANN models achieved the best predictive performances. Applying these models and risk assessment tools to early predict and screen PPD has several implications for public health.

**Keywords** Postpartum depression, Machine learning, Predictive factors, Prediction model

<sup>†</sup>Weijing Qi and Yongjian Wang contributed equally to this work.

\*Correspondence:

Qing Guo

yfguoqing@163.com

Jie Hu

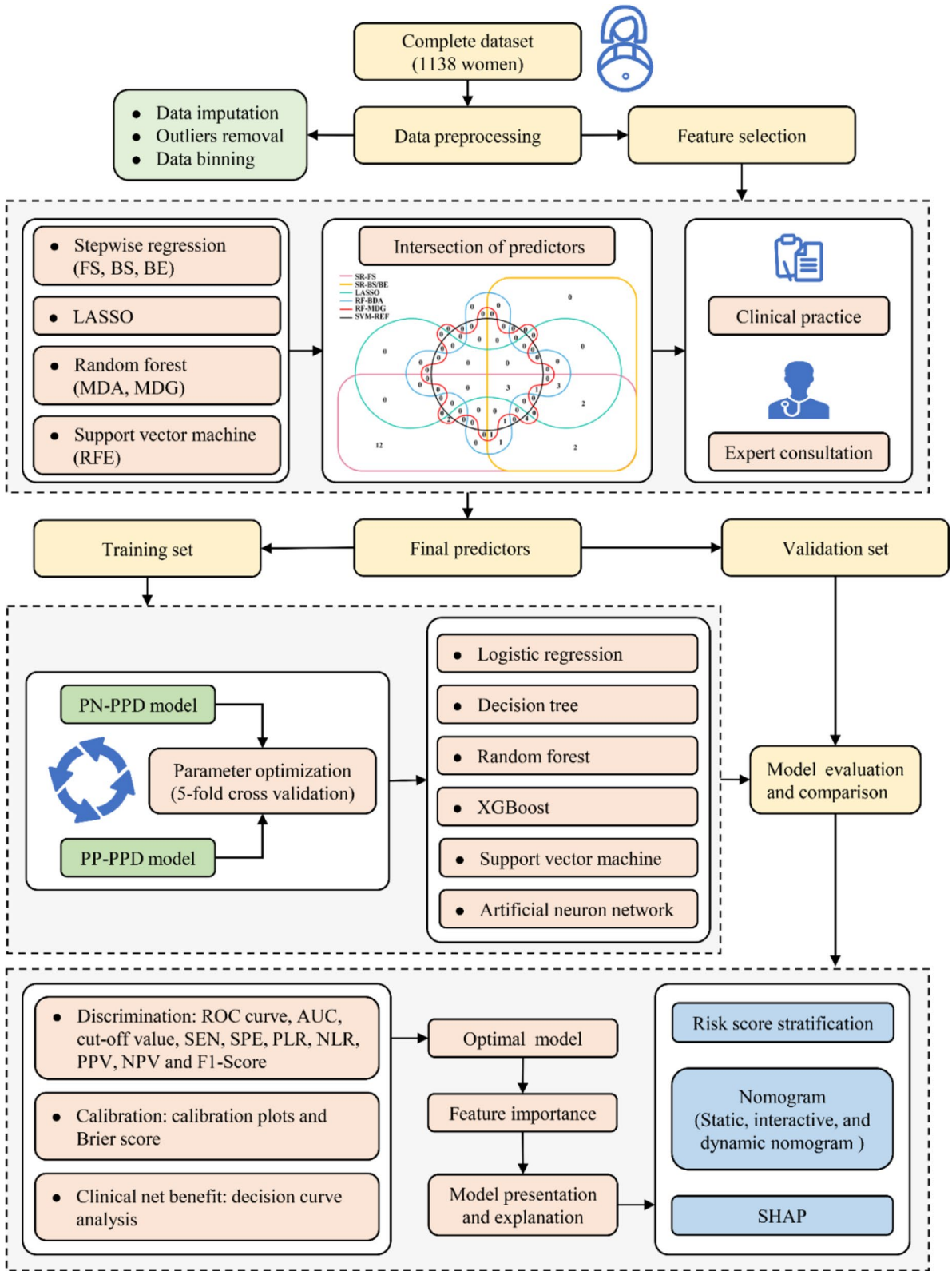
Hujie@hebmh.edu.cn

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



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Graphical Abstract



## Introduction

The physical health of perinatal women has dramatically improved over the past few decades, leading to a substantial decline in miscarriage and mortality rates [25, 38]. However, their mental health has increasingly become a global public health problem [24]. Postpartum depression (PPD) is an apparent depressive symptom or a typical depressive episode in the perinatal period [45]. As the most common type of perinatal psychiatric syndrome, PPD is frequently characterized by persistent low mood, anhedonia, and loss of pleasure [2]. The largest and most inclusive meta-analysis of PPD to date found that the global pooled prevalence of PPD was 17.22% (95% CI 16.00%–18.51%) [43]. Notably, the coronavirus disease 2019 (COVID-19) has evolved into a global pandemic and further exacerbated mental health risks, especially in perinatal women [12]. Therefore, women's postpartum depressive symptoms should receive adequate global public health attention during this period. Previous studies have suggested that PPD is affected by a complex combination of factors, such as demographic, physical, psychological, social, and obstetrics-related factors [8, 48]. However, the underlying mechanism of PPD remains unclear, so it is still an enormous challenge to prevent and intervene PPD.

PPD not only has severe and lasting adverse effects on mothers, infants, and partners [40, 42] but also affects harmonious family relationships, increases medical expenditures, and impedes social development [34, 44]. Although the exact etiology of PPD is unknown, early identification and appropriate intervention by combining biopsychosocial factors to predict the risk of PPD can help prevent the abovementioned adverse effects. Consistently, self-report questionnaires [such as the Edinburgh Postnatal Depression Scale (EPDS) and Beck Depression Inventory (BDI)] are the primary tools for PPD screening, and diagnosis is mainly dependent on the presence of clinical symptoms [19]. In this case, PPD may have already occurred but may not have been noted before the screening. In addition, some perinatal women with PPD are affected by social stigma and are inclined to hide their clinical symptoms [28]. Therefore, developing appropriate predictive models can be helpful in quickly identifying perinatal women with PPD and facilitating supportive care or amelioration of the disease course before overt symptoms develop.

Recently, multiple studies developed prediction models to predict the risk of PPD in women, and most models showed good generalization performance and predictive capability [11]. Nevertheless, only a single type of feature selection method or model construction method was used in some studies [3, 46]. Recent literature on prediction models suggested that results from different

prediction models would be potentially helpful for accuracy improvement [31]. On the other hand, similar to most psychiatric disorders, PPD has a complex etiology involving an interplay of biopsychosocial factors. Traditional modeling approaches have certain limitations in handling complex and multidimensional data [10]. With the advancement of computer technology and data sciences, the emergence of machine learning (ML) algorithms provides a powerful approach for addressing the limitations of traditional methods and has been widely applied to develop diagnostic or prognostic predictive models for improving health care in public health and medical research fields [30]. Unfortunately, some prediction models for PPD are based on complex ensemble learning algorithms, which may not be interpreted biologically [22, 49]. Moreover, few studies have been translated into clinical assessment tools to guide clinical decision-making.

Inspired by these advances and limitations, the purpose of the present study was to (1) identify the predictors of PPD, (2) develop multiple risk prediction models for perinatal women, and (3) design various clinical assessment tools for early screening and personalized care.

## Methods

### Study population

This prospective cohort study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines [13] (Supplementary TRIPOD checklist). Women who underwent obstetric examinations were recruited from a public tertiary maternity hospital between August 2021 and August 2022. To ensure sufficient registry information before and after childbirth, we only recruited women over 18 years of age, had a gestation period  $\leq 21$  weeks, and underwent routine obstetric and laboratory examinations until delivery at this hospital. Women with pregnancy losses (including abortion, miscarriage, or stillbirth), as well as those with preterm deliveries (defined as deliveries occurring before 37 weeks of gestation), were excluded from the study. This exclusion was made to minimize the confounding effects of pregnancy-related complications on the prediction of PPD risk, as these conditions are associated with distinct psychological and physiological stressors. In addition, pregnant women diagnosed with fetal structural or chromosomal abnormalities were also excluded. In the study, all participants provided informed consent before participation and were asked to attend four follow-up visits for data collection: second trimester (gestation weeks 21–24), third trimester (gestation weeks 35–40), 2 weeks, and 6 weeks postpartum.

### Data collection

Candidate predictors associated with PPD were selected based on a literature review [36, 48] and consulting experts in the field. These predictors are easily ascertained and readily available in different clinical scenarios. The actual data were collected prospectively at four-time points and could be classified into four categories:

1. demographic characteristics (age, residence, housing condition, monthly income, education, primiparous women, and planning pregnancy);
2. history of mental illness (women were diagnosed with mental illness before pregnancy or first-degree relatives were diagnosed with mental illness);
3. psychosocial factors (primary caregiver in prenatal (PCPN) and postpartum (PCPP), mother-in-law's care in prenatal (MCPN) and postpartum (MCP) (a 10-point scale ranging from 1 "very poor" to 10 "excellent"), stress-coping style [simplified coping style questionnaire [47] (SCSQ)], personality [Eysenck personality questionnaire [20] (EPQ)] [including psychoticism dimension (EPQ-P), extraversion dimension (EPQ-E), neuroticism dimension (EPQ-N), and melancholic temperament (MT)], social support [perceived social support scale [7] (PSSS)], prenatal anxiety [Beck anxiety inventory [5] (BAI)], prenatal depression [Beck depression inventory [6] (BDI)], marital satisfaction [Enrich marital satisfaction scale [17] (EMSS)], and sleep quality during late gestation [Pittsburgh sleep quality inventory [9] (PSQI)]; and
4. physiological measures [women's blood parameters in the third trimester, including thyroid function tests (TSH, FT3, FT4), blood lipid assays (TG, TC, HDL-C, LDL-C), serum calcium (Ca) and iron (Fe) levels].

According to the number of predictor parameters, we performed a sample size analysis using the "pmsampsize" package in R language and obtained a required sample size of 1014 women. Eventually, 1138 (89.89% overall adherence rate) participants completed all follow-up evaluations and questionnaires. After data collection, all participants were anonymized and assigned internal identification codes. Figure 1 illustrates the steps of participant recruitment and data collection.

### Study outcome

The PPD outcome was assessed using the Edinburgh Postnatal Depression Scale [14] (EPDS) at 6 weeks postpartum. The EPDS is an internationally used 10-item self-report questionnaire to assess the presence and severity of postpartum depressive symptoms. It has satisfactory

psychometric properties and has been translated into several languages. The United Kingdom National Institute for Health and Care Excellence guidelines [32] and the United States preventive services task force [39] recommend the EPDS for screening PPD, but a clear cutoff value has yet to be identified. To avoid missed diagnoses and identify most patients, this study's EPDS cutoff value was set as 11 [27].

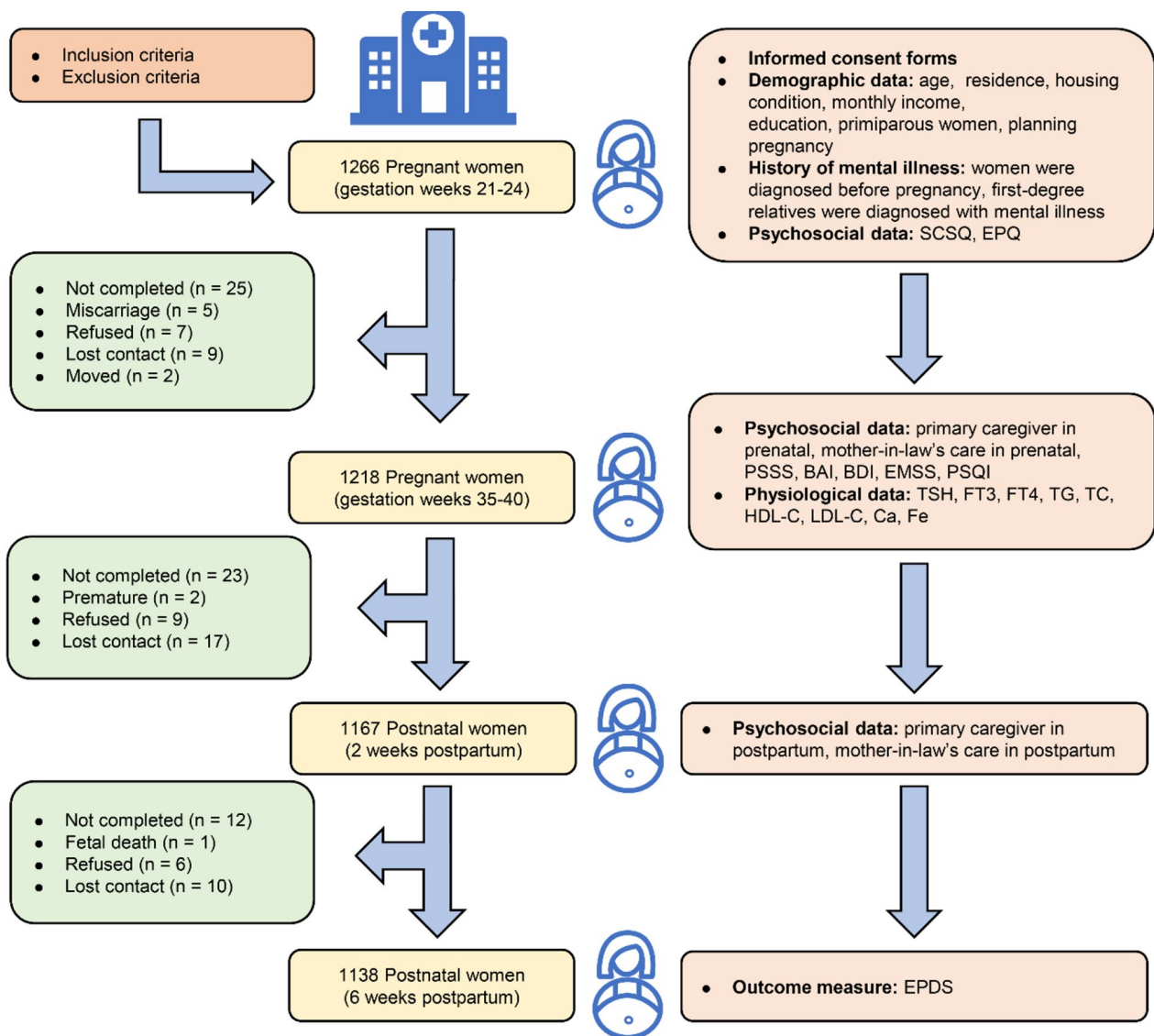
### Feature selection

Before developing the predictive models, seven feature selection methods were applied to the study population to mitigate high correlations among predictors and capture complex relationships between predictors and the outcome variable. These methods were chosen based on their established effectiveness in handling high-dimensional data and identifying the most relevant predictors in predictive modeling. The feature selection methods used were:

1. Stepwise Regression (SR): This method can iteratively add or remove predictors based on statistical significance, making it useful for identifying the most important variables while controlling for multicollinearity. We implemented three variations of stepwise regression: forward selection (FS), backward selection (BS), and bidirectional elimination (BE).
2. Least Absolute Shrinkage and Selection Operator (LASSO): By applying a penalty to the coefficients of less important predictors, LASSO helps to shrink coefficients toward zero, thus identifying a subset of relevant features while avoiding overfitting.
3. Random Forest (RF): We utilized two variations of the RF method: mean decrease accuracy (MDA) and mean decrease Gini impurity (MDG). RF was chosen for its robustness in handling nonlinear relationships and interactions between predictors, as well as its ability to rank predictors based on their importance in predicting the outcome.
4. Support Vector Machine-Recursive Feature Elimination (SVM-RFE): Utilized for its iterative elimination strategy to optimize feature subsets based on model performance, ensuring relevance to the classifier.

We summarized 35 candidate predictors screened by different filtering methods and took their intersection. Under this approach, we chose the predictors with more than five intersections. Finally, we consulted with clinical experts and combined them with clinical reality to determine the final predictors for developing a predictive model of PPD.





**Fig. 1** Flowchart of participants recruitment and data collection. SCSQ simplified coping style questionnaire, EPQ Eysenck personality questionnaire, PSSS perceived social support scale, BAI Beck anxiety inventory, BDI Beck depression inventory, EMSS Enrich marital satisfaction scale, PSQI Pittsburgh sleep quality inventory, TSH Thyroid-stimulating hormone, FT3 serum-free triiodothyronine, FT4 serum-free thyroxine, TG triglyceride, TC total cholesterol, HDL-C high-density lipoprotein-cholesterol, LDL-C low-density lipoprotein-cholesterol, Ca serum calcium, Fe serum iron, EPDS Edinburgh postnatal depression scale

### Model development and evaluation

We calculated the sample size again based on the final predictors to determine the split ratio. Then, the entire dataset was randomly split, stratified by class, into a training and a validation set. The training set was used for model development by six ML algorithms, while the validation set was used for model evaluation. ML algorithms include logistic regression (LR), decision tree (DT), RF, extreme gradient boosting (XGBoost), SVM, and artificial neuron network (ANN). A grid search with fivefold cross-validation was used to obtain optimized

parameters (Supplementary Hyperparameter Tuning Details). Furthermore, optimal models were developed separately for prenatal and postnatal women to predict PPD [PN-PPD model (based on prenatal predictors) and PP-PPD model (based on all predictors)].

We assessed and compared all models' discrimination, calibration, and clinical net benefit performances. Multiple discrimination metrics included sensitivity (SEN, this is also called Recall) and specificity (SPE), the area under the receiver-operating curve (AUC), the cutoff value, the positive likelihood ratio (PLR), the

negative likelihood ratio (NLR), the positive predictive value (PPV, this is also called Precision), the negative predictive value (NPV) and the F1-Score. The AUCs of these models were compared by using the DeLong test. The net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were used to evaluate the additional predictive ability of the PP-PPD model. Calibration was assessed by calibration plots and Brier scores. Clinical usefulness and net benefit were estimated with decision curve analysis (DCA). For interpreting complex optimal models, the SHAP was implemented in Python using the shap package (<https://shap.readthedocs.io/en/latest/>). We also calculated the importance of ranking features from the final model. Figure 2 shows the entire analysis workflow of the present study.

### Statistical analysis

Two independent data entry clerks performed double entry and proofreading to ensure accuracy and reliability. We removed predictors with more than 10% missing values and outliers for data processing. Missing values in less than 10% of variables were imputed using the missForest method. In addition, in the case of sparse data, categories were combined if necessary. In statistical description terms, continuous variables were described as the mean (SD) or median (interquartile range [IQR]) as appropriate. Correlations were determined by Pearson or Spearman analysis. The variance inflation factor (VIF) and tolerance were used to identify collinear independent variables. Univariate analysis was performed using the t-test, chi-square test, or Wilcoxon rank sum test. Statistical significance was defined as a two-sided P value < 0.05. All data were analyzed using R statistics software (version 4.0.3; <https://www.r-project.org>) and Python (version 3.10.8; <https://www.python.org>).

## Results

### Baseline characteristics and correlation analysis

Among the entire dataset, 355 (31.2%) women were above the cutoff values ( $\geq 11$ ) on the EPDS and were regarded as having PPD. The entire cohort's median age was 29 (IQR 27–32), and 55.4% were primiparous women, while 59.0%

were from urban areas. Detailed characteristics of the entire cohort are described in Table 1, and the missing data are summarized in Supplementary Table S1. Women with PPD significantly differed from women without PPD in terms of the MCPN and the MCPD, diagnosis of mental illness before pregnancy, diagnosis of mental illness in first-degree relatives, and SCSQ, EPQ-P, EPQ-E, EPQ-N, MT, PSSS, BAI, BDI, EMSS, and PSQI scores.

Supplementary Fig. S1 displays correlations between all continuous variables. The correlation heatmap showed that TC was highly correlated with LDL-C, and the correlation coefficient was 0.90 ( $P < 0.01$ ). In addition, the correlation coefficient between MCPN and MCPD was 0.50. All variables were analyzed by collinearity analysis (Supplementary Table S2). The VIF of TC was greater than 10 (tolerance less than 0.10), and LDL-C was close to 10 (tolerance close to 0.10), indicating the presence of severe multicollinearity between them [33]. However, MCPD and MCPN did not show multicollinearity.

### Selection of predictor variables

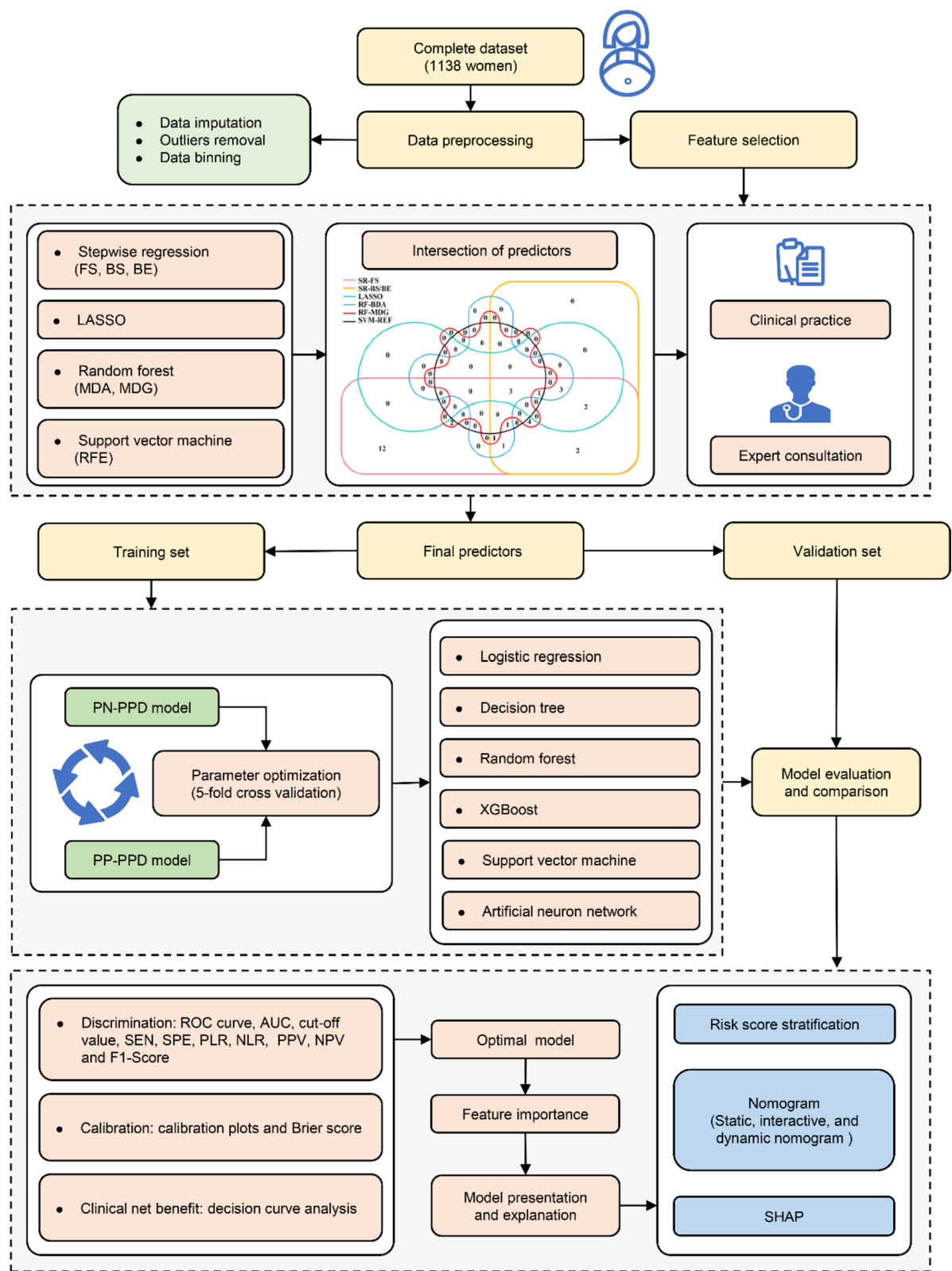
The predictor variables obtained by the seven selection methods are shown in Table 2. The specific parameters of all methods (SR-FS, SR-BS, SR-BE, LASSO, RF-MDA, RF-MDG, and SVM-RFE) are shown in Supplementary Table S3 and Figs. S2–4. Thirty-two, 18, 18, 9, 10, 10, and 5 predictors were identified using seven selection methods. Figure 3 describes the intersection of the predictors selected by the seven methods. The predictors with more than five intersections were chosen as the final predictors, including MCPD, BDI, EPQ-N, BAI, TC, PCPD, EPQ-P, MT, and EMSS. Moreover, primiparous women and FT3 were included in the final predictor set through expert consultation and literature review [26, 29]. Ultimately, 11 predictors were included in developing the model. Notably, MCPD and PCPD were measured after childbirth.

### Model development and comparison

When 11 variables were included, the sample size was again calculated to be 369 women. Therefore, all data were split into training and validation sets at a ratio of 6:4. Supplementary Table S4 provides the descriptive statistics of the two datasets. The training and validation

(See figure on next page.)

**Fig. 2** Analysis workflow for the development and evaluation of models. FS forward selection, BS backward selection, BE bidirectional elimination, LASSO least absolute shrinkage and selection operator, MDA mean decrease accuracy, MDG mean decrease Gini impurity, RFE recursive feature elimination, PN-PPD prenatal prediction model for postpartum depression, PP-PPD postpartum prediction model for postpartum depression, ROC receiver operating characteristic, AUC area under the curve, SEN sensitivity, SPE specificity, PLR positive likelihood ratio, NLR negative likelihood ratio, PPV positive predictive value, NPV negative predictive value, SHAP Shapley additive explanations



**Fig. 2** (See legend on previous page.)

sets were relatively uniformly distributed, in which only the P value of the primiparous women was less than 0.05. The prevalence of PPD was 31.2% in both the training and validation sets. Twelve models were developed by six ML algorithms (PN-PPD and PP-PPD models). The estimates of odd ratios in the LR models are reported in Supplementary Tables S5 and 6 and presented in forest plots (Supplementary Fig. S5 and 6). In addition, Supplementary Figs. S7–14 shows other models' visualization and variable importance.

Table 3 describes the PN-PPD model and the PP-PPD model performance metrics of the validation set. The AUC values of the PN-PPD models ranged from 0.683 to 0.801, and the PP-PPD models ranged from 0.727 to 0.858. According to the DeLong test, the models constructed by the LR and ANN algorithms achieved higher AUC values (0.801 [95% CI 0.758–0.844] in the PN-PPD model and 0.858 [95% CI 0.821–0.895] in the PP-PPD model by LR; 0.858 [95% CI 0.821–0.895] in the PN-PPD model and 0.844 [95% CI 0.805–0.883] in the PP-PPD model by ANN) than the other models, and these AUC values were statistically significant ( $P < 0.05$ ) (Supplementary Tables S7 and 8). In addition, The LR and ANN algorithms demonstrated higher SEN (i.e., recall), PPV (i.e., precision), and F1 scores among other algorithms. For the PN-PPD model, the LR algorithm had a SEN (i.e., recall) of 0.810 (95% CI 0.745–0.874), a PPV (i.e., precision) of 0.532 (95% CI 0.466–0.599), and an F1 score of 0.642, while the ANN algorithm showed a SEN (i.e., recall) of 0.655 (95% CI 0.577–0.733), a PPV (i.e., precision) 0.554 (95% CI 0.478–0.629), and an F1 score of 0.600. For the PP-PPD model, the LR algorithm had a SEN (i.e., recall) of 0.768 (95% CI 0.698–0.837), a PPV (i.e., precision) of 0.669 (95% CI 0.596–0.741), and an F1 score of 0.715, while the ANN algorithm showed a SEN (i.e., recall) of 0.711 (95% CI 0.637–0.786), a PPV (i.e., precision) of 0.660 (95% CI 0.585–0.735), and an F1 score of 0.685.

The discrimination, calibration, and clinical net benefit of the PN-PPD and PP-PPD models on the validation set are shown in Fig. 4. In contrast, the agreement between the observed and predicted events was relatively good with the LR and ANN algorithms and demonstrated a higher net clinical benefit across most ranges of threshold probabilities. On the other hand, compared to the PN-PPD models (except for the decision tree models), the PP-PPD models had higher reclassification and prediction ability (Supplementary Tables S9 and 10). In summary, these results suggest that the LR and ANN algorithms are the optimal ML models for predicting women's PPD.

### Optimal model interpretability

We generated three types of nomograms to provide convenient and personalized risk estimates of PPD. The static and interactive nomograms were assigned in proportion to the effect sizes in the LR model (Supplementary Figs. S15–18). The dynamic nomogram was developed to allow clinicians to enter the values of the variables and then obtain the risk of PPD ([https://yongjianwang.shinyapps.io/PN\\_PPD\\_Nomogram/](https://yongjianwang.shinyapps.io/PN_PPD_Nomogram/) and [https://yongjianwang.shinyapps.io/PP\\_PPD\\_Nomogram/](https://yongjianwang.shinyapps.io/PP_PPD_Nomogram/)) (Supplementary Figs. S19 and 20). Moreover, to further improve the clinical application of the models, we calculated the risk-stratification score separately for prenatal and postnatal screening based on the LR coefficients (Supplementary Tables S11 and 12). Tertiles of the total risk scores (16 to 23 for the PN-PPD model and 6 to 19 for the PP-PPD model) were used to categorize low-risk, intermediate-risk, and high-risk groups for perinatal women.

To explain the complex ANN model, we applied the SHAP value to illustrate how predictors affect women's PPD. The feature importance and interpretation of the ANN model are shown in Fig. 5. The results demonstrated that the EPQ-N, MCPP, BDI, and BAI were significantly more important than other factors, as shown in Fig. 5A and E. In addition, Fig. 5B and F clearly illustrate the strength and direction of every predictor. A higher MCPP score indicated a lower risk of PPD, and greater neuroticism was associated with a greater risk of PPD in the figures. For local interpretability, Fig. 5C, D, G, and H provides four typical relative samples and shows how the ANN models make clinical decisions for individual women. The SHAP value for every predictor as a force contributed to pushing the overall SHAP value (1 = with PPD, 0 = without PPD) higher (red) or pushing it lower (blue), and combined to predict the risk of PPD for individual women. For example, in Fig. 5C, one woman was predicted to suffer PPD due to FT3, EPQ-P, and BAI. The elevated risk was offset by the woman's MT and EPQ-N.

### Discussion

This is the first study to present a novel approach to determine the risk factors for predicting PPD by integrating seven feature selection techniques with real-world perinatal data. Unlike previous models that relied on a limited set of predictors, our study incorporates these methods to systematically identify the most relevant risk factors, ensuring a robust and generalizable model. Additionally, by comparing six ML algorithms, we comprehensively evaluate different modeling strategies, demonstrating that LR and ANN algorithms achieve superior predictive performance. A key innovation is the development of interactive risk assessment tools, including nomograms and web-based risk calculators,



**Table 1** Baseline characteristics of the whole dataset

Characteristic	Total (n = 1138)	PPD (n = 355, 31.2%)	Non-PPD (n = 783, 68.8%)	P-value
Demographic factors				
Age (years), median (IQR)	29 (27–32)	29 (27–32)	29 (27–32)	0.725
Residence, n (%)				0.124
Country	143 (12.6)	50 (14.1)	93 (11.9)	
Town	323 (28.4)	87 (24.5)	236 (30.1)	
City	672 (59.0)	218 (61.4)	454 (58.0)	
Housing condition, n (%)				0.218
Good	48 (4.2)	11 (3.1)	37 (4.7)	
Moderate	498 (43.8)	148 (41.7)	350 (44.7)	
Poor	592 (52.0)	196 (55.2)	396 (50.6)	
Monthly income (yuan), n (%)				0.597
< 6000	377 (33.1)	125 (35.2)	252 (32.2)	
6000–10000	430 (37.8)	129 (36.3)	301 (38.4)	
> 10,000	331 (29.1)	101 (28.5)	230 (29.4)	
Education (years), n (%)				0.213
< 12	142 (12.5)	51 (14.4)	91 (11.6)	
12–16	890 (78.2)	277 (78.0)	613 (78.3)	
> 16	106 (9.3)	27 (7.6)	79 (10.1)	
Primiparous women, n (%)				0.063
Yes	630 (55.4)	211 (59.4)	419 (53.5)	
No	508 (44.6)	144 (40.6)	364 (46.5)	
Planning pregnancy, n (%)				0.287
Yes	631 (55.4)	201 (56.6)	430 (54.9)	
No	507 (44.6)	154 (43.4)	353 (45.1)	
History of mental illness				
Women were diagnosed before pregnancy, n (%)				< 0.001
Yes	16 (1.4)	11 (3.1)	5 (0.6)	
No	1122 (98.6)	344 (96.9)	778 (99.4)	
First-degree relatives were diagnosed with mental illness, n (%)				0.011
Yes	33 (2.9)	17 (4.8)	16 (2.0)	
No	1105 (97.1)	338 (95.2)	767 (98.0)	
Psychosocial factors				
PCPN, n (%)				0.383
Mother-in-law	325 (28.6)	96 (27.0)	229 (29.2)	
Mother	202 (17.7)	73 (20.6)	129 (16.5)	
Husband	531 (46.7)	160 (45.1)	371 (47.4)	
Confinement nurse	80 (7.0)	26 (7.3)	54 (6.9)	
PCPP, n (%)				0.299
Mother-in-law	511 (44.9)	151 (27.0)	360 (46.0)	
Mother	283 (24.9)	100 (20.6)	183 (23.4)	
Husband	140 (12.3)	39 (45.1)	101 (12.9)	
Confinement nurse	204 (17.9)	65 (7.3)	139 (17.7)	
MCPN (score), median (IQR)	8 (6–10)	8 (6–9)	8 (7–10)	< 0.001
MCPN (score), median (IQR)	7 (5–9)	5 (2–7)	8 (6–9)	< 0.001
SCSQ, n (%)				0.006
Positive	1025 (90.1)	307 (86.5)	718 (91.7)	
Negative	113 (9.9)	48 (13.5)	65 (8.3)	
EPQ-P (score), n (%)				< 0.001
< 43.3	129 (11.3)	25 (7.0)	104 (13.3)	

**Table 1** (continued)

Characteristic	Total (n = 1138)	PPD (n = 355, 31.2%)	Non-PPD (n = 783, 68.8%)	P-value
43.3–56.7	730 (64.2)	191 (53.8)	539 (68.8)	< 0.001
> 56.7	279 (24.5)	139 (39.2)	140 (17.9)	
EPQ-E (score), n (%)				
< 43.3	220 (19.3)	98 (27.6)	122 (15.6)	< 0.001
43.3–56.7	494 (43.4)	152 (42.8)	342 (43.7)	
> 56.7	424 (37.3)	105 (29.6)	319 (40.7)	
EPQ-N (score), n (%)				< 0.001
< 43.3	474 (41.7)	80 (22.5)	394 (50.3)	
43.3–56.7	474 (41.7)	152 (42.8)	322 (41.1)	
> 56.7	190 (16.6)	123 (34.7)	67 (8.6)	< 0.001
MT, n (%)				
Yes	66 (5.8)	51 (14.4)	15 (1.9)	
No	1072 (94.2)	304 (85.6)	768 (98.1)	< 0.001
PSSS (score), n (%)				
≤ 60	322 (28.3)	146 (41.1)	176 (22.5)	
> 60	816 (71.7)	209 (58.9)	607 (77.5)	< 0.001
BAI (score), n (%)				
≤ 7	850 (74.7)	204 (57.5)	646 (82.5)	
> 7	288 (25.3)	151 (42.5)	137 (17.5)	< 0.001
BDI (score), n (%)				
≤ 13	1052 (92.4)	280 (78.9)	772 (98.6)	
> 13	86 (7.6)	75 (21.1)	11 (1.4)	< 0.001
EMSS (score), n (%)				
< 30	105 (9.2)	63 (17.8)	42 (5.4)	
30–42	529 (46.5)	178 (50.1)	351 (44.8)	< 0.001
> 42	504 (44.3)	114 (32.1)	390 (49.8)	
PSQI (score), n (%)				
< 5	357 (31.4)	76 (21.4)	281 (35.9)	< 0.001
≥ 5	781 (68.6)	279 (78.6)	502 (64.1)	
Physiological measures				
TSH (μIU/mL), median (IQR)	1.94 (1.37–2.65)	1.93 (1.25–2.62)	1.95 (1.40–2.65)	0.125
FT3 (pmol/L), median (IQR)	4.06 (3.79–4.38)	4.10 (3.83–4.42)	4.04 (3.75–4.37)	0.050
FT4 (pmol/L), median (IQR)	12.93 (11.80–14.22)	13.04 (11.84–14.3)	12.86 (11.79–14.20)	0.300
TG (mmol/L), median (IQR)	2.90 (2.35–3.64)	2.93 (2.35–3.64)	2.90 (2.35–3.64)	0.810
TC (mmol/L), median (IQR)	6.15 (5.46–6.90)	6.14 (5.47–6.88)	6.15 (5.44–6.92)	0.811
HDL-C (mmol/L), median (IQR)	1.96 (1.71–2.22)	1.99 (1.71–2.19)	1.95 (1.71–2.22)	0.727
LDL-C (mmol/L), median (IQR)	3.58 (2.95–4.24)	3.57 (2.92–4.23)	3.60 (2.96–4.25)	0.434
Ca (mmol/L), median (IQR)	2.15 (2.10–2.21)	2.15 (2.10–2.20)	2.15 (2.10–2.21)	0.530
Fe (μmol/L), median (IQR)	13.20 (9.3–18.8)	13.70 (9.22–19.28)	112.9 (9.3–18.48)	0.485

PPD postpartum depression, IQR interquartile range, PCPN primary caregiver in prenatal, PCPP primary caregiver in postpartum, MCPN mother-in-law's care in prenatal, MCPP mother-in-law's care in postpartum, SCSQ simplified coping style questionnaire, EPQ-P psychoticism dimension of the Eysenck personality questionnaire, EPQ-E extraversion dimension of the Eysenck personality questionnaire, EPQ-N neuroticism dimension of the Eysenck personality questionnaire, MT melancholic temperament, PSSS perceived social support scale, BAI Beck anxiety inventory, BDI Beck depression inventory, EMSS Enrich marital satisfaction scale, PSQI Pittsburgh sleep quality inventory, TSH Thyroid-stimulating hormone, FT3 serum free triiodothyronine, FT4 serum free thyroxine, TG triglyceride, TC total cholesterol, HDL-C high density lipoprotein-cholesterol, LDL-C low-density lipoprotein-cholesterol, Ca serum calcium, Fe serum iron

which facilitate immediate clinical application. Notably, by employing SHAP values to enhance interpretability, our study bridges the gap between complex ML models and practical clinical decision-making, allowing

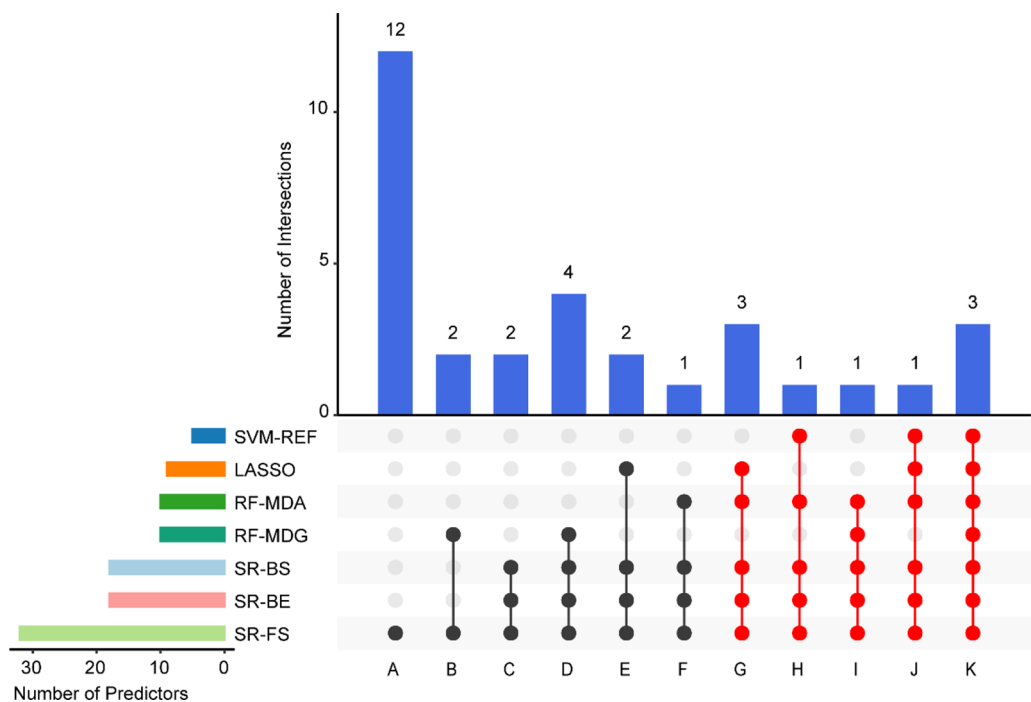
healthcare providers to understand and trust the model's predictions.

We selected 11 features associated with PPD, including MCPP, prenatal depression, neuroticism, prenatal

**Table 2** Predictor variables of PPD using different selection methods

Methods	Predictor variables
SR-FS	age (years), residence, housing condition, monthly income (yuan), education (years), primiparous women, Planning pregnancy, PCPN, PCPP, MCPN, MCPP, women were diagnosed with mental illness before pregnancy, first-degree relatives were diagnosed with mental illness, SCSQ, EPQ-E, EPQ-N, EPQ-P, MT, PSSS, BAI, BDI, EMSS, PSQI, TSH, FT3, FT4, TG, TC, HDL-C, LDL-C, Ca, Fe
SR-BS	residence, primiparous women, Planning pregnancy, PCPP, MCPN, MCPP, ENP-P, ENP-N, MT, PSSS, BAI, BDI, EMSS, FT3, TG, TC, LDL-C, Fe
SR-BE	residence, primiparous women, Planning pregnancy, PCPP, MCPN, MCPP, ENP-P, ENP-N, MT, PSSS, BAI, BDI, EMSS, FT3, TG, TC, LDL-C, Fe
LASSO	primiparous women, MCPP, ENP-P, ENP-N, MT, PSSS, BAI, BDI, EMSS
RF-BDA	MCPP, BDI, EPQ-N, BAI, EMSS, MT, EPQ-P, TC, MCPN, PCPP
RF-MDG	MCPP, BDI, EPQ-N, TSH, Fe, FT3, FT4, TG, TC, LDL-C
SVM-REF	MCPP, BDI, BAI, EPQ-N, PCPP

*SR-FS* stepwise regression-forward selection, *SR-BS* stepwise regression-backward selection, *SR-BE* stepwise regression-bidirectional elimination, *LASSO* least absolute shrinkage and selection operator, *RF-MDA* random forest-mean decrease accuracy, *RF-MDG* random forest-mean decrease Gini, *SVM-RFE* support vector machine-recursive feature elimination, *PCPN* primary caregiver in prenatal, *PCPP* primary caregiver in postpartum, *MCPN* mother-in-law's care in prenatal, *MCPP* mother-in-law's care in postpartum, *SCSQ* simplified coping style questionnaire, *EPQ-P* psychoticism dimension of the Eysenck personality questionnaire, *EPQ-E* extraversion dimension of the Eysenck personality questionnaire, *EPQ-N* neuroticism dimension of the Eysenck personality questionnaire, *MT* melancholic temperament, *PSSS* perceived social support scale, *BAI* Beck anxiety inventory, *BDI* Beck depression inventory, *EMSS* Enrich marital satisfaction scale, *PSQI* Pittsburgh sleep quality inventory, *TSH* Thyroid-stimulating hormone, *FT3* serum free triiodothyronine, *FT4* serum free thyroxine, *TG* triglyceride, *TC* total cholesterol, *HDL-C* high density lipoprotein-cholesterol, *LDL-C* low-density lipoprotein-cholesterol, *Ca* serum calcium, *Fe* serum iron



**Fig. 3** Upset plot of intersections between the predictors. **A** Age (years), housing condition, monthly income (yuan), education (years), primary caregiver in prenatal, women were diagnosed with mental illness before pregnancy, first-degree relatives were diagnosed with mental illness, simplified coping style questionnaire (score), extraversion dimension of the Eysenck personality inventory (score), Pittsburgh sleep quality inventory (score), high-density lipoprotein-cholesterol, and serum calcium. **B** Thyroid-stimulating hormone and serum-free thyroxine. **C** Residence and planned pregnancy. **D** Serum-free triiodothyronine, triglyceride, low-density lipoprotein-cholesterol, and serum iron. **E** Primiparous women and perceived social support scale (score). **F** Mother-in-law's care in prenatal (score). **G** Psychoticism dimension of the Eysenck personality inventory (score), melancholic temperament, and Enrich marital satisfaction scale (score). **H** Primary caregiver in postpartum. **I** Total cholesterol. **J** Beck anxiety inventory (score). **K** Mother-in-law's care in postpartum (score), Beck depression inventory (score), and neuroticism dimension of the Eysenck personality inventory (score). *SR-FS* stepwise regression-forward selection, *SR-BS* stepwise regression-backward selection, *SR-BE* stepwise regression-bidirectional elimination, *LASSO* least absolute shrinkage and selection operator, *RF-MDA* random forest-mean decrease accuracy, *RF-MDG* random forest-mean decrease Gini, *SVM-RFE* support vector machine-recursive feature elimination

**Table 3** Predictive performance metrics of different machine learning algorithms of the validation set

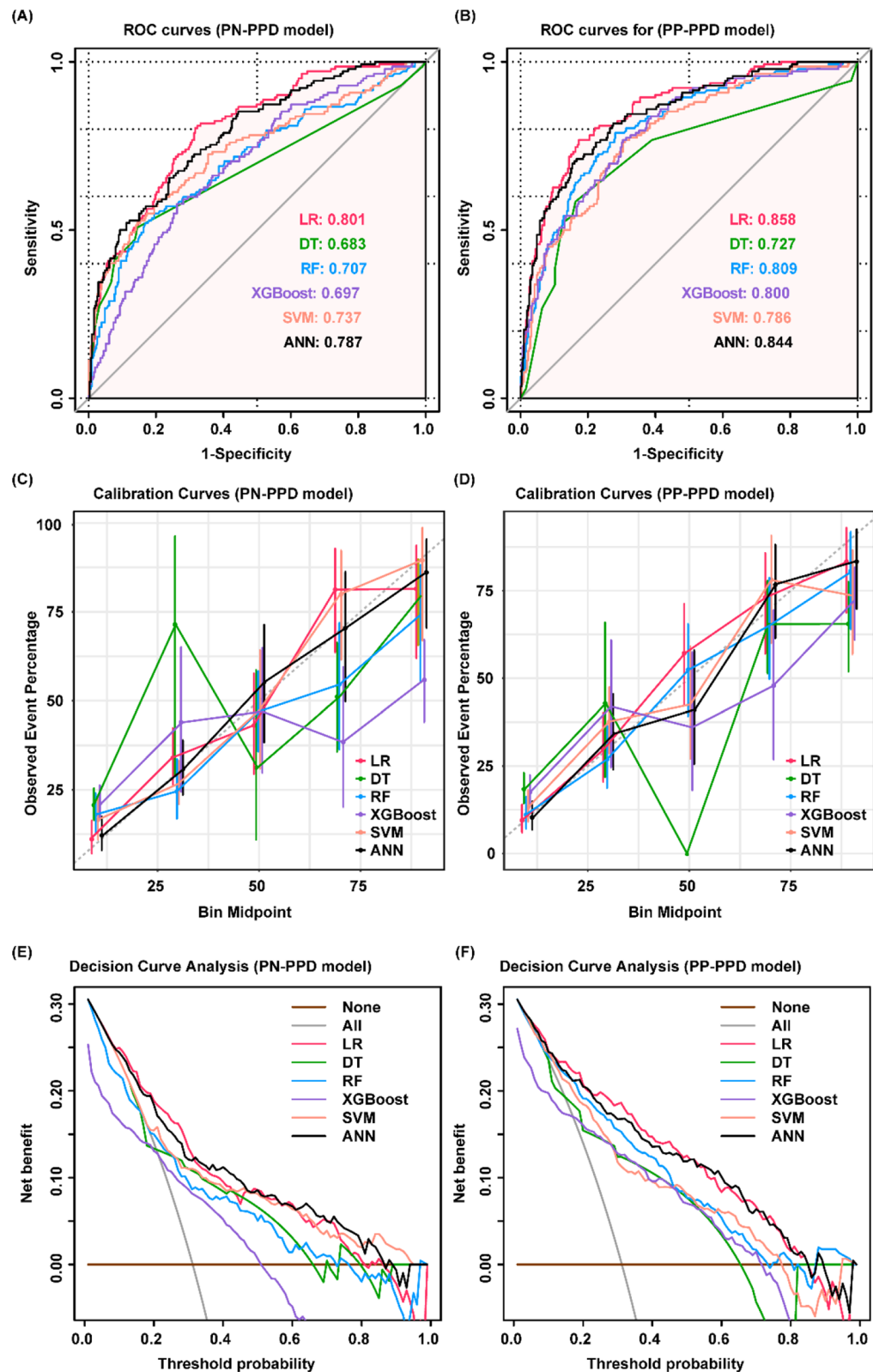
Models	Validation set					
	LR	DT	RF	SVM	XGBoost	ANN
PN-PPD						
AUC (95%CI)	0.801 (0.758–0.844)	0.683 (0.630–0.735)	0.707 (0.653–0.761)	0.737 (0.684–0.790)	0.697 (0.645–0.748)	0.787 (0.743–0.832)
Cutoff value	0.277	0.243	0.411	0.232	0.194	0.280
SEN (95%CI)	0.810 (0.745–0.874)	0.507 (0.425–0.589)	0.549 (0.467–0.631)	0.592 (0.511–0.672)	0.577 (0.496–0.659)	0.655 (0.577–0.733)
SPE (95%CI)	0.677 (0.626–0.729)	0.856 (0.817–0.895)	0.805 (0.761–0.849)	0.792 (0.747–0.837)	0.738 (0.689–0.787)	0.760 (0.713–0.808)
PLR (95%CI)	2.510 (2.098–3.002)	3.527 (2.573–4.834)	2.819 (2.152–3.692)	2.849 (2.205–3.679)	2.204 (1.746–2.783)	2.733 (2.170–3.442)
NLR (95%CI)	0.281 (0.198–0.398)	0.576 (0.484–0.684)	0.560 (0.463–0.677)	0.516 (0.420–0.633)	0.573 (0.467–0.702)	0.454 (0.359–0.574)
PPV (95%CI)	0.532 (0.466–0.599)	0.615 (0.527–0.704)	0.561 (0.479–0.644)	0.564 (0.484–0.643)	0.500 (0.423–0.577)	0.554 (0.478–0.629)
NPV (95%CI)	0.887 (0.847–0.927)	0.793(0.750–0.836)	0.797 (0.753–0.842)	0.810 (0.767–0.854)	0.794 (0.747–0.840)	0.829 (0.786–0.873)
F1 score	0.642	0.556	0.555	0.578	0.536	0.600
Brier score	0.162	0.184	0.191	0.173	0.250	0.141
PP-PPD						
AUC (95%CI)	0.858 (0.821–0.895)	0.727 (0.675–0.780)	0.809 (0.766–0.852)	0.786 (0.741–0.831)	0.800 (0.757–0.844)	0.844 (0.805–0.883)
Cutoff value	0.312	0.242	0.266	0.218	0.047	0.321
SEN (95%CI)	0.768 (0.698–0.837)	0.585 (0.503–0.666)	0.789 (0.722–0.856)	0.768 (0.698–0.837)	0.768 (0.698–0.837)	0.711 (0.637–0.786)
SPE (95%CI)	0.827 (0.786–0.869)	0.837 (0.796–0.878)	0.719 (0.669–0.769)	0.690 (0.639–0.741)	0.696 (0.646–0.747)	0.834 (0.793–0.875)
PLR (95%CI)	4.449 (3.434–5.764)	3.587 (2.693–4.779)	2.805 (2.305–3.415)	2.477 (2.051–2.991)	2.529 (2.090–3.060)	4.281 (3.270–5.605)
NLR (95%CI)	0.281 (0.207–0.380)	0.496 (0.406–0.607)	0.294 (0.212–0.407)	0.337 (0.247–0.458)	0.334 (0.245–0.454)	0.346 (0.266–0.450)
PPV (95%CI)	0.669 (0.596–0.741)	0.619 (0.537–0.702)	0.560 (0.491–0.629)	0.529 (0.461–0.597)	0.534 (0.466–0.603)	0.660 (0.585–0.735)
NPV (95%CI)	0.887 (0.851–0.923)	0.816 (0.774–0.859)	0.882 (0.843–0.922)	0.867 (0.825–0.910)	0.869 (0.827–0.910)	0.864 (0.826–0.903)
F1 score	0.715	0.602	0.655	0.626	0.630	0.685
Brier score	0.137	0.180	0.158	0.168	0.194	0.141

LR logistic regression, DT decision tree, RF random forest, XGBoost extreme gradient boosting, SVM support vector machine, ANN artificial neuron network, AUC area under receiver-operating curve, SEN sensitivity, SPE specificity, PLR positive likelihood ratio, NLR negative likelihood ratio, PPV positive predictive value, NPV negative predictive value

anxiety, TC, PCPP, MT, marital satisfaction, primiparity, and FT3 (Fig. 3). Among them, MCPP, prenatal depression, prenatal anxiety, and neuroticism were determined to be the most significant predictors according to the SHAP value (Fig. 5). Our results indicated that low MCPP was significantly associated with an increased risk of PPD, which might be attributed to poor relationships between mothers-in-law and postpartum women [36]. In China, to help women recuperate after childbirth, mothers-in-law usually play an essential role in postnatal care for both the mother and baby during the postpartum period. However, conflicts between mothers-in-law and postpartum women are common due to different parenting views and lifestyles [4]. Thus, inadequate caregiving as a significant stressor increased PPD risk for Chinese women. This is concordant with our prior studies [37]. Notably, in our study, prenatal anxiety and prenatal depression were common in pregnancy and were significantly associated with PPD. This is supported by a large number of previous studies showing that prenatal depression and anxiety are strong predictors of PPD [1, 48]. One possible explanation is that pregnant women

with depression or anxiety have more prolonged depressive or anxious symptoms, even into the postpartum period [23]. Another possible reason is that women with a history of mental disorders have a higher recurrence rate after delivery [15]. Interestingly, we found that neuroticism was a significant predictor associated with PPD. Originally defined by Eysenck, neuroticism is a key personality trait for affective processing [16]. Specifically, when faced with stress in the postpartum period, neurotic women tend to experience greater nervousness and are more likely to experience worrying and depression [35].

Discrimination, calibration, and clinical net benefit were best in the LR and ANN models (for both PN-PPD and PP-PPD models). Collectively, the newly developed LR and ANN models, which incorporated readily available prenatal and postnatal variables, performed well, as supported by the AUC values of 0.787–0.858 in the validation set. In addition to AUC, we utilize other evaluation metrics, including precision, recall, and F1 score, to evaluate the performance of LR and ANN models. With its relatively simpler structure, the LR model displayed



**Fig. 4** ROC curves, calibration plots and decision curve analysis of PN-PPD and PP-PPD models on the validation set. ROC receiver operating characteristic curve, LR logistic regression, DT decision tree, RF random forest, XGBoost extreme gradient boosting, SVM support vector machine, ANN artificial neuron network, AUC area under the receiver-operating curve



excellent interpretability and high performance across these key metrics. While the ANN model demonstrated a similar AUC and F1 score, it has the advantage of powerful self-learning capabilities, which make it particularly well-suited for capturing complex nonlinear relationships in the data. The higher F1 scores of LR and ANN models indicate that both models strike a reasonable balance between precision and recall, thus ensuring that false positives and negatives are minimized, which is critical in clinical decision-making. For the term of calibration, the agreements between prediction and actuality are shown in the calibration plots. More importantly, the decision curve analysis showed that the LR and ANN models could provide good clinical net benefits to support clinical decision-making. Regarding model applicability, the LR model has the characteristics of a simple structure and strong interpretability. Compared with the LR model, the ANN model has powerful self-learning capabilities and an outstanding advantage in dealing with nonlinear relationships. However, the difference in prediction performance between the two models was insignificant in the study. Furthermore, adding postnatal predictors to the prenatal models can improve prediction performance. Therefore, we recommend combining prenatal and postnatal predictors to enhance the predictive accuracy of PPD. In future research, better prediction results may be achieved with more prenatal and postpartum clinical information.

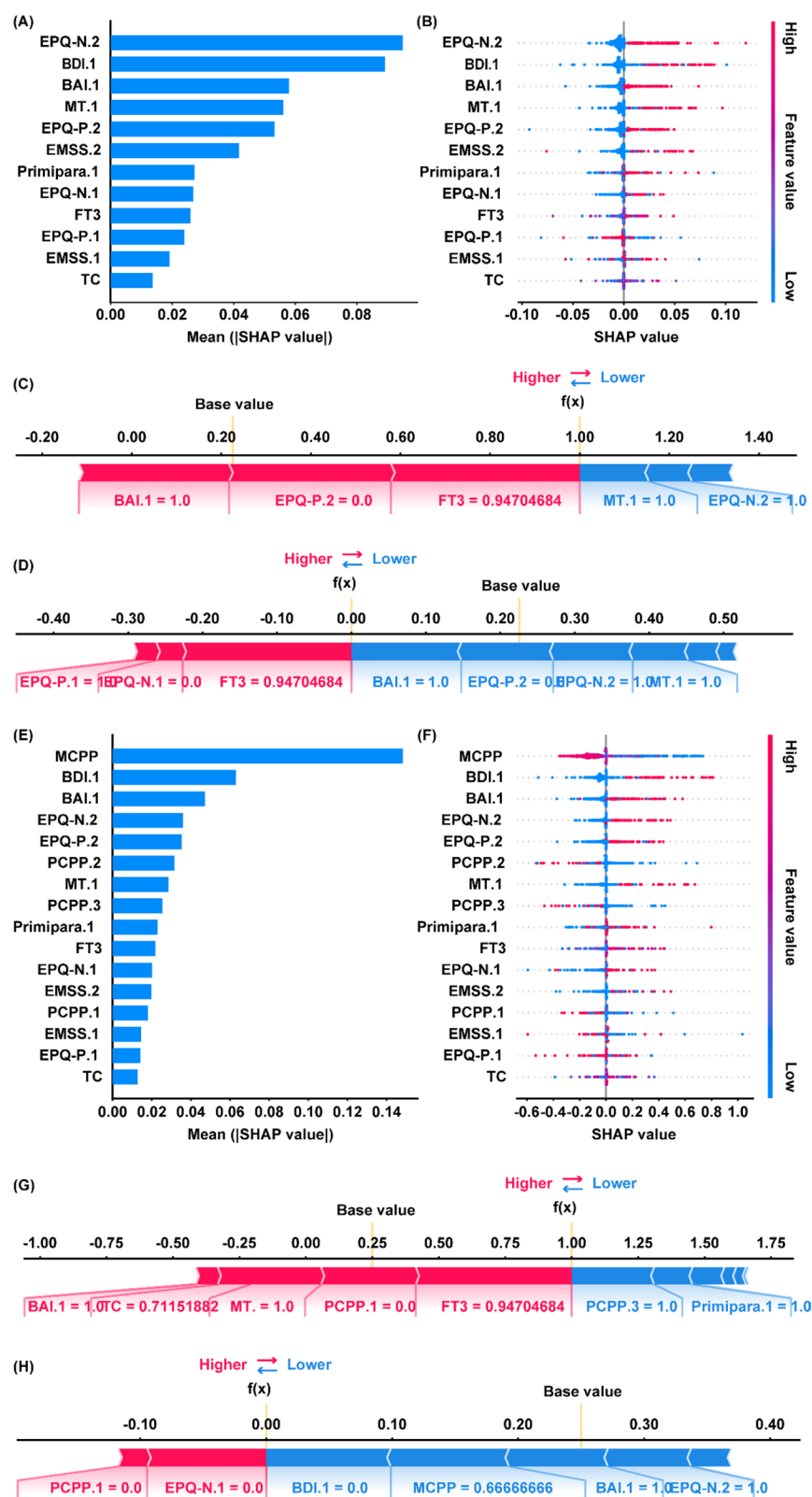
Despite the growing interest in ML models for clinical decision-making, most published prediction models never reach clinical practice. This may be because the models are difficult to understand or because the results lack representativeness and reproducibility [18]. In this study, we primarily included clinically applicable and easy to identify predictors. For clinical use of the optimal model, we designed various risk assessment tools to enable physicians to identify high-risk women immediately. According to the LR algorithm, the nomogram could serve as a tool to evaluate the risk ratio of PPD. On this basis, we developed a website calculator, which may be readily integrated into secondary care to improve

screening efficiency and reduce the burden on physicians. In addition, applying risk scores opens new opportunities to enhance risk stratification and to help prevent PPD. Of note, due to its black-box nature, it is difficult for the ANN model to provide meaningful physician interpretations. Interpretability is generally defined as the ease with which humans can comprehend and explain the process of the ML model's predictions [41]. To address this problem, we applied SHAP values to obtain more readily understandable interpretations. SHAP values are widely accepted and are useful for explaining the relationship between variables and outcomes [21]. It can help physicians better understand the model's decision-making process for appropriate early intervention for women with PPD. Overall, adopting risk assessment tools developed in this study could provide rapid results to physicians.

From a clinical perspective, this study has significant implications for improving PPD screening and intervention strategies. Traditional screening methods for PPD are often time-consuming and require trained personnel, limiting their scalability in resource-constrained settings. The ML-based models developed in this study provide a rapid and automated risk assessment framework, allowing early identification of high-risk individuals. Moreover, by integrating both prenatal and postpartum predictors, these models offer a dynamic approach to monitoring maternal mental health from pregnancy to postpartum recovery. The practical implementation of our risk assessment tools in clinical settings can facilitate targeted interventions, optimize resource allocation, and improve maternal and neonatal outcomes. We suggest that these tools could be embedded in hospital electronic health record systems and routine perinatal care in the future. Clinicians can utilize these assessment tools in real-time during prenatal and postpartum visits for prenatal PPD risk stratification and postpartum PPD monitoring. During routine prenatal screenings, clinicians can input readily available variables into the web-based calculator to generate individualized PPD risk scores. This enables

(See figure on next page.)

**Fig. 5** The feature importance and interpretation of the PP-PPD model (based on the ANN algorithm). **A** and **E** The importance ranking of features based on the mean (|SHAP value|). **B** and **F** A summary plot of the SHAP values for each feature. **C** and **G** SHAP force plot for a woman (PPD). **D** and **H** SHAP force plot for a woman (without PPD). The higher the SHAP value of the feature, the impact of the feature on the model is larger. The red dots in the feature value represent higher values for that individual patient, whereas the blue dots indicate lower feature values. *MCPP* mother-in-law's care in postpartum, *BDI.1* Beck depression inventory (> 13 scores), *BAI.1* Beck anxiety inventory (> 7 scores), *EPQ-N.2* neuroticism dimension of the Eysenck personality questionnaire (> 56.7 scores), *EPQ-P.2* psychoticism dimension of the Eysenck personality questionnaire (> 56.7 scores), *PCPP.2* primary caregiver in postpartum (husband), *MT.1* melancholic temperament (yes), *PCPP.3* primary caregiver in postpartum (confinement nurse); Primopara.1, (yes); *FT3* serum-free triiodothyronine, *EPQ-N.1* neuroticism dimension of the Eysenck personality questionnaire (43.3–56.7 scores), *EMSS.2* Enrich marital satisfaction scale (< 30 scores), *PPPC.1* primary caregiver in postpartum (mother), *EMSS.1* Enrich marital satisfaction scale (30–42 scores), *EPQ-P.1* psychoticism dimension of the Eysenck personality questionnaire (43.3–56.7 scores), *TC* total cholesterol



**Fig. 5** (See legend on previous page.)

early identification of high-risk women, prompting targeted mental health interventions such as counseling or prophylactic therapy. Moreover, pregnant women outside the hospital can use the web calculator to self-assess their PPD risk. The dynamic integration of postnatal predictors allows for ongoing risk reassessment during postpartum checkups, ensuring timely adjustments to care plans. However, several potential barriers to their implementation must be acknowledged. First, embedding these tools into electronic health records requires interoperability with existing hospital systems, which may involve technical and administrative hurdles. Second, healthcare providers may need training to effectively interpret ML-derived risk scores and SHAP-based explanations. Notably, the ethical implications of false positives and negatives in PPD prediction warrant careful deliberation. Overestimation of risk could lead to unnecessary psychological interventions, increasing perinatal women's anxiety and healthcare costs, and underestimation of risk may delay critical interventions. While SHAP values enhance interpretability, clinicians must remain vigilant against over-reliance on model outputs. Therefore, we recommend combining model predictions with clinical judgment and longitudinal symptom monitoring to ensure perinatal women's safety and well-being. Additionally, clinicians should be aware of the potential psychological impact on patients when these assessment tools are used for risk prediction, and appropriate counseling should be provided when necessary.

### Limitations

Several study limitations need to be discussed. First, while the prospective design and robust internal validation via cross-validation strengthen methodological rigor, the absence of external validation on independent cohorts remains a critical constraint. The single-center nature of our dataset introduces potential selection bias and may limit the model's generalizability to populations with distinct demographic profiles, clinical practices, or regional healthcare systems. To address this, future multicenter collaborations will be prioritized to validate and refine the model across diverse settings, ensuring broader clinical applicability. Second, this study primarily focused on second- and third-trimester pregnancies, and first-trimester information was not collected. In future studies, to further improve the accuracy, the number of variables or specific markers could be increased to estimate PPD. Finally, the women in the entire dataset could only be considered representative of the Chinese population with caution, and the risk assessment tools in postpartum were

more likely suitable for women who were cared for by their mothers-in-law. MCPP and PCPP have unique Asian cultural features as predictors of postpartum and need to be considered in future research and clinical application.

### Conclusion

In this study, by combining biopsychosocial risk factors for PPD, we developed and validated ML predictive models to identify the risk of PPD. The LR and ANN models showed excellent and reliable prediction performance. Combining prenatal and postnatal predictors to establish predictive models can significantly improve prediction performance. Various risk assessment tools are easily implemented in practice and could help physicians make clinical decisions easily. Therefore, we believe that our study could be utilized to individually predict the risk of PPD in prenatal and postnatal women, thus assisting clinicians in early precise intervention and effectively reducing or delaying the development of PPD. However, external validation is required to demonstrate the accuracy of the model's predictions in future studies.

### Statement of significance

#### Problem

We need a robust risk-based approach to plan the management of women with postpartum depression, enabling shared decision-making and more personalized care.

#### What is already known

Preventive medical care may reduce medical costs and population burden of postpartum depression compared with treatment post-diagnosis. However, identifying and screening women with high postpartum depression risk and making a standardized preventive and treatment strategy remains a major challenge.

#### What this paper adds

To promote translation into clinical care, this study has been translated into several simple and intuitive risk assessment tools. These tools allow clinicians and nurses to calculate individualized risks of postpartum depression to facilitate shared decision-making on antenatal care and risk-stratified approaches to prevention and intervention.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06289-6>.

Supplementary Material 1.

## Acknowledgements

We are deeply grateful to all the women who participated in this study. To all the people who made efforts for this study, we would like to thank them equally for their collaboration.

## Author contributions

Weijing Qi and Yongjian Wang contributed equally to the article. The two authors designed the study and drafted the manuscript together. Yipeng Wang contributed to the study conception and design and commented on previous versions of the manuscript. Sha Huang, Cong Li, Haoyu Jin, Jinfan Zuo, Xuefei Cui, and Ziqi Wei helped with the recruitment of the participants and collected the data. Qing Guo and Jie Hu conceived and designed the project and are responsible for the overall content. All authors have contributed significantly and agree with the content of the manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (grant number 72074067), Social Science Foundation of Hebei Province (HB22SH006), Natural Science Foundation of Hebei Province (H2022206308), and Humanities and Social Sciences Innovative Talents Support Program of Hebei Medical University (ydkrcjhcz202202). The funding agency had no role in the design of the study, the collection, analysis, and interpretation of the data, or in writing the manuscript.

## Availability of data and materials

Data and code could be obtained by contacting the corresponding author.

## Declarations

### Ethics approval and consent to participate

Ethical approval for the study was obtained from the medical ethics committee of Hebei Medical University (MREC ID No: 2020093) and Shijiazhuang Obstetrics and Gynecology Hospital (MREC ID No: 20200024) following the principles of the Declaration of Helsinki.

### Consent for publication

Not applicable.

### Competing interests

The authors have no conflicts of interest to disclose.

### Author details

<sup>1</sup>Humanistic Care and Health Management Innovation Center, School of Nursing, Hebei Medical University, 361 East Zhongshan Road, Shijiazhuang 050017, Hebei, China. <sup>2</sup>Department of Hepatobiliary Oncology, Tianjin Medical University Cancer Hospital and Institute, National Clinical Research Centre for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Centre for Cancer, Tianjin, China. <sup>3</sup>Shijiazhuang Obstetrics and Gynecology Hospital, 206 East Zhongshan Road, Shijiazhuang 050017, Hebei, China.

Received: 18 October 2024 Accepted: 23 February 2025

Published online: 07 March 2025

## References

- Abdollahi F, Sazlina S-G, Zain AM, Zarghami M, Jafarabadi MA, Lye M-S. Postpartum depression and psycho-socio-demographic predictors. *Asia Pac Psychiatry*. 2014;6(4):425–34. <https://doi.org/10.1111/appy.12152>.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th edn. Washington: American Psychiatric Association; 2013. 21(21), 591–643. <https://doi.org/10.1176/appi.books.9780890425596>.
- Amit G, Girshovitz I, Marcus K, Zhang Y, Pathak J, Bar V, Akiva P. Estimation of postpartum depression risk from electronic health records using machine learning. *BMC Pregnancy Childbirth*. 2021;21(1):630. <https://doi.org/10.1186/s12884-021-04087-8>.
- Ayers JD, Krems JA, Hess N, Aktipis A. Mother-in-law daughter-in-law conflict: An evolutionary perspective and report of empirical data from the USA. *Evol Psychol Sci*. 2022;8(1):56–71. <https://doi.org/10.1007/s40806-021-00312-x>.
- Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: Psychometric properties. *J Consult Clin Psychol*. 1988;56(6):893–7. <https://doi.org/10.1037/0022-006x.56.6.893>.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4(6):561–71. <https://doi.org/10.1001/archpsyc.1961.01710120031004>.
- Blumenthal JA, Burg MM, Barefoot J, Williams RB, Haney T, Zimet G. Social support, type A behavior, and coronary artery disease. *Psychosom Med*. 1987;49(4):331–40. <https://doi.org/10.1097/00006842-198707000-00002>.
- Brummelte S, Galea LAM. Postpartum depression: etiology, treatment and consequences for maternal care. *Horm Behav*. 2016;77:153–66. <https://doi.org/10.1016/j.yhbeh.2015.08.008>.
- Buyse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28(2):193–213. [https://doi.org/10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4).
- Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233–4. <https://doi.org/10.1038/nmeth.4642>.
- Cellini P, Pignoni A, Delvecchio G, Moltrasio C, Brambilla P. Machine learning in the prediction of postpartum depression: a review. *J Affect Disord*. 2022;309:350–7. <https://doi.org/10.1016/j.jad.2022.04.093>.
- Chmielewska B, Barratt I, Townsend R, Kalafat E, van der Meulen J, Gurol-Urganci I, et al. Effects of the COVID-19 pandemic on maternal and perinatal outcomes: a systematic review and meta-analysis. *Lancet Glob Health*. 2021;9(6):e759–72. [https://doi.org/10.1016/S2214-109X\(21\)00079-6](https://doi.org/10.1016/S2214-109X(21)00079-6).
- Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350: g7594. <https://doi.org/10.1136/bmj.g7594>.
- Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry*. 1987;150(6):782–6. <https://doi.org/10.1192/bjp.150.6.782>.
- Di Florio A, Gordon-Smith K, Forty L, Kosorok MR, Fraser C, Perry A, et al. Stratification of the risk of bipolar disorder recurrences in pregnancy and postpartum. *Br J Psychiatry*. 2018;213(3):542–7. <https://doi.org/10.1192/bjp.2018.92>.
- Eysenck HJ. Neuroticism, anxiety, and depression. *Psychol Inq*. 1991;2(1):75–6. [https://doi.org/10.1207/s15327965pli0201\\_17](https://doi.org/10.1207/s15327965pli0201_17).
- Fowers BJ, Olson DH. Enrich marital inventory: a discriminant validity and cross-validation assessment. *J Marital Fam Ther*. 1989;15(1):65–79. <https://doi.org/10.1111/j.1752-0606.1989.tb00777.x>.
- Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. *JAMA Psychiat*. 2018;75(12):1289–97. <https://doi.org/10.1001/jamapsychiatry.2018.2530>.
- Gjerdingen DK, Yawn BP. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. *J Am Board Family Med*. 2007;20(3):280–8. <https://doi.org/10.3122/jabfm.2007.03.060171>.
- Gong Y. Eysenck personality questionnaire revised in China. *Inf Psychol Sci*. 1984;4:11–8. <https://doi.org/10.16719/j.cnki.1671-6981.1984.04.004>.
- Goodwin NL, Nilsson SRO, Choong JJ, Golden SA. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Curr Opin Neurobiol*. 2022;73: 102544. <https://doi.org/10.1016/j.conb.2022.102544>.
- Hochman E, Feldman B, Weizman A, Krivoy A, Gur S, Barzilay E, et al. Development and validation of a machine learning-based postpartum depression prediction model: a nationwide cohort study. *Depress Anxiety*. 2021;38(4):400–11. <https://doi.org/10.1002/da.23123>.
- Hofmeijer-Sevink MK, Batelaan NM, van Megan HJGM, Penninx BW, Cath DC, van den Hout MA, van Balkom AJLM. Clinical relevance of comorbidity in anxiety disorders: a report from the Netherlands Study of Depression and Anxiety (NESDA). *J Affect Disord*. 2012;137(1–3):106–12. <https://doi.org/10.1016/j.jad.2011.12.008>.
- Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. *World Psychiatry*. 2020;19(3):313–27. <https://doi.org/10.1002/wps.20769>.

25. Kuruvilla S, Bustreo F, Kuo T, Mishra CK, Taylor K, Fogstad H, et al. The Global strategy for women's, children's and adolescents' health (2016–2030): a roadmap based on evidence and country experience. *Bull World Health Organ.* 2016;94(5):398–400. <https://doi.org/10.2471/BLT.16.170431>.
26. Lambrinoudaki I, Rizos D, Armeni E, Pliatsika P, Leonardou A, Sygelou A, et al. Thyroid function and postpartum mood disturbances in Greek women. *J Affect Disord.* 2010;121(3):278–82. <https://doi.org/10.1016/j.jad.2009.07.001>.
27. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, DEPRESSion Screening Data (DEPRESSD) EPDS Group. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *BMJ.* 2020;371:m4022. <https://doi.org/10.1136/bmj.m4022>.
28. Li Q, Xue W, Gong W, Quan X, Li Q, Xiao L, et al. Experiences and perceptions of perinatal depression among new immigrant Chinese parents: a qualitative study. *BMC Health Serv Res.* 2021;21(1):739–739. <https://doi.org/10.1186/s12913-021-06752-2>.
29. Liu S, Yan Y, Gao X, Xiang S, Sha T, Zeng G, He Q. Risk factors for postpartum depression among Chinese women: path model analysis. *BMC Pregnancy Childbirth.* 2017;17(1):133–133. <https://doi.org/10.1186/s12884-017-1320-x>.
30. Mahesh B. Machine learning algorithms-a review. *Int J Sci Res.* 2020;9:381–6. <https://doi.org/10.21275/ART20203995>.
31. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* 2019;20(5):e262–73. [https://doi.org/10.1016/s1470-2045\(19\)30149-4](https://doi.org/10.1016/s1470-2045(19)30149-4).
32. NICE. Antenatal and postnatal mental health: Clinical management and service guidance; 2014. <https://www.nice.org.uk/guidance/cg192>
33. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 2007;41(5):673–90. <https://doi.org/10.1007/s11135-006-9018-6>.
34. Pollack LM, Chen J, Cox S, Luo F, Robbins CL, Tevendale HD, et al. Health-care utilization and costs associated with perinatal depression among medicaid enrollees. *Am J Prev Med.* 2022;62(6):e333–41. <https://doi.org/10.1016/j.amepre.2021.12.008>.
35. Puyan   M, Subir   S, Torres A, Roca A, Garc  a-Esteve L, Gelabert E. Personality traits as a risk factor for postpartum depression: a systematic review and meta-analysis. *J Affect Disord.* 2022;298:577–89. <https://doi.org/10.1016/j.jad.2021.11.010>.
36. Qi W, Liu Y, Lv H, Ge J, Meng Y, Zhao N, et al. Effects of family relationship and social support on the mental health of Chinese postpartum women. *BMC Pregnancy Childbirth.* 2022;22(1):65. <https://doi.org/10.1186/s12884-022-04392-w>.
37. Qi W, Zhao F, Liu Y, Li Q, Hu J. Psychosocial risk factors for postpartum depression in Chinese women: a meta-analysis. *BMC Pregnancy Childbirth.* 2021;21(1):174. <https://doi.org/10.1186/s12884-021-03657-0>.
38. Qiao J, Wang Y, Li X, Jiang F, Zhang Y, Ma J, et al. A Lancet commission on 70 years of women's reproductive, maternal, newborn, child, and adolescent health in China. *Lancet.* 2021;397(10293):2497–536. [https://doi.org/10.1016/s0140-6736\(20\)32708-2](https://doi.org/10.1016/s0140-6736(20)32708-2).
39. Siu AL, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, et al. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA.* 2016;315(4):380–7. <https://doi.org/10.1001/jama.2015.18392>.
40. Slomian J, Honvo G, Emonts P, Reginster J-Y, Bruy  re O. Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes. *Womens Health (Lond Engl).* 2019;15:1745506519844044. <https://doi.org/10.1177/1745506519844044>.
41. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2020;10(5): e1379. <https://doi.org/10.1002/widm.1379>.
42. Wang D, Li Y-L, Qiu D, Xiao S-Y. Factors influencing paternal postpartum depression: a systematic review and meta-analysis. *J Affect Disord.* 2021;293:51–63. <https://doi.org/10.1016/j.jad.2021.05.088>.
43. Wang Z, Liu J, Shuai H, Cai Z, Fu X, Liu Y, et al. Mapping global prevalence of depression among postpartum women. *Transl Psychiatry.* 2021;11(1):543. <https://doi.org/10.1038/s41398-021-01663-6>.
44. Wisner KL, Chambers C, Sit DKY. Postpartum depression: a major public health problem. *JAMA.* 2006;296(21):2616–8. <https://doi.org/10.1001/jama.296.21.2616>.
45. Wisner KL, Parry BL, Piontek CM. Postpartum depression. *N Engl J Med.* 2002;347(3):194–9. <https://doi.org/10.1056/nejmcp011542>.
46. Xiao M, Yan C, Fu B, Yang S, Zhu S, Yang D, et al. Risk prediction for postpartum depression based on random forest. *Zhong Nan Da Xue Xue Bao Yi Xue Ban J Central South Univ Med Sci.* 2020;45(10):1215–22. <https://doi.org/10.11817/j.issn.1672-7347.2020.190655>.
47. Xie Y. Reliability and validity of the simplified coping style questionnaire. *Chin J Clin Psychol.* 1998;6(2):114–5. <https://doi.org/10.16128/j.cnki.1005-3611.1998.02.018>.
48. Yim IS, Stapleton LRT, Guardino CM, Hahn-Holbrook J, Schetter CD. Biological and psychosocial predictors of postpartum depression: systematic review and call for integration. *Annu Rev Clin Psychol.* 2015;11:99–137. <https://doi.org/10.1146/annurev-clinpsy-101414-020426>.
49. Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W. Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study. *JMIR Med Inform.* 2020;8(4): e15516. <https://doi.org/10.2196/15516>.

# Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.