Am J Transl Res 2026;18(2):914-931
www.ajtr.org /ISSN:1943-8141/AJTR0169954

Original Article

Construction and validation of a multicenter predictive model for neonatal hypoglycemia risk in infants of mothers with gestational diabetes mellitus

Xiuchen Geng, Mingyan Hou, Shuai Zhang, Wei Yang, Zhijie Geng

Department of Obstetrics, Tianjin Third Central Hospital, Tianjin 300170, China

Received November 12, 2025; Accepted January 12, 2026; Epub February 15, 2026; Published February 28, 2026

Abstract: Objective: This study aimed to develop and validate a predictive model for neonatal hypoglycemia (NH) risk in infants born to mothers with gestational diabetes mellitus (GDM). Methods: A multicenter retrospective cohort study included 3215 GDM mother-infant pairs from five hospitals in Tianjin. Data were split into training, internal, and external validation sets. Risk factors were selected by expert consultation and univariate analysis. Logistic regression, random forest (RF), and radial basis function neural network models were built and evaluated using AUC, sensitivity, and specificity. Results: Fourteen risk factors were identified, with hypothermia (OR = 2.31), insulin treatment during pregnancy (OR = 2.15), and large for gestational age (OR = 2.08) being the strongest. The RF model performed best, with AUC values of 0.896, 0.872, and 0.865 across validation groups. In external validation (threshold = 0.38), sensitivity was 82.19% and specificity 79.38%. Subgroup analysis by maternal age, gestational week, and neonatal sex showed stable performance (AUC 0.848-0.895). A simplified RF model using five key predictors retained 97.34% of performance (AUC = 0.842) and reduced assessment time to 2-3 minutes. Conclusion: The RF model effectively predicts NH risk in GDM newborns with strong generalizability, supporting early clinical identification and intervention. Hypothermia, insulin use, and large for gestational age are core risk factors.

Keywords: Gestational diabetes mellitus, neonatal hypoglycemia, risk prediction model, random forest, multicenter study

Introduction

Gestational diabetes mellitus (GDM), a common metabolic disorder during pregnancy, has seen its incidence rise annually with the global obesity epidemic and delayed childbearing age, becoming a significant public health issue threatening maternal and infant health. Due to abnormal glucose regulation during pregnancy, GDM mothers often expose their fetuses to a hyperglycemic environment in utero, stimulating the proliferation of pancreatic β-cells and excessive insulin secretion, leading to fetal hyperinsulinemia [1, 2]. After birth, the sudden interruption of maternal glucose supply, coupled with persistently high endogenous insulin activity, readily induces neonatal hypoglycemia (NH).

Although some cases present as asymptomatic hypoglycemia, recurrent or severe hypoglycemia can cause irreversible central nervous system damage, increasing the risk of long-term intellectual developmental disabilities and motor dysfunction, posing a potential and lasting threat to neonatal health [3-5]. Therefore, early identification of high-risk individuals for NH in infants of GDM mothers and timely implementation of targeted preventive measures are crucial for reducing adverse outcomes related to hypoglycemia [6]. Currently, clinical prevention and control of NH mainly rely on postpartum blood glucose monitoring, lacking effective prenatal risk assessment tools [7, 8]. Although existing studies have found associations between factors such as pre-pregnancy maternal BMI, history of insulin use, and neonatal birth weight and NH [9-11], these studies are mostly single-center and small-sample explorations with unsystematic risk factor selection and no formation of generalizable quantitative prediction models. Single-center studies, influenced

https://doi.org/10.62347/ROHH6096

by regional, population, and medical practice differences, have limited extrapolation of results and can't meet the clinical needs of different medical institutions.

The development of machine learning algorithms has provided new avenues for risk prediction of multifactorial complex diseases. Algorithms such as Logistic regression, random forest (RF), and radial basis function (RBF) neural network have demonstrated advantages in constructing medical prediction models [12-14]. Therefore, this study planned to conduct a multicenter retrospective cohort study, integrating clinical data of GDM mother-infant pairs from five tertiary hospitals in Tianjin. It systematically identified key risk factors for NH, constructed and validated various machine learning-based prediction models, and compared the performance differences among different models to establish the optimal model. The study results are expected to provide a precise and reliable tool for NH risk assessment in clinical practice, facilitating early warning and individualized prevention and control of NH in infants of GDM mothers, thereby improving maternal and infant outcomes.

Materials and methods

Study design and participants

A multicenter retrospective cohort design was employed in this study, with GDM mothers and their newborns at five tertiary hospitals in Tianjin (including Tianjin Third Central Hospital, the Second Hospital of Tianjin Medical University, etc.) from October 2023 to April 2025 being selected as research participants. All clinical data were derived from the hospital electronic medical record system and the mater nal and child health database. This study was approved by the Ethics Committee of Tianjin Third Central Hospital (IRB2025-040-02).

Inclusion criteria were established according to the *Guidelines for the Diagnosis and Treatment of Hyperglycemia in Pregnancy* (2022) [15], including: Mothers diagnosed with GDM through a 75-g oral glucose tolerance test; Gestational age at delivery ≥ 28 weeks; Newborns with a hospital stay exceeding 24 hours after birth and complete blood glucose monitoring records; Complete and traceable key clinical data for both mothers and newborns. Exclusion criteria included: missing key data, mothers with significant organ dysfunction or major chronic diseases, and newborns with chromosomal abnormalities, congenital malformations, or severe infections.

The sample size estimation took into account the 34 risk factors to be included and the requirements of different machine learning algorithms. According to the principle of Logistic regression analysis, each factor required 5-10 cases of NH. Combined with the reported incidence of hypoglycemia in GDM offspring of about 27% in the literature [16], the basic sample size was calculated to be 1260 cases. Meanwhile, the RF algorithm required 680 cases, and the RBF neural network required 510 cases. Considering an estimated data missing rate of approximately 20%, the final sample size for the modeling group was determined to be 1575 cases.

Allocation was performed using a random number table. Three hospitals were randomly selected from the five, and their eligible cases were allocated: modeling (1575 cases) and internal validation (675 cases) in a 7:3 ratio. The cases from the remaining two hospitals served as the external validation (approximately 965 cases), which was used to assess the generalizability of the models.

Study tools and data extraction

Through a systematic search of Chinese and English databases, this study comprehensively collected relevant literature on NH in offspring of GDM mothers and initially constructed a pool of 34 risk factors, covering three dimensions: demographic characteristics, maternal pregnancy conditions, and neonatal status. To ensure the authority and reliability of the study tools, 15 experts with associate senior professional title or higher in related fields were invited to participate in two rounds of consultation. The experts scored the importance of each item, and the reliability of the consultation results was verified using the expert activity coefficient, authority coefficient (Cr ≥ 0.7), and coordination coefficient (Kendall's W). The *Clinical Data Extraction Form for GDM Mothers and Newborns* was ultimately formed.

Data extraction and quality verification teams were separately established, each composed

of five medical staff from obstetrics and gynecology or neonatology, and all members received unified training. Data extractors systematically searched and extracted relevant data through the hospital electronic medical record system and the maternal and child health database. The extracted content included all clinical records and indicators from the pre-pregnancy, pregnancy, delivery stages of the mothers, and after the birth of the newborns.

10% of the extracted data were randomly selected and compared with the original medical records. If data discrepancies were found, they were immediately fed back and corrected until the overall data error rate was below 2%.

Quality control

Tertiary hospitals with standardized data management were selected as cooperative units to ensure the use of a unified electronic medical record template and blood glucose monitoring standards. Before extraction, the hospital information department assisted in exporting a complete list of qualified cases. A detailed *Data Extraction Operation Manual* was developed by the study to unify the standards.

Data were double-entered into the EpiData 3.1 database by two individuals and compared and corrected. Subsequently, Excel was used for data cleaning, and problematic data were processed through logical verification and outlier identification. For missing values, measures such as imputation, follow-up, or marking and explanation were taken according to the missing rate.

Performance validation of the model in subgroups

Referring to the commonly used dimensions for risk stratification of GDM mother-infant pairs in clinical practice, the total sample was divided into subgroups based on the following three key variables: (1) Maternal age: According to the WHO definition of advanced maternal age, the sample was divided into the < 35 years group (2208) and the ≥ 35 years group (1007). (2) Gestational week at delivery: Referring to the recommendations on the timing of delivery in the *Guidelines for the Diagnosis and Treatment of Gestational Diabetes Mellitus* (2022), the sample was divided into the < 38 weeks group (preterm tendency, 892) and the ≥ 38 weeks group (term, 2323). (3) Neonatal sex: The sample was divided into the male infant group (1609) and the female infant group (1606) to balance the potential impact of sex on neonatal metabolic status.

Based on the RF model, validation was conducted within each subgroup for AUC, sensitivity, specificity, and the Hosmer-Lemeshow goodness-of-fit test. Each subgroup maintained a three-level validation structure to ensure consistency in validation logic with the total sample. Receiver operating characteristic (ROC) curves were plotted for each subgroup, and the χ^2 value and *P* value of the Hosmer-Lemeshow test were calculated, with calibration curves also visualized for the subgroups.

Statistical analysis

SPSS 25.0, R 3.6.2, and other software were used. Measurement data conforming to a normal distribution were reported as mean ± SD, and *t*-test was applied; non-normally distributed data were reported as median (quartile), and non-parametric tests were applied. Count data were reported as frequency (percentage), and χ^2 test was applied. Univariate analysis was conducted with NH occurrence as the dependent variable to screen for potential risk factors with $P < 0.05$.

With NH occurrence as the dependent variable and variables with significance ($P < 0.05$) and clinical relevance from the univariate analysis as independent variables, three prediction models were constructed: logistic regression model: variables were selected using the forward stepwise regression method, with variables with $P < 0.05$ retained in the final model to compute OR and 95% CI; RF model: the optimal parameters (ntree, mtry) were determined through 5-fold cross-validation, and variable importance was assessed based on the reduction in node impurity; RBF neural network model: continuous variables were standardized using Z-score, hidden layers were automatically generated, and training was conducted with the goal of minimizing mean squared error.

Model performance was evaluated. The main evaluation indicators included the area under the curve (AUC), sensitivity, and specificity. The optimal prediction model was selected based

on the performance of each model on the validation set, and its best prediction threshold was determined. The threshold for statistical significance was defined as $\alpha = 0.05$. The diagnostic criteria for NH referred to the 2021 expert consensus [17], which defined NH as meeting any one type of transitional, recurrent, severe, or symptomatic hypoglycemia.

The consistency between the predicted probabilities and the actual incidence rates of the RF model was assessed by the Hosmer-Lemeshow goodness-of-fit test in combination with calibration curve visualization. The three groups were divided into 10 subgroups respectively according to the predicted NH risk probability from low to high. The actual incidence rate of NH (the number of NH cases in the subgroup/the total number of cases in the subgroup) and the mean predicted NH probability (the mean of the predicted probabilities of all individuals in the subgroup) in each subgroup were calculated. The calibration curve was then drawn. The χ^2 value and *P* value were calculated by the Hosmer-Lemeshow test. $P > 0.05$ indicated that there was no statistical difference between the model's predicted probability and the actual incidence rate, and the model had a good fit.

Results

Baseline data of participants

3215 GDM mother-infant pairs were included, and divided into the modeling group (1575 cases), internal validation group (675 cases), and external validation group (965 cases). In **Table 1**, the three groups were basically consistent in demographic characteristics and risk factors for mothers and newborns ($P > 0.05$), with balanced grouping. The average maternal age was 33-35 years old, with a pre-pregnancy BMI indicating overweight; the internal validation group had a higher rate of insulin use (46.37%), and the external group had a higher proportion of metabolic family history (25.80%). The average gestational age of the newborns was about 38.5 weeks, with an average birth weight of about 3290 g and a low incidence of complications. No statistically meaningful distinctions were noted. The three groups exhibited a high degree of consistency in certain indicators (all $P > 0.05$), which may be attributed to several factors. First, the external validation cohort consisted of cases from independent hospitals, minimizing potential selection bias. Second, strict adherence to uniform inclusion and exclusion criteria reduced population heterogeneity. Furthermore, all collaborating hospitals were tertiary-level institutions, where standardized clinical practices and uniform data recording protocols ensured data homogeneity.

Analysis of risk factors for NH in GDM mothers

With "occurrence of NH" as the dependent variable (occurred = 1, not occurred = 0), univariate analysis of NH risk factors was conducted on 1575 GDM mothers. In **Table 2**, 16 factors were found to be associated with the occurrence of NH ($P < 0.05$). Among them, pre-pregnancy BMI ≥ 28 kg/m^2, family history of metabolic diseases, hypertension in pregnancy, hypothyroidism, insulin therapy, use of glucocorticoids, placental abnormalities, and weight gain during pregnancy ≥ 12 kg were identified as maternal-related risk factors. Intrauterine distress, small or large for gestational age, hypothermia, low birth weight or macrosomia, feeding frequency < 6 times, and 1-minute Apgar < 9 were identified as neonatal-related risk factors. Through systematic collinearity testing and independence analysis, factors that did not meet the model inclusion criteria were excluded, resulting in the identification of 14 key risk factors. The specific exclusion criteria were as follows: (1) Collinearity testing was performed using the Variance Inflation Factor (VIF), with a threshold of VIF > 10 indicating severe collinearity. When the VIF values of certain factors exceeded this threshold, they were considered to exhibit strong collinearity. (2) Independence analysis was conducted using Pearson correlation, with a bivariate correlation coefficient of $r > 0.7$ indicating poor independence. Factors reaching this threshold were regarded as highly overlapping in information and lacking sufficient independence. Based on the principle of "prioritizing indicators with clearer clinical significance, greater data completeness, and better accessibility", factors failing to meet these criteria were excluded, ensuring that the variables included in the model were free of significant collinearity and maintained adequate independence. Hypothermia (OR = 2.31), insulin therapy (OR = 2.15), and large for gestational age (OR = 2.08) were strong risk factors for the occurrence of NH, with the remaining factors having OR values between 1.32 and 1.97,

Table 1. Contrast of baseline characteristics of the participants

Category	Indicator	Modeling (n = 1575)	Internal validation (n = 675)	External validation (n = 965)	Statistics	*P*
Demographic risk factors	Maternal age (years)	33.28 ± 4.12	34.46 ± 3.89	34.82 ± 4.05	F = 1.872	0.154
	Pre-pregnancy BMI (kg/m^2)	25.36 ± 3.25	25.45 ± 3.18	25.58 ± 3.31	F = 0.326	0.722
	Family history of metabolic or endocrine disorders	102 (6.48)	94 (13.93)	249 (25.80)	X^2 = 12.364	0.075
Maternal risk factors	Previous diabetes	18 (1.14)	17 (2.52)	46 (4.77)	X^2 = 9.851	0.131
	History of hypertensive disorders in pregnancy	185 (11.74)	178 (26.37)	154 (15.96)	X^2 = 14.286	0.052
	History of hypothyroidism during pregnancy	97 (6.16)	87 (12.89)	111 (11.50)	X^2 = 8.973	0.175
	Previous delivery of a macrosomic infant	32 (2.03)	32 (4.74)	39 (4.04)	X^2 = 7.652	0.264
	Number of pregnancies	2.35 ± 0.87	2.39 ± 0.91	2.43 ± 0.89	F = 0.418	0.658
	Multiple gestation	29 (1.84)	28 (4.15)	22 (2.28)	X^2 = 6.987	0.322
	Weight gain during pregnancy (kg)	9.82 ± 2.35	10.12 ± 2.41	10.41 ± 2.38	F = 1.205	0.3
	Placental abnormalities	167 (10.60)	153 (22.67)	153 (15.85)	X^2 = 13.672	0.064
	Insulin therapy during pregnancy	342 (21.71)	313 (46.37)	168 (17.41)	X^2 = 15.128	0.048
	Use of medications within 24 hours before delivery	218 (13.84)	207 (30.67)	61 (6.32)	X^2 = 16.893	0.031
	Maternal prenatal HbA1c ≥ 6.5%	189 (12.00)	83 (12.30)	118 (12.23)	X^2 = 0.042	0.979
	Mode of delivery: cesarean section	792 (50.30)	342 (50.67)	486 (50.36)	X^2 = 0.035	0.983
	Parity: multiparous	672 (42.67)	286 (42.37)	409 (42.38)	X^2 = 0.018	0.991
	Total labor duration ≥ 12 hours	276 (17.52)	116 (17.19)	165 (17.10)	X^2 = 0.073	0.964
Neonatal risk factors	Intrauterine fetal distress	95 (6.03)	84 (12.44)	78 (8.08)	X^2 = 9.235	0.156
	Gestational age (weeks)	38.52 ± 1.25	38.47 ± 1.31	38.53 ± 1.28	F = 0.187	0.83
	Small for gestational age	89 (5.65)	80 (11.85)	101 (10.47)	X^2 = 11.542	0.089
	Large for gestational age	127 (8.06)	114 (16.89)	134 (13.89)	X^2 = 10.876	0.102
	Parity	1.28 ± 0.45	1.32 ± 0.48	1.35 ± 0.47	F = 0.521	0.594
	Gender (male)	776 (49.27)	347 (51.41)	486 (50.36)	X^2 = 0.689	0.708
	Birth weight (g)	3285.67 ± 425.31	3296.36 ± 418.25	3305.17 ± 421.68	F = 0.298	0.743
	1-minute Apgar score	9.68 ± 0.32	9.69 ± 0.31	9.68 ± 0.33	F = 0.076	0.927
	5-minute Apgar score	9.89 ± 0.15	9.90 ± 0.14	9.90 ± 0.15	F = 0.112	0.894
	Duration of mother-infant skin-to-skin contact (min)	38.75 ± 10.23	39.11 ± 9.87	39.31 ± 10.15	F = 0.158	0.854
	Number of feedings in the first 24 hours after birth	8.2 ± 1.5	8.5 ± 1.4	8.5 ± 1.6	F = 1.025	0.359
	Hypothermia	43 (2.73)	40 (5.93)	37 (3.83)	X^2 = 7.218	0.299
	Respiratory distress syndrome	31 (1.97)	30 (4.44)	14 (1.45)	X^2 = 8.125	0.223
	Meconium aspiration syndrome	17 (1.08)	17 (2.52)	4 (0.41)	X^2 = 9.364	0.154
	Polycythemia	22 (1.40)	21 (3.11)	1 (0.10)	X^2 = 12.875	0.067
	Early-onset neonatal infection	28 (1.78)	13 (1.93)	0 (0.00)	X^2 = 10.542	0.096
	Hemolytic anemia	9 (0.57)	4 (0.59)	0 (0.00)	X^2 = 5.872	0.44
	Time to first feeding > 2 hours	243 (15.43)	104 (15.41)	149 (15.44)	X^2 = 0.001	0.999

Note: BMI: Body mass index.

Table 2. Univariate analysis results of key risk factors for NH occurrence in GDM mothers

Category	Key risk factors	Presence of NH (n = 438)	Absence of NH (n = 1137)	Statistics	*P*	OR (95% CI)
Demographic risk factors	Pre-pregnancy BMI ≥ 28 kg/m^2	156 (35.62)	278 (24.46)	X^2 = 11.326	0.003	1.78 (1.36-2.33)
	Family history of metabolic or endocrine disorders	48 (10.96)	54 (4.75)	X^2 = 8.972	0.007	1.52 (1.18-1.96)
Maternal risk factors	History of hypertensive disorders in pregnancy	89 (20.32)	96 (8.44)	X^2 = 15.682	< 0.001	1.97 (1.54-2.52)
	History of hypothyroidism during pregnancy	52 (11.87)	45 (3.96)	X^2 = 9.215	0.006	1.63 (1.24-2.14)
	Insulin therapy during pregnancy	156 (35.62)	186 (16.36)	X^2 = 22.458	< 0.001	2.15 (1.76-2.62)
	Use of glucocorticoids and other medications within 24 hours before delivery	78 (17.81)	140 (12.32)	X^2 = 10.873	0.004	1.45 (1.13-1.86)
	Placental abnormalities	82 (18.72)	85 (7.47)	X^2 = 12.147	0.002	1.82 (1.43-2.32)
	Weight gain during pregnancy ≥ 12 kg	95 (21.69)	132 (11.61)	X^2 = 8.654	0.008	1.38 (1.09-1.75)
	Maternal prenatal HbA1c ≥ 6.5%	52 (11.87)	121 (10.64)	X^2 = 0.428	0.513	1.13 (0.81-1.57)
	Cesarean section	223 (50.91)	562 (49.43)	X^2 = 0.365	0.546	1.06 (0.87-1.29)
	Multiparity	187 (42.69)	473 (41.59)	X^2 = 0.152	0.696	1.04 (0.85-1.27)
	Total labor duration ≥ 12 hours	75 (17.12)	183 (16.09)	X^2 = 0.284	0.594	1.08 (0.81-1.44)
Neonatal risk factors	Intrauterine fetal distress	68 (15.53)	27 (2.37)	X^2 = 16.932	< 0.001	1.75 (1.36-2.25)
	Small for gestational age	58 (13.24)	31 (2.73)	X^2 = 14.825	< 0.001	1.68 (1.29-2.18)
	Large for gestational age	79 (18.04)	48 (4.22)	X^2 = 18.364	< 0.001	2.08 (1.67-2.59)
	Hypothermia (body temperature < 36.5 °C)	36 (8.22)	7 (0.61)	X^2 = 20.157	< 0.001	2.31 (1.85-2.89)
	Number of feedings within 24 hours after birth < 6 times	85 (19.41)	92 (8.10)	X^2 = 17.685	< 0.001	1.56 (1.23-1.98)
	Time to first feeding > 2 hours	68 (15.53)	165 (14.51)	X^2 = 0.217	0.641	1.08 (0.79-1.46)

Note: BMI: Body mass index; OR: Odds ratio; 95% CI: 95% confidence interval.

Table 3. Core parameters of the Logistic regression model

Variable	β	SE	Wald χ^2	*P*	OR (95% CI)
Constant term	-3.215	0.42	59.872	< 0.001	-
Pre-pregnancy BMI ≥ 28 kg/m^2	0.526	0.19	8.136	0.004	1.692 (1.201-2.385)
Family history of metabolic or endocrine disorders	0.364	0.16	5.021	0.025	1.439 (1.048-1.978)
History of hypertensive disorders in pregnancy	0.618	0.17	12.785	< 0.001	1.855 (1.332-2.583)
Insulin therapy during pregnancy	0.754	0.16	22.364	< 0.001	2.125 (1.603-2.814)
Placental abnormalities	0.489	0.18	7.562	0.006	1.631 (1.168-2.279)
Intrauterine fetal distress	0.592	0.18	10.653	0.001	1.808 (1.287-2.542)
Small for gestational age	0.457	0.19	5.638	0.018	1.579 (1.086-2.297)
Large for gestational age	0.721	0.17	18.932	< 0.001	2.057 (1.536-2.754)
Hypothermia (body temperature < 36.5°C)	0.832	0.21	15.287	< 0.001	2.300 (1.568-3.378)
Number of feedings within 24 hours after birth < 6 times	0.413	0.18	5.521	0.019	1.511 (1.073-2.128)

Note: BMI: Body mass index; OR: Odds ratio; 95% CI: 95% confidence interval.

all of which were independent predictive indicators of NH.

Results of risk prediction model construction

Based on the data of 1575 cases in the modeling group and incorporating 14 key risk factors, the risk prediction models for NH in GDM mothers were constructed using Logistic regression, RF, and RBF neural network algorithms, respectively.

The forward stepwise regression method (Forward: LR) was employed to select from the 14 key risk factors, with the inclusion criterion set at α = 0.05 and the exclusion criterion at α = 0.10. Ultimately, 10 factors were encompassed in the model (**Table 3**). The equation for the Logistic regression model is: logit(P) = -3.215 + 0.526 * (Pre-pregnancy BMI ≥ 28 kg/m^2) + 0.364 * (Family history of metabolic or endocrine diseases) + 0.618 * (History of hypertensive disorders of pregnancy) + 0.754 * (Insulin therapy during pregnancy) + 0.489 * (Placental abnormalities) + 0.592 * (Intrauterine fetal distress) + 0.457 * (Small for gestational age) + 0.721 * (Large for gestational age) + 0.832 * (Neonatal hypothermia) + 0.413 * (Feeding frequency < 6 times within 24 hours after birth). The logit(P) represents the logit transformation of the probability of NH occurrence, that is, logit(P) = ln[P/(1-P)]. The coefficient (β value) of each variable indicates its contribution to logit(P). In the equation, P is the probability of NH in infants of GDM mothers. The values assigned to each risk factor are "yes = 1, no = 0". The goodness-of-fit test for the model showed χ^2 = 8.326 and *P* = 0.402 > 0.05 in the Hosmer-Lemeshow test, indicating a good fit of the model.

The optimal parameters were determined to be ntree = 1500 and mtry = 6 through 5-fold cross-validation using the RF algorithm, at which point the out-of-bag error rate was the lowest (6.82%). The ranking of variable importance showed that the top five factors contributing most to the prediction of NH were: hypothermia, insulin therapy during pregnancy, large for gestational age, intrauterine fetal distress, and history of hypertensive disorders in pregnancy (**Table 4**).

In addition, an RBF neural network model with 14 input nodes, 10 hidden nodes, and 1 output node was constructed. After standardization of the input data, the model was trained using mean squared error as the loss function. The model converged well over 1000 iterations with a learning rate of 0.01, and the error reached its minimum at the 680th iteration (mean squared error = 0.042) without overfitting.

Results of model validation and performance comparison

The three prediction models, namely Logistic regression, RF, and RBF neural network, were validated internally and externally. The results showed that the RF model had the best overall performance, with AUC values of 0.896, 0.872, and 0.865 in the modeling group, internal validation group, and external validation group, respectively. It demonstrated stable predictive performance and strong generalizability. The Logistic regression model was relatively weaker

Table 4. Ranking of risk factor importance in the RF model

No.	Variable	Mean Decrease Gini
1	Hypothermia (Body temperature < 36.5°C)	102.36
2	Insulin therapy during pregnancy	98.75
3	Large for gestational age	92.18
4	Intrauterine fetal distress	85.42
5	History of hypertensive disorders in pregnancy	78.69
6	Placental abnormalities	72.35
7	Pre-pregnancy BMI ≥ 28 kg/m^2	68.54
8	Small for gestational age	65.12
9	Family history of metabolic or endocrine disorders	59.87
10	Number of feedings within 24 hours after birth < 6 times	56.43
11	History of hypothyroidism during pregnancy	52.18
12	Use of glucocorticoids and other medications within 24 hours before delivery	48.76
13	Weight gain during pregnancy ≥ 12 kg	45.32
14	Early-onset neonatal infection	41.29

Note: BMI: Body mass index.

Table 5. Performance evaluation results of the three models on different datasets

Model type	Dataset	AUC	Hosmer-Lemeshow test (χ^2/*P*)	Sensitivity (%)	Specificity (%)	Youden index	PPV (%)	NPV (%)
Logistic regression	Modeling group	0.828	9.752/0.386	78.54	79.82	0.584	68.92	87.15
	Internal validation group	0.801	10.325/0.324	75.12	77.36	0.525	66.45	84.38
	External validation group	0.782	11.087/0.298	72.36	75.18	0.475	65.23	84.69
RF	Modeling group	0.896	7.218/0.512	86.30	82.54	0.688	72.15	92.36
	Internal validation group	0.872	7.854/0.486	83.56	80.12	0.637	69.87	90.54
	External validation group	0.865	8.326/0.453	82.19	79.38	0.616	68.54	89.72
RBF neural network	Modeling group	0.852	8.547/0.438	83.11	81.27	0.644	70.36	90.12
	Internal validation group	0.835	9.128/0.392	80.23	79.45	0.597	68.12	88.45
	External validation group	0.818	9.875/0.357	78.45	77.62	0.561	66.78	87.36

Note: RF: Random forest; RBF: Radial basis function; AUC: Area under the curve; PPV: Positive predictive value; NPV: Negative predictive value.

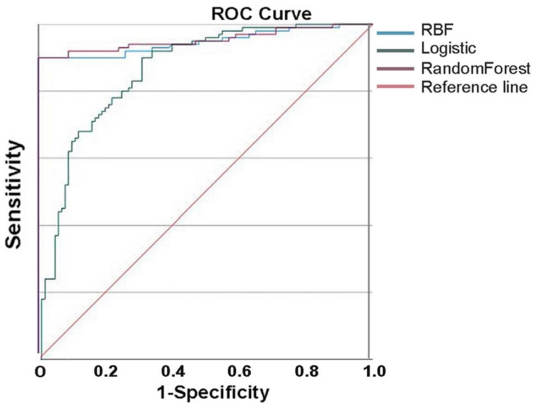

Figure 1. ROC curves of the three models in the external validation group. Note: RBF: Radial basis function; ROC: Receiver operating characteristic.

in extrapolation adaptability. Therefore, the RF model was selected as the optimal prediction model, and its risk determination threshold was set at 0.38. At this threshold, the model achieved a sensitivity of 81.72% and a specificity of 80.05% in the external validation, showing good screening practicability (**Table 5** and **Figure 1**).

Validation results of the model calibration curve

The calibration curves of the RF model in the modeling group, the internal validation group, and the external validation group were all close to the ideal calibration line (Y = X), indicating a good fit (**Figures 2**-**4**). As shown in **Table 6** below, in the Hosmer-Lemeshow test, the modeling group had χ^2 = 7.218, *P* = 0.512; the internal validation group had χ^2 = 7.854, *P* = 0.486; the external validation group had χ^2 = 8.326, *P* = 0.453. All *P* values were greater than 0.05,

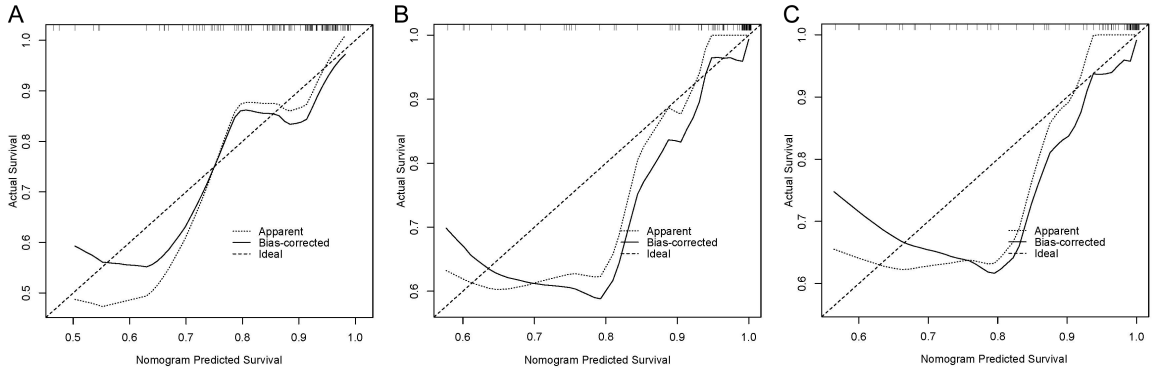

Figure 2. Calibration curve of the random forest (RF) model in the modeling group. Note: (A) Modeling group; (B) Internal validation group; (C) External validation group.

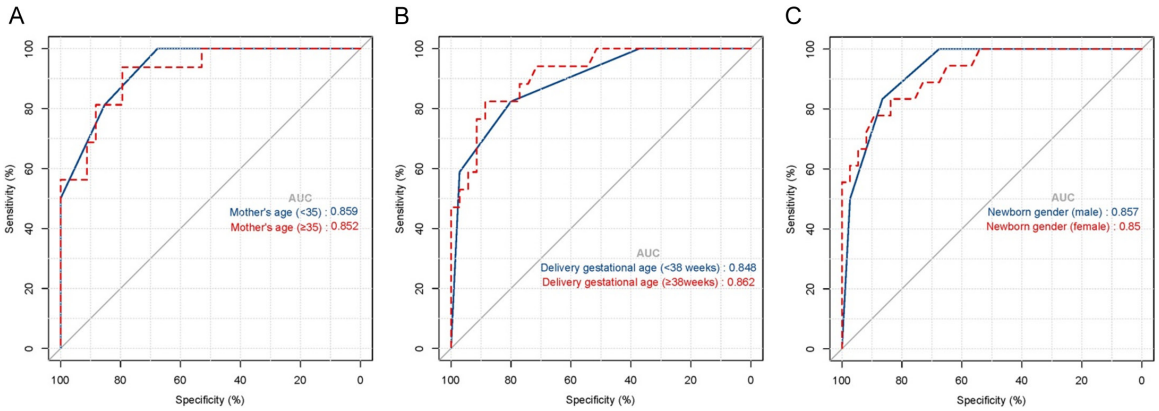

Figure 3. ROC curves of the RF model in different subgroups (external validation group). Note: (A) Represents maternal age; (B) Represents gestational week at delivery; (C) Represents neonatal sex. AUC: area under the curve.

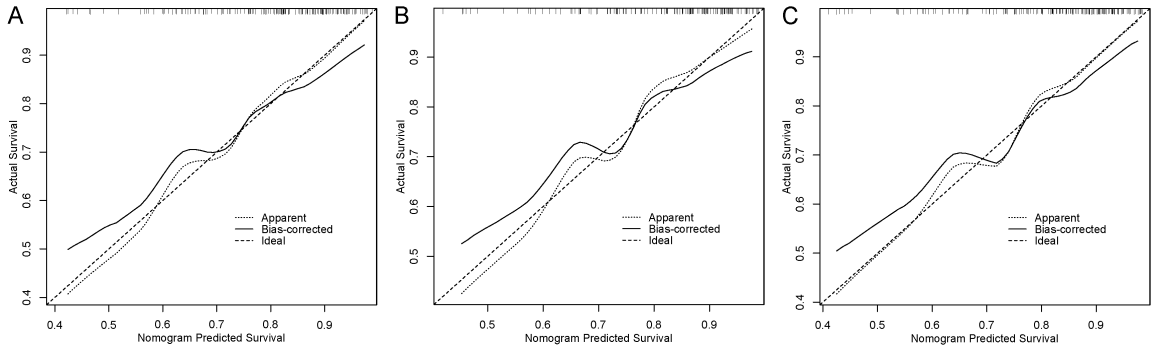

Figure 4. Calibration curves of the random forest (RF) model in different subgroups (external validation group). Note: (A) Represents maternal age; (B) Represents gestational week at delivery; (C) Represents neonatal sex.

indicating that there was no statistical difference between the model's predicted probability and the actual incidence rate. Regarding the calibration slope and intercept, the modeling group had a slope of 0.986 (95% CI: 0.921-1.051) and an intercept of 0.021 (95% CI: -0.032-0.074); the internal validation group had a slope of 0.963 (95% CI: 0.898-1.028) and an intercept of 0.035 (95% CI: -0.028-0.098); the external validation group had a slope of 0.947 (95% CI: 0.882-1.012) and an intercept of 0.042 (95% CI: -0.021-0.105). All values were

Table 6. Calibration curve validation of the random forest (RF) model across datasets

Dataset	Hosmer-Lemeshow test (χ^2/*P*)	Calibration slope (95% CI)	Calibration intercept (95% CI)
Modeling group (n = 1575)	7.218/0.512	0.986 (0.921-1.051)	0.021 (-0.032-0.074)
Internal validation group (n = 675)	7.854/0.486	0.963 (0.898-1.028)	0.035 (-0.028-0.098)
External validation group (n = 965)	8.326/0.453	0.947 (0.882-1.012)	0.042 (-0.021-0.105)

Note: The calibration slope and intercept reflect the agreement between predicted probabilities and observed event rates, with ideal values being a slope of 1 and an intercept of 0. A Hosmer-Lemeshow test *P*-value > 0.05 indicates good model calibration. 95% CI: 95% confidence interval.

Table 7. Performance validation results of the RF model in different subgroups

Stratification dimension	Subgroup	Sample size	Dataset	AUC	Sensitivity (%)	Specificity (%)	Hosmer-Lemeshow test (χ^2/*P*)
Maternal age	< 35 years	2208	Modeling group	0.891	85.76	82.13	6.982/0.543
			Internal validation group	0.868	82.94	79.85	7.516/0.488
			External validation group	0.859	81.67	78.92	8.035/0.457
	≥ 35 years	1007	Modeling group	0.887	84.32	81.56	7.124/0.528
			Internal validation group	0.863	81.89	79.21	7.682/0.467
			External validation group	0.852	80.93	78.15	8.217/0.442
Gestational week at delivery	< 38 weeks	892	Modeling group	0.885	83.95	80.78	7.258/0.516
			Internal validation group	0.859	81.26	78.54	7.831/0.459
			External validation group	0.848	80.12	77.36	8.364/0.435
	≥ 38 weeks	2323	Modeling group	0.895	86.13	82.47	6.873/0.556
			Internal validation group	0.871	83.25	80.13	7.428/0.492
			External validation group	0.862	81.89	79.04	7.982/0.458
Neonatal sex	Male	1609	Modeling group	0.893	85.24	81.89	7.015/0.539
			Internal validation group	0.866	82.37	79.52	7.568/0.485
			External validation group	0.857	81.24	78.63	8.102/0.451
	Female	1606	Modeling group	0.889	84.87	81.34	7.156/0.525
			Internal validation group	0.861	81.75	79.08	7.723/0.463
			External validation group	0.85	80.56	77.89	8.259/0.439

Note: AUC: Area under the curve.

close to the ideal values (slope = 1, intercept = 0), suggesting that the model did not significantly overestimate or underestimate the risk.

Performance evaluation of the RF model in each subgroup

In **Table 7**, the AUC values of the RF model in each subgroup were maintained between 0.848 and 0.895, with sensitivity stable at 80.12%-86.13% and specificity stable at 77.36%-82.47%. The *P* values of the Hosmer-Lemeshow test were all > 0.05, indicating that the model had good discrimination and calibration in subgroups of different ages, gestational weeks, and sexes, with no significant performance fluctuations.

The ROC curves of the RF model in each subgroup are illustrated in **Figure 3**. The ROC curves of all subgroups were > 0.8, further verifying the stability of the model in the subgroups.

The calibration curves of the RF model in each subgroup are illustrated in **Figure 4**. The curves of all subgroups were closely distributed around the ideal line without significant deviation, proving that the model had good consistency between the predicted probability and the actual incidence of NH risk in each subgroup.

Validation of the simplified RF model

Based on the variable importance ranking of the original RF model (**Table 4**), the top 5 core risk factors were selected as input variables for the simplified model, specifically including: neonatal hypothermia (temperature < 36.5°C), maternal insulin treatment during pregnancy, large-for-gestational-age neonate, fetal dis-

Table 8. Comparison of performance between the simplified and full RF models (External validation group)

Model	Number of input variables	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Out-of-bag error rate (%)	Performance retention rate (%)	Clinical assessment duration (minutes)
Full RF model	14	0.865	82.19	79.38	68.54	89.72	6.82	100	5-8
Simplified RF model	5	0.842	80.35	77.61	66.28	88.35	7.25	97.34	2-3

Note: RF: Random Forest; AUC: Area under the curve; PPV: Positive predictive value; NPV: Negative predictive value.

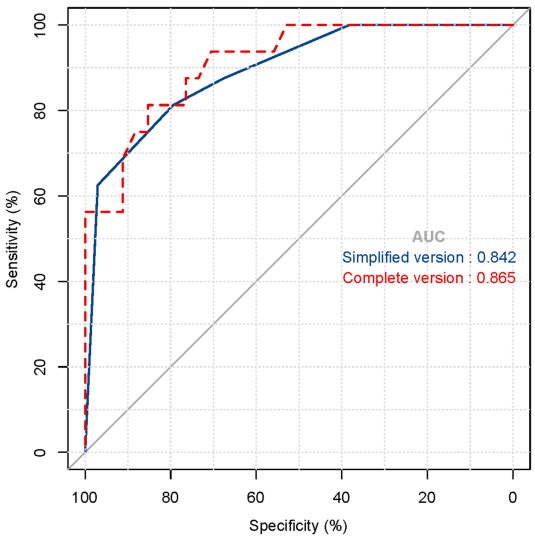

Figure 5. Comparison of ROC curves between the simplified and full random forest (RF) models. Note: AUC: area under the curve.

tress in utero, and maternal history of hypertension during pregnancy. The selection criteria were that the Mean Decrease Gini values of these 5 factors were all > 78, and the OR values in univariate analysis were all > 1.75, making them strong predictors of NH. These factors also covered three dimensions - mother, fetus, and neonate - meeting the convenience requirements for rapid clinical assessment.

The same modeling process as the original model was adopted, with the modeling group (1575 pairs) from the total sample used as the training set. Parameters were optimized through 5-fold cross-validation, ultimately determining ntree = 1200, mtry = 3, with an out-of-bag error rate of 7.25%. During validation, the performance of the simplified model was evaluated in both the internal validation group (675 pairs) and the external validation group (965 pairs), with evaluation indicators including AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and net benefit from decision curve analysis (DCA). Additionally, the time required for model assessment was calculated, simulating the time needed for clinicians to fill in the 5 factors. Combined with the performance retention rate (simplified model AUC/original model AUC × 100%), its clinical promotion value was assessed. The comparison of performance between the simplified and full RF models is shown in **Table 8**. The AUC of the simplified model decreased by only 0.023 compared to the full model, with a performance retention rate of 97.34%. Sensitivity, specificity, PPV, and NPV decreased by 1.84%, 1.77%, 2.26%, and 1.37%, respectively compared to the full model, all within the clinically acceptable range. The out-of-bag error rate increased by only 0.43%, and the clinical assessment time was reduced by more than 50%, balancing performance and convenience.

The comparison of ROC curves between the simplified and full RF models is illustrated in **Figure 5**, with a high degree of overlap between the two curves, demonstrating that the simplified model had discrimination ability close to that of the full model.

The comparison of DCA between the simplified and full RF models is illustrated in **Figure 6**. The net benefit curves of both the simplified and full models were significantly higher than those of "intervene on all" and "intervene on none" strategies, proving that the simplified model could generate net benefits comparable to the full model in clinical decision-making.

The calibration curve of the simplified RF model (external validation group) is illustrated in **Figure 7**. The calibration curve of the simplified model closely fitted the ideal line, with the Hosmer-Lemeshow test $P > 0.05$, indicating good consistency between the predicted probability and the actual incidence rate, without significant overestimation or underestimation.

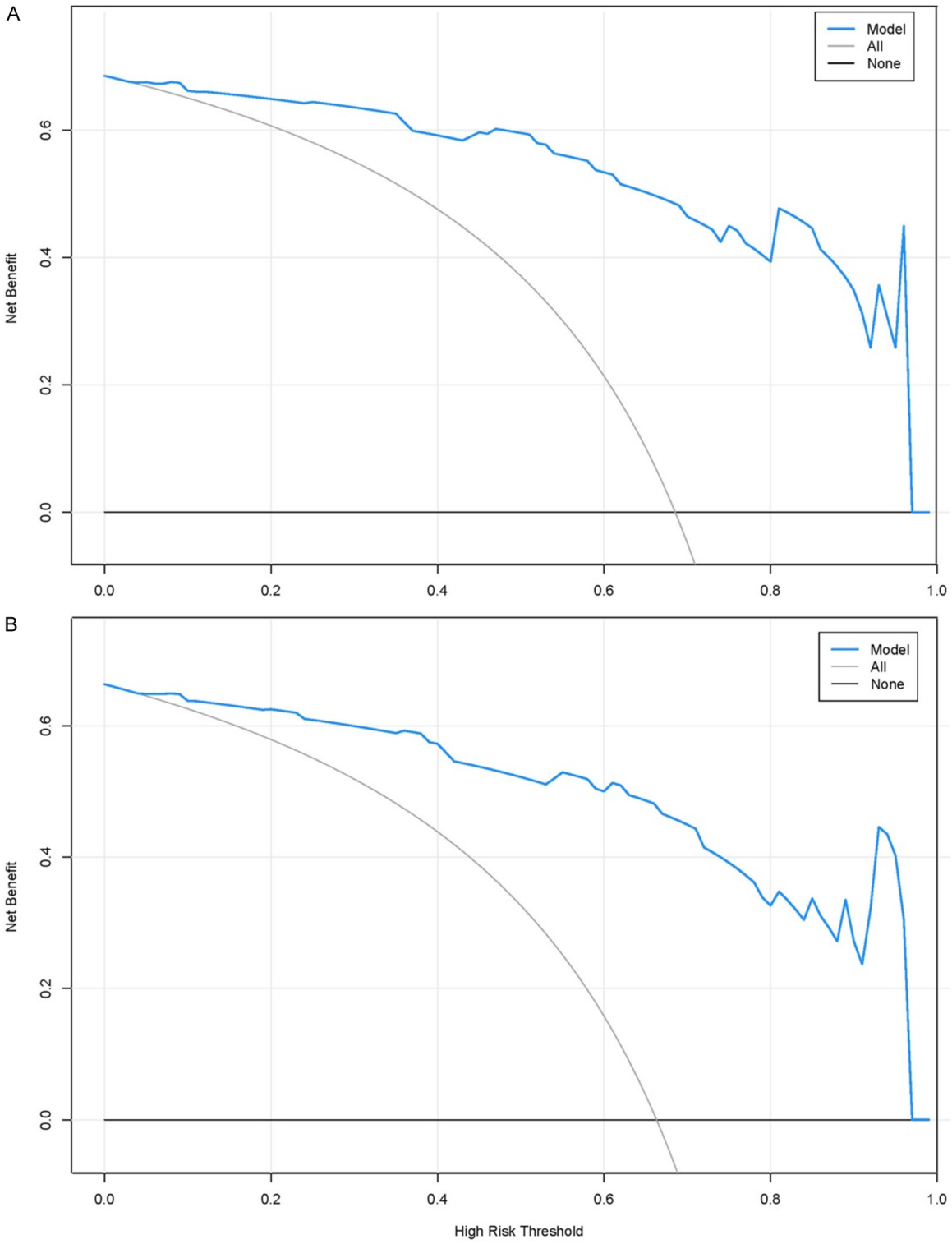

Figure 6. DCA of the simplified and full random forest (RF) models (external validation group). Note: (A) Represents the simplified model; (B) Represents the full model.

Discussion

In this study, a predictive model for the risk of NH in newborns delivered by mothers with GDM was constructed and evaluated based on multi-center data. Through univariate analysis, 14 significant influencing factors were identified, among which hypothermia, insulin treatment

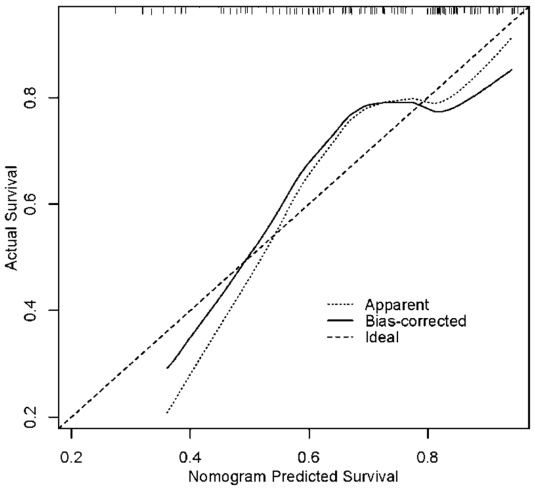

Figure 7. Calibration curve of the simplified random forest (RF) model (external validation group).

during pregnancy, and large for gestational age were the main risk factors, with OR values reaching 2.31, 2.15, and 2.08, respectively. Hypothermia, as the primary risk factor, is likely closely related to the immaturity of the neonatal thermoregulatory center. In a cold environment, the neonatal metabolic rate increases, glucose consumption accelerates, and cold stimulation activates the sympathetic nervous system, further promoting glycogenolysis [18]. Newborns of GDM mothers may already have insufficient glycogen reserves, and under the dual influence of these factors, they are more prone to hypoglycemia. The high risk associated with insulin treatment during pregnancy is directly related to the level of maternal blood glucose control. Insulin, after crossing the placental barrier, may cause fetal hyperinsulinemia. After birth, the maternal glucose supply is interrupted, but the neonatal insulin level remains high, leading to hypoglycemia [19]. Large for gestational age is usually due to maternal hyperglycemia stimulating increased fetal insulin secretion, forming a vicious cycle of hyperinsulinemia and hypoglycemia, which is consistent with the clinical observation of a higher incidence of hypoglycemia in macrosomic infants of GDM mothers [20]. In addition, gestational hypertension and fetal distress, among other factors, affect placental blood flow perfusion, leading to fetal intrauterine hypoxia and nutritional supply disorders, reducing the capacity for glycogen storage and increasing the risk of hypoglycemia [21]. From a clinical perspective, these core risk factors provide clear targets for the precise prevention and management of neonatal hypoglycemia. For neonatal hypothermia, which represents a primary risk factor, immediate postnatal thermal care can be implemented, such as incubator management or skin-to-skin warming, to maintain body temperature above 36.5°C and thereby reduce glucose consumption. For mothers receiving insulin therapy during pregnancy, enhanced monitoring of maternal blood glucose is necessary to prevent excessive insulin exposure that may lead to fetal hyperinsulinemia. Additionally, feeding should be initiated within one hour after birth, with glucose supplementation administered if indicated. In the case of large-for-gestational-age (LGA) infants, the first blood glucose measurement should be performed within 30 minutes after delivery, followed by reassessment every 1-2 hours until stable levels are achieved. These targeted interventions can effectively reduce the risk of hypoglycemia-related neurological injury. The results of model construction and validation indicated that the RF model exhibited the most excellent predictive performance, with an AUC value of 0.865 in the external validation group, which was significantly higher than that of the logistic regression model (0.782) and the RBF neural network model (0.818). This advantage was derived from the ensemble learning characteristics of the RF, which effectively reduced the risk of overfitting in a single model through the collective decision-making mechanism of multiple decision trees and was able to fully capture the complex nonlinear relationships between variables, making it particularly suitable for dealing with the complex scenarios of multifactorial interactions in medical data [22]. The variable importance ranking results were basically consistent with the univariate analysis, further confirming the core predictive value of factors such as hypothermia and insulin treatment, and providing double evidence for the clinical determination of key monitoring indicators. The logistic regression model performed relatively weakly, probably because it assumed a linear relationship between variables and could not fully fit the complex distribution characteristics of clinical data [23]; although the RBF neural network could handle nonlinear relationships, it was more sensitive to sample size and data distri-

bution, and its generalization ability was slightly inferior to that of the RF model. When the risk threshold of the RF model was set at 0.38, it showed a sensitivity of 81.72% and a specificity of 80.05% in the external validation, with both good screening efficiency and diagnostic accuracy. The determination of this threshold provided a quantitative basis for clinical risk stratification management and helped medical staff quickly identify high-risk newborns. By implementing early intervention measures such as enhanced temperature management, increased feeding frequency, and close blood glucose monitoring, the risk of hypoglycemia could be effectively reduced and the clinical outcomes of newborns could be improved. The high sensitivity (82.19%) and specificity (79.38%) of the RF model in this study hold significant clinical value. High sensitivity enables the identification of neonates at potential high risk, minimizing the likelihood of missed diagnoses and associated severe adverse outcomes. Conversely, high specificity helps to reduce unnecessary interventions in low-risk neonates, such as frequent glucose monitoring or unwarranted glucose infusions, thereby decreasing both healthcare resource consumption and maternal-infant stress. Clinically, this model can be integrated into obstetric electronic medical record systems, allowing the automatic input of relevant risk factors immediately after birth. A hypoglycemia risk score can then be rapidly generated, assisting healthcare providers in completing risk stratification within 30 minutes. High-risk neonates can be placed under intensified monitoring protocols, while low-risk infants can continue with routine care, facilitating precise and individualized medical management.

Comparing the RF model developed in this study with the PeRSonal GDM prediction model, which was derived from an Australian multicenter cohort for adverse pregnancy outcomes [24], can further clarify the advantages and positioning of our model. The PeRSonal model focuses on composite adverse outcomes in pregnant women with GDM, including neonatal hypoglycemia, and incorporates 12 routinely collected clinical indicators such as BMI, maternal age, and glucose values from oral glucose tolerance tests. Its temporal validation yielded a C-statistic of 0.68 (95% CI: 0.64-0.72) and a calibration slope of 0.99 (95% CI: 0.75-1.23), indicating acceptable discrimination and good calibration. In contrast, the RF model in this study specifically targets neonatal hypoglycemia among infants of mothers with GDM, achieving an external validation AUC of 0.865 (95% CI not explicitly reported, but markedly higher than that of the PeRSonal model). Both sensitivity (82.19%) and specificity (79.38%) surpass the PeRSonal model's predictive performance for neonatal hypoglycemia within the composite outcome (not separately reported, with overall C-statistic being lower). The main reasons for these differences are as follows: (1) Outcome definition-our study focuses on a single, specific outcome (neonatal hypoglycemia), whereas the PeRSonal model uses a composite of multiple outcomes, which may dilute the predictive signal for any individual endpoint; (2) Targeted risk factors-our model includes core indicators directly related to neonatal hypoglycemia, such as neonatal hypothermia and maternal insulin therapy during pregnancy, while the PeRSonal model incorporates broader maternal baseline characteristics, including ethnicity and family history of diabetes; (3) Modeling algorithm-the RF model captures complex interactions among multiple factors more effectively than the PeRSonal model, whose specific modeling approach is not detailed, thus further improving prediction accuracy for the single outcome. Moreover, decision curve analysis of the PeRSonal model confirmed its clinical utility within a probability threshold of 0.15-0.85. Similarly, the simplified RF model demonstrates net benefit highly consistent with the full model, while reducing clinical assessment time to 2-3 minutes, enhancing its applicability in primary care and emergency settings. Both models confirm that prediction models based on routine clinical data can effectively assist in risk stratification for GDM-related outcomes. However, the RF model developed in this study offers superior clinical reference value for predicting the specific outcome of neonatal hypoglycemia.

The subgroup analysis results of this study showed that the RF model maintained stable predictive performance across key clinical stratification dimensions, including maternal age, gestational week at delivery, and neonatal sex. The AUC values in each subgroup fluctuated

within a range of only 0.047 (0.848-0.895), with no significant decrease in sensitivity and specificity, and the calibration curves all closely fitted the ideal line. These results further strengthened the model's generalizability, providing key support for its application in diverse clinical populations. From a clinical practice perspective, the advanced maternal age group (≥ 35 years) often has characteristics such as reduced metabolic function and increased risk of pregnancy complications, while preterm infants with a gestational week < 38 weeks have insufficient glycogen reserves and weaker thermoregulation abilities. They may affect the accuracy of NH risk prediction. However, in this study, the model's performance remained at a high level in the advanced maternal age group (external validation AUC = 0.852) and the preterm group (external validation AUC = 0.848), indicating that the model had fully captured the association between core risk factors and NH, without significant interference from confounding factors such as age and gestational week. In addition, the performance consistency in the neonatal sex subgroups (male infants AUC = 0.857, female infants AUC = 0.850) also ruled out the impact of sex differences on the model's predictive effect, proving that the model was applicable to all neonates. The value of subgroup analysis also lies in providing a basis for precise stratified intervention. For example, the sensitivity in the < 38 weeks subgroup (80.12%) was slightly lower than that in the term group (81.89%), suggesting that as a high-risk group for NH, the risk decision threshold can be appropriately lowered when applying the model clinically for preterm infants to reduce the false-negative rate. This threshold optimization based on subgroup characteristics can further enhance the clinical adaptability of the model, providing personalized assessment schemes for neonates with different risk features.

In clinical practice, complex predictive models are often difficult to promote due to the large number of input variables and cumbersome operating procedures, especially in scenarios such as obstetric emergencies and primary hospitals, where medical staff need to quickly complete neonatal risk assessments and have higher requirements for model convenience. This study selected five core risk factors to construct a simplified RF model, which not only retained the core predictive efficacy of the full model but also significantly reduced the clinical application threshold. The variable selection for the simplified model followed the principles of strong clinical relevance, easy acquisition, and high predictive contribution. The selected five factors, including neonatal hypothermia and maternal insulin treatment during pregnancy, are not only the top 5 indicators in terms of variable importance in the original model but also routine clinical monitoring indicators. Hypothermia can be obtained through immediate temperature measurement after birth, and the history of insulin treatment and hypertension during pregnancy can be quickly extracted from the mother's prenatal medical records. Large for gestational age and fetal distress can be confirmed through obstetric assessment at delivery, without the need for additional tests, meeting the clinical promotion requirements of no additional medical burden. The performance validation results showed that the AUC of the simplified model only decreased by 0.023 compared with the full model, with a performance retention rate of 97.34%, and the net benefit in the core clinical decision-making interval was highly consistent with the full model. This result indicates that reducing non-core variables did not significantly reduce the model's predictive accuracy, and its ability to distinguish high-risk and low-risk neonates remains at an excellent level. The clinical assessment time for the simplified model was reduced from 5-8 minutes of the full model to 2-3 minutes, greatly improving the operational efficiency of medical staff, especially suitable for scenarios requiring quick decisions during peak delivery times and multiple pregnancies.

From a health economics perspective, the simplified model does not rely on complex computational software or professional statistical knowledge and can be quickly implemented through paper questionnaires or mobile applications, facilitating its promotion in primary hospitals and maternal and child health institutions with relatively limited medical resources. This helps to promote the homogenization of NH risk assessment, enabling more neonates to receive timely targeted interventions and further reducing the incidence of NH-related adverse outcomes. It should be noted that the

core advantage of the simplified model is its convenience, and its performance is slightly lower than that of the full model. Therefore, in scenarios such as neonatal intensive care unit (NICU) and tertiary hospitals where predictive accuracy is extremely important, the full model is still recommended. In primary screening and rapid assessment scenarios, the simplified model can be the preferred tool to achieve precise and convenient stratified application. Future research can further conduct multicenter prospective studies to verify the actual application effects of the simplified model in different medical scenarios and provide more sufficient evidence for its widespread promotion.

Conclusion

Based on a multicenter retrospective cohort study, the RF model was identified as the optimal predictive model through systematic risk factor screening, model construction, and internal and external validation. The model has good predictive performance and generalization ability, and can provide a practical tool for the early identification of NH risk in newborns of mothers with GDM in clinical practice. Hypothermia, insulin treatment during pregnancy, and large for gestational age are the core factors affecting this risk, providing a key direction for clinical intervention. It should be noted that, as a retrospective study, some data may be subject to recall bias or incomplete records; the study data were derived from five hospitals in Tianjin, and the geographical representativeness of the sample is limited, which may affect the applicability of the model in other regions. Moreover, the development of a simplified RF model may introduce potential optimism bias. In this study, the simplified model was constructed by selecting core factors with the highest importance rankings, which ensured retention of predictive performance. However, this selection process was based on the variable importance ranking within the current dataset, potentially leading to overfitting to the characteristics of the present sample. On one hand, several excluded factors might carry additional predictive information in other populations or clinical settings; their omission could result in decreased model performance when applied to heterogeneous external cohorts. On the other hand, the simplification process did not involve pre-validation of the variable selection strategy using an independent external dataset, which may have caused the predictive value of the core factors to be overestimated due to incidental features of the current sample, leading the model to slightly overpredict the risk of neonatal hypoglycemia in real-world clinical application. Future studies should validate the performance of the simplified model through multicenter prospective cohorts across institutions with varying geographic locations and levels of care, thereby mitigating the impact of optimism bias. It is recommended that future prospective studies be conducted to further validate the model's efficacy, expand the geographical scope of the sample sources to enhance the model's universality, and continuously optimize the model by incorporating more dynamic monitoring indicators. Exploring the transformation of the predictive model into a clinically practical tool, such as developing a risk assessment scale or an intelligent prediction system, will provide stronger technical support for the precise prevention and treatment of NH in newborns of mothers with GDM.

Acknowledgements

This work was funded by Tianjin Municipal Health and Wellness Science and Technology Project (TJWJ2023QN047) and Tianjin Key Medical Discipline Construction Project (Grant No. TJYXZDXK-3-021C).

Disclosure of conflict of interest

None.

Address correspondence to: Zhijie Geng, Department of Obstetrics, Tianjin Third Central Hospital, No. 83, Jintang Road, Hedong District, Tianjin 300170, China. Tel: +86-022-84112114; E-mail: nq700129@126.com

References

[1] Zhang M, Huang X, Lin S and Liu B. Association between maternal blood lipids and neonatal hypoglycaemia in pregnancy with gestational diabetes mellitus: a cohort study. Lipids Health Dis 2024; 23: 170.

[2] Zhang X, Rehemutula R, Jin H, Teng Y, Ma J, Mei S, Long Y, Zhao X, Zeng F, Huang Y, Liu F,

Gao X and Zhu C. Risk factors for hypoglycemia among neonates: a prospective cohort study among pregnant people with gestational diabetes mellitus. J Perinat Neonatal Nurs 2024; 38: 65-72.
[3] Zhang Z, Feng Y, Zhang Y, Li X, Li Y, Sun L, Li X, Du H and Zhang J. Predicting neonatal hypoglycemia using AI neural networks in infants from mothers with gestational diabetes mellitus. Med Sci Monit 2024; 30: e944513.
[4] Ward A, Farengo-Clark D, McKenna DB, Safonov A, Good M, Le A, Kessler L, Shah PD, Bradbury AR, Domchek SM, Nathanson KL, Powers J and Maxwell KN. Clinical management of TP53 mosaic variants found on germline genetic testing. Cancer Genet 2024; 284-285: 43-47.
[5] Lee J, Lee NK and Moon JH. Gestational diabetes mellitus: mechanisms underlying maternal and fetal complications. Endocrinol Metab (Seoul) 2025; 40: 10-25.
[6] Liu Y, Wang M and Zhou J. Impact of gestational diabetes mellitus on neonatal birth outcomes. Br J Hosp Med (Lond) 2024; 85: 1-12.
[7] Dinu D, Hagan JL and Rozance PJ. Variability in diagnosis and management of hypoglycemia in neonatal intensive care unit. Am J Perinatol 2024; 41: 1990-1998.
[8] Liu H and Tang T. Pan-cancer genetic analysis of disulfidptosis-related gene set. Cancer Genet 2023; 278-279: 91-103.
[9] Tessema SA, Assele DD, Kebede HB and Mohammed YE. Incidence and risk factors of neonatal hypoglycemia at Hawassa University Comprehensive Specialized Hospital. PLoS One 2025; 20: e0332495.
[10] Kebede SD, Aytenew TM, Agmas K, Asferie WN and Moges N. Incidence of prolonged transitional neonatal hypoglycemia and associated factors among neonatal admissions in South Gondar public hospitals, North-Central Ethiopia: a prospective cross-sectional study. Front Pediatr 2024; 12: 1381867.
[11] He Y, Liu Y, Peng Y, Xu Y, Yan J and Yang H. Association of duration of intrapartum maternal hyperglycemia in well-controlled gestational diabetes mellitus women and risks of neonatal hypoglycemia: a retrospective cohort study. J Matern Fetal Neonatal Med 2025; 38: 2431617.
[12] Cao Y, Yang Y, Chen Y, Luan M, Hu Y, Zhang L, Zhan W and Zhou W. Optimizing thyroid AUS nodules malignancy prediction: a comprehensive study of logistic regression and machine learning models. Front Endocrinol (Lausanne) 2024; 15: 1366687.
[13] Lee D, Lee C, Han K, Goo T, Kim B, Han Y, Kwon W, Lee S, Jang JY and Park T. Machine learning models for pancreatic cancer diagnosis based on microbiome markers from serum extracellular vesicles. Sci Rep 2025; 15: 10995.
[14] Rojas-Carabali W, Cifuentes-González C, Utami A, Agarwal M, Kempen JH, Tsui E, Finger RP, Sen A, Chan A, Schlaen A, Gupta V, de-la-Torre A, Lee B and Agrawal R; PROTON Study Group. Machine learning for predicting recurrent course in uveitis using baseline clinical characteristics. Invest Ophthalmol Vis Sci 2025; 66: 67.
[15] Maternal-Fetal Medicine Committee, Chinese Society of Obstetrics and Gynecology, Chinese Medical Association; Chinese Society of Perinatal Medicine, Chinese Medical Association; Professional Committee of Gestational Diabetes Mellitus, Chinese Maternal and Child Health Association; Wang C, Juan J and Yang H. A summary of Chinese guidelines on diagnosis and management of hyperglycemia in pregnancy (2022). Matern Fetal Med 2023; 5: 4-8.
[16] Arimitsu T, Kasuga Y, Ikenoue S, Saisho Y, Hida M, Yoshino J, Itoh H, Tanaka M and Ochiai D. Risk factors of neonatal hypoglycemia in neonates born to mothers with gestational diabetes. Endocr J 2023; 70: 511-517.
[17] Group of Neonatology; Pediatric Society; Chinese Medical Association. Expert consensus on standard clinical management of neonatal hypoglycemia in China (2021). Zhongguo Dang Dai Er Ke Za Zhi 2022; 24: 1-13.
[18] Hoermann H, Roeper M, Friesl L, Kurz C, Tautz J, Dzietko M, Mayatepek E, Meissner T and Kummer S. Association of neonatal hypothermia with neonatal hypoglycemia. Front Endocrinol (Lausanne) 2025; 16: 1641140.
[19] Abd El Tawwab AE, Mahmoud MS, El-Kady HA, Elashiry AA, Elsayed HG and Seifelnasr MF. Changing paradigms: neonatal outcomes after oral metformin versus insulin in the control of gestational diabetes mellitus (Randomized controlled trial). Clin Ter 2025; 176: 148-149.
[20] Gupta SS, Gupta SS, Chawla R, Gupta KS, Bamrah PR and Gokalani RA. Gestational diabetes mellitus - neonatal and maternal outcomes in women treated with insulin or diet: a propensity matched analysis. Diabetes Metab Syndr 2024; 18: 103145.
[21] Ou L, Zhu Z, Hao Y, Li Q, Liu H, Chen Q, Peng C, Zhang C, Zou Y, Jia J, Li H, Wang Y, Su B, Lai Y, Chen M, Chen H, Feng Z, Zhang G and Yao M. 1,3,6-Trigalloylglucose: a novel potent anti-helicobacter pylori adhesion agent derived from aqueous extracts of Terminalia chebula Retz. Molecules 2024; 29: 1161.

[22] Loe A, Murray S and Wu Z. Random forest for dynamic risk prediction of recurrent events: a pseudo-observation approach. Biostatistics 2024; 26: kxaf007.
[23] Pantanowitz L, Pearce T, Abukhiran I, Hanna M, Wheeler S, Soong TR, Tafti AP, Pantanowitz J, Lu MY, Mahmood F, Gu Q and Rashidi HH. Nongenerative artificial intelligence in medicine: advancements and applications in supervised and unsupervised machine learning. Mod Pathol 2025; 38: 100680.
[24] Cooray SD, Boyle JA, Soldatos G, Allotey J, Wang H, Fernandez-Felix BM, Zamora J, Thangaratinam S and Teede HJ. Development, validation and clinical utility of a risk prediction model for adverse pregnancy outcomes in women with gestational diabetes: the PeRSonal GDM model. EClinicalMedicine 2022; 52: 101637.