Supplementary information

Prediction of birth weight in pregnancy with gestational diabetes mellitus using an artificial neural network

Menglin ZHOU[1], Jiansheng JI[1,2], Ni XIE[1,3], Danqing CHEN[1]

[1]*Department of Obstetrics, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310006, China*
[2]*Department of Gynaecology and Obstetrics, People's Hospital of Quzhou, Quzhou 324000, China*
[3]*Community Healthcare Center of Caihe Street, Hangzhou 310020, China*

Materials and methods

Subjects and related data extraction

Subjects were 2462 pregnant women with GDM (single live births) who underwent routine antenatal visits and delivery in Women's Hospital of Zhejiang University School of Medicine from January, 2019 to August, 2020. From medical record database, certain data of these women and their neonate were collected, including maternal age, height, educational level, gravidity, parity, pre-pregnancy weight, gestational weight gain (GWG), conception method, fundal height, maternal abdominal circumference, gestational age at birth, results of oral glucose tolerance test (OGTT) during the 24–28th week in gestation, common complications status (including hypertensive diseases, intrahepatic cholestasis of pregnancy, hypothyroidism diseases and hepatitis B virus carrying status), interval from last fetal B-ultrasound scan to birth, fetal biparietal diameter (fBPD), fetal femur length (fFL), fetal abdominal circumference (fAC), fetal head circumference (fHC), amniotic fluid volume and fetal presentation.

Prenatal ultrasound examinations

Women undergoing routine prenatal visits in our hospital received a two-dimension transabdominal sonography examination every 4–6 weeks, and it would be more frequent when it was close to the expected date of confinement. And a few women in labor had no time to complete an antepartum sonography scan, so that the interval from their last sonography to birth might be as long as 4 weeks.

The Antenatal ultrasound examinations were performed with 2-dimentional transabdominal sonography, and the method of measurement was the same as the previous study (Cesnaite et al., 2020): the fBPD was measured on the transverse plane of the fetal head at the level of the thalami and cavum septum pellucidum, the markers were placed on the contour of the head in the widest part of the contour, perpendicular to the center line, with superior marker placed over the bone overline and the lower marker placed not reaching the bone path; the fHC was measured in the same plane as the fBPD, the markers were placed on the cranial bone; the fAC was measured on the axial plane of the upper abdomen, with the marker placed on the outer contour of the fetal skin; the fFL was measured on the longest diaphyseal axis, with the markers placed on the ends of diaphysis.

Baseline characteristics of GDM patients and their fetuses

Among 2462 subjects, 80% subjects (1970) were randomly selected as training dataset, and the other 20% as testing dataset. Between the two datasets, there were no significant differences in maternal demographic and clinical characteristics(including age, educational level, pre-pregnancy weight, gravidity, parity, conception mode, gestational weight gain, gestational age at delivery, fundal height, abdominal circumference and OGTT results), maternal common complications status (hypertensive diseases, intrahepatic cholestasis of pregnancy, hypothyroidism diseases and hepatitis B virus carrying status), fetal ultrasonic data (including examination time interval to delivery, fBPD, fFL, fAC, fHC, amniotic fluid volume and cephalic presentation) and birth weight, which suggested the plausibility of the artificial neural network (ANN) derived from training dataset being tested in our testing dataset (Table 1).

Table 1 Clinical and general characteristic of subjects in training dataset and testing dataset

Characteristics	**Training group (*n*=1970)**	**Testing group (*n*=492)**	***P***
Maternal demographic and clinical characteristics			
Age (years)	32 (29, 35)	32 (29, 35)	0.45
Educational level			0.783
Lower than college (%)	14.3	13.6	
College (%)	70.1	71.7	
Higher than college (%)	15.6	14.7	
Height (cm)	160.2±5.0	160.1±5.2	0.482
Pre-pregnancy weight (kg)	55.4±8.4	55.2±8.3	0.551
Gravidity ≤2 (%)	67.9	67.4	0.834
Primipara (%)	57.1	59.3	0.386
Assisted conception (%)	5.6	6.9	0.265
Gestational weight gain (kg)	12 (10, 15)	12 (10, 15)	0.478
Gestational age (d)	275 (269, 281)	275 (269, 280)	0.891
Fundal height (cm)	34.0±2.3	33.9±2.2	0.386
Abdominal circumference (cm)	98.5±6.2	98.6±6.2	0.497
OGTT results in 24-28 weeks			
FBG (mmol/L)	4.5 (4.3, 4.9)	4.6 (4.3, 4.9)	0.187
1h-BG	10.2 (9.3, 10.8)	10.2 (9.2, 10.9)	0.942
2h-BG	8.9 (8.3, 9.6)	8.9 (8.1, 9.6)	0.598
Maternal complications			
Hypertensive diseases (%)	8.1	7.3	0.574
Hypothyroidism diseases (%)	6.1	6.5	0.707
Hepatitis B virus carrier (%)	4.7	6.1	0.213
ICP (%)	3.5	3.3	0.823
Fetal sonography data			
Time interval to delivery (d)	3 (2, 7)	3 (2, 8)	0.79
Biparietal diameter (cm)	9.3 (9, 9.5)	9.3 (9, 9.5)	0.947
Femur length (cm)	7.1 (6.9, 7.3)	7.1 (6.9, 7.3)	0.867
Head circumference (cm)	33 (32.3, 33.5)	32.9 (32.3, 33.4)	0.335
Abdominal circumference (cm)	34.4 (33.3, 35.3)	34.5 (33.4, 35.3)	0.681
Amniotic fluid volume$^{\eta}$			0.478
Level I (%)	7.8	7.5	
Level II (%)	83.6	85.5	
Level III (%)	8.6	6.9	
Cephalic presentation (%)	95.4	94.9	0.67
Birth weight (g)	3262.9±498.1	3258.6±475.7	0.865

$^{\eta}$The amniotic fluid volume was graded into three levels: Level I, amniotic fluid index <8 cm or maximum vertical pocket depth <3 cm; Level III, amniotic fluid index >18 cm or maximum vertical pocket depth >5 cm; Level II, the other conditions.
OGTT, oral glucose tolerance test; BG, blood glucose; FBG, fasting blood glucose; ICP, intrahepatic cholestasis of pregnancy.

Data pre-processing

There existed much difference in the magnitudes and units among selected features, resulting in possible imbalanced weights assigned during the training process. A normalization method was performed to reduce the possible influence, converting the eigenvalues of the sample to the same magnitude [0,1]:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where x is a specific value of a certain variable, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the certain variable among all samples, respectively, and $\tilde{x}$ is the final normalized value.

Establishment of the ANN model

By comprehensive consideration of obstetrical views, Pearson's correlation and distance correlation, features considered more likely to determine fetal weight were selected to participate in the establishment of the neural network model. Traditionally, eighty percent of the sample were randomly selected as the training group, and the other twenty percent as the testing group. With training dataset, a two-layer feed-forward ANN with tanh hidden neurons was established (Fig. 1).

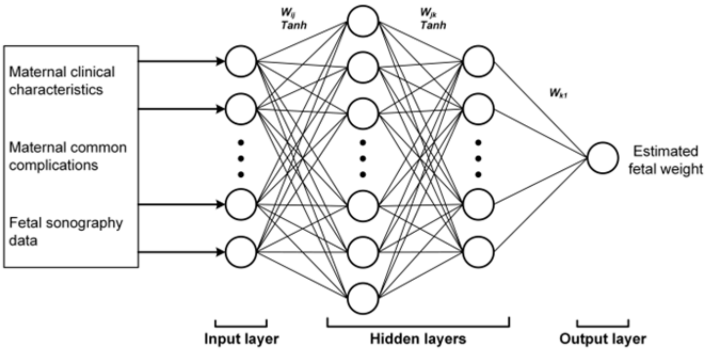

Fig. 1 A two-layer feed forward artificial neural network for estimating fetal weight. Input layer included features of maternal clinical characteristics, maternal common complications and fetal sonography data; hidden layers were calculated with tanh algorithm from last layers, and output layer without tanh transformation.

Training of the ANN model

Using the method of 10-fold cross validation, key hyper-parameters were determined. In the final ANN, imputing features were parity (primipara or multipara), height, pre-pregnancy weight, gestational weight gain, fasting blood glucose level in OGTT, gestational age at delivery, fundal height, maternal abdominal circumference, hypertensive diseases status, ultrasonic examination time interval to delivery, fBPD, fFL, fAC, fHC and amniotic fluid volume. These features of training dataset were imputed into a two-layer feed-forward ANN with tanh hidden neurons, trained with adaptive moment estimation back propagation algorithm; with optimization of hyper-parameters in the process of 10-fold cross validation, loss in each cross-validating dataset descended to the lowest level (about 150 g) at the epoch of about 7800 (Fig. 2).

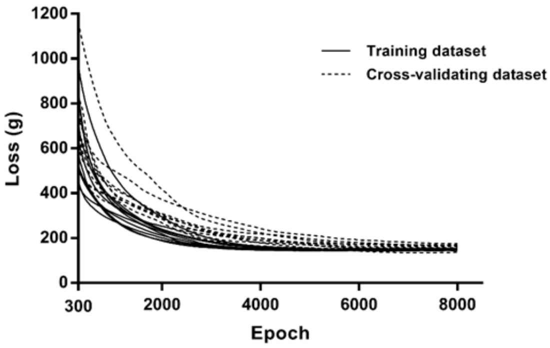

Fig. 2 Ten-fold cross validation results of training artificial neural network. The total training dataset were randomly and evenly divided into ten subsets, nine of which was used to train and the other one for cross-validation, for ten times by turns. Loss in each cross-validating dataset with optimal hyper-parameters descended to the lowest level at the epoch of about 7800.

Performance evaluation

The performance of proposed model was mainly evaluated by mean absolute error (MAE) and mean absolute percentage error (MAPE), calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \qquad (2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \times 100\% . \qquad (3)$$

And the quality of the model was also evaluated by some common accuracy-evaluating percentages, including the percentages of an absolute error between the actual and estimated fetal weight <250 g and <100 g, respectively, and the percentages of the ratio of the absolute error to actual fetal weight <10% and <5%, respectively.

Statistical analysis

To test differences of the features between training dataset and testing dataset, for measurement data in normal distribution or skewed distribution, *Student's* T test or *Mann-Whitney* U test was used, respectively; for enumeration data, Chi-square test was used. The paired T test was used for testing differences in the absolute errors and the absolute percent errors between the proposed model and other formula estimates. For the differences in proportions of absolute errors<100 g and<200 g and absolute percent errors<10% and<5%, we used McNemar test.

All the statistical analyses were performed using the software programs of IBM SPSS (ver. 22.0; IBM Corp., Armonk, NY, USA) and STATA (version 14; Stata Corp., College Station, TX, USA). The statistical significance was set at an α level of 0.05 with a two-sided test, and Bofferoni correction was used in the comparisons involved in multiple tests.

Assessment of the ANN estimation in different birth weight interval by certain cut-off points

First, a comparison of estimated error was performed in two different birth weight intervals, with 3500 g as the cut-off point. Consistent with the smoothed curves, the absolute estimated error of higher actual weight (>3500 g) was much higher than that of lower actual weight (<3500 g) (201.0±160.1 g vs. 133.3±137.5 g, *P*<0.001); and the real error of higher actual weight was much lower (−124.7±225.4 g vs. 39.3±187.5 g, *P*<0.001). When 2500 g and 4000 g were set as cut-off points, that was, the weight data was divided into low birth weight, normal birth weight and high birth weight, the high birth weight was greatly underestimated (MAE: 304.3±161.3 g; mean real error: −277.8±205.5 g; *P*<0.001 for both when compared with zero), while there were no significant differences of absolute estimated error and real estimated error between low birth weight and normal birth weight (*P*>0.05 for both) (Table S1).

Reference

Cesnaite, G., Domza, G., Ramasauskaite, D., Volochovic, J., 2020. The accuracy of 22 fetal weight estimation formulas in diabetic pregnancies. *Fetal Diagn Ther*, **47**(1):54-59. https://dx.doi.org/10.1159/000500452

Table S1 Estimated error in different weight intervals

	Weight interval	MAE (g)	MRE (g)
With 2500 g and 4000 g as the cut-off points	<2500 g	110.2±113.1	−1.7±159.2
	2500–4000	149.3±145.5	1.9±208.6
	≥4000 g	304.3±161.3***	−277.8±205.5***
With 3500 g as the cut-off point	<3500 g	133.3±137.5	39.3±187.5
	≤3500 g	201.0±160.1$^{\#\#\#}$	−124.7±225.4$^{\#\#\#}$
All dataset	NA	153.5±147.9	−9.7±213.6

MAE, mean absolute error; MRE, mean real error (estimated weight – actual weight). ****P*<0.001 for the comparison with the interval of birth weight ≥2500 and<4000 g; $^{\#\#\#}$*P*<0.001 for the comparison with the interval of birth weight<3500 g.