Similarity Report

Original Document:

Ranjbar
et al. BMC Pregnancy and Childbirth (2023) 23:803
https://doi.org/10.1186/s12884-023-06128-w
RESEARCH

Machine learning-based approach
for predicting low birth weight

Amene Ranjbar
1
, Farideh Montazeri
2
, Mohammadsadegh Vahidi Farashah
3
, Vahid Mehrnoush
2
,
Fatemeh Darsareh
2*
 and Nasibeh Roozbeh
2

Abstract
Background Low birth weight (LBW) has been linked to infant mortality. Predicting LBW is a valuable preventative
tool and predictor of newborn health risks. The current study employed a machine learning model to predict LBW.
Methods This study implemented predictive LBW models based on the data obtained from the “Iranian Mater-
nal and Neonatal Network (IMaN Net)” from January 2020 to January 2022. Women with singleton pregnancies above the gestational age of 24 weeks were included. Exclusion criteria included multiple pregnancies and fetal anomalies. A predictive model was built using eight statistical learning models (logistic regression, decision tree classification, random forest classification, deep learning feedforward, extreme gradient boost model, light gradi-
ent boost model, support vector machine, and permutation feature classification with k-nearest neighbors). Expert opinion and prior observational cohorts were used to select candidate LBW predictors for all models. The area under the receiver operating characteristic curve (AUROC), accuracy, precision, recall, and F1 score were measured to evaluate their diagnostic performance.
Results We found 1280 women with a recorded LBW out of 8853 deliveries, for a frequency of 14.5%. Deep learning
(AUROC: 0.86), random forest classification (AUROC: 0.79), and extreme gradient boost classification (AUROC: 0.79) all
have higher AUROC and perform better than others. When the other performance parameters of the models men-
tioned above with higher AUROC were compared, the extreme gradient boost model was the best model to predict LBW with an accuracy of 0.79, precision of 0.87, recall of 0.69, and F1 score of 0.77. According to the feature impor-
tance rank, gestational age and prior history of LBW were the top critical predictors.
Conclusions Although this study found that the extreme gradient boost model performed well in predicting LBW, more research is needed to make a better conclusion on the performance of ML models in predicting LBW.

Keywords Low birth weight, Fetal weight, Birth weight, Machine learning, X gradient boost model

Background

Birth weights less than 2500 g are called low birth weight (LBW). LBW has been linked to infant mortality and its consequences [1]. Predicting LBW is thus a valuable pre
-

ventative tool and predictor of newborn health risks. Previous research has found that maternal demographics, preexisting health conditions, and prenatal care level are all linked to LBW [2, 3]. Thus, pinpointing which pregnant patients are most likely to have a baby with LBW during the preconception or early pregnancy stages is

BMC Pregnancy and Childbirth

*Correspondence:
Fatemeh Darsareh
famadarsareh@yahoo.com
[1] Fertility and Infertility Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran
[2] Mother and Child Welfare Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran
[3] Amirkabir University of Technology, Tehran, Iran

critical for saving neonatal lives and reducing potentially avoidable medical costs through direct clinical and health policy interventions. There are some documented studies on using ML in perinatal care and maternal health. Previous LBW prediction studies achieved good performance in predicting LBW; however, all previous studies recommended more studies due to study limitations such as small sample size or limited feature selection [4–7]. In this study, we aimed to evaluate the performance of eight different ML algorithms in predicting LBW.

Methods

The findings of this retrospective cohort study are based on birth records obtained from the "Iranian Maternal and Neonatal Network (IMaN Net)," a legitimate national system, from January 2020 to January 2022. IMaN Net is a comprehensive system for registering maternal and newborn information on the outcomes of each delivery, which is completed daily by midwives in all birth centers and hospitals throughout Iran in an integrated manner. All patients' personal information was deidentified and not disclosed.

Women with singleton pregnancies above the gestational age of 24 weeks who gave birth during a study period were included. The target population in this study was divided into LBW ("d 2499 g) and not LBW ("e 2500 g), which is the national standard definition [8]. Exclusion criteria included multiple pregnancies and fetal anomalies.

A predictive model was built using eight statistical learning models, including logistic regression, decision tree classification, random forest classification, deep learning feedforward, extreme gradient boost classification (XGBoost), light gradient boost (LGB), support vector machine (SVM), and permutation feature classification with k-nearest neighbors (KNN). Expert opinion and prior observational cohorts were used to select candidate LBW predictors for all models [9, 10]. Predictor factors included maternal age, educational level, maternal occupation, place of residence, inadequate prenatal care (less than three prenatal care visits), smoking, drug addiction, maternal anemia, cardiovascular disease, chronic hypertension, hepatitis, COVID-19, overt diabetes, gestational diabetes and thyroid dysfunction, parity, preeclampsia, fetal gender, method of childbirth, previous history of LBW, supplementary and vitamins intake were obtained from patient medical records. We used Chi-square test to evaluate the association between predicting factors mentioned above and LBW. Then we performed ML analysis approach. We followed the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View to report our findings. The programming language Python was chosen to create the machine learning model. Scikit-learn was used to implement the ML algorithm. Scikit-learn is a machine-learning library written in Python. It includes an extensive collection of cutting-edge machine-learning algorithms for both supervised (including the multi-output classification and regression algorithm) and unsupervised problems [11].

Internal validation was carried out with the help of k-fold cross-validation. The cases were randomly assigned to either the "training set" (70%) or the "test set" (30%) using a random number generator. The original dataset kept the rate of LBW and non-LBW groups in the training and test sets constant. Using the training set, we arranged the parameters of the prediction models and evaluated their performance using the “test set”. The average performance was calculated by repeating these ten times.

Metrics, including area under the receiver operating characteristic curve (AUROC), accuracy, precision, recall, and F1 score, were used to assess the predictive power of the models. The accuracy metric calculates how often a model is correctly predicted across the entire dataset. Precision measures how many of the model’s “positive” predictions were correct. The model’s recall estimates how many positive class samples in the dataset were correctly identified. The F1 score combines precision and recall by using their harmonic mean, and maximizing the F1 score implies maximizing both precision and recall simultaneously. As a result, researchers have chosen the F1 score to evaluate their models in conjunction with accuracy. We used AUROC as the primary performance metric because it is a widely used index to describe the ML model’s ability to predict outcomes. The metrics were scaled from 0 to 1, with higher values indicating a better model [12].

Results

Of 8850 eligible cases, we found 1280 women with a recorded LBW, for a frequency of 14.5%. The demographic and clinical characteristics of study population is given in Table 1. As it shown, maternal age, living residency, gestational age, parity, access to prenatal care, maternal anemia, chronic hypertension, preeclampsia, drug addiction, COVID-19, previous LBW, and newborn gender was linked to LBW.

In this study, we attempt to evaluate parameters and feature selection based on performance parameters using various ML algorithms. A plot ROC chart, as shown in Fig. 1, and calculate AUROC as a plot that allows the user to visualize the tradeoff between the classifier’s sensitivity. Deep learning (AUROC: 0.86), random forest classification (AUROC: 0.79), and XGBoost classification

(AUROC: 0.79) all have higher ROC_AUC and perform better than others, as shown in Fig. 1.

Other performance parameters for each algorithm are shown in Table 2. Other performance parameters indicate that the XGBoost classification performs more than all. Random forest classification and deep learning feedforward are also very close. When the accuracy, precision, recall, and F1 score of the models mentioned above with higher AUROC were compared, the XGBoost model was the best model to predict LBW with an accuracy of 0.79, precision of 0.87, recall of 0.69, and F1 score of 0.77. The confusion matrix of the XGBoost classification model is shown in Fig. 2.

Figure 3 presents an analysis of the importance of variables in the XGBoost algorithm. As the feature importance rank was identified, gestational age and previous history of LBW were the top critical predictors.

Discussion

With the exponential growth in the quantity and dimension of healthcare data in recent years, ML approaches for dealing with complex and high-dimensional data have been introduced [13–15]. In this study, we aimed to evaluate the performance of eight different ML algorithms in predicting LBW. According to our findings, the XGBoost classification model had a more significant diagnostic performance parameter with an AUROC of 0.79, accuracy of 0.79, precision of 0.87, recall of 0.69, and F1 score of 0.77. XGBoost classification is a supervised machine learning algorithm based on a distributed gradient-boosted decision tree [16]. It can produce consistent

Table 1 Demographic and clinical factors associated with low birth weight

Outcome	Non-LBW (n = 7570)	LBW (n = 1280)	P-value
Maternal age			< 0.001
13–19	137 (1.8)	36 (2.8)	
20–35	6247 (82.5)	995 (77.7)	
Above 35	1186 (15.7)	249 (19.5)	
Education			0.348
Illiterate	480 (6.3)	73 (5.7)	
Primary	2344 (31.0)	378 (29.5)	
High-school/Diploma	3463 (45.8)	609 (47.6)	
Advanced	1283 (16.9)	220 (17.2)	
Occupation			0.299
Housewife	6798 (89.8)	1146 (89.5)	
Worker/employee	772 (10.2)	134 (10.5)	
Living residency			0.045
Urban	5056 (66.8)	823 (64.3)	
Rural	2517 (33.2)	457 (35.7)	
Gestational age			< 0.001
24–36+6	429 (5.7)	801 (62.6)	
37–41	7141 (94.3)	479 (37.4)	
Parity			< 0.001
Primiparous	2056 (27.1)	439 (34.3)	
Multiparous	5517 (72.9)	841 (65.7)	
Access to prenatal care			0.030
Yes	7343 (97.0)	1255 (98.0)	

No	230 (3.0)	25 (0.2)	
Maternal anemia			0.047
No	7364 (97.2)	1233 (96.3)	
Yes	209 (2.8)	47 (3.7)	
Chronic hypertension			0.005
No	7501 (99.0)	1256 (98.1)	
Yes	72 (1.0)	24 (1.9)	
Cardiovascular disease			0.803
No	7492 (98.9)	1267 (99.0)	
Yes	81 (1.1)	13 (1.0)	
Diabetes			0.276
No	6420 (84.8)	1094 (85.5)	
Yes	1153 (15.2)	186 (14.5)	
Preeclampsia			< 0.001
No	7196 (95.0)	1083 (84.6)	
Yes	377 (5.0)	197 (15.4)	
Drug addiction			< 0.001
No	7530 (99.4)	1251 (97.7)	
Yes	42 (0.6)	29 (2.3)	
Previous low birth weight			< 0.001
No	7479 (98.8)	1089 (85.1)	
Yes	94 (1.2)	191 (14.9)	

Data are presented as n (%)

Table 1 (continued)

Outcome	Non-LBW (n = 7570)	LBW (n = 1280)	P-value
COVID-19			0.020
No	7465 (98.6)	1250 (97.7)	
Yes	108 (1.4)	30 (2.3)	
Thyroid dysfunction			0.999
No	6778 (89.5)	1146 (89.5)	
Yes	795 (10.5)	134 (10.5)	
Hepatitis			0.079
No	7543 (99.6)	1279 (99.1)	
Yes	30 (0.4)	1 (0.1)	
Newborn gender			< 0.001
Male	3942 (52.1)	599 (46.8)	
Female	3631 (47.9)	681 (53.2)	
Supplementary intake			0.078
No	4 (0.1)	5 (0.4)	
Yes	7569 (99.9)	1275 (99.6)	

results while minimizing overfitting by employing a parallel tree-boosting strategy. Furthermore, XGBoost can use the importance score to determine the importance of each feature. Previous studies evaluating different ML machines for predicting LBW will also have promising results. According to Ahmadi et al., the random forest model performed well in terms of diagnostic performance, with an accuracy of 0.95, recall of 0.72, and AUROC of 0.89 [5]. Another study by Desiani et al. found that naive Bayes had excellent diagnostic performance, with an accuracy of 0.85 and a recall of 0.72 [17]. However, both studies were limited by a small sample size (less than 1000 participants).

Recent studies with larger sample sizes also demonstrated good performance. For example, in a survey by Eliyati et al., with a sample size of 12,500 study participants, SVM showed high diagnostic performance in predicting LBW with an accuracy of 0.93 [18]. Ren et al. used a more extensive study in this field, with a sample size of 266,687 birth records over six years. According to their findings, the XGBoost classification model had the highest recall score of 0.85, but the AUROC score was only 0.61 [19].

Although our study did not have the largest sample size of any study in this field, we believe that using hospital records made our feature selection rich enough to make a reasonable conclusion on identifying LBW risk factors. In our study, we surveyed maternal age, educational level, place of residence, inadequate prenatal care (fewer than three prenatal care visits), drug addiction, maternal anemia, cardiovascular disease, chronic hypertension, pyelonephritis, hepatitis, COVID-19, overt diabetes, gestational diabetes and thyroid dysfunction, parity, preeclampsia, and history of LBW. Among all the potential predisposing factors of LBW, gestational age and previous history of LBW were the top critical predictors. In line with previous findings [20, 21], gestational age is the highest predictor of LBW. Being born too soon (premature birth) is the most common cause of LBW. The prior history of LBW was another weighted factor in predicting LBW. It has been reported that the risk of LBW recurs between pregnancies. Women with a

Fig. 1 AUROC of ML models

Table 2 Performance parameters of models with the highest AUROC

Row	Algorithms	Accuracy	Precision	Recall	F_1 Score
1	Random Forest Classification	0.78	0.85	0.70	0.77
2	XGBoost Classification	0.79	0.87	0.69	0.77
3	Deep Learning- Feed Forward	0.78	0.84	0.70	0.76

Fig. 2 XGBoost classification confusion matrix

Fig. 3 Feature importance of the XGBoost classification in the prediction of low birth weight previous LBW baby have been identified as potential carriers of the recurrent risk and have a higher recurrence risk of LBW in their subsequent pregnancy than those with a previous normal birth weight baby [22]. Other factors, such as maternal comorbidities, sociodemographic characteristics, and fetal gender, were not among the weighted factors in predicting LBW, in contrast to previous studies. In one study, Bekele et al. found that fetal gender, marriage to birth interval, mother's occupation, and mother's age were all weighted factors in predicting LBW [23]. Another study found that maternal pre-pregnancy weight, mother's age, number of doctor visits during the first trimester, and previous preterm labor were the most significant risk factors for LBW [4]. The differences in findings could be attributed to study design, the type of ML models used, geographical differences, or imbalances in each study's dataset. It should be noted, however, that clinicians can use the key findings of each study to identify high-risk pregnant patients early in their pregnancy using early prenatal care screening tools.

Although we used a large dataset with a lot of maternal and neonatal information, a significant variable, like body mass index, was missing in most of the birth records, so we couldn't use this factor in our selection features, which is a significant limitation of the study.

Conclusion

Using ML approaches to predict LBW yielded promising results. As a result, this study might add to the current perinatal care literature. Although this study found that the XGBoost model performed well in predicting LBW, more research is needed to make a better conclusion on the performance of ML models in predicting LBW.

Abbreviations

LBW Low birth weight
ML Machine learning
XGBoost Extreme gradient boost
LGB Light gradient boost
SVM Support vector machine
KNN K-nearest neighbors

Acknowledgements

All of the authors acknowledged Hormozgan University of Medical Sciences.

Authors' contributions

F.D. and N.R. wrote the proposal. F.M. and V.M. contributed significantly to data collection. The findings were analyzed and interpreted by M.F. F.D., the primary contributor to the manuscript's commenting and editing. A.R. assessed the manuscript's scientific content critically. The final manuscript for submission was read and approved by all authors.

Funding

None.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
This study complied with the Declaration of Helsinki and was performed according to ethics committee approval. The Ethics and Research Committee of the Hormozgan University of Medical Sciences approved the study. The records of all patients who provided informed consent for using their data for research purposes were analyzed. For those under the age of 18 the informed consent was taken from their gurdian. Statistical analysis was performed with patient anonymity following ethics committee regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 August 2023 Accepted: 14 November 2023

References

1. Chen Y, Li G, Ruan Y, Zou L, Wang X, Zhang W. An epidemiological survey on low birth weight infants in China and analysis of outcomes of full-term low birth weight infants. BMC Pregnancy Childbirth. 2013;13:242. https:// doi. org/ 10. 1186/ 1471- 2393- 13- 242. PMID:24370213;PMCID: PMC3877972.
2. Heaman MI, Martens PJ, Brownell MD, Chartier MJ, Derksen SA, Helewa ME. The Association of inadequate and intensive prenatal care with maternal, fetal, and infant outcomes: a population-based study in Manitoba, Canada. J Obstet Gynaecol Can. 2019;41(7):947–59. https:// doi. org/ 10. 1016/j. jogc. 2018. 09. 006. Epub 2019 Jan 11 PMID: 30639165.
3. Cunningham FG, Leveno KJ, Bloom SL, Hauth JC, Rouse DJ, Spong CY. Williams obstetrics. New York: McGraw-Hill; 2010. p. 804- 831. 23.
4. Senthilkumar D, Paulraj S. Prediction of low birth weight infants and its risk factors using data mining techniques. Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management; IMEOM ’15. Dubai: IEOM Society; 2015.
5. Ahmadi P, Alavimajd H, Khodakarim S, Tapak L, Kariman N, Amini P, Pazhuheian F. Prediction of low birth weight using random forest: a comparison with logistic regression. Arch Adv Biosci. 2017;8(3):36–43. https:// doi. org/ 10. 22037/ jps. v8i3. 15412.
6. Borson N, Kabir M, Zamal Z, Rahman R. Correlation analysis of demographic factors on low birth weight and prediction modeling using machine learning techniques. Proceedings of the 4th World Conference on Smart Trends in Systems, Security and Sustainability; WorldS4 ’20. London: Institute of Electrical and Electronics Engineers; 2020. p. 169–73.
7. Faruk A, Cahyono ES. Prediction and classification of low birth weight data using machine learning techniques. Indones J Sci Technol. 2018;3(1):18–28. https:// doi. org/ 10. 17509/ ijost. v3i1. 10799.
8. International statistical classification of diseases and related health problems, 10th revision. World Health Organization; 2004. Availabe at: https:// apps. who. int/ iris/ bitst ream/ handle/ 10665/ 42980/ 92415 46530_ eng. pdf? seque nce= 1& isAll owed=y.
9. Schimmel MS, Bromiker R, Hammerman C, Chertman L, Ioscovich A, Granovsky-Grisaru S, Samueloff A, Elstein D. The effects of maternal age and parity on maternal and neonatal outcome. Arch Gynecol Obstet. 2015;291(4):793–8. https:// doi. org/ 10. 1007/ s00404- 014- 3469-0.
10 Sharifi N, Dolatian M, FathNezhadKazemi A, Pakzad R, Yadegari L. The relationship of the structural and intermediate social determinants of health with low birth weight in Iran: a systematic review and meta-analysis. Sci J Kurdistan Univ Medical Sci. 2018;23(2):21–36. https:// doi. org/ 10. 29252/ sjku. 23.2. 21.
11. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: machine

learning in Python. J Mach Learn Res. 2011;12:2825–30. https:// doi. org/
10. 1145/ 27869 84. 27869 95.

12. Yen SJ, Lee YS. Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. Proceedings of the 2016 International Conference on Intelligent Computing; ICIC ’06. Kunming: Springer; 2006. p. 731–40.
13. Boujarzadeh B, Ranjbar A, Banihashemi F, Mehrnoush V, Darsareh F, Saffari M. Machine learning approach to predict postpartum haemorrhage: a systematic review protocol. BMJ Open. 2023;13(1):e067661. https:// doi. org/ 10. 1136/ bmjop en- 2022- 067661. PMID:36657750;PMCID: PMC9853215.
14. Mehrnoush V, Ranjbar A, Farashah MV, Darsareh F, Shekari M, Jahromi MS. Prediction of postpartum hemorrhage using traditional statistical analysis and a machine learning approach. AJOG Glob Rep. 2023;3(2):100185. https:// doi. org/ 10. 1016/j. xagr. 2023. 100185. PMID:36935935;PMCID: PMC10020099.
15. Darsareh F, Ranjbar A, Farashah MV, Mehrnoush V, Shekari M, Jahromi MS. Application of machine learning to identify risk factors of birth asphyxia. BMC Pregnancy Childbirth. 2023;23(1):156. https:// doi. org/ 10. 1186/ s12884- 023- 05486-9. PMID:36890453;PMCID:PMC9993370.
16. Chen T, He T. xgboost: eXtreme gradient boosting. The Comprehensive R Archive Network; 2017. https:// cran. micro soft. com/ snaps hot/ 2017- 12- 11/ web/ packa ges/ xgboo st/ vigne ttes/ xgboo st. pdf.
17. Desiani A, Primartha R, Arhami M, Orsalan O. Naive Bayes classifier for infant weight prediction of hypertension mother. J Phys Conf Ser. 2019;1282(1):012005. https:// doi. org/ 10. 1088/ 1742- 6596/ 1282/1/ 012005.
18. Eliyati N, Faruk A, Kresnawati ES, Arifieni I. Support vector machines for classification of low birth weight in Indonesia. J Phys Conf Ser. 2019;1282(1):012010. https:// doi. org/ 10. 1088/ 1742- 6596/ 1282/1/ 012010.

19. Ren Y, Wu D, Tong Y, López-DeFede A, Gareau S. Issue of data imbalance on low birthweight baby outcomes prediction and associated risk factors identification: establishment of benchmarking key machine learning models with data rebalancing strategies. J Med Internet Res. 2023;25:e44081. https:// doi. org/ 10. 2196/ 44081. PMID:37256674;PMCID: PMC10267797.
20. Loreto P, Peixoto H, Abelha A, Machado J. Predicting low birth weight babies through data mining. Proceedings of the 2019 World Conference on Information Systems and Technologies; WorldCIST '19; March 27–29, 2018. Naples: Springer; 2019. pp. 568–77.
21. Khan W, Zaki N, Masud MM, Ahmad A, Ali L, Ali N, Ahmed LA. Infant birth weight estimation and low birth weight classification in United Arab Emirates using machine learning algorithms. Sci Rep. 2022;12(1):12110. https:// doi. org/ 10. 1038/ s41598- 022- 14393-6.
22. Mvunta MH, Mboya IB, Msuya SE, John B, Obure J, Mahande MJ. Incidence and recurrence risk of low birth weight in Northern Tanzania: a registry based study. PLoS ONE. 2019;14(4):e0215768. https:// doi. org/ 10. 1371/ journ al. pone. 02157 68.
23. Bekele WT. Machine learning algorithms for predicting low birth weight in Ethiopia. BMC Med Inform Decis Mak. 2022;22(1):232. https:// doi. org/ 10. 1186/ s12911- 022- 01981-9.

Compared Document:

 Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Digital Object Identifier 10.1109/ACCESS.2017.Doi Number Fetal Weight Estimation Via Ultrasound Using Machine Learning MIAO FENG 1 , LI WAN 2 , ZHI LI 1 , LINBO QING 1 , XIAORONG QI 2 1 College of Electronics and Information Engineering, Sichuan University, Chengdu, 610065, China 2 Department of Gynecology and Obstetrics, Development and related disease of women and children key laboratory of Sichuan Province, Key Laboratory of Birth Defects and Related Diseases of Women and Children, Ministry of Education , West China Second Hospital, S ichuan University, Chengdu, 610041, P. R. China Corresponding author: Linbo Qing (e-mail: qing_lb@scu.edu.cn) and Xiaorong Qi (e-mail: qixiaorong11@163.com). This work was supported in part by the Key Research and Development Project of Science & Technology Department of Sichuan Province under Grant 2019YFG0192. ABSTRACT Accurate fetal weight estimation is important for both fetuses and their mothers. The *low* birth weight (LBW, birth weight ð< 2 500g) and high birth weight (HBW, birth weight ð³ 4000g) fetuses and their mothers are linked to both short and long-term health outcomes, like high perinatal mortality rate, various complications and chronic disease in life. Because of the imbalanced small data sets and body size heterogeneities between different fetal weight groups, it is difficult for the commonly used regression formulas to get a satisfying performance, especially for HBW and LBW fetuses. The aim of this study is to propose *a* machine learning solution to improve the fetal weight estimation accuracy and to help the clinicians identify potential risks before delivery. A clinical data set of 7875 singleton fetuses were analyzed. The synthetic minority over-sampling technique (SMOTE) was employed to solve the imbalanced learning problem. Then the *support* vector machine (SVM) algorithm was utilized for fetal weight classification. Finally, the deep belief network (DBN) was employed to estimate the fetal weight based on different ultrasound parameters. The estimation result of the proposed model showed mean absolute percent error (MAPE) of 6.09 ð± 5.06% and mean absolute error (MAE) of 198.55 ð± 158.63g. It demonstrated that our model outperformed the commonly-used regression formulas, especially for the HBW and LBW fetuses. INDEX TERMS Deep belief network, fetal weight estimation, synthetic minority over-sampling technique, *support* vector machine, ultrasound. I. INTRODUCTION Fetal weight is an essential factor to predict the short and long- term health consequences [1]. According to birth weight (BW), the neonates are defined by the World Health Organization (WHO) as three groups, namely *low* birth weight (LBW, BW ð< 2500g), *normal* birth weight (NBW, 2500g ð£ BW ð<

4000g) and high birth weight (HBW, BW ð³ 4000 g) which is also called macrosomia [2]. *Low* birth weight is connected with fetal and neonatal mortality and inhibited growth, it can also cause long-term diseases in their childhood, such as mental retardation and learning disabilities [3], [4]. Macrosomia can cause perinatal asphyxia and death, moreover, for maternities, *the* risk of caesarean section, prolonged labour, abnormal haemorrhage, and perineal trauma increases [5], [6]. In the long term, macrosomia is more likely to be associated with obesity, diabetes, and heart disease [4]. Therefore, it is significant to estimate fetal weight accurately during pregnancy and identify *low* birth weight fetuses or macrosomia correctly. Once the risk has been identified, the maternal or neonatal morbidity and mortality can be reduced by taking appropriate clinical decisions and precautions [7]. A recognized method for fetal weight estimation is ultrasound measurement since it is non-invasive, non- hazardous, and relatively accurate [8], [9]. Various regression formulas based on different combinations of ultrasound parameters have been introduced. Dudley [10] evaluated regression formulas delivered from 11 different methods and claimed that there was no consistently superior formula for ultrasound fetal weight estimation. All the regression formulas have a problem, which is that they perform well among NBW fetuses, but is likely to be less accurate when applied to the entire fetal weight ranges [11]. One particular reason is the imbalanced sample size, the real clinical data set consists primarily of NBW fetuses with just a small part of HBW fetuses and LBW fetuses, therefore, using a regression This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 formula to estimate the birth weight of an HBW or LBW fetus is much more difficult than estimating the birth weight of an NBW fetus. Previous studies have reported that except for ultrasound parameters, there are multiple variables related to fetal birth weight, such as fetal sex, maternal age, height, large gestational weight gain, and gestational diabetes [12], [13]. It’s hard for a simple traditional regression formula to reflect the complex multi-dimensional and nonlinear relation between all these variables and the fetal weight. Besides, *most* of the fixed regression formulas are derived from a typical clinical population, so it may be less accurate to apply these formulas to other populations [10]. Recently, artificial neural networks (ANN) has been applied by many researches to predict fetal weight to overcome the problems of traditional regression methods [14], [15], [16]. Comparison of estimated fetal weight (EFW) accuracy showed that these ANN models both significantly outperformed the commonly-used regression formulas. However, the case numbers of these methods were small, neural networks are more suitable for finding patterns in large sample data, and enlarging the neonatal cases may improve *the* performance of these ANN models. Although the mean accuracy of all fetal weight ranges has improved significantly, the estimation is still less accurate when body weight is below 2500g and above 4000g due to the imbalanced data sets. What’s more, it is revealed that the fetal density, morphologic feature or body configurations are different among the three fetal weight groups [8], [11]. However, *most* of the methods were designed for the entire fetal weight ranges by using a single ANN model which ignore those differences between each group. Thus it is necessary to classify fetuses into different groups and determine different significant variables for each group. All these ANN models used the back propagation (BP) network as the learning algorithm to train the ANN. Nevertheless, the BP network suffers from an uncontrolled convergence speed and local optima, it also needs large numbers of tag data. Deep belief network (DBN) [17] is a kind of deep learning model which is composed of multiple layers of restricted boltzmann machines (RBMs). DBN has a greedy layer-wise unsupervised pre-training process as well as a top-down fine- tuning procedure for optimizing the model’s performance. The training process of DBN is faster than ANN, because the RBM is trained by just comparing the divergence. Besides, the pretraining procedure helps to find latent variables behind the data, which can be regarded as weights initialization of a BP network. Because of these advantages, DBN can avoid the problems that BP network have [18], [19]. *In* *this* study, we proposed a classification-based birth weight prediction model, which is built upon DBN networks. We collected 7875 singleton fetuses from West China Second University Hospital, the synthetic minority over-sampling technique (SMOTE) [20] was utilized to enlarge the training data *of* LBW and HBW fetuses to overcome the imbalanced learning problem. Then, the *support* vector machine (SVM) algorithm was utilized to classify the fetuses into two groups: BW ð< 4000g (LBW, NBW) and BW ð³ 4000g (HBW). Finally, these two groups were trained in two different DBN models respectively where each DBN model had different significant input parameters. The experimental results demonstrate that the proposed method outperforms the regression formulas and is an effective way to estimate the fetal birth weight. This paper also provides a possible method for predicting fetal weight in gestation period. We hope that the proposed method can help clinicians to assess fetal growth before delivery and provide valuable information in delivery management and clinical decision-making. II. METHODS A. DATA This study was a retrospective review of delivery records in West China Second University Hospital between January 2016 and December 2017. As one of the medical centers for women and children in China. The patients are from all over China with different provinces and ethnic groups which can fully reflect the diversity of patients, thus making the data from this hospital represent the general population of China. All the fetuses were examined by ultrasound which was performed by skilled obstetric residents within seven days prior to delivery [11]. As *shown* *in* Fig. 1, the women who had missing ultrasound parameters or without ultrasound information within seven days of delivery were excluded, twins or multiple births were also removed. In total, 7875 women with singleton fetus were analyzed in our study. FIGURE 1. Flow chart of patient selection. Table I shows the maternal and fetal characteristics of the 7875 singleton neonates. There were 7200 (91.43%) NBW fetuses, 485 (6.16%) HBW fetuses, and 190 (2.41%) LBW fetuses. The age of these women was between 18 and 48 years old, with an average age of 30.81 ð± 3.97 years. The mean gestational age at delivery was 39.50 ð± 0.79 weeks (36- 40weeks), the average actual birth weight was 3331.59 ð± 409.18g (930-5120g). 13708 women interviewed 3447 women had missing ultrasound parameters 54 non-singleton births 2332 women without ultrasound information within 7 days before delivery 7875 singleton livebirths This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 TABLE I. Maternal and Fetal Characteristics of Neonates (n=7875). Maternal and fetal characteristics Values Maternal ages(years), mean ð± SD 30.81 ð± 3.97 Gestation at delivery (weeks), mean ð± SD 39.50 ð± 0.79 Actual birth weight (g), mean ð± SD 3331.6 ð± 409.2 2500 ð£ BW ð< 4000g, number 7200 BW ð³ 4000g, number 485 BW ð< 2500g, number 190 SD = standard deviation; BW = birth weight. *In* this study, the following six parameters before delivery were adopted in constructing the estimation models, i.e. four ultrasonographic parameters including biparietal diameter (BPD), head circumference (HC), femur length (FL), abdominal circumference (AC) , and two maternal parameters, namely maternal fundal height (FUH) and maternal abdominal circumference (MAC). The architecture of the proposed *birth* weight estimation model is illustrated in Fig. 2, before fetal size classification, we employed SMOTE algorithm to solve the imbalanced learning problem. The fetuses were classified into group I (BW < 4000g) and group II (BW "e 4000g), then two DBN models were employed to estimate the birth weight of group I and group II separately. In this paper, we classify all the fetal samples into two categories instead of three because the case number of LBW samples are too small to achieve a satisfying classification performance, meanwhile we mainly focus on improving the prediction accuracy of NBW and HBW fetuses, so we treat the LBW samples and the NBW samples as the same class. FIGURE 2. Architecture *of* the *birth* weight estimation model. B. SMOTE-BASED DATA AUGMENTATION As *shown* in Table I, the sample numbers for HBW and LBW are insufficient, resulting *in* an imbalanced learning problem. Machine learning networks typically expect large amounts of balanced data sets. Therefore, when applied to complex data sets with imbalanced class distributions, these networks tend to provide inaccurate performance and cannot properly represent the distributive characteristics of the data [21]. SMOTE algorithm is an excellent synthetic data augmentation algorithm which has been applied in various areas, it can prevent overfitting problems compared with those methods that simply replicate the minority samples [20]. The illustration of SMOTE *is* *shown* in Fig. 3. FIGURE 3. Illustration of synthetic minority over-sampling technique. The steps of SMOTE are as follows: (1) We define the given training data set as S , the set min S is *the* minority class examples (min S S ðÎ), for each minority class sample min i S x ðÎ , obtain its k-nearest neighbors within the minority class. (2) Choosing neighbors from the k-nearest neighbors randomly *based* on the required over-sampling amount, assuming that the selected neighbor is ˆ i min xS ðÎ . (3) Synthetic samples are generated according to (1): multiply the difference between the sample i x and its nearest neighbor ˆ i x by *a* random number ðh (0< ðh <1), then add it to the sample i x to create a new synthetic sample new x . ˆ () new i i i x x x x ðh ð= ð+ ð- (1) FIGURE 4. Distribution comparison of data with and without SMOTE- based data augmentation. In this paper, the numbers of the nearest neighbors for each minority class sample was set to 5. The data distribution before and after the data augmentation *is* *shown* in Fig. 4, there were 9000 samples covering 7200(80%) NBW fetuses, 1000 (11.11%) HBW fetuses, and 800(8.89%) LBW fetuses after data augmentation. All these augmented data were only put into *the* training set, then we divided the other data without data augmentation into training and testing set randomly. Thus there is only real clinical data in the testing set. *As* a result, 6900 fetuses (76.66%) were included as the training set while 2100 fetuses (23.34%) made up the testing set. Majority class samples Minority class samples Synthetic sample 2 x 1 x 3 x 4 x ˆ i x new x i x This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 C. SVM-BASED FETAL SIZE CLASSIFICATION The body configurations of different fetal size may be different and therefore, different groups of fetal weight may be affected by different parameters. As illustrated in Fig. 2, in order to reduce the effect of body size heterogeneities between different fetal weight groups, SVM algorithm *was* used to classify fetal size before fetal weight regression. The fetuses were classified into two groups, group I (BW< 4000g) and group II (BW ð³ 4000g). To determine which parameters were highly correlated with different fetal birth weight ranges, we utilized the Pearson correlation coefficient analysis to investigate *the* most significant *parameters* for each fetal weight range. The Statistical Product and Service Solutions (SPSS

22.0) was employed to perform this statistical analysis. The Pearson correlation coefficients between different fetal weight ranges and parameters *are* *shown* in Table II. It shows that for HBW range, BPD and AC *have* a higher correlation coefficient with the actual birth weight than the other parameters, so BPD and AC are selected as the classification features for SVM. TABLE II. Pearson Correlation Coefficient Analysis For Three Fetal Weight Ranges and All Parameters. Fetal weight ranges Pearson correlation coefficient BPD HC FL AC MAC FUH LBW 0.706** 0.757** 0.785** 0.838** 0.081 0.168 NBW 0.480** 0.496** 0.523** 0.690** 0.333** 0.387** HBW 0.199** 0.194** 0.185** 0.305** 0.158** 0 **, the parameter with a significance of p < 0.01; LBW = *low* birth weight (BW < 2500g); NBW = *normal* birth weight (2500g ð£ BW ð< 4000g); HBW = high birth weight (BW ð³ 4000g). D. DBN-BASED FETAL WEIGHT ESTIMATION *In* our study, DBN *was* used to build the fetal *birth* weight estimation model, according to the classification result of SVM model, two DBN models with different input parameters were designed to predict the fetal birth weight. DBN is a multilayer structure consists of a series of individual RBMs [22], [23]. The RBM is a bipartite connectivity graph, *as* *shown* in Fig. 5, v and h represent the visible layer and the hidden layer of RBM respectively [24]. For one particular RBM, the energy of the joint configuration of (,)vh is defined as: (,)exp((,)) exp() i ijjij ij ijij p v hE v h b va hw v h ðqðqðµð-ð= ð+ð+ ðå ðå ðåðå (2) whe (, ,)w b aðqð= is the model parameter set, ij w is the weight between i v and j h , i b and j a are the bias for the layer v and layer h , respectively. The activation probability of the jth hidden unit is: 1 (,) 1 exp() j ijij i p h v w va ðqð= ð+ð-ð- ðå (3) The activation probability of the ith visible unit is: 1 (,) 1 exp( ) i ij j i j p v h w h b ðq ð= ð+ ð- ð- ðå (4) where the activation function is sigmoid equation. The gradient of log-likelihood function *was* used to optimize the ðq , it can be described asÿ log ( ) i j i j data model ij pv v h v h w ðq ð¶ ð=ð- ð¶ (5) log ( ) jj data model j pv hh a ðq ð¶ ð=ð- ð¶ (6) log ( ) ii data mo pv vv b ðq ð¶ ð=ð- ð¶ (7) where data and model represent the expectations with respect to the training data distribution and the model distribution respectively. The data can be obtained by calculating the conditional probability distributions from training set, while computing the model is intractable, as a simple and efficient solution, the contrastive divergence (CD) *was* used to approximate model [25]. FIGURE 5. RBM structure. The architecture of the DBN *model* *is* *shown* in Fig. 6. The DBN model consists of n RBM layers and a BP output layer, in our study, n is set to be 2. In the input layer, we use a sequence ð{ ð} 1 1 2 , , ..., m V x x x ð= to represent the input parameters, where i x represents each parameter and m is the number of the input parameters. As *shown* in Table II, all parameters are significantly correlated with NBW fetuses (p < 0.01) while BPD, HC, FL, AC, MAC are significantly correlated with HBW fetuses (p < 0.01). So for Group I (BW < 4000g), BPD, HC, FL, AC, FUH , and MAC are selected as the input parameters, while the input parameters of Group II (BW ð³ 4000g) are BPD, HC, FL, AC, and MAC. In each RBM layer, i V and i H is the visible layer and hidden layer of this RBM layer respectively, and i H is also regarded as the visible layer of the next RBM layer. There is a BP network on the top of the RBMs, which is regarded as the output layer, y represents the predicted fetal birth weight. i W , i B , and i A are the connection weights and bias between two layers. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017
FIGURE 6 . Architecture of the DBN model. There are two steps *in* the training process of a DBN model: Step I : pre training In this step, each RBM layer is trained separately with the greedy unsupervised algorithm. Step II: fine-tuning In this step, a supervised BP network is utilized to fine tune the whole model, the input of this BP network is n H , *which* is the output vector of the last RBM layer. We evaluated *the* performance of the model by using mean absolute error (MAE) and mean absolute percent error (MAPE), the equations are represented as follows: 1 1 ˆ n ii i MAE y y n ð= ð=ð- ðå (8) 1 ˆ 1 100% n ii i i yy MAPE ny ð= ð- ð=ð´ ðå (9) where n is the number of fetuses, i y and ˆ i y are th actual *birth* weight and estimated birth weight of the th i fetus. III. RESULTS A. PERFORMANCE COMPARISON OF IMBALANCED AND BALANCED DATA SETS Table III shows the estimation performance of the proposed model, the MAPE and MAE for the fetuses of all the fetal weight ranges are 6.09 ð± 5.06% and 198.55 ð± 158.63g. Table IV provides the result *of* the prediction model without SMOTE based data augmentation. It can be observed that the prediction performance of minority groups HBW and *LBW* has been improved after data augmentation in the proposed model, especially for the HBW group, there is a 1.17% decline in MAPE and 49.02g decline in MAE respectively. Thus, it can be concluded that using the balanced data set to predict the fetal weight is more effective than just using the original imbalanced data set. TABLE III. The Result of The Proposed Birth Weight Estimation Model (in testing set, n= 2100). Fetal weight ranges MAPE±SD (%) MAE±SD (g) ABW 6.09±5.06 198.55±158.63 LBW 11.09±10.38 245.94±174.20 NBW 5.65±4.49 179.70±147.79 HBW 5.85±5.04 244.84±217.70 ABW = all birth weight; LBW = *low* birth weight (BW<2500g); NBW = *normal* birth weight (2500g ð£ BW<4000g); HBW = high birth weight (BW ð³ 4000g); MAPE = mean absolute percent error; MAE = mean absolute error; SD = standard deviation. TABLE IV. The Result of Estimation Model Without SMOTE-based Data Augmentation (in testing set, n= 2100). Fetal weight ranges MAPE±SD (%) MAE±SD (g) ABW 6.21±4.23 202.69±141.96 LBW 11.72±9.21 260.82±201.00 NBW 5.76±4.23 183.30±133.85 HBW 7.02±4.33 293.86±168.10 B. PERFORMANCE COMPARISON OF CLASSIFIED AND UNCLASSIFIED DATA SETS To investigate the

effectiveness of SVM based *classification* in the proposed model, we validated *the* performance of the estimation model in the unclassified data set. Table V illustrates the result, compared with the result *shown* in Table III, it was demonstrated that using SVM algorithm to classify the fetuses into different fetal size groups can improve the prediction performance, especially for HBW fetuses. TABLE V. The Result of Estimation Model in The Unclassified Data Sets (in testing set, n= 2100). Fetal weight ranges MAPE±SD (%) MAE±SD (g) ABW 6.17±5.17 201.22±155.33 LBW 11.89±11.01 248.94±184.56 NBW 5.76±4.50 188.08±144.32 HBW 7.10±4.77 297.07±206.54 C. ACCURACY COMPARISON OF DIFFERENT FETAL WEIGHT ESTIMATION METHODS We compared the fetal weight estimation accuracy of the proposed model with that of the commonly used regression formulas *shown* in Table VI, including the formulas of Hadlock et al. (1985) [26], Shepard et al. (1982) [27], Hsieh et al. (1987) [28], and Woo et al. (1985) [29]. We calculated the MAPE and MAE for each regression formula method in the same testing set (n=2100) as our proposed method. The result is illustrated in Table VII, compared with the accuracy of regression formulas, the proposed method increased the predictive accuracy MAPE for 0.25% to 21.01% and MAE for 12.49g to 681.24g. Fig. 7 and Fig. 8 illustrate that the proposed approach is superior to regression formulas, it can reach a statistically higher degree of accuracy. TABLE VI. Five Published Regression Formulas For Fetal Weight Estimation. TABLE VIII. Estimation Results of The Regression Formulas For Different Fetal Weight Ranges (in testing set, n= 2100). TABLE VII. The Estimation Results of Regression Formula Methods (in testing set, n= 2100). Methods MAPE±SD (%) MAE±SD (g) Hadlock1 (1985) 6.34±4.76 211.90±163.40 Hadlock2 (1985) 6.42±4.93 211.04±159.65 Shepard (1982) 7.42±5.83 240.00±179.13 Hsieh (1987) 7.29±5.68 235.64±174.30 Woo (1985) 27.10±10.45 879.79±301.31 FIGURE 7. The mean absolute percent error (MAPE) of the proposed model as compared to other methods. FIGURE 8. The mean absolute error (MAE) of the proposed model as compared to other methods. For the regression formulas, the fetal weight estimation accuracy of different fetal birth weight ranges *is* *shown* in Table VIII. As compared with the result of the proposed model *shown* in Table III, it is demonstrated that our model outperforms the regression formulas among NBW and HBW fetuses, while for LBW fetuses, the estimation results of Hadlock1 (1985) and Hadlock2 (1985) are slightly better than those of the proposed model. IV. DISCUSSION Most of the EFW models were designed for all fetal weight ranges by using a single formula or prediction model, it may have a high estimation accuracy among NBW group, but the accuracy of HBW and LBW was still not satisfactory. We considered one of the reasons was that the *sample* size *of* the minority weight ranges was small, in our study, the data augmentation algorithm SMOTE *was* used to solve this problem. Bernstein and Catalano claimed that for the HBW fetuses, their soft tissue mass in the limbs increased a lot, but the measurements of AC, HC, and FL may do not account for this, which would cause fetal weight underestimation [30]. Therefore, we hypothesized that when the fetal weight grows to more than 4000g, the growth trend is no longer reflected in the changes of ultrasound parameters, but in the changes of the maternal parameters such as FUH and MAC. So the FUH and MAC before delivery were adopted as the input *parameters* of the DBN models. Chuang et al. [8] also hypothesized other possible reasons, such as the density, the fetal morphologic feature might be different between each birth weight groups. In order to reduce the influence of body size heterogeneities between different fetal weight groups, SVM algorithm was utilized to classify fetuses into two groups: birth weight below 4000g and over or equal to 4000g, we analyzed these two groups and found *the* most significant *parameters* for each group. Then they were trained in two different DBN *models* with the corresponding input parameters. We compared the References Formulas Hadlock1 (1985) 10 log ()1.326 0.003260.01070.04380.158BWACFLHCACFLð=ð-ð´ ð´ ð+ð´ ð+ð´ ð+ð´ Hadlock2 (1985) 10 log ( ) 1.3596 0.00386 0.0064 0.00061 0.0424 0.1 74 BW AC FL HC BPD AC AC FL ð= ð- ð´ ð´ ð+ ð´ ð+ ð´ ð´ ð+ ð´ ð+ ð´ Shepard (1982) 10 log ( ) 1.7492 0.166 0.046 2.6 BW BPD AC AC BPD ð= ð- ð+ ð´ ð+ ð´ ð- ð´ ð´ Hsieh (1987) 22 10 log ( ) 2.7193 1.745 0.001 7.6742 0.0001 0.1432 9 .4962 0.001 BW BPD FL AC BPD FL AC BPD ð= ð+ ð´ ð´ ð´ ð- ð´ ð´ ð´ ð- ð´ ð+ ð´ ð´ ð´ Woo (1985) 10 log ( ) 1.14 0.16 0.05 2.8 0.001 0.04 4.9 BPD AC BPD AC FL FL AC ð= ð+ ð´ ð+ ð´ ð- ð´ ð´ ð´ ð+ ð´ ð- ð´ ð´ ð´ Fetal weight ranges Hadlock1 (1985) Hadlock2 (1985) Shepar Hsieh (1987) LBW- MAPE±SD (%) 10.19±7.09 10.91±7.44 12.38±8.39 12.13 ± 8.13 LBW- MAE±SD (g) 230.49±163.95 247.38±173.23 282.05±196.26 273.64±187.14 NBW- MAPE±SD (%) 5.81±4.31 6.01±4.57 7.24±5.61 7.13±5.47 NBW- MAE±SD (g) 190.87±141.22 195.27±144.16 233.32±173.06 230.11±169.15 HBW- MAPE±SD (%) 9.26±5.22 7.92±5.05 6.63±4.99 6.34±4.80 HBW- MAE±SD (g) 391.38±225.98 334.82±217.65 279.51±212.74 267.34±204.96 prediction results of the classified and unclassified models. As the estimation accuracy *shown* in Table III and Table V, it is demonstrated that classify the fetuses into different groups and predict the *birth* weight using different significant parameters result in higher accuracy observably. *In* *our* study, we investigated the effectiveness of SMOTE- based data augmentation, it *is* *shown* in Table IV that if the original imbalanced dataset was utilized directly as the training set *of* the prediction model, the minority groups HBW and LBW fetuses would be seriously biased to the majority group NBW fetuses. *As* a result, the prediction accuracy of HBW and LBW fetuses were relatively low. After data augmentation based on SMOTE, a balanced training dataset was reconstructed, *the* performance of HBW groups had realized a significant increase. For the LBW group, there was a slight rise on the performance, because the *sample* size of this group was too small to get a satisfying improvement. It also proved that the increase of the sample number of HBW and LBW fetuses can improve the estimation performance, future study is suggested to collect more actual HBW and LBW fetuses to improve the prediction accuracy. From Table III and Table VIII, we *found* that the proposed method outperformed the regression formulas almost in all fetal ranges. There were also *a* lot of EFW researches based on ANN model. Wu et al. [16], Chuang et al. [8], and Cheng et al. [14] both adopted ANN to estimate birth weight, they all claimed that ANN model outperformed the regression formulas, however, their researches were all based on small study groups with a fetal case of 109, 1353 and 2127 respectively. Wu et al. [16], Chuang et al. [8] also claimed that the EFW was less accurate for LBW and HBW fetuses. *In* our study, there were a total of 7875 cases which was significantly larger than the above studies. The sufficient balanced sample cases are more conducive to the DBN to analyze the complex relationship between input and output, resulting in higher accuracy. In addition, DBN can solve the problem of uncontrolled convergence speed and local optima of ANN. There are several limitations of the study that should be recognized. First, those women *who* gave birth to twins or multiple births were excluded, and it may prevent us from fully understanding the birth weight of both singleton, twins, and multiple births. Future study may take these patients into consideration. Secondly, there are various parameters we can get from an ultrasound measurement, such as occipito-frontal diameter (OFD), femur length (FL), head circumference (HC), biparietal diameter (BPD), abdominal circumference (AC), and fetal gender. *In* *our* study, we selected BDP, HC, FL, AC, FUH, and MAC as the input *parameters* of the proposed model according to the expert's clinical experience and Pearson correlation analysis. However, studies showed that different ultrasound parameters were associated with birth weight at different degrees, and different combinations of parameters may affect the prediction results. Cheng et al. [14] cross- validate the significance among ultrasound parameters by using Spearman correlation analysis, it showed that compared with other factors, AC is most correlated with birth weight. In addition, several maternal characteristics, such as pre-existing or gestational diabetes, prolonged gestation, maternal age, significant weight gain during pregnancy, and body mass index are also linked to fetal weight [4], [31]. Future studies are warranted to improve the accuracy of EFW by adding more effective parameters and the rational combination *of* maternal and fetal parameters is the key to establish a better estimation model. Thirdly, the proposed model improved the estimation accuracy at the extreme weight ranges (HBW, LBW) due to the balanced data sets produced by SMOTE- based data augmentation. However, there are also drawbacks in SMOTE algorithm, such as variance and over generalization, the variant of SMOTE such as Borderline- SMOTE can be utilized to improve the estimation performance [32]. Besides, these augmented samples cannot replace the real neonates. To solve this problem, further studies are suggested to collect more data from real HBW and LBW fetuses born in hospitals. Finally, this work was a retrospective study from a single center, in the further work, a multicenter prospective study is planned to expand the sample size to test the accuracy of the proposed model. V. CONCLUSION *In* *this* study, we proposed a novel fetal weight estimation model which combined SVM based classification with DBN to improve *the* performance of EFW in all fetal weight ranges, we also solved the imbalanced learning problem by utilizing SMOTE based data augmentation. It was demonstrated from the result that the proposed model outperformed the regression formulas. Our study revealed that DBN is a promising approach for fetal weight estimation, it also proved that classify fetuses into different groups and predict *birth* weight using different significant parameters are effective. We believe that the proposed method may help clinicians to assess fetal growth during the pregnancy and provide valuable information in delivery management and clinical decision- making. With respect to the future work, we intend to collect more new-born fetuses of all fetal weight ranges from multiple centers, explore additional effective parameters and optimize the structure of DBN to improve the prediction accuracy. APPENDIX Abbreviations and full names: BW: Birth weight LBW: *Low* birth weight (birth weight ð< 2 500g) NBW: Normal birth weight (2500g ð£ *birth* weight < 4000g) HBW: High birth weight (birth weight ð³ 4000g) SMOTE: Synthetic minority over-sampling technique SVM: *Support* vector machine DBN: Deep belief network MAPE: Mean absolute percent error MAE: Mean absolute error WHO: the World Health Organization ANN: Artificial neural networks Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 EFW: Estimated fetal weight BP: Back propagation RBMs: Restricted boltzmann machines BPD: Biparietal diameter HC: Head circumference FL: Femur length AC: Abdominal circumference FUH: Maternal fundal height

MAC: Maternal abdominal circumference CD: Contrastive divergence OFD: Occipito-frontal diameter ACKNOWLEDGMENT We are grateful for Dr. Xiaorong Qi and Mrs. Li Wan at the Department of Obstetrics and Gynecology of West China Second University Hospital for clinical data organization. We would like to also thank all doctors for the ultrasound measurements. REFERENCES [1] Badshah S, Mason L, McKelvie K, Payne R, Lisboa PJ. Risk factors for *low* *birth* weight in the public hospitals at Peshawar. NWFP-Pakistan BMC Public Health. 2008 Jun 4;8:197. [2] Berard J, Dufour P, Vinatieretal D. Fetal macrosomia: risk factors and outcome. A study of the outcome concerning 100 cases > 4500g. Eur J Obstet Gyn R B 1998;77(1):51– 9. [3] Chiarotti A, Puopolo M, Gissler M, Sihvonen E, Hemminki K. Effects of socio environmental factors on neurocognitive performance in premature or low- birth weight preschoolers. Ann Ist Super Sanita. 2001;37:553– 9. [4] Mengesha H G, Wuneh A D, Weldearegawi B, et al. *Low* *birth* weight and macrosomia in Tigray, Northern Ethiopia: who are the mothers at risk?. Bmc Pediatr 2017; 17(1):144. [5] Haram K, Pirhonen J, Bergsjo P. Suspected big baby: a difficult clinical problem in obstetrics. Acta Obstet Gynecol Scand 2015; 81(3): 185–194. [6] Henriksen T. The macrosomic fetus: a challenge in current obstetrics. Acta Obstet Gynecol Scand 2008;87(2):134–145. [7] Phillips A M, Galdamez A B, Ounpraseuth S T, et al. Estimate of fetal weight by ultrasound within two weeks of delivery in the detection of fetal macrosomia. Aust NZ J Obstet Gyn 2014; 54(5):441-444. [8] Chuang L, Hwang J Y, Chang C H, et al. Ultrasound estimation of fetal weight with the use of computerized artificial neural network model. Ultrasound Med Biol 2002; 28(8):991-996. [9] Farmer, R M, Medearis, A L, Hirata, G I, et al. The use of a neural network for the ultrasonographic estimation of fetal weight in the macrosomic fetus . Am J Obstet Gynecol 1992; 166(5):1467-1472. [10] Dudley N J. A systematic review of the ultrasound estimation of fetal weight. Ultrasound in Obst Gyn 2005; 25(1):80-89. [11] Sabbagha R E, Minogue J, Tamura R K, et al. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, and small-for-gestational- age fetuses. Am J Obstet Gynecol 1989; 160(4):854-860. [12] Koyanagi A, Zhang J, Dagvadorj A, et al. Macrosomia in 23 developing countries: *an* analysis of a multicountry, facility-based, cross-sectional survey. Lancet 2013; 381(9865):476-483. [13] Oliver M, Mcnally G, Leader L. Accuracy of sonographic prediction of birth weight. Aust NZ J Obstet Gyn 2013; 53(6):584–588. [14] Cheng Y C, Yan G L, Chiu Y H, et al. Efficient fetal size classification combined with artificial neural network for estimation of fetal weight. Taiwan J Obstet Gyne 2012; 51(4):545-553. [15] Mohammadi H, Jahanseir M, Allahmoradi Z, et al. Ultrasound Estimation of Fetal Weight by Artificial Neural Network in Normal Pregnancies. 2010 international conference on measurement and control engineering:575-579 [16] Wu J, Yang T, Lin J, et al. Estimation of fetal weight on the basis of neural network. Journal of Biomedical Engineering 2005; 22(5):922-925. [17] Hinton G, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. Neural computation, 2006, 18(7): 1527-1554. [18] Tan Q , Huang W , Li Q . An intrusion detection method based on DBN in ad hoc networks[M]// Wireless Communication and Sensor Network: *Proceedings* of the *International* Conference on Wireless Communication and Sensor Network (WCSN 2015). [19] Shang C , Yang F , Huang D , et al. Data-driven soft sensor development based on deep learning technique. Journal of Process Control, 2014, 24(3):223-233. [20] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002; 16(1):321-357. [21] He H, Garcia E A. Learning from Imbalanced Data. IEEE T Knowl Data En 2009; 21(9):1263-1284. [22] Hinton, G. E.; Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. Science 2006, 313, 504" 507. [23] Bengio, Y. Learning Deep Architectures for AI[J]. Foundations and Trends? in Machine Learning, 2009, 2(1):1-127. [24] Ackley D H , Hinton G E , Sejnowski T J . A Learning Algorithm for Boltzmann Machines *[M]// Connectionist models and their implications: readings from cognitive science. Ablex Publishing Corp. 1988. Citation information: DOI 10.1109/ACCESS.2019.2925803, IEEE Access VOLUME XX, 2017 [25] Hinton G E . Training Products of Experts by Minimizing Contrastive Divergence. Neural computation, 2002, 14(8): 1771-1800. [26] Hadlock F P, Harrist R B, Sharman R S, et al. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. Am J Obstet Gynecol 1985; 151(3):333-337. [27] Shepard M J, Richards V A, Berkowitz R L, et al. An evaluation of two equations for predicting fetal weight by ultrasound. Am J Obstet Gynecol 1982; 142(1):47-54. [28] Hsieh F J, Chang F M, Huang H C, et al. Computer- assisted analysis for prediction of fetal weight by ultrasound-comparison of biparietal diameter (BPD), abdominal circumference (AC) and femur length (FL). Taiwan Yi Xue Hui Za Zhi 1987; 86(9):957-964. [29] Woo J S, Wan C W, Cho K M. Computer-assisted evaluation of ultrasonic fetal weight prediction using multiple regression equations with and without the fetal femur length. J Ultras Med 1985; 4(2):65-67. [30] Bernstein I M, Catalano P M. Influence of fetal fat on the ultrasound estimation of fetal weight in diabetic mothers. Obstet Gynecol 1992; 79(4):561-563. [31] Jara T W D. Factors Associated with Macrosomia among Neonates Delivered at Debre Markos Referral Hospital, Northwest Ethiopia, 2014: A Case Control Study. Journal of Diabetes & Metabolism 2014; 5(12). [32] B.X. Wang and N. Japkowicz, “Imbalanced Data Set Learning with Synthetic Samples,” Proc. IRIS Machine Learning Workshop, 2004.