MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231063

Prognosticating Fetal Growth Restriction and Small for Gestational Age by Medical History

Herdiantri SUFRIYANA[a,b], Fariska Zata AMANI[c], Aufar Zimamuz Zaman AL HAJIRI[d], Yu-Wei WU[a,e] and Emily Chia-Yu SU[a,e,f,1]
[a] *Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taiwan*
[b] *Department of Medical Physiology, Faculty of Medicine, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[c] *Department of Obstetrics and Gynecology, Faculty of Medicine, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[d] *Faculty of Medicine, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[e] *Clinical Big Data Research Center, Taipei Medical University Hospital, Taiwan*
[f] *Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taiwan*
ORCiD ID: Herdiantri Sufriyana https://orcid.org/0000-0001-9178-0222, Fariska Zata Amani https://orcid.org/0000-0002-8887-0304, Aufar Zimamuz Zaman Al Hajiri https://orcid.org/0000-0003-2302-1640, Yu-Wei Wu https://orcid.org/0000-0002-5603-1194, Emily Chia-Yu Su https://orcid.org/0000-0003-4801-5159

Abstract. This study aimed to develop and externally validate a prognostic prediction model for screening fetal growth restriction (FGR)/small for gestational age (SGA) using medical history. From a nationwide health insurance database (n=1,697,452), we retrospectively selected visits of 12-to-55-year-old females to healthcare providers. This study used machine learning (including deep learning) and 54 medical-history predictors. The best model was a deep-insight visible neural network (DI-VNN). It had area under the curve of receiver operating characteristics (AUROC) 0.742 (95% CI 0.734 to 0.750) and a sensitivity of 49.09% (95% CI 47.60% to 50.58% at with 95% specificity). Our model used medical history for screening FGR/SGA with moderate accuracy by DI-VNN. In future work, we will compare this model with those from systematically-reviewed, previous studies and evaluate if this model's usage impacts patient outcomes.

Keywords. Fetal growth restriction, small for gestational age, machine learning, deep learning, electronic health records, risk prediction

1. Introduction

Fetal growth restriction (FGR) and small for gestational age (SGA) are a condition with different diagnostic measures [1]. Preventing FGR/SGA may reduce neonatal

[1] Corresponding Author: Emily Chia-Yu Su. email: emilysu@tmu.edu.tw

morbidity and its costs [2], but an effective prevention needs accurate screening [3]. Current methods using ultrasound or biomarkers were inaccessible in most settings [4, 5]. Meanwhile, a health insurance claim database abundantly records medical history of which association [6] with FGR/SGA allows its proactive screening, particularly in countries with universal health coverage [7]. However, studies have yet to develop a screening method for FGR/SGA using medical history. Both statistical and computational machine learning (ML) could predict pregnancy outcomes using medical history [8,9]. We aimed to develop and externally validate a prognostic prediction model for screening FGR/SGA using medical history in nationwide insured women.

2. Methods

We followed a pre-registered protocol [10]. Data are available from the *badan penyelenggara jaminan sosial (BPJS) kesehatan* in Indonesia (https://e-ppid.bpjs-kesehatan.go.id/). This study was under a project of deep-insight visible neural network (DI-VNN) for predicting diseases. Its ethical review was exempted (the Taipei Medical University Joint Institutional Review Board, TMU-JIRB no.: N202106025).

We retrospectively selected 12-to-55-year-old females that visited all levels of care of a nationwide health insurance in Indonesia (August 2019 [11]: access no.: 5064/I.2/0421). Post-delivery visits were excluded. An event outcome was defined by the International Classification of Disease version 10 (ICD-10) codes prefixed by O365 (mothers) and P05 (newborns). The candidate predictors were medical histories of diagnoses and procedures. We identified 54 of them, including latent candidate predictors (multiple pregnancies, varicella, urinary tract infection, and placenta previa).

Model 1 was a statistical ML of ridge regression (RR) using latent candidate predictors pre-selected by multivariate analyses. Models 2 to 4 were computational ML using 10-fold, cross-validated principal components (PCs) with top proportions of variance explained to pursue ≥20 events per candidate predictor for: (1) elastic net regression (PC-ENR); (2) random forest (PC-RF); and (3) gradient boosting machine (PC-GBM). The fifth model was DI-VNN. Unlike the protocol [10], we did not limit candidate predictors by a false discovery rate; otherwise, network architecture could not be constructed. All model was recalibrated by logistic regression or a general additive model using locally weighted scatterplot smoothing (GAM-LOESS).

This study applied internal (~64%) and external (~36%) validations (random sampling ~16%; geographical and temporal splitting ~20%). For calibration split, we used ~20% of the internal validation. Before recalibration, we applied: (1) five-fold cross-validation for hyperparameter tuning and (2) bootstrapping for final training. Unlike the protocol [10], we chose 100 repetitions for bootstrapping, due to sample size.

We evaluated the models in optimal range of predicted probabilities among all. A model was well-calibrated if it approximated: (1) similar regression-reference line; (2) distinguished distributions of events and nonevents; (3) intercept of ~0 and slope of ~1; and (4) the Brier score of ≤0.05. We defined the best clinical utility among well-calibrated models by decision curve analysis based on its net benefit at 95% specificity. Eventually, we evaluated the best discrimination among well-calibrated models by ROC curve, sensitivity at 95% specificity, and the AUROC (≥0.5). AUROCs by external validation was shown for comparison. We computed 95% confidence interval (CI) inferred by bootstrapping and cross-validation. The analytical codes are available at https://github.com/herdiantrisufriyana/fgr_sga.

3. Results

From the database (*n*=1,697,452), we selected 12-to-55-year-old females (*n*=169,746) that had visited (*n*=507,319) primary, secondary, and tertiary care. The visits had no overlap between internal and external validation sets. To characterize subjects in the internal validation set (Table 1), we included uncensored outcomes (*n*=26,576).

Table 1. Subject characteristics for association tests and internal validation set.

Variable		Not FGR/SGA [a] (n=26,459)	FGR/SGA [a] (n=117)	*P* value
Pregnancy episode [b]	First pregnancy, [c] no. (%)	25,096 (94.85)	109 (93.16)	(reference)
	Second pregnancy, [c] no. (%)	1,363 (5.15)	8 (6.84)	*P*=.41
Maternal age	Mean (SD), year	29 (6)	28 (6)	*P*=.006**
Insurance class	First, no. (%)	3,604 (13.62)	21 (17.95)	(reference)
	Unspecified, no. (%)	87 (0.33)	1 (0.85)	*P*=.51
	Second, no. (%)	9,226 (34.87)	50 (42.74)	*P*=.78
	Third, no. (%)	13,542 (51.18)	45 (38.46)	*P*=.03*
Marital status	Married, no. (%)	16,831 (63.61)	77 (66)	(reference)
	Single, no. (%)	2,397 (9.06)	20 (17)	*P*=.02*
	Unspecified, no. (%)	7,117 (26.90)	20 (17)	*P*=.05
	Divorce/widow, no. (%)	114 (0.43)	77 (66)	*P*=.97
Occupation segment of the householder	Central-government employee, no. (%)	7,683 (29.04)	20 (17.1)	(reference)
	Private employee, no. (%)	9,611 (36.32)	57 (48.7)	*P*=.002**
	Private employer or self-employed, no. (%)	7,871 (29.75)	35 (29.9)	*P*=.06
	Local-government employee, no. (%)	1,278 (4.83)	5 (4.3)	*P*=.42
	Unemployed, no. (%)	16 (0.06)	5 (4.3)	*P*=.98

[a], Subject per pregnancy (no censored delivery); [b], Excluding non-pregnancy; [c], The 1st and 2nd pregnancies within database period; FGR, fetal growth restriction; SGA, small for gestational age; SD, standard deviation.

The well-calibrated models were (Figure 1a): PC-ENR, PC-GBM, and DI-VNN. The PC-GBM was the most well-calibrated (intercept -0.00098, 95% CI -0.13098 to 0.12902; slope 0.95, 95% CI 0.46 to 1.44; Brier score 0.0063). The net benefits of the three models were higher than those of treat-all and treat-none prediction (Figure 1b). The DI-VNN was the best clinical utility (net benefit 0.0023, 95% CI 0.0022 to 0.0024).

The discrimination ability differed among the well-calibrated models by ROCs and AUROCs (Figure 1c). Based on the internal calibration split, the best model was also DI-VNN (AUROC 0.742, 95% CI 0.734 to 0.750; sensitivity 49.09%, 95% CI 47.60% to 50.58%). Using external validation, the AUROC of the DI-VNN (0.561, 95% CI 0.558 to 0.564) was considerably robust (i.e., the 95% CI >0.5).

4. Discussion

For predicting FGR/SGA, the previous models required ultrasound or biomarkers and a specific gestational age: (1) 25~42 weeks [12]; (2) 10~14 weeks [13]; and (3) 30~35 weeks [14]. Meanwhile, an effective prevention of FGR should be given ≤16 weeks' gestation [3]. The DI-VNN were competitive to those models. External validation estimated the DI-VNN to outperform the previous model [13]. Evaluation of the previous models [12,13,14] also used training set only, which might be overoptimistic.

Our models did not require ultrasound or biomarkers to predict FGR/SGA in advance, including the DI-VNN. We could apply our models to any pregnant women without a specific gestational age. Unlike the previous studies, we externally validate

the future performance of the DI-VNN. However, the accuracy of DI-VNN was only moderate although similar accuracies were achieved by the previous models. The impact of DI-VNN on patient outcomes is also still unclear. Yet, many previous studies in medicine have yet to evaluate the impacts of their prediction models [15].

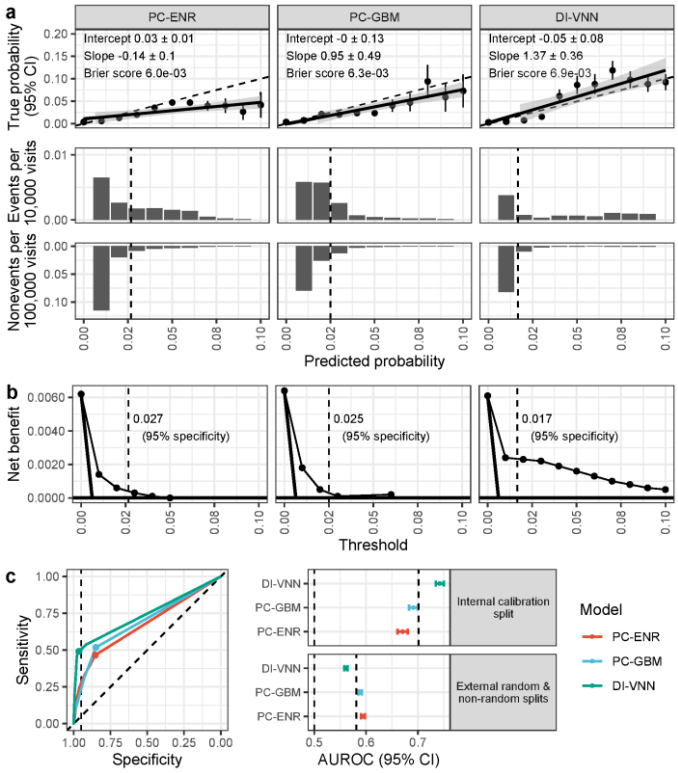

Figure 1. Model calibration (a), clinical utility (b), and discrimination (c). Solid lines show regressions over the estimates of true probabilities. The dotted lines show a threshold of 95% specificity (a, b, c) or AUROCs of 0.5 and the average per facet (c, AUROC comparison plot). AUROC area under the curve of receiver operating characteristics; CI, confidence interval; DI-VNN, deep-insight visible neural network; ENR, elastic net regression; GBM, gradient boosting machine; PC, principal components.

5. Conclusions

FGR/SGA could be screened using medical history with moderate accuracy by DI-VNN. In future work, this model should be compared with those from systematically-reviewed, previous studies and evaluated for its impacts on patient outcomes.

Acknowledgments

The Indonesian BPJS Kesehatan gave access to the dataset. This study was funded by: (1) the Postdoctoral Accompanies Research Project from the National Science and Technology Council (NSTC) in Taiwan (grant no.: NSTC111-2811-E-038-003-MY2)

to Herdiantri Sufriyana; and (2) the Ministry of Science and Technology (MOST) in Taiwan (grant nos.: MOST110-2628-E-038-001 and MOST111-2628-E-038-001-MY2), and the Higher Education Sprout Project from the Ministry of Education (MOE) in Taiwan (grant no.: DP2-111-21121-01-A-05) to Emily Chia-Yu Su.

References

[1] Fetal Growth Restriction: ACOG Practice Bulletin, Number 227. Obstet Gynecol. 2021 Feb;137(2):e16-e28, doi: 10.1097/aog.0000000000004251.
[2] Ho T, Zupancic JAF, Pursley DM, Dukhovny D. Improving Value in Neonatal Intensive Care. Clin Perinatol. 2017 Sep;44(3):617-25, doi: 10.1016/j.clp.2017.05.009.
[3] Roberge S, Nicolaides K, Demers S, Hyett J, Chaillet N, Bujold E. The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. Am Journal Obstet Gynecol. 2017 Feb;216(2):110-20.e6, doi: 10.1016/j.ajog.2016.09.076.
[4] Pedroso MA, Palmer KR, Hodges RJ, Costa FDS, Rolnik DL. Uterine Artery Doppler in Screening for Preeclampsia and Fetal Growth Restriction. Rev Bras Ginecol Obstet. 2018 May;40(5):287-93, doi: 10.1055/s-0038-1660777.
[5] Luntsi G, Ugwu AC, Nkubli FB, Emmanuel R, Ochie K, Nwobi CI. Achieving universal access to obstetric ultrasound in resource constrained settings: A narrative review. Radiography (Lond). 2021 May;27(2):709-15, doi: 10.1016/j.radi.2020.10.010.
[6] Selvaratnam RJ, Wallace EM, Flenady V, Davey MA. Risk factor assessment for fetal growth restriction, are we providing best care? Aust N Z J Obstet Gynaecol. 2020 Jun;60(3):470-3, doi: 10.1111/ajo.13147.
[7] Wagstaff A, Neelsen S. A comprehensive assessment of universal health coverage in 111 countries: a retrospective observational study. Lancet Glob Health. 2020 Jan;8(1):e39-49, doi: 10.1016/s2214-109x(19)30463-2.
[8] Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, Wu YW, Su EC. Comparison of Multivariable Logistic Regression and Other Machine Learning Algorithms for Prognostic Prediction Studies in Pregnancy Care: Systematic Review and Meta-Analysis. JMIR Med Inform. 2020 Nov;8(11):e16503, doi: 10.2196/16503.
[9] Sufriyana H, Wu YW, Su EC. Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. EBioMedicine. 2020 Apr;54:102710, doi: 10.1016/j.ebiom.2020.102710.
[10] Sufriyana H, Wu YW, Su EC-Y. Human and machine learning pipelines for responsible clinical prediction using high-dimensional data. Protocol Exchange. 2021 2022/02/04. doi: 10.21203/rs.3.pex-1655/v1.
[11] Ariawan I, Sartono B, Jaya C. Sample dataset of the BPJS kesehatan 2015-2016. Jakarta BPJS Kesehatan. 2019.
[12] Shlossman P, Scisione A, Manley J, et al. Doppler assessment of the intrafetal vasculature in the identification of intrauterine growth retardation. Which vessel is 'best' or is a combination better? Am J Obstet Gynecol. 1998;178(S1):S88, doi: 10.1016/j.ajog.2012.01.022.
[13] Carbone JF, Tuuli MG, Bradshaw R, Liebsch J, Odibo AO. Efficiency of first-trimester growth restriction and low pregnancy-associated plasma protein-A in predicting small for gestational age at delivery. Prenat Diagn. 2012 Aug;32(8):724-9, doi: 10.1002/pd.3891.
[14] Valiño N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 30-34 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound Obstet Gynecol. 2016 Feb;47(2):194-202, doi: 10.1002/uog.14928, doi: 10.1002/uog.14928.
[15] Zhou Q, Chen ZH, Cao YH, Peng S. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. NPJ Digit Med. 2021 Oct;4(1):154, doi: 10.1038/s41746-021-00524-2.