Technology and Health Care 28 (2020) S181–S186
DOI 10.3233/THC-209018
IOS Press

Predictive models of hypertensive disorders in pregnancy based on support vector machine algorithm

Lin Yang[a,1,*], Ge Sun[a,1], Anran Wang[b,1], Hongqing Jiang[c,1], Song Zhang[a,*], Yimin Yang[a], Xuwen Li[a], Dongmei Hao[a], Mingzhou Xu[d] and Jing Shao[e]
[a]*College of Life Science and Bioengineering, Beijing University of Technology, Intelligent Physiological Measurement and Clinical Translation, Beijing International Base for Scientific and Technological Cooperation, Beijing, 100024, China*
[b]*Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, 100020, China*
[c]*Haidian Maternal and Children Health Hospital, Beijing, 100080, China*
[d]*Beijing Aerospace ChangFeng Co. Ltd., Beijing, 100071, China*
[e]*Beijing Yes Medical Devices Co. Ltd., Beijing, 100152, China*

Abstract.
BACKGROUND: The risk factors of hypertensive disorders in pregnancy (HDP) could be summarized into three categories: clinical epidemiological factors, hemodynamic factors and biochemical factors.
OBJECTIVE: To establish models for early prediction and intervention of HDP.
METHODS: This study used the three types of risk factors and support vector machine (SVM) to establish prediction models of HDP at different gestational weeks.
RESULTS: The average accuracy of the model was gradually increased when the pregnancy progressed, especially in the late pregnancy 28–34 weeks and $\geqslant$ 35 weeks, it reached more than 92%.
CONCLUSION: Multi-risk factors combined with dynamic gestational weeks' prediction of HDP based on machine learning was superior to static and single-class conventional prediction methods. Multiple continuous tests could be performed from early pregnancy to late pregnancy.

Keywords: Support vector machine algorithm, machine learning, model research

1. Introduction

Hypertensive disorders in pregnancy (HDP) has been a major cause of increased morbidity and mortality among pregnant women and perinatal infants [1]. There are various risk factors of HDP, mainly related to clinical epidemiology, hemodynamics and biochemistry [2–5]. Using epidemiological factors alone

[1]Lin Yang, Ge Sun, Anran Wang and Hongqing Jiang are co-first authors.
[*]Corresponding authors: Lin Yang and Song Zhang, College of Life Science and Bioengineering, Beijing University of Technology, Intelligent Physiological Measurement and Clinical Translation, Beijing International Base for Scientific and Technological Cooperation, Beijing, 100024, China. Tel.: +86 13426181228; +86 13901218968; E-mails: yanglin@bjut.edu.cn; zhangsong@bjut.edu.cn.

to predict HDP, the detection rate was only about 30%. Studies have shown that maternal factors and uterine artery Doppler were better than maternal factors alone in predicting early-onset preeclampsia [6]. A joint assessment of multiple risk factors is needed. In order to improve the prediction rate, researchers have carried out a variety of combinations of different biomarkers. Myers et al. used maternal historical medical information and biochemical information placental growth factor (PlGF), soluble endoglin (sEng) and soluble fms-like tyrosine kinase-1 (sFlt-1) to establish a joint prediction model. The sensitivity of the model to predict pre-eclampsia was 45% at 14–16 weeks of gestation when the false positive rate was 5% [7]. Jaana et al. also used maternal characteristics and combined screening markers to develop a predictive risk model for early-onset pre-eclampsia. The best detection rate was 47% [8].

Current studies have focused on exploring changes in static parameters at a specific time during the first and second trimesters of pregnancy, and prospectively assessing the risk of HDP [9–12]. However, all kinds of static and single-class conventional prediction methods did not achieve very good model accuracy and effect. Some prediction models had relatively high detection rates for early-onset preeclampsia. Nevertheless, the detection rate was not ideal for late-onset preeclampsia or gestational hypertension.

With the extensive establishment of the hospital database, data mining and machine learning are gradually applied to clinical research. HDP prediction models at different gestational weeks could be established by effectively applying hospital medical record data and combining machine learning algorithms.

2. Materials and methods

2.1. Subjects and specimens

In this study, the whole gestation period was divided into five stages: pregnancy $\leqslant$ 13 weeks, pregnancy 14–20 weeks, pregnancy 21–27 weeks, pregnancy 28–34 weeks, pregnancy $\geqslant$ 35 weeks [13]. The subjects of this study were 507 pregnant women at the Beijing Obstetrics and Gynecology Hospital from 2006 to 2008, and 183 pregnant women at the Beijing Haidian Maternal and Children Health Hospital from 2015 to 2016.

The inclusion criteria were: (a) Pregnant women without chronic hypertension, heart disease, anemia or other chronic diseases; (b) Pregnant women without long-term oral drugs; (c) No fetal malformation.

From the first maternity examination of pregnant women, the radial artery pulse wave waveform of each pregnant woman who went to the hospital for prenatal examination was collected and tracked. After the delivery, according to the final diagnosis of doctors, the subjects were divided into the HDP group or normal group.

Retrospective methods were used to collect the maternal medical records of pregnant women at different gestational weeks, including outpatient medical records, admission records, discharge records and delivery records. The demographic data of pregnant women, blood routine and biochemical examination during pregnancy were collected. The basic information of the pregnant women is shown in Table 1.

As shown in Table 2, this study selected easily accessible clinical predictors, which were all high-risk factors for HDP with statistical differences. Most of the clinical epidemiological parameters were obtained by prenatal examination. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) of pregnant women was obtained from prenatal examination data at various gestational weeks. The characteristic parameters of pulse wave were obtained by detecting the pulse wave of radial artery. Radial artery pulse wave detection at the Beijing Obstetrics and Gynecology Hospital was obtained by MP HDP detection instrument developed by Beijing Yes Medical Devices Co. Ltd. The eight-channel PowerLab

Table 1
Basic information of the pregnant women

Parameters	HDP group	Normal group	P
Number	300	390	–
Detection times	561	1102	–
Age	29.9 ± 4.2	29.6 ± 3.5	0.37
Height (cm)	162.1 ± 5.2	162.2 ± 5.1	0.93
Pre-pregnancy weight (kg)	62.9 ± 12.1	55.1 ± 7.6	0.00

Note: $P < 0.05$ has significant difference.

Table 2
Risk factors of HDP

Type	Risk factors
Clinical epidemiological factors	Pre-pregnancy BMI, pregnant BMI, a multiple pregnancy, history of spontaneous abortion, history of HDP.
Hemodynamic factors	SBP, DBP, PP, MAP, Waveform Area Parameters, CI, TPR.
Biochemical factors	Platelet, Hematocrit, MPV, Cr, UA.

Notes: BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; PP, pulse pressure. MAP, mean arterial pressure; CI, Cardiac output index; TPR, Total peripheral resistance; MPV, Mean platelet volume; Cr, Creatinine; UA, Uric acid. Between the HDP group and normal group, P values of all risk factors were less than 0.05. $P < 0.05$ has significant difference.

data acquisition system, LabChart 8 software and strain gauge pressure sensor were used to collect radial artery pulse wave at the Beijing Haidian Maternal and Child Health Hospital. Biochemical parameters were obtained by blood routine examination and biochemical examination. SPSS 23.0 software was used for statistical basic analysis and support vector machine (SVM) was used to construct the predictive model of HDP in MATLAB 2018a.

2.2. Statistical analyses

In 1995, Vapnik and Cortes proposed a new pattern recognition algorithm: support vector machine (SVM). SVM has special advantages in pattern recognition of small samples, non-linear and high-dimensional data [14]. It can find the best compromise between the complexity of the model and the learning ability to obtain the best generalization ability in the case of limited training data. Compared with other traditional pattern recognition methods, SVM requires relatively less sample data.

The algorithm used was non-linear SVM. Input: Training Data Set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$, where $x_i \in R^n$, $y_i \in \{-1, +1\}$, $i = 1, 2, \ldots, N$; Output: classification decision function.

(a) Selecting appropriate kernel function $K(x, z)$ and parameter C to construct and solve optimization problems. $\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i$, s.t. $\sum_{i=1}^{N} \alpha_i y_i = 0$, where $0 \leqslant \alpha_i \leqslant C$, $i = 1, 2, \ldots, N$. Finding the optimal solution: $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*)^T$.

(b) Selecting a positive component $0 < \alpha_j^* < C$ of α^* and calculate $b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i \cdot x_j)$.

(c) Constructing decision function: $f(x) = sign\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x \cdot x_i) + b^*\right)$.

In this study, Libsvm toolbox developed by Prof. Lin Chih-Jen of National Taiwan University was selected as a pattern recognition classifier. Using RBF kernel function to solve the multi-classification problem. At different gestational weeks, the risk factors of HDP were selected as input of the SVM classifier. After 5-fold cross-validation of the training sets, the effect of models at all gestational weeks was obtained.

Table 3
Basic information of the selected pregnant women

Parameters	HDP group	Normal group	P
Number	300	390	–
Detection times	300	390	–
Age	29.9 ± 4.2	29.6 ± 3.5	0.37
Height (cm)	162.1 ± 5.2	162.2 ± 5.1	0.93
Pre-pregnancy weight (kg)	62.9 ± 12.1	55.1 ± 7.6	0.00

Note: $P < 0.05$ has significant difference.

Table 4
Performance of prediction models for HDP with different types of risk factors

Factors	Training set AC-avg
E	75.10%
H	77.03%
B	78.76%
E + H	82.63%
E + B	83.98%
H + B	84.94%
E + H + B	87.64%

Notes: E, Epidemiology; H, Hemodynamics; B, Biochemistry.

Table 5
Training models and test results of each gestational week

Models	Training set AC-avg
Pregnancy $\leqslant$ 13 weeks sub-model	80.73%
Pregnancy 14–20 weeks sub-model	81.93%
Pregnancy 21–27 weeks sub-model	84.42%
Pregnancy 28–34 weeks sub-model	92.10%
Pregnancy $\geqslant$ 35 weeks sub-model	92.95%

3. Results

The total number of pregnant women included in this study was 690, who were divided into the HDP group (300 women) and normal group (390 women). Detection times were randomly selected for each woman during the different gestational weeks, as can be seen in Table 3.

There are two batches of data in the data collection process. Five hundred and seven subjects gave birth at Beijing Obstetrics and Gynecology Hospital from 2006 to 2008 and 183 subjects gave birth at Beijing Haidian Maternal and Children Health Hospital from 2015 to 2016. The ratio was 3 to 1. In Table 4, the training data set which belonged to the first batch of data were cross-validated five times by combining three kinds of risk factors. As shown in Table 4, the predictive value of these factors of HDP was explored.

As shown in Table 4, with the increase of risk factors of HDP, the predictive effect improved. In the five sub-models provided in Table 1, 5-fold cross-validation of the training sets which randomly accounted for 75% of the total number of each sub-model were carried out respectively (Table 5).

As shown in Table 5, the average accuracy of the model gradually increased when the pregnancy progressed, and the overall accuracy was more than 80%. Especially in late pregnancy, the average accuracy of 28-34 weeks and $\geqslant$ 35 weeks models could reach more than 92%. The highest average accuracy of the model without considering gestational weeks was only 87.64% (Table 4).

4. Discussion

Nicolaides et al. established risk models based on maternal factors, uterine artery pulsation index, and more at 11–13 weeks, 19–24 weeks, 30–34 weeks and 35–37 weeks [15–18]. Lim et al. recorded the blood vessel activity factor and uterine artery pulsation index of pregnant women in the four stages of 11–14, 18–22, 28–32 and 34 weeks of pregnancy, and measured the central artery pressure of pregnant women by radial artery pulse wave to explore the changes of arteriosclerosis [19]. Gómez et al. tracked the uterine artery pulsation index at 11–41 weeks [20]. This study aimed to establish a reference range of pulsation index based on gestational weeks. We furthermore aimed to accurately assess the risk of HDP, which required comprehensive measurements of multiple risk factors for pregnant women at all gestational weeks.

In this study, support vector machine (SVM) was used to establish the predictive model of HDP. Combined prediction of epidemiology, hemodynamics and biochemistry was helpful to improve the model accuracy, and the discriminant effect was better than that of single prediction. The prediction accuracy of the model considering gestational weeks increased gradually with the progress of pregnancy, which was better than the model without considering this factor.

5. Conclusion

Based on the literature, the risk factors of HDP are summarized as clinical epidemiological factors, hemodynamic factors and biochemical factors. The dynamic risk assessment model of HDP by SVM was established with the change of gestational weeks. Continuous testing could be carried out from early to late pregnancy for early detection of risk and intervention. The predictive model proposed in this study could be applied to the risk assessment of all kinds of HDP. In future studies, the model needs to be refined according to different diseases.

Acknowledgments

This work was supported by National Key R&D Program of China (no. 2019YFC0119700), Bill & Melinda Gates Foundation (no. OPP1148910), Beijing Natural Science Foundation (no. 7172015) and Intelligent Physiological Measurement and Clinical Translation, Beijing International Base for Scientific and Technological Cooperation.

Conflict of interest

None to report.

References

[1] Magee LA, Pels A, Helewa M, Rey E, von Dadelszen P. SOGC Hypertension Guideline Committee. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy: executive summary. J Obstet Gynaecol Can. 2014; 36: 575-576.

[2] Pennington KA, Schlitt JM, Jackson DL, Schulz LC, Schust DJ. Preeclampsia: Multiple approaches for a multifactorial disease. Disease Models & Mechanisms. 2012; 5: 9-18.

[3] Sunsaneevithayakul P, Titapant V, et al. Relation between gestational weight gain and pregnancy outcomes. J Obstet Gynaecol Res. 2014; 40(4): 995-1001.
[4] Gongora MC, Wenger NK. Cardiovascular complications of pregnancy. International Journal of Molecular Sciences. 2015; 16(10): 23905-23928.
[5] Åafak Ã, Başer E, et al. Predictivity of mean platelet volume in severe preeclamptic women. Clinical & Experimental Hypertension Part B Hypertension in Pregnancy. 2016; 35(4): 474-482.
[6] Poon LC, Stratieva V, Piras S, Piri S, Nicolaides KH. Hypertensive disorders in pregnancy: combined screening by uterine artery Doppler, blood pressure and serum PAPP at 11–13 weeks. Prenat Diagn. 2010; 30(3): 216-223.
[7] Myers JE, Kenny LC, Mccowan LM, et al. Angiogenic factors combined with clinical risk factors to predict preterm pre-eclampsia in nulliparous women: a predictive test accuracy study. Bjog An International Journal of Obstetrics & Gynaecology. 2013; 120(10): 1215-1223.
[8] Jaana N, Teemu K, et al. Performance of first trimester biochemical markers and mean arterial pressure in prediction of early-onset pre-eclampsia. Metabolism-clinical and Experimental. 2017; 75: 6-11.
[9] Cahit B, Janina F, et al. Maternal serum anti-Müllerian hormone at 11–13 weeks' gestation in the prediction of preeclampsia. The Journal of Maternal-fetal & Neonatal Medicine. 2015; 28(8): 865.
[10] Lehnen H, et al. Predictive Value of the sFlt-1: PlGF Ratio in Women with Suspected Preeclampsia. The New England Journal of Medicine. 2016; 374(1): 13-22.
[11] Katja M, et al. Prediction of pre-eclampsia and its subtypes in high-risk cohort: hyperglycosylated human chorionic gonadotropin in multivariate models. Bmc Pegnancy and Childbirth. 2018; 18(279).
[12] Sana S, et al. Evaluation of Soluble TNF-like weak inducer of apoptosis (sTWEAK) Levels to Predict Preeclampsia in Early Weeks of Pregnancy. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2019; 234: 165-170.
[13] Meah VL, Cockcroft JR, et al. Cardiac output and related haemodynamics during pregnancy:a series of meta-analyses. Heart. 2016; 102: 518-526.
[14] Kil DH, Shin FB. Pattern recognition and prediction with applications to signal processing. New York: Springer-Verlag, 1998.
[15] O'Gorman N, Wright D, et al. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11–13 weeks gestation. American Journal of Obstetrics & Gynecology. 2016; 214(1): 103.e1-103.e12.
[16] Gallo DM, Wright D, et al. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 19–24 weeks' gestation. American Journal of Obstetrics & Gynecology. 2016; 214(5): 619.e1-619.e17.
[17] Tsiakkas A, Saiid Y, et al. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 30–34 weeks' gestation. American Journal of Obstetrics & Gynecology. 2016; 215(1): 87.e1-87.e17.
[18] Andrietti S, Silva M, et al. Competing-risks model in screening for pre-eclampsia by maternal factors and biomarkers at 35–37 weeks' gestation. Ultrasound in Obstetrics & Gynecology. 2016; 48(1): 72-79.
[19] Lim WY, Saw SM, et al. A cohort evaluation on arterial stiffness and hypertensive disorders in pregnancy. BMC Pregnancy and Childbirth. 2012; 12(1): 160-167.
[20] Gómez O, Figueras F, et al. Reference ranges for uterine artery mean pulsatility index at 11–41 weeks of gestation. Ultrasound in Obstetrics & Gynecology. 2008; 32(2): 128-132.