19th World Congress in Fetal Medicine

Machine learning algorithms for the prediction of preterm preeclampsia

Torres J, Martinez RJ, Espino S, Mateu P, Solis JM, Estrada G, Rojas L, Villafan JR.
Instituto Nacional de Perinatologia, Mexico City, Mexico

Objective

Our objective was to use statistical machine learning to analyze clinical and laboratory data obtained during 11-13.6 weeks of pregnancy and to use them to develop a prediction model for preterm preeclampsia.

Methods

This is a prospective cohort of non-selected population in Mexico City developed exclusively for research purposes in a first-level clinic of women recruited between April 2019 and October 2021. Inclusion criteria were age >18 years and singleton pregnancy at the 11-13.6-week scan. The protocol was approved by the Ethics and Research Internal Review Board of the National Institute of Perinatology (2021-1-38) and was conducted ethically under the World Medical Association Declaration of Helsinki. All enrolled women authorized and provided signed informed consent. The primary outcome was the detection rate (DR) of the machine learning model in the prediction of pPE in the Mexican population. Secondary outcomes were the DR of the logistic regression model in the prediction of pPE and the performance comparison between both models. 75/35 training and validation data sets were created. We performed an elastic net, a regularized regression method that linearly combines the ridge and lasso regression. The elastic net typically outperforms classic regression methods by performing both shrinkage and automatic selection of predictors. Due to the automatic selection of predictors achieved by penalty, no previous subset selection, which has typically been used in previous methods, must be performed, thereby reducing the variance and instability of the prediction model. Automatic selection of predictors performed in elastic net results in a simpler, sparse model that includes only a subset of variables, thereby allowing for better interpretation of the model. Ten-fold cross-validation was used to determine the shrinkage parameter of the elastic net. The second model was created by nested logistic regression adding anthropometric variables, serum, and ultrasound biomarkers to a previous maternal history model using a stepwise method for variable selection. The performance of the models created was assessed at 5% and 10% false positive rates (FPR) and compared using the area under the curve (AUC), and the calibration of ML algorithm was evaluated by Hosmer-Lemeshow goodness-of-fit test. Both models were centered at 37 weeks for the prediction of preterm (pPE).

Results

Description of the cohort and characteristics of the study population. A total of 3067 pregnant women were enrolled in the original cohort and 247 women (8.05%) were excluded due to incomplete data. 2820 pregnant women were included in the final analysis. Among the included women, 115 (4.07%) developed PE, of which 72 (2.6%) were delivered before 37 weeks of gestation. Development of Machine Learning Model for Preterm Preeclampsia Prediction The prediction model was developed using elastic nets. During cross-validation, 36 elastic net models were built based on the training data in each run. Elastic net modelling performed variable selection along with regularization, each resulting in a specific elastic net model containing a subset of input features. During the modeling process, the 10-fold cross-validation optimization model was performed and applied to the test set, respectively. The results of this analysis showed that PlGF, MAP, UtA PI, BMI, antiphospholipid syndrome, PE in previous pregnancy, preexisting diabetes, smoker, spontaneous pregnancy, other drugs, PAPP-A, lupus, chronic hypertension, and maternal age were the most important input variables to predict pPE. We runed fifty iterations comparing each model performance. The optimal stable training model had an AUC of 0.870; the validation model showed an AUC of 0.905 with a detection rate (DR) of 0.735 and a FPR of 10%, the Hosmer-Lemeshow test was non-significant ($p = 0.114$), indicating that the model predicted the probability of pPE and the observed probability of pPE fit well. For comparison, we fitted a prediction model via nested logistic regression using the same covariates. Similar performance was obtained, which we expect is because most of the risks that have the highest impact on the prediction (antiphospholipid syndrome, history of preeclampsia, preexisting diabetes, spontaneous pregnancy, lupus, chronic hypertension, and maternal age) are categoric variables. The logistic model performance had AUC of 0.920, DR of 0.735 and FPR of 10%.

Conclusion

One of the advantages of the elastic net is its ability to perform automatic choice of predictors among all available variables, based on a patient's characteristics, the prediction model calculates the score for each patient that reflects her chance of experiencing preeclampsia. This eliminates the need for the suboptimal approach of a priori choosing predictors for the model. The findings that are reported here demonstrate that machine learning is a well-suited method for the prediction of preeclampsia.