Machine-learning prediction of pre-eclampsia using first trimester maternal characteristics and biomarkers

Zuria Ansbacher[1], Argyro Syngelaki[2], Hamutal Meiri[3],

Kypros H. Nicolaides[2], Yoram Louzoun[1]

1. Department of Mathematics, Bar Ilan University, Ramat Gan, Israel
2. Fetal Medicine Research Institute, King's College Hospital, London, United Kingdom
3. The ASPRE Consortium and TeleMarpe, Tel Aviv Israel

Corresponding Author: Prof. Kypros H. Nicolaides, The Fetal Medicine Research Institute, 16-20 Windsor Walk, London SE5 8BB, United Kingdom, Phone: +44 20 3299 3246, mail: kypros@fetalmedicine.com

Running Title: Machine-learning-based prediction of pre-eclampsia

Keywords: pre-eclampsia, first trimester, placental growth factor, mean arterial pressure, uterine artery Doppler, machine learning, artificial intelligence, neural network, prior risk, posterior risk, ASPRE study.

ABSTRACT

Objective: To evaluate the accuracy of predicting the risk to develop pre-eclampsia (PE) according to first-trimester maternal characteristics, medical history and biomarkers using artificial intelligence and machine learning methods. The prediction is performed using raw values that are not standardized.

Methods: The data were derived from prospective non-intervention screening for PE at 11-13 weeks' gestation in two maternity hospitals in the UK. The data were divided into three subsets. The first set composed of 35,437 subjects was used to develop the training process, the second set of 5,000 subjects was utilized to optimize the machine learning hyperparameters and a third set of 20,352 subjects was coded and used for the validation. An artificial neural network was used to predict from the demographic characteristics and medical history the prior risk that was then combined with biomarker values to determine the risk of all cases of PE, and preterm PE with delivery at <37 weeks' gestation. An additional network was trained without the race input. Biomarkers included uterine artery pulsatility index (UtA-PI), mean arterial blood pressure (MAP), placental growth factor (PlGF), and pregnancy-associated placental protein A (PAPP-A). All markers were entered using raw values not converted into standardized multiples of the median (MoMs). The prediction accuracy was estimated using the area under the receiver operator characteristic (ROC) curve (AUC). We further computed the detection rate for 10%, 20%, and 40% false positive rates (FPR). We used a non-parametric test to compare the expected AUC when we randomly scrambled the labels and kept the predictions. For the general prediction, we performed 10,000 permutations of the labels. When the AUC was higher than the one obtained in all 10,000 permutations, we reported a p value of <0.0001. For the race-specific analysis, we performed 1,000 permutations. When the AUC was higher than all permutations, we reported a p value of <0.001.

Results: The detection rate of preterm PE, at 10% FPR, was 45% in screening by maternal factors and this increased to 73% with addition of biomarkers with AUC above 0.9. The race information was important for this prediction; when the race input was removed from the predictor, the detection rate decreased to 37-43% in screening by maternal factors and 55-60% with addition of biomarkers. The accuracy of prediction of all cases of PE was lower than that of preterm PE. The AUC was 0.76 in screening by maternal factors and 0.82 with addition of biomarkers; the respective detection rates, at 10% FPR, were 35% and 48%.

Conclusion: The performance of screening for PE of a non-linear approach with no need for a population-based normalization is similar to that of logistic regression.

CONTRIBUTION

What are the novel findings of this work?

Non-linear classifiers can be used in combination with maternal risk factors and non-normalized first trimester biomarkers to predict preterm pre-eclampsia (PE) with high accuracy. The incidence of PE and properties are race dependent and ignoring the race information significantly reduces the prediction accuracy in general, and further so for non-White populations.

What are the clinical implications of this work?

This work opens the way to a transferable PE prediction for women different than the now-standard approach. This will allow for wider usage of first trimester preterm PE prediction.

INTRODUCTION

Pre-eclampsia (PE) is a major cause of maternal and fetal morbidity and mortality.[1,2] First-trimester screening for PE by a combination of maternal characteristics and medical history with measurements of mean arterial pressure (MAP), uterine artery pulsatility index (UtA-PI), serum placental growth factor (PlGF), and serum pregnancy-associated plasma protein-A (PAPP-A) could predict about 75% of preterm PE, with delivery at <37 weeks' gestation and 40-45% of term PE, at 10% false positive rate (FPR).[3-5] Treatment of the high-risk group with aspirin (150 mg/day) from 12 to 36 weeks of gestation reduces the rate of preterm PE by about 60%.[6]

The method of screening for PE developed by the Fetal Medicine Foundation, the competing risks approach, assumes that every woman has a personalized distribution of gestational age at delivery with PE; whether she experiences preeclampsia or not before a specified gestational age depends on competition between delivery before or after the development of preeclampsia.[5] The distribution of biomarkers is specified conditionally on the gestational age at delivery with PE, where the values of UtA-PI, MAP, PlGF and PAPP-A are expressed as a multiple of the median (MoM) after adjustment for various maternal factors and gestational age that were found to provide a substantive contribution to the $\log_{10}$ transformed values[7-10]. However, MoM-based methods require detailed information on the distributions of all measures in a large enough cohort for the prediction, that are often lacking in many populations. In addition, when applied to biochemical markers, the conversion to MoM has to be adjusted to different batches, manufacturers, and analyzers, which has to be repeatedly renewed. Recently, artificial intelligence, machine learning, and deep learning methods have attracted strong interest around the world, and these methods have been already tested for their use in the diagnosis and prediction of many prenatal complications, such as Down Syndrome, various structural anomalies identified by ultrasound or Autism Spectrum Disorder.[11-14] In such studies, learning from dataset patterns enabled artificial intelligence and machine learning methods to identify interactions between variables and outcomes, not accessible by linear methods.[14,15]

The objective of this study was to examine the potential value of neural networks for the prediction of PE by a combination of maternal factors and biomarkers obtained at 11-13 weeks gestation without using MoMs.

METHODS

Study population

The data were derived from prospective screening for adverse obstetric outcomes in women attending for their routine first-trimester hospital visit in pregnancy at King's College Hospital and Medway Maritime Hospital, UK. These visits, which were held at 11^{+0} -13^{+6} weeks' gestation, included first, recording of maternal characteristics and medical history,[3] second, transabdominal ultrasound for measurement of the left and right UtA-PI by color Doppler and calculation of the mean PI,[18] third, measurement of MAP by validated automated devices and standardized protocol,[19] and fourth, measurement of serum concentration of PLGF and PAPP-A using a DELFIA Xpress system (PerkinElmer Life and Analytical Sciences, Waltham, MA, USA) or Cobas e411 system (Roche Diagnostics, Penzberg, Germany). The women gave written informed consent to participate in the study, which was approved by the NHS Research Ethics Committee.

The inclusion criteria for this study were singleton pregnancy undergoing first-trimester combined screening for aneuploidy and subsequently delivering a phenotypically normal live birth or stillbirth at ≥24 weeks' gestation. We excluded pregnancies with aneuploidies and major fetal abnormalities and those ending in termination, miscarriage or fetal death before 24 weeks.

Outcome measures were preterm PE, with delivery at <37 weeks' gestation and term PE with delivery at ≥37 weeks. Data on pregnancy outcomes were collected from the hospital maternity records or the general medical practitioners of the women. The obstetric records of all women with pre-existing or pregnancy-associated hypertension were examined to determine if the condition was PE, as defined by the American College of Obstetricians and Gynecologists (ACOG).[2] According to this definition, diagnosis of PE requires the presence of new-onset hypertension (blood pressure ≥ 140 mmHg systolic or ≥ 90 mmHg diastolic) at ≥ 20 weeks gestation or chronic hypertension and either proteinuria (≥ 300 mg/24 h or protein-to-creatinine ratio > 30 mg/mmol or ≥ 2+ on dipstick testing) or evidence of renal dysfunction (serum creatinine > 97 μmol/L), hepatic dysfunction (transaminases ≥ 65 IU/L) or hematological dysfunction (platelet count < 100 000/μL).

Machine learning

The input data were Z-scored. Categorical parameters were translated to a one-hot representation and not normalized. For the prediction we used a feed-forward neural network with 2 hidden layers. The activation function, which is commonly used in neural networks was a Rectified Linear Unit (Relu); this is defined as y = max (0, x). Dropout was applied to the second layer. An Adam optimizer was used[17]. The loss function was binary-cross-entropy with logits (weighted). All the machine learning was performed with the Pytorch. The following hyperparameters were tuned, by a grid search, implemented via NNI - an automatic tool for hyperparameters tuning, which optimizes Machine Learning performance (https://github.com/microsoft/nni) to determine the batch size, the learning rate, the dropout rate, the sizes (number of neurons) in each hidden layer, the activation function, and the weight decay. The tuning was done separately for the prior-and-posterior risks-based predictions. The tuning was done on an internal validation set, and the results are reported on a test set, not available at the time of the tuning. The dataset was split into a first subset of training prepared from the data of 35,437 subjects, internal validation of a subset of 5,000 subjects, and a test subset of 20,352 subjects. While the data of the outcome of the training sub-set were disclosed, for the validation of the final data subset, the outcome data were coded and unknown to the team in Israel that conducted the machine learning analysis. The tuning was performed on the Area Under Curve (AUC) made of the sensitivity and the specificity on the internal validation.

Statistical analysis

We used two methods for evaluation. The prediction accuracy was estimated using the ROC curve AUC. We further computed the detection rate (the Recall) as a function of total fraction defined as positive. Since the total fraction of PE in the population is low, our goal was to minimize the fraction of women defined to be positive but maximize the recall. The p values reported are the probability that the results are random. We used a non-parametric permutation test and compared the expected AUC to the one obtained when we randomly scrambled the labels of each sample, but kept its predicted score. For the general prediction, we performed 10,000 permutations of the labels. When the AUC was higher than the one obtained in all 10,000 permutations, we reported a p-value of <0.0001. For the race-specific analysis, we performed 1,000 permutations. When the AUC was higher than all permutations, we reported a p-value of <0.001.

Experimental setup

We performed multiple tests and in all tests we used the same training/validation and test division. We did the prediction either using or ignoring the race input. When we ignored the race input, the prediction was performed on the entire dataset, but the test was done on each race separately.

In all cases, we analyzed the following combinations that were tested independently: (a) PE vs.no PE, (b) preterm PE vs. no PE, (c) preterm PE vs. no PE OR term PE, and (d) preterm PE vs. term PE. When preterm PE was compared to no PE, the term PE cases were ignored in both the training and the test data subsets.

RESULTS

Characteristics of the study population

The study population of 60,789 pregnancies included 1,736 (2.9%) subjects that developed PE. The characteristics of the study population are summarized in Table 1. In women who developed PE, compared to those who did not, there was a higher body mass index and interpregnancy interval, a larger proportion of women of black race, a higher incidence of chronic hypertension, diabetes mellitus Type 1, systemic lupus erythematosus or antiphospholipid syndrome, family history of PE, and conception through assisted fertility method and a lower incidence of smoking.

Performance of screening for pre-eclampsia

The data were separated into training, internal validation, and external test validation sets. The training set was the input of an artificial neural network to predict three independent tasks (Figure 1 for experiment flowchart): first, PE vs. no PE, second, preterm PE vs. no PE (data of term PE were omitted), and third, preterm PE vs. everything else (no PE plus term PE). The internal validation set was used for tuning hyper-parameters to maximize the AUC of the internal validation. The trained model was then applied to the test sub-set, and we report here these results.

The detection rates, at FPR of 10%, 20%, and 40%, on the test sets are described in Table 2, and the AUC ROC are presented in Figure 2. This prediction was performed separately on prior risk data (all the demographic and medical and pregnancy history data) and the posterior data with addition of biomarker values. The posterior risk included data of PAPP-A, PlGF, MAP and UtA-PI. Consistently, the accuracy increased when the posterior values were added. Yet, even without using those, the artificial neural network predicts PE in

noteworthy success. Specifically, the AUC for preterm PE increased from 0.8 to 0.915 when the posterior information was added and detection rate, at 10% FPR, increased from 45% to 73%.

The results on the influence of the mother race on test accuracy are shown in Table 3. The races of the study population were White, Black, South Asian, East Asian, and Mixed. However, the number of positive cases for South Asian, East Asian, and Mixed are too low and are thus not reported. Removing the race consistently significantly reduced the accuracy of all predictors. For example, the AUC of preterm PE that was based on prior information dropped to 0.75 and the detection rate at 10% FPR was reduced to 37%. When comparing populations, the accuracy for the white population was higher than the black population, consistently on all classifiers.

DISCUSSION

Main findings

In this first trimester screening study the approach and methodology of artificial intelligence and machine learning with the assistance of neural network algorithms was used for predicting the risk for subsequent development of PE. There were two main findings: first, at 10% FPR, the prediction of preterm PE was 45% in screening by maternal characteristics and medical history and this increased to 73% after addition of biomarkers, and second, inclusion of race in the prediction algorithm was important, because when this was not included the detection rate, at 10% FPR, of combined screening was reduced to 55-60%.

Comparison with results of previous studies and implications for clinical practice.

First trimester prediction of preterm PE is important because treatment of the high-risk group with aspirin (150 mg/day from 12 to 36 weeks' gestation) reduces the rate of early PE with delivery at <32 weeks by about 90% and preterm PE by about 60%.[4,6,20] Consequently, early prediction and prevention of PE was been adopted in the guidelines of the International Society for the Study of Hypertension in Pregnancy[1] and the International Federation of Gynecology and Obstetrics (FIGO).[21]

The predictive performance for preterm PE using artificial intelligence and machine learning methods was similar to that achieved by the competing risk model. [3-5,22,23] The advantage of the machine learning approach is use of raw biomarker data without the need for conversion into MoMs, which would simplify the implementation of screening. Additionally, calculators from the machine learning approach can be easily and rapidly introduced through an automated way with use of cloud-based or any other on-line tools.

Strengths and limitations

The main strength of the study was the large population derived from prospective screening for PE, recording all the important factors in maternal demographic characteristics and medical history known to be associated with PE, measurement of MAP and UtA-PI with use of standardized protocols and appropriately trained practitioners, and measurement of PlGF and PAPP-A within 30 minutes of collection with automated machines that are calibrated on a daily basis.

The limitation of the study is that the prediction algorithm has not been tested in other populations. For example, our finding of the large influence of race on the accuracy of the prediction for PE demonstrates a limitation to our study as the race element introduces a bias towards the predominating race in the study population. Consequently, adjustments to the algorithm are necessary when testing other populations.

Conclusion

A novel automated machine learning approach was found useful and accurate in the first trimester prediction of preterm PE. The study demonstrated the importance of taking into account race in the prediction of PE.

Conflict of Interest: None

Funding Resources: ZA was sponsored by Bar Ilan University Graduate program. The screening study for preeclampsia was supported by grants from the Fetal Medicine Foundation (UK Charity No: 1037116).

REFERENCES

1. Magee LA, Brown MA, Hall DR, Gupte S, Hennessy A, Karumanchi SA, Kenny LC, McCarthy F, Myers J, Poon LC, Rana S, Saito S, Staff AC, Tsigas E, von Dadelszen P. The 2021 International Society for the Study of Hypertension in Pregnancy classification, diagnosis & management recommendations for international practice. *Pregnancy Hypertens* 2021; ***27***: 148–169.

2. American College of Obstetricians and Gynecologists. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstet Gynecol* 2020; **135**, e237–e260.

3. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. *Am J Obstet Gynecol* 2015; **213**: 62.e1-62.e10.

4. O'Gorman N, Wright D, Syngelaki A, Akolekar R, Wright A, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation. *Am J Obstet Gynecol* 2016; **214**: 103.e1-103.e12.

5. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. *Am J Obstet Gynecol* 2020; **223**: 12-23.e7.

6. Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, de Paco Matallana C, Akolekar R, Cicero S, Janga D, Singh M, Molina FS, Persico N, Jani JC, Plasencia W, Papaioannou G, Tenenbaum-Gavish K, Meiri H, Gizurarson S, Maclagan K, Nicolaides KH. Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. *N Engl J Med* 2017; **377**: 613-622.

7. Tayyar A, Guerra L, Wright A, Wright D, Nicolaides KH. Uterine artery pulsatility index in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 689-97.

8. Wright A, Wright D, Ispas CA, Poon LC, Nicolaides KH. Mean arterial pressure in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 698-706.

9. Wright D, Silva M, Papadopoulos S, Wright A, Nicolaides KH. Serum pregnancy-associated plasma protein-A in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **46**: 42-50.

10. Tsiakkas A, Duvdevani N, Wright A, Wright D, Nicolaides KH. Serum placental growth factor in the three trimesters of pregnancy: effects of maternal characteristics and medical history. *Ultrasound Obstet Gynecol* 2015; **45**: 591-598.

11. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018; **4**: e00938.

12. He F, Lin B, Mou K, Jin L, Liu J. A machine learning model for the prediction of down syndrome in second trimester antenatal screening. *Clin Chim Acta* 2021; **521**: 206-211.

13. Matthew J, Skelton E, Day TG, Zimmer VA, Gomez A, Wheeler G, Toussaint N, Liu T, Budd S, Lloyd K, Wright R, Deng S, Ghavami N, Sinclair M, Meng Q, Kainz B, Schnabel JA, Rueckert D, Razavi R, Simpson J, Hajnal J. Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. *Prenat Diagn* 2022; **42**: 49-59.

14. Caly H, Rabiei H, Coste-Mazeau P, Hantz S, Alain S, Eyraud JL, Chianea T, Caly C, Makowski D, Hadjikhani N, Lemonnier E, Ben-Ari Y. Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD. *Sci Rep* 2021; **11**: 6877.

15. Zhang H, Mo J, Jiang H, Li Z, Hu W, Zhang C, Wang Y, Wang X, Liu C, Zhao B, Zhang J, Zhang K. Deep learning model for the automated detection and histopathological prediction of meningioma. *Neuroinformatics* 2021; **19**: 393-402.

16. Deo RC. Machine learning in medicine. *Circulation* 2015; **132**: 1920-1930.

17. Zhang, Zijun. Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), pp. 1-2. IEEE, 2018.

18. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11+0 to 13+6 weeks in the prediction of pre-eclampsia. *Ultrasound Obstet Gynecol* 2007; **30**: 742-749.

19. Poon LC, Zymeri NA, Zamprakou A, Syngelaki A, Nicolaides KH. Protocol for measurement of mean arterial pressure at 11-13 weeks' gestation. *Fetal Diagn Ther* 2012; **31**: 42-48.

20. Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. *Am J Obstet Gynecol* 2018; **218**: 287-293.e1.

21. Poon LC, Shennan A, Hyett JA, Kapur A, Hadar E, Divakar H, McAuliffe F, da Silva Costa F, von Dadelszen P, McIntyre HD, Kihara AB, Di Renzo GC, Romero R, D'Alton M, Berghella V, Nicolaides KH, Hod M. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention. *Int J Gynaecol Obstet* 2019; **145**: 1-33.

22. Tan MY, Syngelaki A, Poon LC, Rolnik DL, O'Gorman N, Delgado JL, Akolekar R, Konstantinidou L, Tsavdaridou M, Galeva S, Ajdacka U, Molina FS, Persico N, Jani JC, Plasencia W, Greco E, Papaioannou G, Wright A, Wright D, Nicolaides KH. Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. *Ultrasound Obstet Gynecol* 2018; **52**: 186-195.

23. Wright D, Tan MY, O'Gorman N, Poon LC, Syngelaki A, Wright A, Nicolaides KH. Predictive performance of the competing risk model in screening for preeclampsia. *Am J Obstet Gynecol* 2019; **220**: 199.e1-199.e13.

Table 1. Characteristics of the study population.

Characteristic	Normal (n=59,139)	Pre-eclampsia (n=1,736)	p-value
Maternal age (year)	31.0 (26.6, 34.8)	31.2 (26.7, 35.2)	0.112
Maternal weight (kg)	67.0 (59.2, 78.0)	74.0 (63.9, 87.2)	<0.0001
Maternal height (cm)	165 (160, 169)	164 (159, 168)	<0.0001
Gestational age (day)	89.0 (86.0, 92.0)	89.0 (86.0, 92.0)	0.019
Race			<0.0001
White	43,963 (74.3%)	993 (57.2%)	
Black	9,790 (16.6%)	599 (34.5%)	
South Asian	2,641 (4.5%)	83 (4.8%)	
East Asian	1,230 (2.1%)	24 (1.4%)	
Mixed	1,515 (2.6%)	37 (2.1%)	
Medical history			
Chronic hypertension	630 (1.1%)	215 (12.4%)	<0.0001
Diabetes mellitus type 1	228 (0.4%)	12 (0.7%)	<0.0001
Diabetes mellitus type 2	294 (0.5%)	26 (1.5%)	
SLE/APS	113 (0.2%)	9 (0.5%)	0.006
Smoking	5,667 (9.6%)	101 (5.8%)	<0.0001
Family history of PE	2,257 (3.8%)	136 (7.8%)	<0.0001
Method of conception			<0.0001
Spontaneous	57,258 (96.8%)	1,644 (94.7%)	
In vitro fertilization	1,408 (2.4%)	72 (4.2%)	
Use of ovulation drugs	473 (0.8%)	20 (1.2%)	
Parity			<0.0001
Nulliparous	27,303 (46.2%)	1,008 (58.1%)	
Parous, no previous PE	30,179 (51.0%)	494 (28.5%)	
Parous, previous PE	1,657 (2.8%)	234 (13.5%)	
Interpregnancy interval (year)	3.0 (2.0, 4.9)	3.85 (2.3, 6.7)	<0.0001
Biomarkers			
Biomarkers			
Mean arterial pressure (mm Hg)	86.3 (81.1-91.8)	93.8 (87.8-99.8)	<0.0001
Uterine artery pulsatility index	1.7 (1.3, 2.0)	1.9 (1.5, 2.4)	<0.0001
Placental growth factor (pg/mL)	35.3 (25.9, 49.6)	28.1 (20.4, 40.7)	<0.0001
Pregnancy associated plasma protein-A (IU/L)	2.7 (1.7, 4.2)	2.3 (1.4, 3.8)	<0.0001

Values given in n (%) or median (interquartile range)

PE = preeclampsia; IQR = interquartile range; SLE = systemic erythematosus lupus; APS = antiphospholipid syndrome.

Table 2. Performance of screening for pre-eclampsia on the test set.

Values reported are the probability that the results are random, and were calculated by 10,000 random permutations. Value of <0.0001 means that the reported AUC is higher than all random realizations.

Method of screening	AUC	P value	Detection rate (%)		
			FPR 10%	FPR 20%	FPR 40%
All PE vs. no PE (n=60,789)					
Maternal factors	0.758	<0.0001	35.1	52.4	76.0
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	0.800	<0.0001	48.1	64.4	82.7
Preterm PE vs. no PE (n=59,551) [a]					
Maternal factors	0.799	<0.0001	41.7	64.2	80.8
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	0.902	<0.0001	73.5	85.4	94.0
Preterm PE vs. everything else (n=60,789)					
Maternal factors	0.791	<0.0001	41.1	60.9	82.1
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	0.897	<0.0001	68.2	84.8	94.0

AUC = area under the operator characteristic curve, FPR = fase positive rate, MAP = Mean arterial pressure, PAPP-A = Pregnancy-associated plasma protein-A, PLGF = Placental growth factor, UtA-PI = Uterine artery pulsatility index.

Table 3. Performance of screening for pre-eclampsia on test set, according to maternal race. The analysis were separated to races: White and Black (the other races do not have enough data), and in addition, on the whole data together. The P values were calculated here by 1000 random permutations.

Method of screening	**Race**	**AUC**	**P value**	**Detection rate (%)**		
				FPR 10%	**FPR 20%**	**FPR 40%**
All PE vs. no PE (n=60,789)						
Maternal factors	All	0.742	<0.001	32.8	49.9	74.3
	White	0.739	<0.001	30.5	50.0	73.9
	Black	0.722	<0.001	33.5	45.8	70.0
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	All	0.799	<0.001	43.7	60.3	81.1
	White	0.733	<0.001	39.9	57.2	77.4
	Black	0.826	<0.001	43.3	62.1	85.7
Preterm PE vs. no PE (n=59,551)						
Maternal factors	All	0.755	<0.001	37.1	55.0	77.5
	White	0.773	<0.001	44.4	58.3	77.8
	Black	0.691	<0.001	33.9	43.5	67.7
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	All	0.885	<0.001	60.9	81.5	93.4
	White	0.868	<0.001	58.3	80.6	90.3
	Black	0.882	<0.001	56.5	75.8	95.2
Preterm PE vs. everything else (n=60,789)						
Maternal factors	All	0.753	<0.001	37.7	55.0	75.5
	White	0.769	<0.001	43.1	58.3	79.2
	Black	0.69	<0.001	37.1	40.3	67.7
Maternal factors, PAPP-A, PlGF, MAP, UtA-PI	All	0.880	<0.001	60.3	80.1	93.4
	White	0.865	<0.001	58.3	79.2	90.3
	Black	0.872	<0.001	54.8	72.6	95.2

Figure Legends

Figure 1. Flowchart of the learning process. The data are divided to training, validation and test. The training dataset is used for the training process of the machine, and the validation set is used as external data to check the machine performance. Different combinations of hyper-parameters are checked in this process, and the parameters, which optimize the performance on the validation set are chosen to the final model. Afterwards, the trained model is applied to the test set.

Figure 2. AUC ROC Curve for 3 screenings: (A) PE vs. no PE (B) Preterm PE vs. no PE (C) Preterm PE vs. everything else, i.e. no PE or term PE. Performances of posterior data are presented by the blue curve, and of prior data by the orange curve.