Review

A review on Machine Learning Deployment Patterns and Key Features in the Prediction of Preeclampsia

Louise Pedersen *, Magdalena Mazur-Milecka, Jacek Ruminski, Stefan Wagner *

Posted Date: 11 June 2024

doi: 10.20944/preprints202406.0725.v1

Keywords: deployment pattern; machine learning; prediction; preeclampsia; risk assessment; review

 doi:10.20944/preprints202406.0725.v1

Review

A Review on Machine Learning Deployment Patterns and Key Features in the Prediction of Preeclampsia

Louise Pedersen [1,*], Magdalena Mazur-Milecka [2], Jacek Ruminski [2] and Stefan Wagner [1,*]

[1] Aarhus University, Department of Electrical and Computer Engineering, Denmark
[2] Gdansk University of Technology, Department of Biomedical Engineering, Poland
* Correspondence: ltp@ece.au.dk (L.P.); sw@ece.au.dk (S.W.)

Abstract: Previous reviews have investigated machine learning (ML) models used to predict the risk of developing preeclampsia but have not described how the ML models are intended to be deployed throughout pregnancy or feature performance. The aim of this study is to provide an overview of the existing ML models and their intended deployment patterns and performance along with identified features of high importance. This review used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 guidelines. PubMed, Engineering Village, and the Association for Computing Machinery were searched between January and February 2024. A total of 86 studies were found of which 14 were included. Out of 12 studies, eight showed the intent to use the ML model as a single-use, two intended a dual-use, and two intended multiple-use. A total of seven studies listed the features of the highest importance. Systolic and diastolic blood pressure were listed along with mean arterial pressure to be of high importance. Out of four studies intending to use the ML model more than a single-use, three of them were conducted in the years 2023 and 2024, whereas the remaining study is from 2011. No ML model emerged as superior across the subgroups of PE. Utilizing body mass index and either mean arterial pressure or diastolic blood pressure and systolic blood pressure may benefit the performance. The deployment patterns are mainly single use being within the gestation weeks 11+0 to 14+1.

Keywords: deployment pattern; machine learning; prediction; preeclampsia; risk assessment; review

1. Introduction

Preeclampsia (PE) is a pregnancy-related disorder that affects 2-8% of all pregnancies worldwide, contributing to severe morbidity of the women and the baby. Together with eclampsia, it is responsible for 10-15% of maternal deaths in countries of low- and middle-income [1]. When diagnosed the only cure is delivery of the baby and placenta [2]. In women with an increased risk of PE, early administration of aspirin has shown promise in reducing preterm PE (onset before 37 gestational weeks) by up to 62% when the treatment is initiated before gestational week 16 [3]. Consequently, there is considerable interest in risk assessment of PE before week 16 of gestation, to minimize the incidence of preterm PE and thereby the severe morbidity and mortality rates.

The Fetal Medicine Foundation (FMF) has developed a competing risk model for PE, which is widespread as a decision support tool for first-trimester screening for PE [2,4]. The competing risk model combines maternal factors, mean arterial pressure (MAP), pulsatility index of the blood flow in the uterine arteries (UtA-PI), placental growth factor (PlGF), and pregnancy-associated plasma protein A (PAPP-A) [5]. The full feature list for FMF is provided in Appendix A. While typically used as a one-step model, FMF can also be used as a two-step model. The first step involves maternal factors and MAP with a 50% screen-positive rate (SPR) followed by the second step involving UtA-PI and PlGF. Completing the first-trimester screening in two steps with 50% of the pregnant population included in the second step yielded comparable results [6]. This approach reduces the number of women in need of UtA-PI and PlGF measurements. Given the measurements of UtA-PI

 doi:10.20944/preprints202406.0725.v1

and PlGF, there is a need for extra equipment and specially trained healthcare professionals [6]. Reducing the pregnant population in need of UtA-PI and PlGF measurements, the expenses associated with the prediction of PE will likewise be reduced, which will be beneficial to countries of low- and middle-income.

A further development is to investigate the use of machine learning (ML), given its increasing utilization in healthcare, including obstetrics [7]. As highlighted in recent reviews conducted by Hackelöer et al. and Ranjbar et al., the use of ML has been investigated within the prediction of PE risk [4,7]. Multiple models have been tested along with different feature selections, where the features of maternal factors (ethnicity, age, obstetric history, hypertension, family history, diabetes, systemic lupus erythematosus, antiphospholipid syndrome, conception method, and body mass index (BMI) or weight and height), PAPP-A, PlGF, and UtA-PI are emerging as the standardized feature set, that researchers develop upon [8]. Bertini et al.'s review identified the features with important value in risk assessment of PE listed among their included studies, though only one study's features were mentioned [9].

To our knowledge, existing reviews have not explored how the existing ML models are intended to be deployed during pregnancy. Furthermore, no reviews investigated whether the ML models are intended to be of single-use or multiple-use. The features identified by the ML models to be of important predictive value in the PE risk assessment have likewise not been detected in more than one systematic review by Bertini et al.

This review aims to address these gaps by investigating the existing ML models of PE risk assessment and their intended deployment pattern and performance. In this context, the review wants to clarify if the ML models were intended to be deployed as single use, dual use, or multiple use during pregnancy. Additionally, this review seeks to provide an overview of which features included in the ML models have proven to be of high predictive importance to that exact model.

The review questions:

1. Which ML models have been included in the prediction of PE?
2. Which ML model demonstrates the highest predictive capability?
3. Which features are integrated into the individual ML models?
4. Which features did the individual ML model identify to be of high predictive value?
5. When are the individual ML models intended to be used during pregnancy?
6. How frequently are the individual ML models intended to be deployed throughout pregnancy?

2. Materials and Methods

2.1. Study Design

This review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [10].

2.2. Eligibility Criteria

Inclusion criteria encompassed records written in English, with full-text accessibility, and employing ML for predicting PE. Records unrelated to the topic, such as those focusing on pathology or postpartum applications of ML, were excluded. Similarly, records lacking ML testing, non-transparent feature selection for ML training, or using ML to detect the presence of PE were excluded. Records using extensive blood tests in predicting PE were excluded from this review based on the increased expenses associated with blood tests. Reviews were likewise excluded.

2.3. Search Strategy

A comprehensive search strategy was implemented using truncation and the Boolean operator "OR" to identify relevant articles. The search was refined using the Boolean operator "AND" to focus on the review's topic. The combination of search terms was as follows:

(pregn* OR obstetrics) AND (early OR surveillance OR monitor*) AND (detect* OR program OR predict* OR intervention OR screen*) AND (Artificial intelligence OR AI OR machine learning OR

 doi:10.20944/preprints202406.0725.v1

deep learning OR internet of things) AND (first trimester OR intelligent OR automat*) AND (preeclampsia [Title/Abstract])

The search was conducted on three different databases: PubMed, Engineering Village, and Association for Computing Machinery (ACM) between January 2024 and February 2024. The selection process is documented in a PRISMA flow diagram. No restrictions were imposed regarding the year of publication or country. Additionally, no filters or limits were used within the search databases regarding the Eligibility criteria. The ACM database was set to search for records within "The ACM Guide to Computing Literature" to include as many records as possible.

2.4. Selection Process

The screening of identified records was performed by two reviewers, who independently assessed relevant records based on headline and abstract content. Subsequently, a thorough eligibility screening was conducted, wherein the reviewers went through the full text to exclude records not meeting the predefined eligibility criteria and scope of this review. When facing disagreements about a record's inclusion or exclusion, the reviewers discussed the record and its suitability for the scope of the review to obtain consensus.

2.5. Data Collection

Data extraction was carried out by two reviewers who worked independently at two separate organisations. Extracted data were listed using a customized form, which included the following categories:

1. Study characteristics: Study type, year of publication, and country.
2. Participant information: Number of participants and the incidence of PE cases used for training, validation, and test sets in the ML models.
3. Features: Variables used for training the ML model.
4. ML models employed in the study.
5. Best performance: Identifying the best-performing ML model and its prediction of PE subgroups. For those studies, where the prediction of PE has not been specified other than predicting PE, it has been denoted as predicting "All PE" within this review to compare across studies. The performance is evaluated using performance metrics (AUC, ROC, accuracy, sensitivity, recall, specificity, precision, F1-score, Brier score, screen positive rate (SPR), true positive (TP), true positive rate (TPR), detection rate (DR), false detection rate (FDR), false negative rate (FNR), and false positive (FP)). Among the listed terms, sensitivity, recall, and TPR refer to the same metric value, describing the prediction of positive cases from all the positive cases within the dataset [11].
6. Top predictive features: The top five features identified by the individual ML model to be of high importance for predicting PE among its included features.
7. The intended use of the ML model: Is either reported or interpreted from the study. Including the number of times the ML model is intended to be used and which gestational week within the pregnancy, if this has been denoted by the authors.

2.6. Risk of Bias

A standardized methodology for evaluating bias risk in the included studies and for addressing missing information was not employed. Instead, two independent reviewers evaluated each study and documented any identified bias.

3. Results

The search strategy resulted in 86 records. The total number of records was 22 included in the full-text eligibility screening after removing duplicates and screening titles and abstracts. As illustrated in **Error! Reference source not found.**, a total of 14 studies met the inclusion criteria and were included in this review.

 doi:10.20944/preprints202406.0725.v1

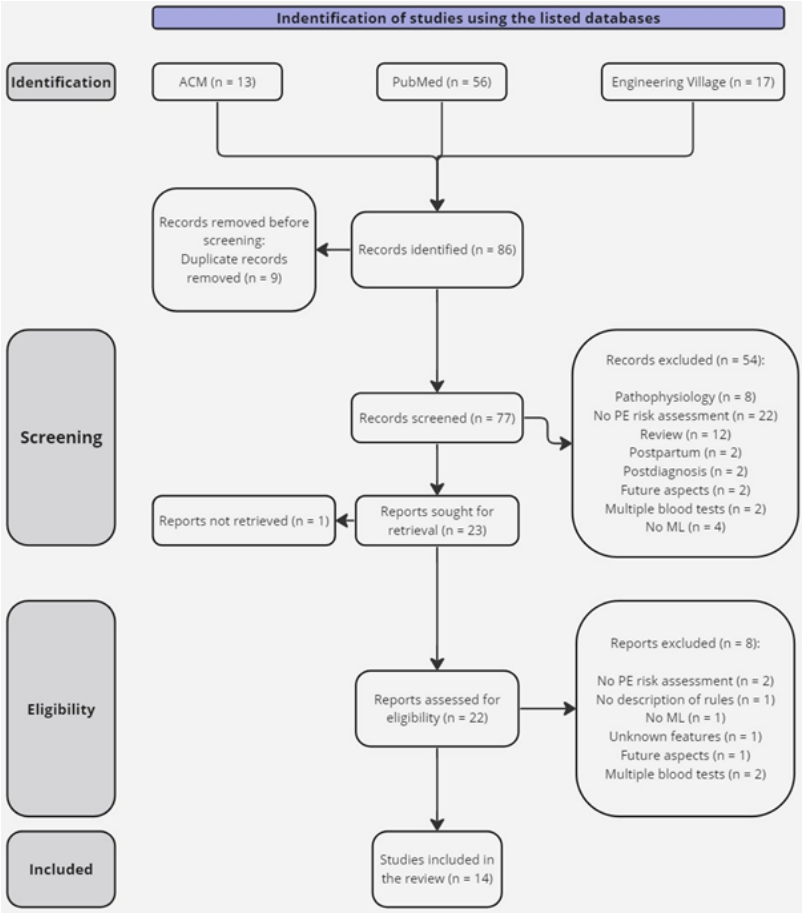


Figure 1. Prisma flow diagram describing the data collection.

A summary of the extracted data from the included studies is presented in **Error! Reference source not found.**.

Table 1. Included studies in the review are listed with additional information regarding the study type, developed machine learning (ML) models, features used and of high importance, identified bias, and the utilization of the models. The following abbreviations were used: Random Forest (RF), AdaBoost classification trees (AdaBoost CT), neural networks (NN), support vector machines (SVM), stochastic gradient boosting (Stoch. GBoost), Extreme gradient Boost (XGBoost), K-nearest neighbours (KNN), decision tree (DT), Receiver operating characteristic curve (ROC), Area under the Receiver operating characteristic curve (AUC), false-positive rate (FPR), detection rate (DR), true-positive rate (TPR), screen-positive rate (SPR), false detection rate (FDR), false negative rate (FNR), positive predictive value (PPV), negative predictive value (NPV), multiples of median (MoM), decision tree (DT), placental growth factor (PlGF), mean arterial pressure (MAP), Uterine artery pulsatility index (UtA-PI), pregnancy-associated plasma protein A (PAPP-A), Antiphospholipid syndrome (APS), blood pressure (BP), and body mass index (BMI). The color coding within the "Best performing ML"-column indicates the performance level among the included studies; Green: high performance value, Yellow: medium performance value, and Red: low performance value.

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
A predictive Bayesian network model for home management of preeclampsia [12]	Velikova M. et al. (2011)	Retrospective research (Netherlands)	Training set: Using incidence rates and prior probabilities from literature, for risk factors, and gynecologist estimated measurements and research studies was used for measurements. Test set: 417 (7.9% PE)	112 features from 10 checkups including: Age, smoking, obese, chronic hypertension, parity history of preeclampsia, blood pressure, hemoglobin, protein, and creatinine	Temporal Bayesian Network Model	**Temporal Bayesian Network Model**: All PE: GA week 12: True-positive: 82% False-positive: 54% GA week 16: True-positive: 73% False-positive: 39%	Not specified	Not listed all 112 features used for training the model. Data was not consistent, which lead to some missing values. Small test set. Not divided PE into subgroups. Retrospective study.	Two times: GA week 12 and 16 Intended to be multiple times: not specified which gestations weeks
Machine Learning Approach for Pre-Eclampsia Risk Factors Association [13]	Martínez-Velasco A. et al. (2018)	Retrospective cohort (Italy)	Training and validation set: 1,634 (16.46% PE)	25 features including poverty status, highest education, pregnancy in weeks, and water retention	RF AdaBoost CT Stoch. GBoost Glmnet MAR-Splines Linear Discriminant Analysis Bayesian GLM NN with Feature Extraction SVM Radial Kernel SVM Linear Kernel KNN Single C5.0 Tree Boosted Logistic Regression C4.5-like Trees	**RF**: All PE: Not specified: ROC: 0.85 Accuracy: 85% Sensitivity: 68% Specificity: 86% Precision: 20% F1: 0.31	1. Gestation weeks completed 2. Poverty 3. Water retention/edema 4. Toxemia 5. Highest educational degree	Not tested on a new dataset. Missing values were replaced by the specific features mean value. Not divided PE into subgroups. Retrospective study.	One time: Not specified when
Preeclampsia Prediction Using machine learning and Polygenic Risk Scores From Clinical and Genetic Risk Factors in Early and Late Pregnancies [14]	Kovachev a VP. Et al. (2024)	Retrospective study (United States)	Training and validation set: 1,125 (7.8% PE)	Routinely collected features at first hospital visit: Demographic, smoking/drug use/alcohol use before pregnancy, BMI, systolic and diastolic BP. Additional feature: a hypertension genetic risk score.	Logistic Regression XGBoost	**XGBoost** without the additional feature: All PE: Before gestation week 14: AUC: 0.74 Accuracy: 91% Sensitivity: 97% Specificity: 26% Precision: 41% Before birth: AUC: 0.91 Accuracy: 93% Sensitivity: 97% Specificity: 43% Precision: 57%	Shapley: 1. History of PE 2. Mean diastolic BP (<14 weeks) 3. Mean systolic BP (first prenatal visit) 4. History of renal disease 5. BMI	Not divided PE into subgroups. Retrospective study. Not specified how the data set was used for training and what the performance is based on.	Two times: one model for first prenatal visit and one model for before the delivery admission (not specified further)

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
An interpretable longitudinal preeclampsia risk prediction using machine learning[1] [15]	Eberhard BW. Et al. (2023)	Cohort Retrospective (United States)	Training set: 98,241 Test set: 22,511 External validation set: 7,705 Total: 120,752 (5.7% PE)	Maternal risk factors, medications, insurance, vital signs, and procedural information data. All information that is routinely collected at week 14, 20, 24, 28, 32, 36, and 39.	XGBoost Deep NN Elastic Net RF Linear Regression	**XGBoost**: External validation set: All PE: Gestation week 14 AUC: 0.63 Specificity: 78% Sensitivity: 62% PPV: 13% NPV: 95% Accuracy: 77% F1-score: 0.24 Gestation week 20: AUC: 0.64 Specificity: 79% Sensitivity: 64% PPV: 12% NPV: 0.96% Accuracy: 78 F1-score: 0.25 Gestation week 24: AUC: 0.67 Specificity: 85% Sensitivity: 37% PPV: 13% NPV: 96% Accuracy: 82% F1-score: 0.19 Gestation week 28: AUC: 0.69 Specificity: 84% Sensitivity: 40% PPV: 13% NPV: 96% Accuracy: 81% F1-score: 0.2 Gestation week 32: AUC: 0.71 Specificity: 83% Sensitivity: 44% PPV: 14% NPV: 96% Accuracy: 81% F1-score: 0.21 Gestation week 36: AUC: 0.76 Specificity: 85% Sensitivity: 49% PPV: 17% NPV: 96%	Shapley: 1. Diastolic and systolic BP 2. Maternal age 3. Insurance 4.Interpregnancy interval 5. chronic and gestational hypertension	Not divided PE into subgroups. Retrospective study.	Multiple times: week 14, 20, 24, 28, 32, 36, 39, and on admission. They made a model for each time point.

[1] This study is a pre-print and has not been peer-reviewed

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
						Accuracy: 83% F1-score: 0.25 Gestation week 39: AUC: 0.86 Specificity: 88% Sensitivity: 66% PPV: 25% NPV: 98% Accuracy: 86% F1-score: 0.36 On admission: AUC: 0.9 Specificity: 88% Sensitivity: 75% PPV: 28% NPV: 98% Accuracy: 87% F1-score: 0.41			
Dynamic gestational week prediction model for pre-eclampsia based on ID3 algorithm [16]	Li Z. et al. (2023)	Case-control retrospective (China)	Training set: 1,272 (18% PE) Test set: 546 (26% PE) Total: 1818 (20.4% PE)	Maternal risk factors. Dynamic parameters include among others: gestational week, BMI, systolic and diastolic BP, pulse, MAP, hematocrit, platelet count, creatinine, uric acid, and PlGF.	Iterative Dichotomiser algorithm	**Iterative Dichotomiser algorithm**: All PE: Macro average: Precision: 76% Recall: 73% F1-score: 75% Weighted average: Precision: 88% Recall: 89% F1-score: 89%	Not specified	Not clarified which data set is used to evaluate the performance. Not clarified how the data sets are constructed with a study population of 932. Retrospective study. Not divided the prediction into PE subgroups.	Multiple times: At prenatal visits at different gestational weeks. (not specified further)
Development of a prediction model on preeclampsia using machine learning-based method: a retrospective cohort study in China [17]	Liu M. et al. (2022)	Cohort Retrospective study (China)	Training set: 9,945 Test set: 1,105 Total: 11,050 (1.3% PE)	Maternal risk factors, MAP, PAPP-A, β-human chorionic gonadotropin, and UtA-PI. Collected at first prenatal visit and at 6 weeks of gestation.	Deep Artificial NN DT Logistic Regression RF SVM Linear kernel	**RF**: All PE: AUC: 0.86 Accuracy: 74% Precision: 82% Recall: 42% F1-score: 0.56 Brier score: 0.17	Not specified	Low number of PE cases. Not divided PE into subgroups. Retrospective study.	One time: first prenatal visit (not specified further)
Novel electronic health records	Li Y-x. et al. (2021)	Retrospective cohort study (China)	Total: 3,759 (5.08% PE)	38 features: demographics (age, BMI, mean BP), pregnancy history (parity,	RF SVM Linear versus radial kernel XGBoost	**XGBoost**: All PE: All features: AUC: 0.96 Accuracy: 92 %	1. Fasting plasma glucose 2. Mean BP 3. BMI	Not divided PE into subgroups. Retrospective study.	One time: early second trimester (not

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
applied for prediction of pre-eclampsia: Machine-learning algorithms [18]				PE history), medical conditions (diabetes, hypertension history), and laboratory tests (hemoglobin, platelet counts, MAP) at early second trimester A simple model using 8 questions with 18 binary features (hypertension history, parity, diabetes) and 2 continuous features (maternal age and BMI)	Logistic Regression	F1-score: 0.57 Simple model: AUC: 0.84 Accuracy: 83% F1-score: 0.34	4. Maternal abdominal circumference5. Serum uric acid	Not specified the sizes of the training, internal validation, and temporal validation sets.	specified further)
Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia [19]	Sufriyana H. et al. (2020)	Retrospective case-control study (Indonesia)	Internal validation set: Cases 3,054, controls 17,921 External validation with geographic split: Cases 145, controls 1,177 External validation with temporal split: Cases 119, controls 785 Total: 23,201 (14,3% PE)	Health insurance data: Demographic (age, family role, labor-type) and diagnoses (causes of disease and organ-related diseases) from one year before PE development and during gestation. For those with event times in 2015, diagnoses within 2 years prior the event was included together with the feature of time to event.	Logistic Regression DT Artificial NN RF SVM Ensemble algorithm	**RF**: All PE: External validation with geographical split: AUC: 0.76 Precision: 82% with FPR of 10% External validation with temporal split: AUC: 0.70 Precision: 78% with FPR of 10% External validation in subgroup geographical split (approximation from study figure): AUC 12-24 months before PE: 0.77 AUC 9-<12 months before PE: 0.88 AUC 6-<9 months before PE: 0.78 AUC 2 days – 6 months before PE: 0.75 External validation in subgroup temporal split (approximation from study figure): AUC 12-24 months before PE: 0.76 AUC 9-<12 months before PE: 0.86 AUC 6-<9 months before PE: 0.68	Not specified	Demands health information that might not be available in the same databases. Retrospective study	Not specified

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
Ethnicity as a Factor for the Estimation of the Risk for Preeclampsia : A Neural Network Approach. [20]	Neocleous KC et al. (2010)	Prospective study (England)	Training set: 6793 (1,7% PE) Test set: 36 (44% PE) Verification set: 9 (56% PE) Total: 6838 (1.99% PE)	MAP, Uterine pulsatility index (UPI), PAPP-A, weight, ethnicity, height, smoking, alcohol, drugs, conception, crown rump length, mother had PE, medical condition, previous PE, and gestation age	NN	AUC 2 days – 6 months before PE: 0.67 **NN**: All PE: Training set with ethnicity: Cases predicted: 45% PE cases predicted: 84% Training set without ethnicity: PE cases predicted: 85% Test set with ethnicity: Cases predicted: 72% PE cases predicted: 94% Test set without ethnicity: PE cases predicted: 100% Verification set with ethnicity: Cases predicted: 78% PE cases predicted: 100% Verification set without ethnicity: PE cases predicted: 100%	Not specified	Did not use common performance metric values to evaluate the model.	Not specified
Performance of a machine learning approach for the prediction of pre-eclampsia in a middle-income country [21]	Torres-Torres J. et al. (2023)	Prospective cohort study (Mexico)	Training set: 1,068 Validation set: 914 Test set: 1,068 Total: 3,050 (4.07% PE)	Maternal characteristics (age, smoking, other drugs (heroin or cocaine), alcohol intake, BMI, congenital heart disease, hypothyroidism, polycystic ovary syndrome, and PE in previous pregnancy), MOM of: MAP, UtA-PI, and PlGF.	Elastic Net	**Elastic Net**: All PE: AUC: 0.78 DR: 50% at 10% FPR Early-onset (<34 gestation weeks): AUC: 0.96 DR: 88% at 10% FPR Pre-term PE (<37 gestation weeks): AUC: 0.90 DR: 77% at 10% FPR	Regularization Coefficient: 1. PlGF 2. MAP 3. UtA-PI 4. BMI 5. APS	Only including high risk patients who did not adhere to aspirin treatment	One time: first trimester (not specified further)
Validation of machine-learning model for first-trimester prediction of	Gil MM. et al. (2024)	Validation using prospective cohort data (Spain)	Training set: 30,352 Validation set: 10,000 Test set: 20,352	Maternal risk factors, MAP, UtA-PI, PlGF, and PAPP-A. Using the raw data and not MoM.	Feed-Forward NN with two hidden layers compared to FMF	**NN**: All PE: AUC: 0.85 DR: 56% at 10% SPR (without PAPP-A)	Not specified by Gil et al. According to the developer of	Small number of PE cases. 6% of the patients took aspirin. Similar or	One time: first prenatal visit specified by Ansbacher-Feldman et al. [23]. (not

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
pre-eclampsia using cohort from PREVAL study [22]			External test set (PREVAL): 10,110 (2.27% PE)			Early-onset PE (<34 gestation weeks): AUC: 0.92 DR: 84% at 10% SPR (without PAPP-A) Pre-term PE (<37 gestation weeks): AUC: 0.91 DR: 78% at 10% SPR (without PAPP-A)	the ML model Ansbacher-Feldman et al.[23] using Shapley: 1. MAP 2. UtA-PI 3. PlGF 4. Racial origin 5. Chronic hypertension	less detection rate compared to FMF.	specified further)
Predictive Performance of machine learning-Based Methods for the Prediction of Preeclampsia -A Prospective Study [24]	Melinte-Popescu A-S et al. (2023)	Prospective case-control study (Romania)	Training set: 163 Test set: 70 Total: 233 (50% PE)	Maternal risk factors (age, BMI, community (urban or rural), personal history of renal disease, obesity, and hyperglycemia) and MoM of: MAP, UtA-PI, PAPP-A, PlGF, and Placental protein-13 collected at first trimester.	DT Naïve Bayes SVM with Linear Kernel RF	**Naïve Bayes**: All PE: AUC: 0.98 Accuracy: 99% Precision: 96% TPR: 96% FNR: 4% PPV: 96.4% FDR: 4% F1: 0.98 Recall: 96% **DT**: Early-onset (<34 gestation weeks): AUC: 0.95 Accuracy: 94% Precision: 93% TPR: 93% FNR: 7% PPV: 75% FDR: 25% F1: 0.86 Recall: 75% **RF**: Late-onset PE (>34 gestation weeks): AUC: 0.84 Accuracy: 88% Precision: 93% TPR: 67% FNR: 33% PPV: 92,9% FDR: 7% F1: 0.93 Recall: 93% **DT**: Moderate PE (Not specified): AUC: 0.80 Accuracy: 82% Precision: 85%	Not specified	Small dataset.	One time: First prenatal visit (not specified further)

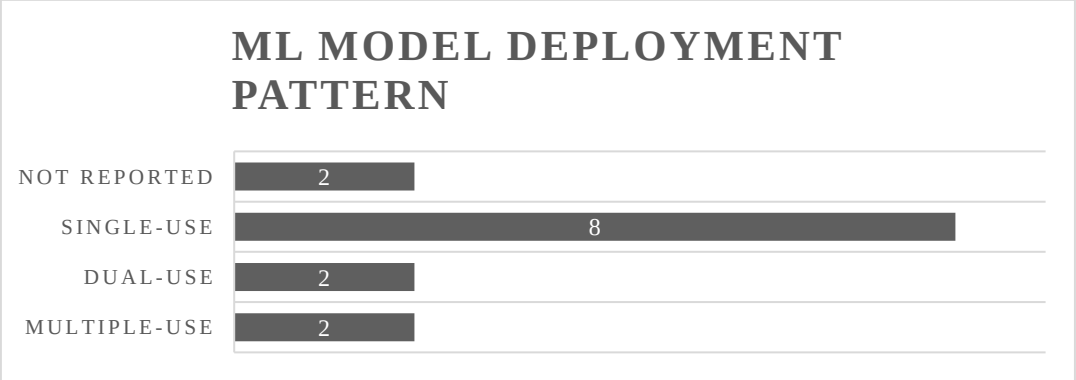
 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
						TPR: 75% FNR: 25% PPV: 92% FDR: 8% F1: 0.88 Recall: 92% **RF**: Severe PE (when certain criteria are present): AUC: 0.76 Accuracy: 77% Precision: 33% TPR: 86% FNR: 14% PPV: 33% FDR: 67% F1: 0.33 Recall: 33%			
Early prediction of preeclampsia via machine learning [25]	Marić I. et al. (2020)	Retrospective cohort study (United States)	Total: 5,245 (10.7 % PE)	Maternal characteristics (age, height, weight), mean and max of systolic and diastolic BP, history of PE, other medical diseases (diabetes, autoimmune conditions), urine glucose and protein, platelet count, and medications (aspirin, insulin).	Elastic Net Gradient Boosting Multiple Logistic Regression	**Elastic Net**: All PE: AUC: 0.79 TPR: 45% FPR: 8% Early-onset (<34 gestation weeks): AUC: 0.89 TPR: 72% FPR: 9%	All PE: 1. Hypertension 2. History of PE 3. insulin 4. Mean systolic BP 5. Number of babies Early-onset (<34 gestation weeks): 1. Hypertension 2. Number of babies 3. History of PE 4. Protein 3+ 5. Anemia	Data set contains missing values. Retrospective study. Not specified the sizes of the training set and test set.	One time: week 16 of gestation
Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: A population	Sandström A. et al. (2019)	Retrospective cohort study (Sweden)	Total: 62,562 (4.4% PE)	Gestational length, age, BMI, MAP, capillary glucose, protein in urine, hemoglobin, infertility duration, region of birth, smoking,	RF Backward selection model on multivariable logistic regression Multivariable regression model using FMF variables	**Multivariable regression model**: Early-onset (<34 gestation weeks): AUC: 0.68 Sensitivity: 31% for 10% FPR. Preterm PE (<37 gestation weeks): AUC: 0.68	Not specified	Missing values in the data set. Not specified the sizes of the training and test sets. No external validation.	One time: first prenatal visit (not specified further)

 doi:10.20944/preprints202406.0725.v1

Study (reference)	Author (Year of publication)	Study type (country)	Type of dataset: Participants (PE %)	Features used for training. (full list is available in Appendix A)	ML models included in the study	Best performing ML model: Group of PE predicted: Time of use: listed performance metrics	The ML models top five included features with high predictive value	Bias of the study	Deployment pattern of the ML model: when in pregnancy is the model described to be deployed
based cohort study [26]				alcohol, family history, and diseases		Sensitivity: 29% for 10% FPR. **Backward selection model**: Term PE (≥37 gestation weeks): AUC: 0.67 Sensitivity: 28% with 10% FPR		Retrospective study.	

 doi:10.20944/preprints202406.0725.v1



illustrates the intended use of the ML models tested within the included studies. With Neucleous et al. and Sufriyana et al. not specifying when the ML models were intended to be used, these were listed as "not reported". The remaining studies' ML models were categorized according to their deployment patterns: single-use, dual-use, or multiple-use prediction models. This classification was done based on the information provided within the respective studies.

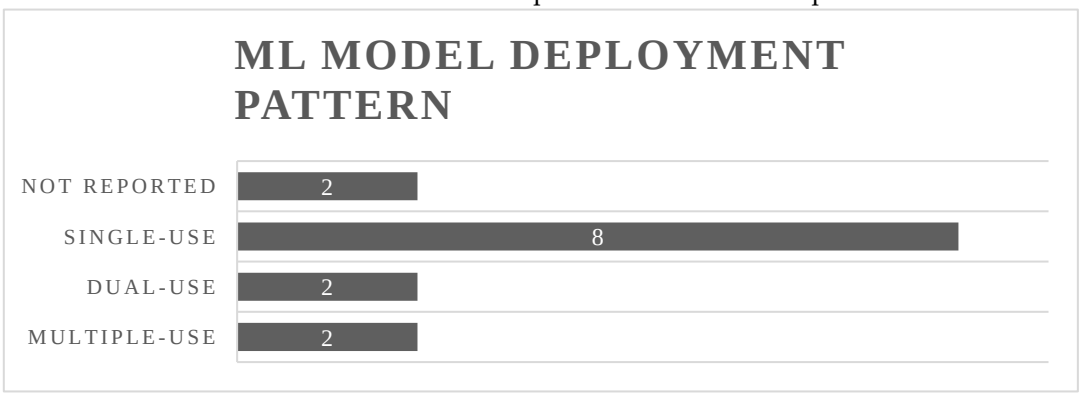


Figure 2. Illustration of the intended use of the prediction models as given in the studies or interpreted by the reviewers.

3.1. Performance of Machine Learning Models

Error! Reference source not found. displays the ML models used within the included studies, alongside those that exhibited the highest performance within them. It is observed in **Error! Reference source not found.** that certain ML models excelled in predicting different subgroups of PE, thus reflecting their best performances in **Error! Reference source not found.** across all included subgroups in the studies.

 doi:10.20944/preprints202406.0725.v1

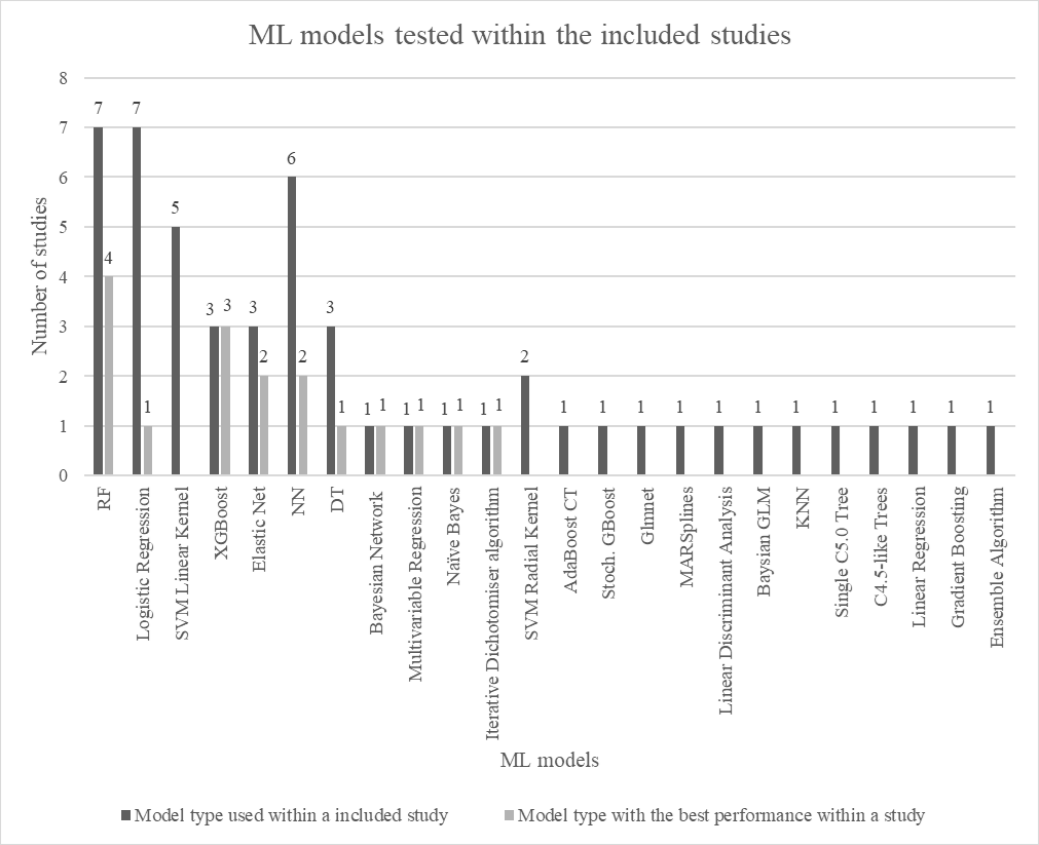


Figure 3. ML models tested within the included studies. The blue indicates how many studies included the ML model within their study. The orange indicates the ML model with the best performance within the included studies.

Considering AUC and recall values, which emerged across 10 out of 14 studies leading to the most used performance metrics, Torres-Torres et al. achieved the highest AUC of 0.96 as well as a DR of 88% at a FPR of 10% in predicting early-onset PE (<34 weeks of gestation), utilizing Elastic Net Regression [21]. Torres-Torres et al. did not report a recall value, hence the highest recall value for early-onset PE was achieved by Gil et al at 84%. For preterm PE (<37 weeks of gestation), Gil et al. attained the highest AUC of 0.91 and the highest recall value of 78% at a SPR of 10% [22], incorporating a Feed-Forward NN [23]. As Gil et al. refer to their DR to be the same as recall, this is included in this performance comparison [22,23]. Melinte-Popescu et al. reported the highest AUC value of 0.84 and recall value of 93% for late-onset PE (>34 weeks of gestation) using RF. Furthermore, in predicting all cases of PE, Melinte-Popescu et al. attained the highest AUC of 0.98 using Naïve Bayes. For term PE (>37 weeks of gestation) Sandström et al. obtained an AUC of 0.67 at a FPR of 10% deploying a Backward Selection model on Multivariable Logistic Regression [26].

4. Discussion

4.1. Best-Performing Machine Learning Model

RF, Logistic Regression, NN, SVM with a linear kernel, Elastic Net, Decision Tree, and XGBoost were the most used ML models. Considering the AUC and recall values, no single type of ML model emerged as superior across the different subgroups of PE (early-onset PE, late-onset PE, preterm PE, term PE, and all PE). Especially concerning the same data set, Melinte-Popescu et al. achieved the highest AUC for LO-PE and all PE using two different types of ML models. Despite RF and Logistic Regression being the predominant models only four out of seven and one out of seven studies identified RF and Logistic Regression as the best-performing model, respectively. XGBoost, on the other hand, demonstrated the best performance in three out of three studies, outperforming RF in

 doi:10.20944/preprints202406.0725.v1

two of those. However, regarding achieving the highest AUC, XGBoost did not attain the highest among any of the included studies. However, RF and Logistic Regression had the highest AUC for LO-PE and term PE, respectively.

Based on the results of the best ML models, we could hypothesize that using multiple models for identifying subgroups of PE could be beneficial. However, the compared models use different features as well as different data sets of dissimilar sizes. Especially, the rate of PE cases in the different datasets ranges from 1.3% to 50% among the included studies' total populations. As only three of the 14 studies reported the rate of PE cases in their training set, a comparison by the training set for all studies was not possible. Nonetheless, identifying the ML model's performance metrics, the rate of PE cases does not seem to influence their performance results. In fact, Melinte-Popescu et al. had the highest incidence of PE cases of 50% in their population of 233 and achieved the highest performance for late-onset PE and all PE. Yet, Li Y-X et al. had an incidence of 5% PE in their population of 3,759 and achieved an AUC of 0.96 and an accuracy of 92. This is 0.02 less in the AUC and 7% less in accuracy than Melinte-Popescu et al. With seven studies having their rate of PE cases in between 5% and 50% these studies all had AUC and accuracy values less than Li Y-X et al. Indicating, that there is no correlation between the rate of PE cases and the performance of the ML model within the included studies. Yet, they are all based on different features and therefore this might be the factor influencing the performance outcome. The population size and rate of PE cases might also be influencing the outcome, but is not visible within this review, as the studies do not use the same features. Selecting one model that will perform with high prediction on different data sets is challenging as there is no ML model that outperforms others on every single data set even though the data sets are similar [27]. Making Gil et al.'s performance noteworthy, as their model was originally developed by Ansbacher-Feldman et al. on another population employing raw input data similar to that used in the FMF algorithm [23].

4.2. Feature Selection

Torres-Torres et al, Gil et al., and Melinte-Popescu et al. used features similar to FMF (such as maternal age, MAP, UtA-PI, PlGF, and PAPP-A) (Appendix A). Notably, neither Melinte-Popescu et al. nor Torres-Torres et al. included racial origin as a feature, as was done by Gil et al. where it was rated to be the fourth highest predictive feature. Gil et al.'s ML model incorporated the use of aspirin and raw input data instead of MoM values. While Melinte-Popescu et al. and Torres-Torres et al. added more than one feature diverse from FMF and used BMI instead of weight and height. Torres-Torres et al. rated BMI to be the fourth-highest predictive feature of their ML model. As BMI is calculated based on weight and height, including all three features can potentially cause correlation. Collinearity makes it challenging to identify the individual feature's effect on the outcome and impacts the development of the model [28]. Therefore, the choice of features needs to take this factor into account.

Among the 14 included studies seven of them highlighted features of high predictive importance. Within six of the seven studies, BP measurements were listed in the top five. Systolic BP occurred one time more frequently than MAP and diastolic BP. MAP is calculated based on both diastolic and systolic BP, with diastolic being the primary contributor. Regardless of whether it is systolic BP, diastolic BP or MAP, all pressure-related parameters show significance in PE risk assessment. However, a systematic review conducted by Bertini et al. highlighted systolic BP to be of particularly high importance to the ML models [9]. Yet, the best-performing ML models identified within this review all used MAP instead of systolic and diastolic BP. No study was identified to compare the ML model's performance regarding MAP versus systolic and diastolic BP. Therefore, we have no basis for asserting which method of BP measurement is superior. However, such a comparison could be beneficial in the future development of ML models.

Li Y-x et al. identified that a questionnaire involving features such as maternal age, BMI, and medical conditions (Appendix A**Error! Reference source not found.**) can achieve an AUC of 0.84 [18]. Utilizing a ML model based on a questionnaire is arguably more cost-efficient and less intrusive compared to models that use several blood tests and involve healthcare professionals for ultrasound

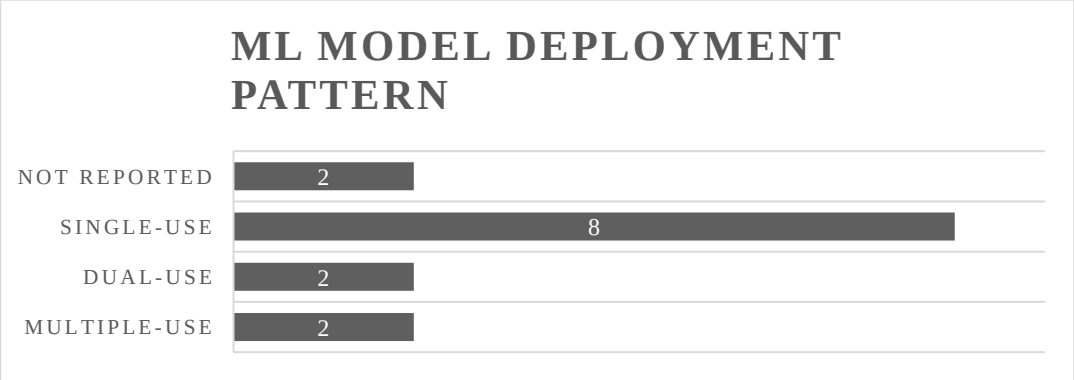
 doi:10.20944/preprints202406.0725.v1

and blood pressure measurements. Across the seven studies listing their top five predictive features, 14 features were identified to be suitable for a questionnaire. This involves the features of BMI, maternal abdominal circumference, insulin, chronic hypertension, racial origin, antiphospholipid syndrome, water retention/edema, PE family history, number of babies, poverty, edema, highest education, insurance, and renal disease. Concerning maternal abdominal circumference, the expecting mother will be able to answer this if she is provided with a measuring tape. Yet, including this measurement alongside BMI needs to be done with caution. The reason is that these features might be collinear as they both depend on the person's weight and height. With collinear features, the model's performance can potentially be affected. These features are not all currently included in the FMF algorithm, nor has the combination of these features been tested within a single ML model along with the FMF algorithm's maternal characteristics. However, incorporating these features into a questionnaire for the expecting mother appears relevant to clarify the potential of a ML model based on a questionnaire in PE risk assessment as a preliminary step.

Sufriyana et al. is the only study using features from the expecting mother's health insurance record dated months before the development of PE. These features are based on the recorded diagnosis within their health insurance and listed as the codes from the International Classification of Diseases 10th Revision (Appendix A**Error! Reference source not found.**) [19]. The proposed approach achieved the highest AUC when using data collected 9-<12 months before developing PE. Achieving an AUC of approximately 0.88 (geographical split) and 0.86 (temporal split) by only using data from 9-<12 months before the development of PE. This time period is defined by Sufriyana et al. to correspond to endometrial maturation [19]. This result indicates a potential for using patient health record data as part of a prediction model for PE. Additionally, using available record data in a ML model is a cost-effective approach, though the records might be diverse among hospitals leading to potential bias.

4.3. Machine Learning Deployment Pattern

In eight out of 14 studies, ML models were utilized as a single-use application, indicating their prevalent usage and testing. Nevertheless,



suggests a growing interest in implementing ML models for multiple uses, with proposed strategies by Eberhard et al. and Li et al., both conducted in 2023. As identified in Table 1, three out of four studies intending to use the ML model more than once were conducted in the years 2023 and 2024, whereas the remaining study is from 2011. The included studies were conducted in the time period of 2010 to 2024. Velikova et al. was the sole study from 2011 to 2023 investigating the multiple use of a ML model in the PE risk assessment at different gestational weeks. Yet, they only provided the risk prediction for week 12 and week 16 within their study. Additionally, Velikova et al. aimed to create a model which could be used as a decision support tool for home monitoring, though this was not tested within this study. However, three out of five studies conducted in the years 2023 to 2024 used the ML model more than once or created a model for each time point. This indicates a potential shift in the research field of PE risk assessment using ML models. Yet, none of the included studies have investigated the proposed adaptive ML model as mentioned in Hackelöer et al.'s review, which aims to monitor the development of PE. The BP progression along with gestations weeks was

 doi:10.20944/preprints202406.0725.v1

investigated by Lazdam et al. and Macdonald-Wallis et al. [29,30]. They identified differences in the progression of diastolic and systolic BP within pregnant women developing PE as early as weeks 12 to 21 of gestation[29,30]. Eberhard et al. likewise indicate in their study, that BP's importance to the ML model increases as gestation age progresses [15]. This suggests that an adaptive and multiple-use ML model including the BP progression will be beneficial in the PE risk assessment and PE development from week 12 of gestation. With home-monitoring as suggested by Velikova et al. would be a valued contribution, as the associated problems from BP changes appear only days later [12]. This use could potentially enhance predictive accuracy by reducing the number of false positives and lead to more personal care within obstetrics concerning PE treatment. An adaptive and multiple-used ML model will therefore both predict the risk of developing PE before gestational week 16 as well as help detect the development of PE at an early stage.

Three of the 14 studies do not indicate when the model is intended to be used, whereas the remaining indicates the first time to be either "first prenatal visit", "week 16 of gestation", "early second trimester", or "first trimester". Compared to the FMF algorithm the earliest predictive algorithm is to be used at gestation week 11+0 to 14+1, where the first prenatal visit usually takes place. The first trimester ends by gestation week 12, so the first prenatal visit can likewise be in the early stages of the second trimester. Hence, the different definitions of the first intended use are within the same time period except Marić et al.'s study being utilized at week 16 of gestation. Yet, according to Rolnik et al. and van Doorn et al. should the aspirin treatment be initiated before week 16, making the prediction at week 16 of gestation on the last time point possible for this initiation.

Concerning using the ML model later in the pregnancy, only two studies specified the exact gestation weeks where it is intended to be used. These are week 16 of gestation in Velikova et al.'s study, and weeks 20, 24, 28, 32, 36, 39, and on admission in Eberhard et al.'s study. Whereas the remaining two studies either did not specify any information or used the definition of "before the delivery admission". Resulting in no similar frequency of use within these studies. The use of a ML model more than once has been identified to be a new and growing part of the research area of PE risk assessment, which reflects the lack of a common frequency of usage patterns.

4.4. Limitations

No standardized method was used in the bias assessment, which is a limitation of this study. Using a standardized method would have clarified the included studies' different risks of bias in a systematic manner. Additionally, discrepancies in subgroups of PE and the absence of a common performance metric hinder a comparative analysis of performance among all the included studies. Not all studies use the same performance metrics, which made it unable to get every study into consideration in being the best-performing ML model across the studies. Comparing the performances of different ML models that are all trained and tested on different data sets on diverse populations as well as developed on different feature sets is a limitation of this review. Such a comparison might have caused bias as different feature combinations and population groups result in different outcomes. Furthermore, in five out of 14 studies, only one ML type was tested, biasing this review's findings concerning the best-performing ML model within each study. With only one ML model listed within a study, this automatically becomes the best-performing model without any comparison.

4.5. Future Research

The ML models within this review were trained and tested on the collected data being either retrospective or prospective. Five out of 14 studies were prospective studies, leaving nine studies being retrospective. Performing retrospective studies means that the data can include some missing values, which Sandström et al. experienced. This could potentially have affected the development of the ML models, as they had to use mean values for the missing elements. Similarly, retrospective studies do not make it possible to investigate different features, which were not collected at that time. Hindering potential feature selection. Yet, four of the studies using prospective data were not tested on new prospectively collected data, in the sense of predicting the risk with the developed ML model

 doi:10.20944/preprints202406.0725.v1

at the time of data collection. Only Gil et al. performed the risk assessment at the time of data collection, yet the clinicians and participants were not informed about the outcome. Prospective validation of the models would be of high importance in the context of implementing it in practice, as Torres-Torres et al. likewise point out. As a ML model is intended to be a decision support tool in the PE risk assessment just as the FMF algorithm is today. Future research would benefit from testing their ML model on prospective data with an unknown outcome at the prediction time. This will highlight their model's performance in the intended use in a clinical setting.

Out of 14 studies, the data sets used for their ML models were only available online for one study, whereas five other studies reported that it could be made available if contacted. The authors of the five studies have been contacted to attain access to their data set in order to replicate their results. Out of these five studies none replied. Two studies reported that access could be gained by getting approval or contacting other parts than the authors. The remaining six studies did not report anything on the data sets' accessibility.

5. Conclusions

In conclusion, the analysis of studies investigating the risk assessment capabilities of ML models for PE reveals a diverse landscape of models and parameters used to evaluate them. RF, Logistic Regression, NN, and SVM were frequently used ML models. While AUC and recall emerged as common performance metrics No single ML model proved consistently superior across different subgroups of PE, nor even within the same studies. Instead, using different ML models has shown potential in the prediction of early-onset PE, preterm PE, late-onset PE, term PE, and all PE.

BP was identified as being the most predictive feature in the risk assessment of PE. Highlighting diastolic and systolic BP measurements to be of important value for a ML model alongside MAP. The BP parameter that will benefit the ML model's performance the most is unknown. BMI was likewise identified as a predictive feature, though including this together with weight and height will potentially cause a correlation in the ML.

ML models being deployed as a dual- or multiple-use have been investigated in recent studies, suggesting an increased interest in the multiple-use, though eight of the studies intended their ML models to be of single-use. Furthermore, no frequency in the dual- or multiple-use is identified to be repeated among the studies. Incorporating features such as BP progression throughout gestation may enhance the predictive accuracy of ML models for PE risk assessment and limit the number of women being falsely predicted to be at high risk of PE. Among the studies including when their ML models were intended to be deployed for the first time, only one study intended on week 16 of gestation. The remaining studies intended to use it within the timeframe of gestation week 11+0 to 14+1, making aspirin treatment possible to be initiated on time.

Limitations of this review include comparing the studies even though they are trained and tested on diverse data sets, population groups, and divergent feature selection schemes. Additionally, ML model performance is listed in different subgroups of PE risk assessment and there is an absence of a common predictive metric. Five studies only tested a single ML model which arguably affected the results concerning which ML models had the best performance. Thus, not all studies could be compared and taken into consideration.

Author Contributions: Conceptualization, L.P. and S.W.; methodology, L.P. and M.M-M.; formal analysis, L.P.; investigation, L.P. and M.M-M; writing—original draft preparation, L.P.; writing—review and editing, M.M-M, S.W, and J.R.; visualization, L.P.; supervision, S.W. and J.R.; funding acquisition, S.W. and J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This review was funded by Innovation Fund Denmark (IFD), UEFISCDI Romania, and NCBR Poland in the framework of the ERA PerMed (EU Grant 779282), JTC 2021; project WODIA -Personalized Medicine Screening and Monitoring Programme for Pregnant Women Suffering from Preeclampsia and Gestational Hypertension.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

 doi:10.20944/preprints202406.0725.v1

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

This appendix contains the features used to train the machine learning models in the different studies included in this review. The features were identified within the paper of the study, or the supplemental documents provided with the paper. The intent is to get a clear understanding of the different features used to train machine learning models along with their performances illustrated in Table 2.

Table 2. Features used within the Fetal Medicine Foundation (FMF) are listed together with the features used to train the individual machine learning models within each of the included studies. The following abbreviations were used: mean arterial pressure (MAP), placental growth factor (PlGF), Uterine artery pulsatility index (UtA-PI), pregnancy associated plasma protein A (PAPP-A), blood pressure (BP), and body mass index (BMI), multiples of median (MoM), and preeclampsia (PE).

Study	Features used in the machine learning model
FMF competing risk model [5]	Maternal factors: Age Heigh Weight Racial origin Conception method Smoking Chronic hypertension Diabetes mellitus Systemic lupus erythematosus Antiphospholipid syndrome Mother of the pregnant woman's history of PE Parity Previous had PE Gestational age at prior birth Birthweight of the baby in last pregnancy Years between birth Estimated conception data MAP UtA-PI PlGF PAPP-A
A predictive Bayesian network model for home management of preeclampsia [12]	Values taken at each of the following gestational week: 12, 16, 20, 24, 28, 32, 36, 38, 40, and 42: Age Smoking Obese Chronic hypertension Parity-history PE Treatment Systolic BP Diastolic BP Hemoglobin Creatinine Protein/creatinine
Machine learning approach for preeclampsia risk factors association [13]	Duration of completed pregnancy in weeks. Toxemia Education (completed years of schooling) Highest completed year school or degree Pregnancy outcome Labor force status Poverty

 doi:10.20944/preprints202406.0725.v1

	Water retention/edema Race Anemia Sex Birth order Birth weight One-minute and five-minute APGAR scores Month of pregnancy when prenatal care began Number of prenatal visits Weight gained during pregnancy Medical risk factors for the pregnancy Obstetric procedures performed Delivery complications Congenital anomalies and abnormalities Mother's marital status Number of live births now living The parents' age Hispanic origin State/country of birth
Preeclampsia Prediction Using machine learning and Polygenic Risk Scores From Clinical and Genetic Risk Factors in Early and Late Pregnancies [14]	Maternal age at delivery Self-reported race Relf-reported ethnicity (Hispanic or non-Hispanic) Hospital (tertiary or community) Gravidity Parity Gestational age at delivery Gestational age at preeclampsia diagnosis Last BMI before pregnancy BMI at delivery Maximal diastolic BP during pregnancy Maximal systolic BP during pregnancy Family history of chronic hypertension Family history of preeclampsia Interpregnancy interval In vitro fertilization Multiple gestation Smoking before pregnancy Drugs of abuse before pregnancy Drugs of abuse during pregnancy Alcohol use before pregnancy High-risk pregnancy Maximal BMI before pregnancy Mean BMI in the period 0-14 gestational weeks Systolic BP at first prenatal visit Diastolic BP at first prenatal visit History of pregestational diabetes History of kidney disease before pregnancy History of gestational diabetes in a prior pregnancy History of a prior high-risk pregnancy History of autoimmune disease History of preeclampsia in a prior pregnancy Family history of hypertension Family history of PE Minimal platelet count in the period 0-14 gestational weeks and in pregnancy before preeclampsia diagnosis or delivery Maximal uric acid in the period 0-14 gestational weeks and in pregnancy before preeclampsia diagnosis or delivery Presence of proteinuria in the period 0-14 gestational weeks and in pregnancy before preeclampsia diagnosis or delivery Systolic BP polygenic risk score Small for gestational age or intrauterine growth restriction Last BMI during pregnancy before preeclampsia diagnosis or delivery Maximal BMI before pregnancy Prescription of antihypertensive medication during pregnancy Diagnosis of gestational hypertension during pregnancy

 doi:10.20944/preprints202406.0725.v1

Performance of a machine learning approach for the prediction of pre-eclampsia in a middle-income country [21]	Maternal age Nulliparity Spontaneous pregnancy Induction of ovulation In-vitro fertilization Gestation age at screening Smoker Alcohol intake Other drugs (heroin or cocaine) Pre-existing diabetes Chronic hypertension Lupus Antiphospholipid syndrome Polycystic ovary syndrome Hypothyroidism Congenital heart disease PE in a previous pregnancy Fetal growth restriction in a previous pregnancy Mother of the patient had PE BMI MAP MAP (MoM) UtA-PI UtA-PI (MoM) PlGF PlGF (MoM) PAPP-A Gestational age at delivery
Validation of machine-learning model for first-trimester prediction of pre-eclampsia using cohort from PREVAL study. Based on the machine learning model trained by Ansbacher-Feldman et al. [23]	Maternal age Maternal weight Maternal height Gestation age at screening Racial origin Medical history: Chronic hypertension Diabetes type I Diabetes type II Systemic lupus erythematosus/antiphospholipid syndrome Smoker Family history of PE Method of conception: Spontaneous In-vitro fertilization Use of ovulation drugs Obstetric history: Nulliparous Parous, no previous PE Parous, previous PE Interpregnancy interval Aspirin MAP UtA-PI Serum concentration of pregnancy-associated plasma protein-A (PAPP-A) Serum concentration of PlGF
An interpretable longitudinal preeclampsia risk prediction using machine learning [15]	Maternal age Self-reported race Self-reported ethnicity (Hispanic or non-Hispanic) Private insurance Public insurance Alcohol use history Smoking history Illicit drugs history

 doi:10.20944/preprints202406.0725.v1

Gravidity
Parity
In vitro fertilization
Nulliparous
Interpregnancy interval
Multiple gestation
Maximal systolic BP:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal diastolic BP:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal heart rate:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal BMI:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal weight:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Family history of chronic hypertension
Family history of preeclampsia
Family history of diabetes
Family history of heart disease
Family history of hyperlipidemia
Family history of stroke
Past history of diabetes
Past history of gestational diabetes
Past history of cesarean delivery
Past history of preterm birth
Past history of gynecologic surgery
Past history of asthma
Past history of chronic hypertension
Past history of gestational hypertension

 doi:10.20944/preprints202406.0725.v1

Past history of high-risk pregnancy
Past history of hyperemesis gravidarum
Past history of migraine
Past history of obesity
Past history of PE
Past history of pregnancy related fatigue
Past history of sexually transmitted disease
Chronic hypertension
Anemia during pregnancy
Headaches during pregnancy
Autoimmune disease
High risk pregnancy
Hyperemesis gravidarum
Pregnancy related fatigue
Oligohydramnios:
At week 39 and admission
Proteinuria:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal aspartate transferase:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal white blood count:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal alanine transaminase:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal serum calcium:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal serum creatinine:
0-14 weeks
0-20 weeks
0-24 weeks

 doi:10.20944/preprints202406.0725.v1

0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal eosinophils:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal serum glucose:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal hemoglobin:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal lymphocytes:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Maximal platelets:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Minimal red blood count:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks
0-36 weeks
0-39 weeks
0 weeks - admission
Antihypertensive medications:
0-14 weeks
0-20 weeks
0-24 weeks
0-28 weeks
0-32 weeks

 doi:10.20944/preprints202406.0725.v1

	0-36 weeks 0-39 weeks 0 weeks – admission
Predictive Performance of machine learning-Based Methods for the Prediction of Preeclampsia-A Prospective Study [24]	Maternal age BMI Medium: Urban Rural Parity: Nulliparity Multiparity Smoking status during pregnancy The use of assisted reproductive technologies Personal or family history of PE Personal history of hypertension Personal history of renal disease Personal history of diabetes Personal history of systemic lupus erythematosus/antiphospholipid syndrome Hyperglycemia in pregnancy Obesity Interpregnancy interval MAP (MoM) UtA-PI (MoM) PAPP-A (MoM) PLGF (MoM) Placental protein-13 (MoM)
Dynamic gestational week prediction model for pre-eclampsia based on ID3 algorithm [16]	Static parameters: Multiple births Spontaneous miscarriage history History of hypertension in pregnancy History of diabetes mellitus Family history of hypertension Preconception BMI Dynamic parameters: Gestational week BMI during pregnancy Systolic BP Diastolic BP Pulse pressure MAP Pulse waveform area parameters Cardiac output Cardiac index Total peripheral resistance Hematocrit Mean platelet volume Platelet count Alanine aminotransferase Aspartate aminotransferase Creatinine Uric acid PlGF
Development of a prediction model on preeclampsia using machine learning-based method: a retrospective cohort study in China [17]	Maternal age Height Weight BMI Parity Method of conception Previous diagnosis of hypertension History of diabetes mellitus History of gestational diabetes History of PE

 doi:10.20944/preprints202406.0725.v1

Novel electronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms [18]	History of fetal growth restriction MAP β-human chorionic gonadotropin PAPP-A Gestational age at screening Chronic hypertension Left uterine artery PI Right uterine artery PI Mean uterine artery PI All features: Maternal age BMI Mean BP Maternal abdominal circumference Gravidity Parity PE in a previous pregnancy Prior cesarean delivery Pregnancy interval Nulliparity Multifetal gestations Assisted reproductive technology Pre-pregnancy diabetes Heart disease Thyroid disease Renal disease Autoimmune diseases Mental disorder Uterine leiomyoma Adenomyosis Uterine malfunctions History of seizure disorder Family history of hypertension Hemoglobin White blood cell count Platelet counts Direct bilirubin Total bilirubin Alanine aminotransferase Γ-glutamyl transferase Total protein Albumin Globulin Fasting plasma glucose Total bile acid Creatinine Serum urea nitrogen Serum uric acid Baseline risk features: Nulliparity Multifetal gestations PE in a previous pregnancy Pre-gestational diabetes BMI Maternal age Assisted reproductive technology Kidney diseases Autoimmune diseases Questionnaire features: Family history of hypertension Nulliparity Prior cesarean delivery Pregnancy interval Multifetal gestations

 doi:10.20944/preprints202406.0725.v1

	Assisted reproductive technology
	Gravidity
	Parity
	Pre-gestational diabetes
	Heart disease
	Thyroid disease
	Renal disease
	Autoimmune diseases
	Mental disorder
	Uterine leiomyoma
	Adenomyosis
	Uterine malfunctions
	History of seizure disorder
	Maternal age
	BMI
	Maternal age
	Height
	weight
	Blood pressure:
	Mean systolic
	Mean diastolic
	Maximum systolic
	Maximum diastolic
	Race
	Ethnicity:
	Hispanic
	Non-Hispanic
	unknown
	Gravida:
	Nulliparous
	Multiparous
	Number of babies
	Medical history:
	PE
	Assisted reproductive treatment
	Chronic hypertension
	Diabetes (type I or type II)
	Obesity
Early prediction of preeclampsia via machine learning [25]	Renal disease
	Autoimmune conditions:
	Systemic lupus erythematosus
	Discoid lupus erythematosus
	Systemic sclerosis
	Rheumatoid arthritis
	Dermatomyositis
	Polymyositis
	Undifferentiated connective tissue disease
	Celiac disease
	Antiphospholipid syndrome
	Sexually transmitted diseases (human papillomavirus, chlamydia, genital herpes)
	Hyperemesis gravidarum
	Headache
	Migraine
	Poor obstetrics history
	Poor obstetrics history
	Medical history at 17 weeks of gestation:
	Gestational diabetes
	Anemia
	High-risk pregnancy
	Routine prenatal laboratory results:
	Protein from urine
	Glucose from urine
	Platelet count
	Red blood cells

 doi:10.20944/preprints202406.0725.v1

	White blood cells Creatinine Hemoglobin Hematocrit Monocytes Lymphocytes Eosinophils Neutrophils Basophils Blood type with Rh Uric acid Rubella Varicella Hepatitis B Syphilis Chlamydia Gonorrhea Intake of medication: Aspirin Nifedipine Aldomet Labetalol Insulin Glyburide Prednisone Azathioprine Plaquenil Heparin Levothyroxine Doxylamine Acyclovir
Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: A population based cohort study [26]	Multivariable regression model: Family history of PE Country of birth Method of conception Gestational length Maternal age Height Weight Smoking in early pregnancy Pre-existing diabetes mellitus Chronic hypertension Systemic lupus erythematosus MAP Backward selection model and RF model: Gestational length first examination in weeks Maternal age BMI MAP Capillary glucose Protein in urine Hemoglobin Previous miscarriage Previous ectopic pregnancy Infertility duration Family situation: Single Living together with partner Other Region of birth: Sweden Nordic countries (except Sweden) Europe (except of Nordic countries) Africa

 doi:10.20944/preprints202406.0725.v1

Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia [19]	North America South America Asia Oceania Smoking 3 months before pregnancy Smoking at registration Snuff 3 months before pregnancy Snuff at registration Alcohol consumption three months before registration Alcohol consumption at registration Family history of PE Family history of hypertension Infertility: Without treatment Ovary simulation In-vitro fertilization Cardiovascular disease Endocrine disease Pre-existing diabetes Thrombosis Psychiatric disease systemic lupus erythematosus Epilepsy Chronic hypertension Morbus Chron/Ulcerous colitis Lung disease or asthma Chronic kidney disease Hepatitis Gynecological disease or operation Recurrent urinary tract infections Blood group Demographic: Age Marriage Family role Member strata Member type International Classification of Diseases 10th Revision coded diagnoses: A codes B codes C codes D codes E codes F codes G codes H codes I codes J codes K codes L codes M codes N codes Infection-related codes: G0, H00, H01, H10, H15, H16, H20, H30, H60, H65, H66, H67, H68, H70, I0, J0, J1, J2, J40, J41, J42, J85, J86, K12, K2, K35, K36, K37, K5, K65, K67, K73, K80, K81, L0, M00, M01, M02, N7 Immune-related codes: B20, D8, E10, G35, G61, G70, I0, J30, J31, J32, J35, J45, L2, L50, M04, M05, M06, M15, M16, M17, M18, M19, M3, M65, N00, N01, N03, N04 Nervous system-related codes: A8, C7, G Eye-related codes: C69, H0, H1, H2, H3, H4, H5 Ear-related codes: C30, D02, H6, H7, H8, H9

 doi:10.20944/preprints202406.0725.v1

Ethnicity as a Factor for the Estimation of the Risk for Preeclampsia: A Neural Network Approach [20]	Heat-related codes: C38, I2, I3, I4, I5 Respiratory system-related codes: A1, C0, C3, J Digestive system-related codes: A0, C0, C1, K0, K1, K3, K4, K5, K6 Skin and subcutaneous-related codes: B0, B1, B8, C43, C44, L Musculoskeletal system-related codes: C40, C41, M Urinary system-related codes: C64, C65, C66, C67, C68, N0, N1, N2, N3 Reproduction system-related codes: A5, A60, A61, A62, A63, A64, C51, C52, C53, C54, C55, C56, C57, C58, N7, N8 Liver and pancreas-related codes: B15, B16, B17, B19, C22, C23, C24, C25, K7, K8 Breast-related codes: C50, N6 Vascular-related codes: I1, I7, I8 MAP Uterine Pulsatility index PAPP-A Ethnicity Weight Height Smoking Alcohol consumption Previous PE Conception: Spontaneous Ovulation drug In-vitro fertilization Medical condition of pregnant woman Drugs taken by the pregnant woman Gestation age Crown rump length Mother had PE

References

1. L. Duley, "The Global Impact of Pre-eclampsia and Eclampsia," *Seminars in Perinatology*, vol. 33, no. 3. pp. 130–137, Jun. 2009. doi: 10.1053/j.semperi.2009.02.010.
2. L. A. Magee, K. H. Nicolaides, and P. von Dadelszen, "Preeclampsia," New England Journal of Medicine, vol. 386, no. 19, pp. 1817–1832, May 2022, doi: 10.1056/NEJMra2109523.
3. R. van Doorn et al., "Dose of aspirin to prevent preterm preeclampsia in women with moderate or high-risk factors: A systematic review and meta-analysis," PLoS One, vol. 16, no. 3 March, Mar. 2021, doi: 10.1371/journal.pone.0247782.
4. M. Hackelöer, L. Schmidt, and S. Verlohren, "New advances in prediction and surveillance of preeclampsia: role of machine learning approaches and remote monitoring," Archives of Gynecology and Obstetrics, vol. 308, no. 6. Springer Science and Business Media Deutschland GmbH, pp. 1663–1677, Dec. 01, 2023. doi: 10.1007/s00404-022-06864-y.
5. D. Wright, A. Syngelaki, R. Akolekar, L. C. Poon, and K. H. Nicolaides, "Competing risks model in screening for preeclampsia by maternal characteristics and medical history," Am J Obstet Gynecol, vol. 213, no. 1, pp. 62.e1-62.e10, Jul. 2015, doi: 10.1016/j.ajog.2015.02.018.
6. D. Wright, D. M. Gallo, S. Gil Pugliese, C. Casanova, and K. H. Nicolaides, "Contingent screening for preterm pre-eclampsia," Ultrasound in Obstetrics and Gynecology, vol. 47, no. 5, pp. 554–559, May 2016, doi: 10.1002/uog.15807.
7. A. Ranjbar, F. Montazeri, S. R. Ghamsari, V. Mehrnoush, N. Roozbeh, and F. Darsareh, "Machine learning models for predicting preeclampsia: a systematic review," BMC Pregnancy Childbirth, vol. 24, no. 1, p. 6, Jan. 2024, doi: 10.1186/s12884-023-06220-1.

8. V. B. Brunelli and F. Prefumo, "Quality of first trimester risk prediction models for pre-eclampsia: A systematic review," BJOG, vol. 122, no. 7, pp. 904–914, Jun. 2015, doi: 10.1111/1471-0528.13334.
9. A. Bertini, R. Salas, S. Chabert, L. Sobrevia, and F. Pardo, "Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review," Frontiers in Bioengineering and Biotechnology, vol. 9. Frontiers Media S.A., Jan. 19, 2022. doi: 10.3389/fbioe.2021.780389.
10. M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," Syst Rev, vol. 10, no. 1, Dec. 2021, doi: 10.1186/s13643-021-01626-4.
11. C. Ao, W. Zhou, L. Gao, B. Dong, and L. Yu, "Prediction of antioxidant proteins using hybrid feature representation method and random forest," Genomics, vol. 112, no. 6, pp. 4666–4674, Nov. 2020, doi: 10.1016/j.ygeno.2020.08.016.
12. M. Velikova, P. J. F. Lucas, and M. Spaanderman, "A Predictive Bayesian Network Model for Home Management of Preeclampsia," in Peleg, M., Lavrač, N., Combi, C. (eds) Artificial Intelligence in Medicine. AIME 2011. Lecture Notes in Computer Science(LNAI,volume 6747), Springer, Berlin, Heidelberg, 2011, pp. 179–183. doi: 10.1007/978-3-642-22218-4_22.
13. A. Martínez-Velasco, L. Martínez-Villaseñor, and L. Miralles-Pechuán, "Machine learning approach for pre-eclampsia risk factors association," in GOODTECHS 2018 – 4th EAI International Conference on Smart Objects and Technologies for Social Good, Association for Computing Machinery, Nov. 2018, pp. 232–237. doi: 10.1145/3284869.3284912.
14. V. P. Kovacheva, B. W. Eberhard, R. Y. Cohen, M. Maher, R. Saxena, and K. J. Gray, "Preeclampsia Prediction Using Machine Learning and Polygenic Risk Scores from Clinical and Genetic Risk Factors in Early and Late Pregnancies," Hypertension, vol. 81, no. 2, pp. 264–272, Feb. 2024, doi: 10.1161/HYPERTENSIONAHA.123.21053.
15. B. W. Eberhard, R. Y. Cohen, J. Rigoni, D. W. Bates, K. J. Gray, and V. P. Kovacheva, "An Interpretable Longitudinal Preeclampsia Risk Prediction Using Machine Learning.," medRxiv, Aug. 2023, doi: 10.1101/2023.08.16.23293946.
16. Z. Li et al., "Dynamic gestational week prediction model for pre-eclampsia based on ID3 algorithm," Front Physiol, vol. 13, Oct. 2022, doi: 10.3389/fphys.2022.1035726.
17. M. Liu et al., "Development of a prediction model on preeclampsia using machine learning-based method: a retrospective cohort study in China," Front Physiol, vol. 13, Aug. 2022, doi: 10.3389/fphys.2022.896969.
18. Y. xin Li et al., "Novel electronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms," Pregnancy Hypertens, vol. 26, pp. 102–109, Dec. 2021, doi: 10.1016/j.preghy.2021.10.006.
19. H. Sufriyana, Y. W. Wu, and E. C. Y. Su, "Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia," EBioMedicine, vol. 54, Apr. 2020, doi: 10.1016/j.ebiom.2020.102710.
20. C. Neocleous, K. Nicolaides, K. Neokleous, and C. Schizas, "Ethnicity as a Factor for the Estimation of the Risk for Preeclampsia: A Neural Network Approach," Springer-Verlag, 2010.
21. J. Torres-Torres et al., "Performance of a machine learning approach for the prediction of pre-eclampsia in a middle-income country," Ultrasound in Obstetrics & Gynecology, Sep. 2023, doi: 10.1002/uog.27510.
22. M. M. Gil et al., "Validating a machine-learning model for first-trimester prediction of pre-eclampsia using the cohort from the PREVAL study," Ultrasound in Obstetrics & Gynecology, Jan. 2023, doi: 10.1002/uog.27478.
23. Z. Ansbacher-Feldman, A. Syngelaki, H. Meiri, R. Cirkin, K. H. Nicolaides, and Y. Louzoun, "Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers," Ultrasound in Obstetrics and Gynecology, vol. 60, no. 6, pp. 739–745, Dec. 2022, doi: 10.1002/uog.26105.
24. A. S. Melinte-Popescu, I. A. Vasilache, D. Socolov, and M. Melinte-Popescu, "Predictive Performance of Machine Learning-Based Methods for the Prediction of Preeclampsia—A Prospective Study," J Clin Med, vol. 12, no. 2, Jan. 2023, doi: 10.3390/jcm12020418.
25. I. Marić et al., "Early prediction of preeclampsia via machine learning," Am J Obstet Gynecol MFM, vol. 2, no. 2, May 2020, doi: 10.1016/j.ajogmf.2020.100100.
26. A. Sandström, J. M. Snowden, J. Höijer, M. Bottai, and A. K. Wikström, "Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: A population based cohort study," PLoS One, vol. 14, no. 11, Nov. 2019, doi: 10.1371/journal.pone.0225716.
27. G. James, D. Witten, T. Hastie, and R. Tibshirani, "Statistical Learning," in An Introduction to Statistical Learning with Applications in R, Second Edition., Springer, 2023, pp. 15–58.

 doi:10.20944/preprints202406.0725.v1

28. G. James, D. Witten, T. Hastie, and R. Tibshirani, "Linear Regression," in An Introduction to Statistical Learning with Applications in R, Second Edition., Springer, 2023, pp. 59–128.
29. M. Lazdam et al., "Unique Blood Pressure Characteristics in Mother and Offspring After Early Onset Preeclampsia," pp. 1338–1345, 2012, doi: 10.1161/HYPERTENSIONAHA.112.198366.
30. C. Macdonald-Wallis, D. A. Lawlor, A. Fraser, M. May, S. M. Nelson, and K. Tilling, "Preeclampsia Blood Pressure Change in Normotensive, Gestational Hypertensive, Preeclamptic, and Essential Hypertensive Pregnancies," 2012, doi: 10.1161/HYPERTENSIONAHA.