

Review

Comparison of Multivariable Logistic Regression and Other Machine Learning Algorithms for Prognostic Prediction Studies in Pregnancy Care: Systematic Review and Meta-Analysis

Herdiantri Sufriyana^{1,2}, MSc, MD; Atina Husnayain^{1,3}, MPH; Ya-Lin Chen^{1,4}, PharmD; Chao-Yang Kuo¹, MSc; Onkar Singh^{5,6}, MSc; Tso-Yang Yeh⁷; Yu-Wei Wu^{1,8}, PhD; Emily Chia-Yu Su^{1,8}, PhD

¹Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

²Department of Medical Physiology, College of Medicine, University of Nahdlatul Ulama Surabaya, Surabaya, Indonesia

³Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

⁴School of Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan

⁵Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Science, Academia Sinica, Taipei, Taiwan

⁶Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

⁷School of Dentistry, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

⁸Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan

Corresponding Author:

Emily Chia-Yu Su, PhD

Graduate Institute of Biomedical Informatics

College of Medical Science and Technology

Taipei Medical University

250 Wu-Xing Street

Taipei, 11031

Taiwan

Phone: 886 2 663 82736 ext 1515

Email: emilysu@tmu.edu.tw

Abstract

Background: Predictions in pregnancy care are complex because of interactions among multiple factors. Hence, pregnancy outcomes are not easily predicted by a single predictor using only one algorithm or modeling method.

Objective: This study aims to review and compare the predictive performances between logistic regression (LR) and other machine learning algorithms for developing or validating a multivariable prognostic prediction model for pregnancy care to inform clinicians' decision making.

Methods: Research articles from MEDLINE, Scopus, Web of Science, and Google Scholar were reviewed following several guidelines for a prognostic prediction study, including a risk of bias (ROB) assessment. We report the results based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Studies were primarily framed as PICOTS (population, index, comparator, outcomes, timing, and setting): Population: men or women in procreative management, pregnant women, and fetuses or newborns; Index: multivariable prognostic prediction models using non-LR algorithms for risk classification to inform clinicians' decision making; Comparator: the models applying an LR; Outcomes: pregnancy-related outcomes of procreation or pregnancy outcomes for pregnant women and fetuses or newborns; Timing: pre-, inter-, and peripregnancy periods (predictors), at the pregnancy, delivery, and either puerperal or neonatal period (outcome), and either short- or long-term prognoses (time interval); and Setting: primary care or hospital. The results were synthesized by reporting study characteristics and ROB and by random effects modeling of the difference of the logit area under the receiver operating characteristic curve of each non-LR model compared with the LR model for the same pregnancy outcomes. We also reported between-study heterogeneity by using τ^2 and I^2 .

Results: Of the 2093 records, we included 142 studies for the systematic review and 62 studies for a meta-analysis. Most prediction models used LR (92/142, 64.8%) and artificial neural networks (20/142, 14.1%) among non-LR algorithms. Only 16.9% (24/142) of studies had a low ROB. A total of 2 non-LR algorithms from low ROB studies significantly outperformed

LR. The first algorithm was a random forest for preterm delivery (logit AUROC 2.51, 95% CI 1.49-3.53; $I^2=86\%$; $\tau^2=0.77$) and pre-eclampsia (logit AUROC 1.2, 95% CI 0.72-1.67; $I^2=75\%$; $\tau^2=0.09$). The second algorithm was gradient boosting for cesarean section (logit AUROC 2.26, 95% CI 1.39-3.13; $I^2=75\%$; $\tau^2=0.43$) and gestational diabetes (logit AUROC 1.03, 95% CI 0.69-1.37; $I^2=83\%$; $\tau^2=0.07$).

Conclusions: Prediction models with the best performances across studies were not necessarily those that used LR but also used random forest and gradient boosting that also performed well. We recommend a reanalysis of existing LR models for several pregnancy outcomes by comparing them with those algorithms that apply standard guidelines.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42019136106; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=136106

(*JMIR Med Inform* 2020;8(11):e16503) doi: [10.2196/16503](https://doi.org/10.2196/16503)

KEYWORDS

machine learning; pregnancy complications; prognosis; clinical prediction rule; meta-analysis; systematic review

Introduction

Background

Pregnancy is a common health condition that requires long-term rigorous care to anticipate adverse outcomes. Most pregnancy outcomes are identified after delivery; however, these are results of interactions among multiple factors occurring for many weeks beforehand. The number of factors and their interactions along with the time intervals make predictions of pregnancy outcomes very complicated. Multiple or multivariable logistic regression (LR) is widely used to deal with similar multifactorial problems in health outcome research [1]. Applied to medicine, statistics, and machine learning (computer science), this algorithm fits multiple parameters in a prediction model by assuming that predictors are linearly and additively related to an outcome [2]. Nevertheless, nonlinear problems commonly occur in human physiology because of complex interactions, such that a linear model might not be capable of adequately predicting outcomes [3]. With the growth of machine learning applications in health care, applying other algorithms may scale up the solution space for accurate predictions of pregnancy outcomes long before giving birth.

Despite improvements in maternal and neonatal mortality, conditions still differ between developing and developed countries or regions [4]. The most common causes of maternal deaths are hemorrhage, hypertension, and sepsis [5], whereas the causes of neonatal deaths are mostly due to prematurity, birth asphyxia, and infections [6]. Postpartum hemorrhage and sepsis are further compounded by multiple causes and risk factors [7,8], and hypertension in pregnancy or prematurity is associated with multiple mechanisms [9,10]. The aforementioned diseases and complications cannot be very easily predicted by a single epidemiological predictor, a single measure by a medical device, or a single biomarker. Furthermore, interactions among multiple predictors also might not be captured by a single machine learning algorithm including LR. Therefore, a prediction study may need to compare multiple machine learning algorithms to develop a prognostic prediction model that uses multiple predictors.

Machine learning algorithms have long been applied for clinical prediction purposes. A support vector machine demonstrated a

summary of receiver operating characteristics (ROCs) of >90% for breast cancer prognostic prediction [11]. To predict therapeutic outcomes in depression, the pooled estimated accuracy of machine learning algorithms was 0.82 (95% CI 0.77-0.87) [12]. However, the difference in the logit area under the ROC curve (AUROC) was 0.00 (95% CI -0.18 to 0.18) between LR and machine learning in studies with a low risk of bias (ROB) [13]. A similar conclusion was found for predicting intracerebral hemorrhage ($P=.49$) outlined in a systematic review [14]. These previous results imply that (1) machine learning algorithms may or may not perform better than traditional modeling by LR and (2) applying only a single algorithm may cause an investigator to lose the chance to obtain a model with optimal predictive performance using the same predictors. Meanwhile, a unique interaction should exist between a set of predictors and a pregnancy outcome. A particular predictive algorithm may work best to capture this predictor-outcome interaction. Prediction tasks are even more challenging in pregnancy care because they demand more prognostic instead of diagnostic predictions. Yet, unlike the common nature of other long-term conditions in health care (eg, diabetes mellitus), the onset, time to event, and target population in pregnancy care are rather apparent. However, unpredictable events leading to disabilities and death in a population such as pregnant women or newborns are also not easily accepted as in other populations (eg, patients with cancer and older adults). Thus, clinicians should apply several prediction models with satisfactory predictive performances throughout the pregnancy period. Clinicians and investigators would benefit from knowing whether an LR or other algorithms have a better chance of achieving satisfactory predictive performances for a particular pregnancy outcome. However, no previous systematic review in pregnancy care has reviewed multiple machine learning algorithms and compared their predictive performances, including LR, to predict pregnancy outcomes.

This review will allow investigators and clinicians in pregnancy care to consider the development or application of prediction models throughout the pregnancy period. This review demonstrates which algorithms have shown robust predictive performances for a particular pregnancy outcome using a similar set of predictors. Investigators in pregnancy care may also consider whether a reanalysis by another predictive algorithm

is needed by using existing data previously analyzed by an algorithm including LR. Beyond the algorithm issue, the development of machine learning models also requires an adequate methodology and interpretable results [15]. Biased conclusions should be avoided when describing machine learning predictive performances [11,16]. Standard guidelines are important when investigating and reviewing machine learning applications in clinical prediction modeling [15,17].

Objectives

By applying the standard guidelines, we aim to review machine learning models and compare their predictive performances between LRs and other machine learning algorithms. In this review, we focus on machine learning models either developed or validated for making prognostic predictions in pregnancy care intended to inform clinicians' decision making.

Methods

Protocol and Registration

We reported this study based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [18] and conducted the review based on several guidelines related to prediction studies. The review objective was defined according to a standard of key items [19]. Our eligibility criteria were composed of items elaborated with 2 guidelines for developing and reporting a prediction model and a guideline for assessing the applicability. These included transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [20] and another that focuses on machine learning modeling in biomedical research (hereafter referred to as guidelines for developing and reporting machine learning predictive models in biomedical research [MLP-BIOM]) [15]. Applicability was assessed using assessments that were a part of the *prediction model risk of bias assessment tool* (PROBAST) [17,21]. Data were extracted based on the checklist for critical appraisal and data extraction for systematic reviews of prediction modeling studies (CHARMS), which also describes items for the review objective. Our review protocol was registered with PROSPERO (CRD42019136106).

Eligibility Criteria

Before defining the eligibility criteria, we decided to view the LR as one of many algorithms in the machine learning field with respect to its use in statistics and data science. A prediction model development consisted of several elements: predictor selection, parameter fitting, and hyperparameter optimization [2]. In this review, the term *prediction model* refers to all those elements, whereas the term *prediction algorithm* refers to a parameter-fitting method. Using the same set of predictors, we would expect different predictive performances if the parameters of a model are fitted using different algorithms. A prediction algorithm in machine learning is a way for the computer to learn from data by fitting the parameters with respect to predicting a class measured by hyperparameters from the human user [22]. Several optimization algorithms have been developed to reduce the human role in determining these hyperparameters, such as sequential search, random search, and Bayesian optimization [23]. However, that is beyond the scope of this review.

By focusing on prediction algorithms, we defined eligibility criteria to screen studies by the title, abstract, and full text. We also assessed the applicability by examining the full text. These were the candidates we selected for the qualitative analysis. Key items of population, index, comparator, outcomes, timing, setting (PICOTS) [19] and additional items [15,20] composed the eligibility criteria. The first item of these criteria was a review question framed using PICOTS. The key items consisted of the following:

1. Population: men or women in procreative management, pregnant women, and fetuses or newborns.
2. Index: multivariable prognostic prediction models applying non-LR algorithms for risk classification tasks intended to inform clinicians' decision making.
3. Comparator: multivariable prognostic predictions applying an LR algorithm, excluding a scoring system in which the parameters determined by humans instead of using LR, for risk classification tasks intended to inform clinicians' decision making.
4. Outcomes: pregnancy-related outcomes of procreative management or pregnancy outcomes for pregnant women or fetuses or newborns.
5. Timing: with predictors being measured at the pre-, inter-, and peripregnancy periods and outcomes being assessed at the pregnancy, delivery, and either puerperal or neonatal period, short- and long-term prognoses were applied.
6. Setting: primary care or hospital.

Additional items were the availability of several reporting components as required by TRIPOD and MLP-BIOM. These components included (1) data sources, (2) outcomes, (3) evaluation metrics, (4) predictors, (5) descriptive statistics, (6) event sample sizes, (7) modeling methods or algorithms, and (8) model validation.

After briefly screening studies by eligibility criteria, we conducted an applicability assessment by thoroughly examining the full texts. Using PROBAST guidelines, we assessed the applicability according to the review question framed by PICOTS. Low, high, or unclear criteria were determined for applicable, not applicable, or unclear applicability, respectively. The assessment covered 3 domains of participants, predictors, and outcomes. Only those fulfilling *low* criteria were selected for the qualitative analysis.

For the quantitative analysis, studies had to report the AUROC. Studies were selected from those applicable for the qualitative analysis. If there were at least three LR models and a non-LR model from any studies for an outcome, all studies with that outcome were included in the meta-analysis. This was determined based on the requirement of a minimum number of data points to calculate the variance as part of the meta-analytical procedure. If studies did not report the AUROC, we estimated the sensitivity and specificity using the trapezoidal rule (see *Summary Measures* and *Synthesis of Results* sections).

Information Sources

We searched the MEDLINE, Scopus, Web of Science, and Google Scholar databases up to May 2020. There was no limit on the publication period. However, considering the limitations

of the search interface in Google Scholar, we only retrieved results from the last year with keywords in the abstract or the entire period with those keywords in the title. We also limited the publication period to the last 10 years for search results by keywords including “logistic regression multivariable prediction.” This was because we estimated that there would be enormous amounts of studies applying LR because we applied a broad range of outcomes in this study. In contrast, we might lack studies using other machine learning models, although the outcomes were broad.

Search

The initial search filter was limited to the title, abstract, keywords, or Medical Subject Heading (MeSH; MEDLINE only) using “machine learning” AND pregnancy. We also used “machine learning AND ([pregnancy outcome from initial search] NOT pregnancy).” Keywords for pregnancy outcomes were used based on MeSH to generalize a variety of terms for pregnancy outcomes from selected studies. If the MeSH term contained “pregnancy,” then we used the alternative entry terms in the webpage recorded for this MeSH term. If all entry terms also contained “pregnancy,” then we used the term without negating “pregnancy.” In addition, we also substituted the “machine learning” part with one of the keywords consisting of “decision tree,” “artificial neural network,” “support vector machine,” “random forest,” “artificial intelligence,” “deep learning,” and “logistic regression multivariable prediction.” All keywords are described in [Multimedia Appendix 1](#). These search terms were applied to all databases.

Study Selection

Duplicate records from multiple databases were removed. We refined the search results in the title or abstract using EndNote X8 (Clarivate Analytics) by “(supervised NOT unsupervised) OR prediction OR classification.” Records were screened by HS and AH, and the results were assessed by HS, AH, YC, CK, OS, TY, and YW. Disagreements were resolved by discussion with the last author (ES). Study selection was conducted in brief and thorough assessments. These brief assessments were intended to select studies by checking eligibility criteria from TRIPOD and MLP-BIOM in the title, abstract, and briefly in the full-text article. A thorough assessment of the applicability from PROBAST was conducted later before the ROB assessment.

Data Collection Process

We extracted data based on the CHARMS checklist, which includes (1) outcomes, (2) study design, (3) data sources, (4) data source design, (5) setting, (6) type of study, and (7) modeling methods or algorithms, and (8) predictive performance. Outcomes were pooled as distinct MeSH terms. Study and data source designs were classified into prospective, retrospective, nested case-control, case-control, and cross-sectional. We defined the type of study based on the model validation, which might be development, validation, or both. Eligible studies were described as developing prediction models by applying LR, non-LR, or both algorithms. Predictive performances were only taken from studies that were eligible for the meta-analysis (see *Eligibility Criteria* section). If there

were multiple models developed within a study using the same algorithm, we retrieved the AUROC from the best performing one among the models. If both LR and non-LR algorithms were applied in a study, we selected the predictive performances of the best models applying either the LR or non-LR algorithm. Model performances derived from external validation were preferred if available.

ROB Within and Across Studies

We used PROBAST to assess the ROB [17,21]. The ROB in individual studies was assessed as low, high, or unclear in 4 domains of participants, predictors, outcomes, and analyses. In addition, 20 signaling questions were answered for each study in a transparent and accountable form. Across studies, we described the proportion of low, unclear, or high ROB. ROB were compared for each domain. We also summarized the answers for each signaling question.

Summary Measures

We compared AUROCs from studies that reported this metric. Logit transformation was applied to the AUROCs. We computed logit AUROC differences between each non-LR and LR algorithm across studies. Summary measures from any eligible studies with all, low, or high ROB were pooled by random effects modeling, as previously described [24]. Assuming that selected studies were random samples from a larger population, we chose a random effects model that attempted to generalize findings beyond the included studies using that assumption [25]. Despite this, we did not conduct random effects modeling for all selected studies considering the broad range of target populations, outcomes, and algorithms. Meanwhile, we conducted this review within a narrower field compared with a previous systematic review of machine learning in medicine [13]. Therefore, we only applied random effects modeling to the predictive performances of selected studies using a particular pregnancy outcome. These studies consisted of a minimum number of non-LR and 3 LR models from any studies. This minimum number was considered to obtain a minimum number of data points of logit AUROCs to compute the interval estimates in a random effects model. We depicted the AUROCs using forest plots; thus, one can see which prediction algorithm may have a better chance of obtaining optimal predictive performance for a particular pregnancy outcome.

Pooled estimates of pairwise differences in logit AUROCs were described by points and the 95% CI [26]. A positive difference in logit AUROCs means that the non-LR algorithm had a higher logit AUROC than that of the LR algorithm. The difference was significant if 0 was not included within the 95% CI. The number of pairwise comparisons (k) for each random effects model was reported. We also reported variance across studies (τ^2) and I^2 as absolute and relative values of between-study heterogeneity, respectively.

If a study did not report the AUROC, we estimated this metric based on sensitivity and specificity. As a specificity of 0% means a sensitivity of 100% and *vice versa*, the AUROC could be estimated from the reported sensitivity and specificity using a common rule to calculate the area of the trapezoid (Equation 1). Before we subtracted the AUROC of a non-LR algorithm

from that of an LR algorithm, we applied a logit transformation (Equation 2).

$$\text{AUROC} = 0.5 \times (1 - \text{specificity}) \times \text{sensitivity} + \text{specificity} \times \text{sensitivity} + 0.5 \times (1 - \text{sensitivity}) \times \text{specificity} \quad (1)$$

$$\text{Logit}(\text{AUROC}) = \log(\text{AUROC} / (1 - \text{AUROC})) \quad (2)$$

We used RStudio 1.2 (RStudio) with R 3.6.1 and an additional package, metafor 2.4.0, for random effects modeling. We applied the restricted maximum likelihood estimator method [27]. These are common tools and recommended modeling methods for meta-analyses [28].

Synthesis of Results

We described the characteristics of the studies consisting of population, study design, timing, and setting. This was described as the number of algorithms used for prediction modeling. The algorithms were categorized into LR, non-LR, or both algorithms. We also show the proportion of each characteristic compared with all characteristics within the same algorithm category.

ROBs within studies were described for the number of low, high, or unclear ROB studies. This was reported for overall assessment results and by domain in studies that used LR, non-LR, or both algorithms. ROBs across studies were described for the proportion of studies in which the answer to each signaling question led to low, high, or unclear ROB studies. We intended to show what makes most studies considered to have high ROBs.

Meta-analytical results were described by a forest plot faceted by outcome. Each facet showed comparisons of differences in logit AUROCs for each random effects model of non-LR versus LR algorithms. This demonstrated which algorithms tended to

outperform LR for each pregnancy outcome. Comparisons that included non-LR high ROB studies were color coded. The best predictive performance for each outcome was reported. Between-study heterogeneity for each random effects model was also reported.

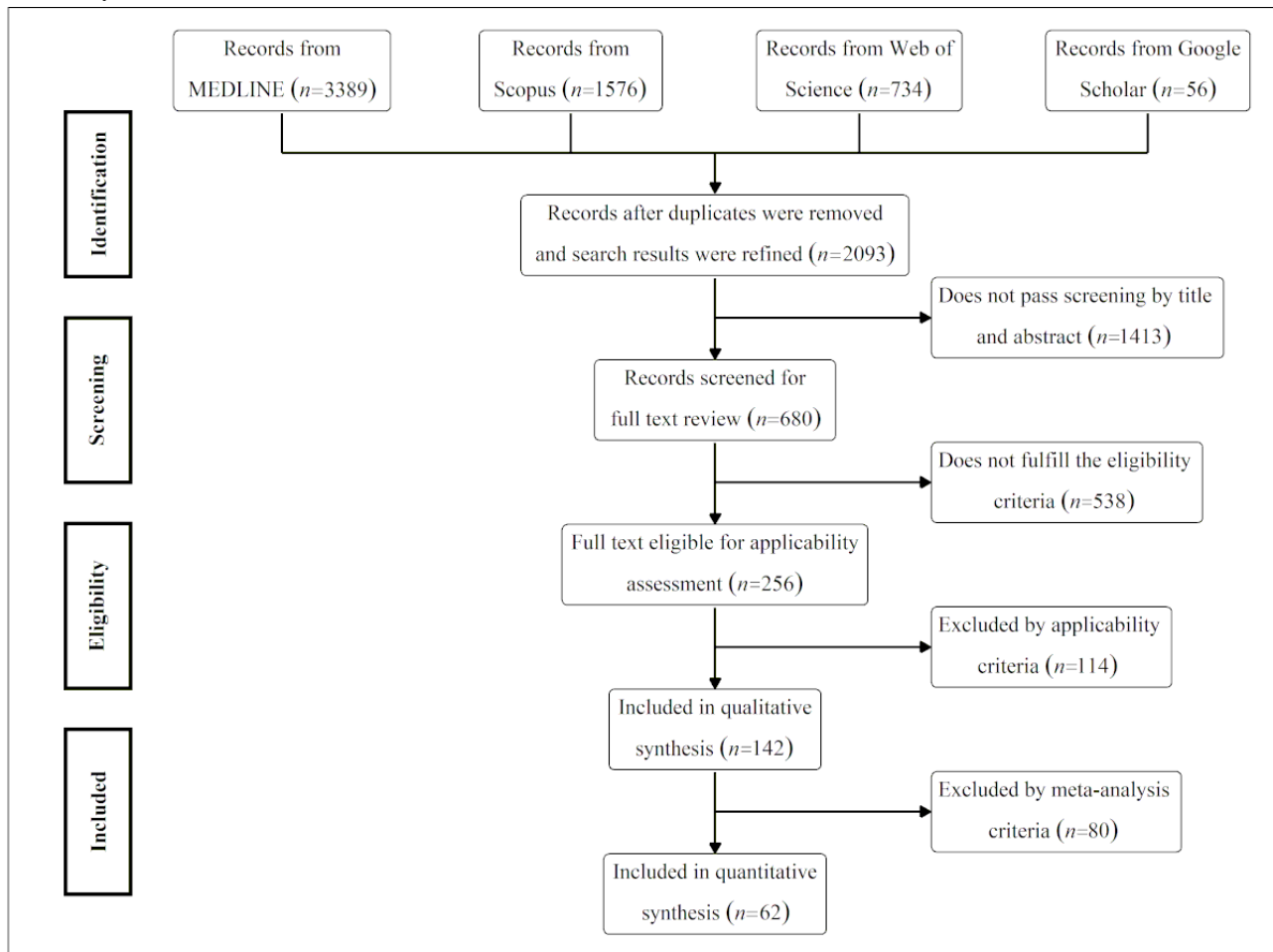
We described predictors in the prediction models from studies in the meta-analysis. For each outcome in the meta-analysis, we selected only random effects models in which an algorithm significantly outperformed the other. This was determined by the 95% CI of the difference in logit AUROCs between a non-LR and an LR model for an outcome. If any, we only selected those that included only non-LR low ROB studies. Only predictors in the final model were included. This was intended to elucidate predictor-outcome interactions that characterized an algorithm if it outperformed the others for a particular outcome.

Results

Study Selection

We found 2093 records from 4 literature databases (Figure 1). The search filters consisted of 144 combinations of keywords from 8 machine learning terms and 18 MeSH terms for pregnancy outcomes recursively derived from the keywords “machine learning AND pregnancy” (Multimedia Appendix 1). We refined the search results, identified research articles (not including conference abstracts or theses), and removed duplicates. After screening and eligibility assessment, we included 142 studies for the qualitative analysis, of which 62 were used for the quantitative analysis. A detailed description of the eligible criteria, process of study selection, and list of studies for the full-text review are given in Multimedia Appendix 1.

Figure 1. Study selection workflow.



Characteristics of the Studies

Briefly, we collected studies that either developed or validated a prediction model applying either LR (77/142, 54.2%) [29-105] or non-LR machine learning algorithms (50/142, 35.2%; Table 1) [106-155]. Overall, 15 studies applied both LR and non-LR algorithms (15/142, 10.6%) [156-170]. The cohort population of the studies in this review consisted of every type of population, study design, timing, and setting that we desired to discuss in this review. More studies discussed fetuses or newborns than pregnant women in non-LR prediction studies (26/50, 52% vs 11/50, 22%). Meanwhile, the opposite occurred in LR studies that focused more on pregnant women than fetuses or newborns (50/77, 65% vs 19/77, 25%). Most used data sets were from retrospective cohorts for LR (53/77, 69%) [29-36,38-42,47,49-54,56-61,64-66,68-71,76-80,83,87-90,92,94,95,97,100-105], non-LR prediction studies (27/50, 54%) [107,108,111-113,116,117,121,122,127,130-138,140,142,143,148,150,151,153,154], or both (9/15, 60%) [157-160,163-165,167-170]. A retrospective cohort is one of the recommended study designs for prognostic purposes instead of diagnostic prediction [21].

This corresponds to our review question that warrants prognostic predictions in pregnancy care intended to inform clinicians' decision making.

Only a few studies had prediction timing up to the puerperal or neonatal period for LR (2/77, 3%) [74,85], non-LR (3/50, 6%) [114,129,149], or both algorithms (2/15, 13%) [162,168]. This is because some predictors were assessed after delivery, whereas our review question demanded those be assessed up to delivery. We also considered studies using data sets from either primary care or hospital settings because the data are applicable for clinicians' decision making on a daily basis. As applicability was already included in the eligibility assessment before the qualitative analysis, eligible studies were not found to use data sets from either primary care or hospital settings, such as from a house-to-house survey or a screening program. Most used data sets were from hospital settings, whereas only a few of those were from primary care settings in the LR (6/77, 8%) [65,69,73,77,78,87], non-LR (6/50, 12%) [119,122,132,135,148,153], or both algorithms (1/15, 7%) [162]. A detailed description of this is also given in Multimedia Appendix 1.

Table 1. Characteristics of eligible studies.

Variable	Number of studies (percentage based on column total)			
	LR ^a (n=77), n (%)	Non-LR (n=50), n (%)	Both (n=15), n (%)	Total (n=142), n (%)
Population				
Pregnant women	50 (65)	11 (22)	6 (40)	67 (47.2)
Fetuses or newborns	19 (25)	26 (52)	7 (47)	52 (36.6)
Men or women in procreative management	8 (10)	13 (26)	2 (13)	23 (16.2)
Study design				
Retrospective	53 (69)	27 (54)	9 (60)	89 (62.7)
Nested case-control	4 (5)	14 (28)	2 (13)	20 (14.1)
Prospective	13 (17)	4 (8)	0 (0)	17 (12)
Cross-sectional	3 (4)	3 (6)	3 (20)	9 (6.3)
Case-control	4 (5)	2 (4)	1 (7)	7 (4.9)
Timing				
At delivery	28 (36)	26 (52)	7 (46.7)	61 (42.9)
At pregnancy	34 (44)	21 (42)	5 (33.3)	60 (42.3)
Mixed timing	13 (17)	0 (0)	1 (6.7)	14 (9.9)
Puerperal or neonatal period	2 (3)	3 (6)	2 (13.3)	7 (4.9)
Setting				
Hospital	61 (79)	43 (86)	9 (60)	113 (79.6)
Both	10 (13)	1 (2)	5 (33)	16 (11.3)
Primary care	6 (8)	6 (12)	1 (7)	13 (9.2)

^aLR: logistic regression.

LR and Other Machine Learning Algorithms

Most studies applied an LR (92/142, 64.8%) to develop a prediction model (Table 2). Meanwhile, an artificial neural network was mostly applied by non-LR studies (20/142, 14.1%). Studies that applied LR and non-LR algorithms mostly compared LR with an artificial neural network (5/15, 33%) [161,163,165,166,170] and decision tree (5/15, 33%) [156,159,167-169], but decision trees tended to be paired with an LR compared with an artificial neural network (5/7, 71% vs 5/20, 25%).

The characteristics of study populations showed that pregnant women and fetuses or newborns were the populations of most

studies developed using LR and non-LR models, respectively. Among pregnant women, the LR algorithm was mostly applied to develop predictions for outcome categories of obstetric labor (13/77, 17%) [36,46,47,54,57,62,64,70,83,86,91,97,103], pregnancy-induced hypertension (12/77, 16%) [30,31,43,48,55,65,66,68,76,81,93,105], and gestational diabetes (7/77, 9%) [33,45,49,84,94,100,104]. Among fetus or newborn populations, non-LR algorithms were mostly applied to develop predictions for outcome categories of premature birth (12/50, 24%) [111,112,115,116,118,119,121,122,125,130,141,143] and fetal distress (9/50, 18%) [113,124,128,137,138,145,146,152,155]. In addition, more non-LR algorithms (13/20, 65%) were applied for the outcome category of *in vitro* fertilization than for the LR algorithm.

Table 2. Machine learning algorithm and category of outcome.

Variable	Number of studies (percentage based on column total)			
	LR ^a (n=77), n (%)	Non-LR (n=50), n (%)	Both (n=15), n (%)	Total (n=142), n (%)
Machine learning algorithm				
Logistic regression	77 (100)	N/A ^b	15 (100)	92 (64.8)
Artificial neural network	N/A	15 (30)	5 (33)	20 (14.1)
Support vector machine	N/A	9 (18)	1 (7)	10 (7.0)
Deep neural network	N/A	8 (16)	1 (7)	9 (6.3)
Random forest	N/A	7 (14)	1 (7)	8 (5.6)
Decision tree	N/A	2 (4)	5 (33)	7 (4.9)
Gradient boosting	N/A	3 (6)	2 (13)	5 (3.5)
Naïve Bayes	N/A	4 (8)	0 (0)	4 (2.8)
Ensemble of algorithms	N/A	2 (4)	0 (0)	2 (1.4)
Category of outcome				
Premature birth	9 (12)	12 (24)	3 (20)	24 (16.9)
In vitro fertilization	7 (9)	13 (26)	2 (13)	22 (15.5)
Obstetric labor	13 (17)	1 (2)	2 (13)	16 (11.3)
Pregnancy-induced hypertension	12 (16)	4 (8)	0 (0)	16 (11.3)
Fetal distress	1 (1)	9 (18)	0 (0)	10 (7.0)
Gestational diabetes	7 (9)	2 (4)	1 (7)	10 (7.0)
Cesarean section	4 (5)	3 (6)	2 (13)	9 (6.3)
Fetal development	4 (5)	1 (2)	0 (0)	5 (3.5)
Small-for-gestational-age infant	3 (4)	1 (2)	1 (7)	5 (3.5)
Others	17 (22)	4 (8)	4 (27)	25 (17.6)

^aLR: logistic regression.

^bN/A: not applicable.

ROB Within and Across Studies

ROB is described for each eligible study in [Multimedia Appendix 1](#) [29-170]. Among the 142 eligible studies, there were 24 (16.9%) low ROB studies [38,61-63,71,98,104,110,113,115,117-119,128,134,141,142,145,147,149,155,157,158,169], 117 (82.4%) high ROB studies [29-37,39-60,64-70,72-97,99-103,105-109,111,112,114,116,120-123,125-127,129-133,135-140,143,144,146,148,150-154,156,159-168,170], and 1 (0.7%) unclear ROB study

([Table 3](#)) [124]. Among the low ROB studies, the categories of outcomes were premature birth (7/24, 30%) [38,63,115,118,119,141,169], fetal distress (5/24, 21%) [71,113,128,145,155], *in vitro* fertilization (4/24, 17%) [61,110,134,158], gestational diabetes (2/24, 8%) [104,157], cesarean section (CS; 2/24, 8%) [117,142], obstetric labor (1/24, 4%) [62], pregnancy-induced hypertension (1/24, 4%) [147], central nervous system malformations (1/24, 4%) [149], and others (1/24, 4%) [98].

Table 3. Risk of bias within studies.

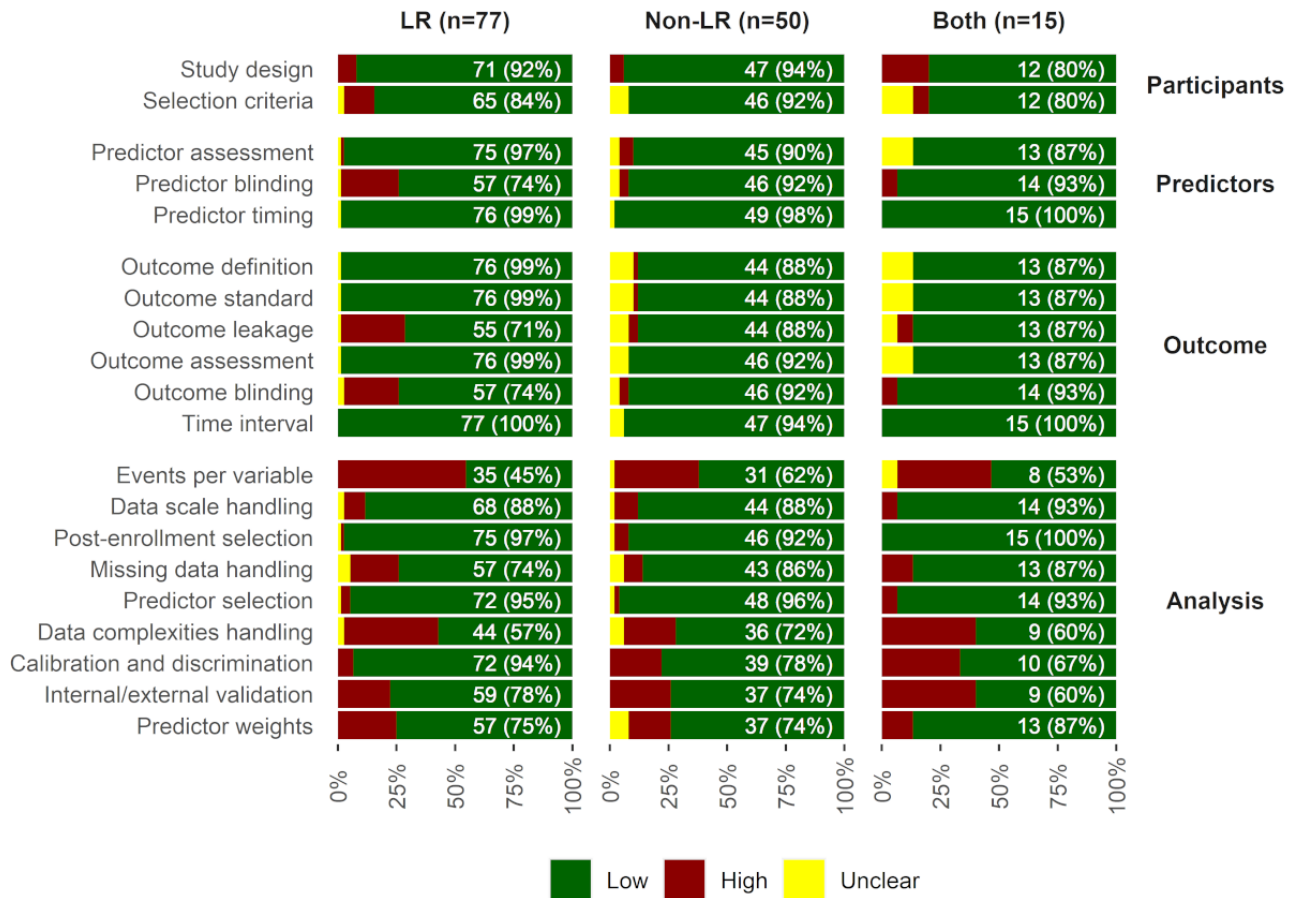
Assessment by domain	Studies by algorithm			
	LR ^a (n=77), n (%)	Non-LR (n=50), n (%)	Both (n=15), n (%)	Total (n=142), n (%)
Participants				
Low	60 (78)	44 (88)	11 (73)	115 (80.9)
High	15 (19)	3 (6)	4 (27)	22 (15.5)
Unclear	2 (3)	3 (6)	0 (0)	5 (3.5)
Predictors				
Low	54 (70)	43 (86)	12 (80)	109 (76.8)
High	20 (26)	5 (10)	1 (7)	26 (18.3)
Unclear	3 (4)	2 (4)	2 (13)	7 (4.9)
Outcome				
Low	51 (66)	40 (80)	11 (74)	102 (71.8)
High	24 (31)	4 (8)	2 (13)	30 (21.1)
Unclear	2 (3)	6 (12)	2 (13)	10 (7.1)
Analysis				
Low	8 (10)	15 (30)	3 (20)	26 (18.3)
High	69 (90)	35 (70)	12 (80)	116 (81.7)
Unclear	0 (0)	0 (0)	0 (0)	0 (0.0)
Overall				
Low	7 (9)	14 (28)	3 (20)	24 (16.9)
High	70 (91)	35 (70)	12 (80)	117 (82.4)
Unclear	0 (0)	1 (2)	0 (0)	1 (0.7)

^aLR: logistic regression.

ROB is also described across the studies in [Table 3](#) and [Figure 2](#). The corresponding signaling questions for each term and the answers for each study are described in [Multimedia Appendix 1](#). Low ROB studies were the fewest in the analysis domain (26/142, 18.3%), consisted of the LR (8/77, 10%) [[38,61-64,71,96,98,104](#)], non-LR (15/50, 30%) [[63,110,113,115,117-119,124,128,134,141,142,145,147,155](#)], and both algorithms (3/15, 20%) [[157,158,169](#)]. In the analysis domain,

the fewest low ROB studies that achieved the minimum events per variable (EPV) consisted of LR (35/77, 45%) and non-LR (31/50, 62%) prediction studies. More calibration and discrimination tests were conducted using LR (72/77, 94%) than by non-LR (39/50, 78%) prediction studies. In contrast, more non-LR prediction studies appropriately handled missing data (43/50, 86%) compared with LR prediction studies (57/77, 74%).

Figure 2. Signaling questions with respect to ROB domains across studies. Bars from low/high/unclear ROB are stacked to be 100%. Domains are described on the right-hand side. The number on the bar is the number of low ROB studies (total LR/non-LR/both at top) based on a single signaling question summarized as a term on the left-hand side. LR: logistic regression; ROB: risk of bias.

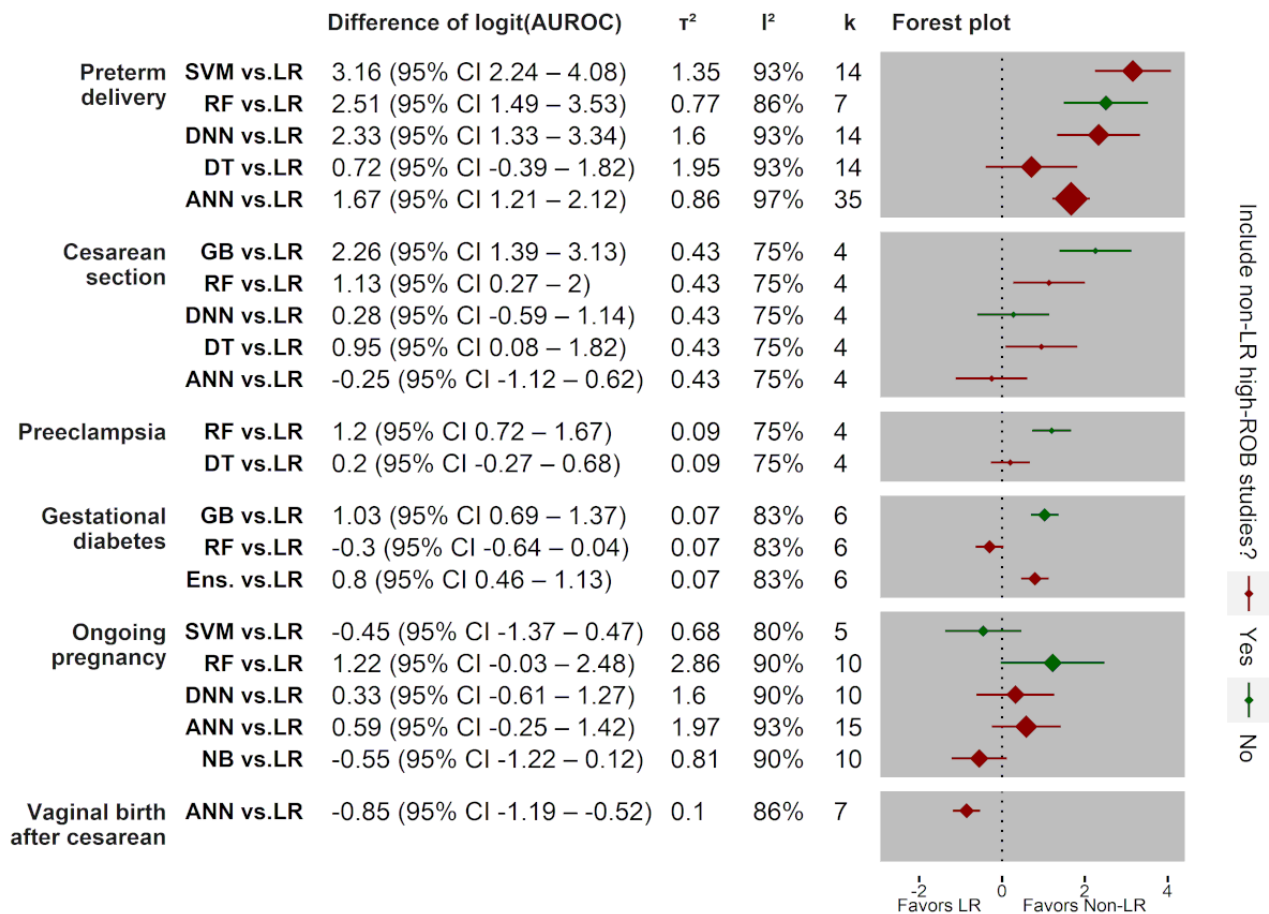


Comparison of the Predictive Performance

There were 62 studies in the meta-analysis that had outcomes that were predicted by at least one non-LR and 3 LR models (see *Summary Measures* section). Overall, 21 random effects models of the predictive performance by non-LR versus LR models are shown in a forest plot (Figure 3). Forest plots of logit AUROC differences for each random effects model are described (Multimedia Appendix 1). With respect to candidate studies (n) included in the final random effects models, we

developed 5 random effects models for preterm delivery (20/62, 32%) [32,44,60,63,75,87,96,111,112,115,118,119,121,125,130,141,143,156,163,169], 5 for CS (7/62, 11%) [79,90,106,117,142,166,167], 2 for pre-eclampsia (6/62, 10%) [31,48,65,76,123,147], 3 for gestational diabetes (9/62, 15%) [33,45,84,94,100,104,108,139,157], 5 for ongoing pregnancy (13/62, 21%) [73,78,99,110,132,134-136,148,150,153,158,170], and 1 for vaginal birth after CS (7/62, 11%) [36,47,57,64,83,97,165].

Figure 3. Forest plot of random effects models for differences in logit AUROCs from a non-LR with any LR prediction models. Plots were grouped by outcome. The lines indicate the 95% CI with diamonds whose sizes were determined by the number of pairwise comparisons (k). Absolute and relative values of between-study heterogeneities are denoted by τ^2 and I^2 , respectively. Colors of the boxes and lines were determined based on the existence of high ROB studies among those using non-LR algorithms. ANN: artificial neural network; AUROC: area under the receiver operating characteristic curve; DNN: deep neural network; DT: decision tree; Ens: ensemble of multiple algorithms; GB: gradient boosting; LR: logistic regression; NB: naïve Bayes; RF: random forest; ROB: risk of bias; SVM: support vector machine.



To determine the final random effects model for each comparison, we identified studies that were responsible as the source of heterogeneity and removed those AUROCs from the random effects model. We excluded a non-LR [121] and an LR study [84] that developed a prediction model for preterm delivery and gestational diabetes, respectively. This is because their AUROCs were outliers compared with those for the same outcome and algorithm. We also excluded 3 LR studies [32,63,87]. In those studies, preterm delivery was defined as delivering within 1 to 2 weeks of preterm labor presentation. Meanwhile, the majority of studies for this outcome defined preterm delivery as that before 37 weeks of gestation.

The non-LR models significantly outperformed the LR models in preterm delivery (4/5 non-LR models), CS (3/5 non-LR models), pre-eclampsia (1/2 non-LR models), and gestational diabetes (2/3 non-LR models). From those that examined preterm delivery, a prediction model did not include a non-LR high ROB study [115] compared with those from 7 LR studies [32,44,60,63,75,87,96]. This model applied a random forest (differences in logit AUROC 2.51; 95% CI 1.49-3.53). The same algorithm was applied to a prediction model from a non-LR low ROB study in pre-eclampsia [147]. For random effects modeling, this model also significantly outperformed those from 4 LR studies (1.2, 95% CI 0.72-1.67) [31,48,65,76].

Meanwhile, prediction models from non-LR low ROB studies of Saleem et al [142] and Artzi et al [157] significantly outperformed those from the corresponding LR studies as an aggregate for CS (2.26, 95% CI 1.39-3.13) and gestational diabetes (1.03, 95% CI 0.69-1.37). Interestingly, the models were developed using a gradient boosting algorithm that used multiple decision trees similar to a random forest.

In contrast, a prediction model using a non-LR algorithm significantly underperformed compared with those using an LR in a random effects model (-0.85, 95% CI -1.19 to -0.52). This applied an artificial neural network to predict vaginal birth after a CS [165]. This model underperformed compared with those from 7 LR studies [36,47,57,64,83,97,165]. However, the non-LR study was a high ROB study.

A random effects model developed for comparison of artificial neural networks and LR to predict preterm delivery had the highest heterogeneity by I^2 (97%; $k=35$). This number means that 97% of the total variability among 35 data points of differences in logit AUROCs was caused by between-studies heterogeneity instead of sampling error within each study [171]. This is reasonable because a higher variance occurs with a larger number of comparisons within a random effects model. In contrast, a random effects model with the smallest number of

comparisons ($k=4$) also had the lowest heterogeneity by I^2 (75%). This random effects model was developed to analyze comparisons of non-LR and LR algorithms for either CS or pre-eclampsia. Nevertheless, a diverse target population and hyperparameter optimization conceivably caused the heterogeneity of the predictive performance, although the same outcome was predicted using the same data set and machine learning algorithm. The lowest I^2 in this meta-analysis remains classified as substantial heterogeneity instead of moderate or unimportant; thus, performing random effects instead of fixed effect modeling is recommended to address this issue [172].

However, I^2 only indicates that the difference in logit AUROCs substantively varies across studies but does not tell how much this metric varies [173]. To interpret the absolute heterogeneity for the difference in logit AUROCs, we needed to consider the observed AUROC of a non-LR model for each of the random effects models. The observed AUROCs were described for each of the original studies in this meta-analysis in [Multimedia Appendix 1](#).

A random effects model developed for comparison of random forests and LR to predict ongoing pregnancy had the highest absolute value of heterogeneity ($\tau^2=2.86$). In this random effects model, random forests were applied to develop predictions in 2 studies that reported AUROCs of 0.740 (95% CI 0.710-0.770) [158] and 0.9820 [134]. We simulated a sequence of logit AUROCs to identify equivalent differences in AUROCs to approximate a difference of the logit value in the random effects model (1.22, 95% CI -0.03 to 2.48). AUROC differences of 0.206 and 0.026 were equivalent to a difference in the logit AUROC of 0.91, compared with those aggregated from LR models for the random forest models of Blank et al [158] and Mirroshandel et al [134], respectively. Using τ^2 , one can calculate the 95% prediction interval (PI) of the logit AUROC difference, as previously described [173]. This estimates the potential AUROC of the random forest to predict ongoing pregnancy with respect to an LR using different populations. For this random effects model, the 95% PI of the logit AUROC difference ranged from -4.75 to 7.19. This is equivalent to 0.257 lower and >0.73 higher than AUROCs of any LRs in the random effects model for the random forest model of Blank et al [158]. For the random forest model of Mirroshandel et al [134], the 95% PI was equivalent to 0.018 lower and 0.943 higher than the AUROCs of any LRs in the random effects model. This is a reasonably wide PI for the highest τ^2 in this meta-analysis, although the non-LR study had a low ROB. This is because ROB only reflects the risk of a predictive performance that differs from the true value of the training sample. However, the ROB does not reflect the difference if the predictive performance is compared with other samples across different populations.

For the random effects model with the lowest τ^2 and including a non-LR low ROB study, the random effects model had a logit AUROC difference of 1.03 (95% CI 0.69-1.37) for a prediction model of gestational diabetes using gradient boosting. The prediction study reported an AUROC of 0.875 (95% CI 0.868-0.885) [157]. The 95% PI of the logit AUROC difference estimated an equivalent AUROC that ranged from 0.0096 lower

to 0.425 higher than the AUROCs of any LR in the random effects model. The gradient boosting model from this study is likely to outperform an LR to predict gestational diabetes.

In addition, we may need to know the τ^2 meaning for the random effects model with the highest I^2 and larger numbers of comparisons (k). This random effects model had an AUROC difference of 1.67 (95% CI 1.21-1.94; 95% PI -2.08 to 5.42; $k=35$) for a prediction model of preterm delivery using an artificial neural network. Overall, 5 non-LR studies were included in this random effects model. The remaining studies reported AUROCs of 0.88 [111], 0.94 [118], 0.945 [125], 0.9115 [163], and 0.911 (95% CI 0.862-0.960) [130]. Considering only the lowest (0.862) and highest (0.960) that covered all of the AUROCs, the artificial neural network model may have AUROCs of 0.119 lower and 0.864 higher than those of any LR. The AUROC interval was also as wide as that of the random effects model with the highest τ^2 .

Descriptive Analysis of Predictors

A random effects model was selected for each outcome except for ongoing pregnancy, which fulfilled our criteria to describe the predictors. For each outcome in the meta-analysis, we selected random effects models in which either a non-LR algorithm significantly outperformed the LR or it was significantly underperformed by the LR. This was determined by the 95% CI of the difference in the logit AUROCs between the non-LR and LR models for an outcome. If any, we only selected those including only non-LR low ROB studies. The random effects models were random forest versus LR for preterm delivery, gradient boosting versus LR for CS, random forest versus LR for pre-eclampsia, gradient boosting versus LR for gestational diabetes, and artificial neural network versus LR for vaginal birth after a CS. As we only extracted the AUROC of either the best LR or non-LR model, only predictors and outcomes of that model were considered if there were multiple models for different subtypes of the outcome in a study.

For preterm delivery, Despotovic et al [115] developed a random forest model using a previously published standardized electrohysterogram (EHG) data set [174]. This data set was also used by other studies in this meta-analysis to predict the same outcome using different algorithms [118,125,130,141,143,169]. All predictors were features extracted from the multichannel EHG obtained at around 22 and 32 weeks of gestation to predict delivery after 39 and 34 to <37 weeks of gestation for term and preterm delivery, respectively. Compared with their counterparts, LR models used predictors consisting of maternal demographics or lifestyle [44,60,75,96,163], medical or obstetric histories [44,75,96,156,163], clinical predictors from obstetrical examinations [44,163], EHG [169], and biomarkers [75]. These were obtained before pregnancy [60,96,156,163], at 11 to 14 weeks of gestation [75], 18 to 34 weeks of gestation [44,163,169], or near events within 1 to 2 weeks [44]. The LR models were developed to predict preterm delivery at 20 to <37 weeks of gestation [44,75,96,163,169] and any delivery at <37 weeks of gestation (predictors could be taken before pregnancy) [60,156].

For CS, Saleem et al [142] developed a gradient boosting model using a previously published standardized cardiotocogram (CTG) data set [175]. This data set was also used by Fergus et al [117] in this meta-analysis to predict the same outcome using a deep neural network. All predictors were features extracted from the CTG data set obtained at first- and second-stage labor for a maximum of 90 min preceding delivery to predict a CS. Compared with their counterparts, LR models used predictors consisting of maternal characteristics [79,90,166], medical histories [167], obstetric histories [90,166,167], and clinical predictors from obstetric examinations [90,166,167], ultrasound measures [79], routine laboratory tests [90], and medications [90]. These were obtained before [90,166,167] and during pregnancy [79,90,166,167]. The LR models were developed to predict CS [166,167], emergency CS [79], and CS in pregnant women with gestational hypertension or mild pre-eclampsia at term [90].

For pre-eclampsia, Sufriyana et al [147] developed a random forest model that used a nationwide health insurance data set. The predictors consisted of maternal demographics and medical histories but excluded obstetric ones. These were obtained before and during pregnancy up to 2 days before the events (pre-eclampsia or eclampsia of any severity and timing). Meanwhile, the LR counterparts used maternal demographics or lifestyle [31,65,76], medical histories [31,65,76], obstetric histories [31,65,76], family histories [31,76], clinical or obstetric examinations [31,65], ultrasound measures [65], routine laboratory tests [76], and biomarkers [48,65]. These predictors were obtained before pregnancy [31], at 11 to 13 weeks of gestation [65], and at <20 weeks of gestation [48]. LR models were developed to predict pre-eclampsia of any severity and timing [31,48,65,76]. The predictors were taken before pregnancy, and this disorder occurs after 20 weeks of gestation by definition.

For gestational diabetes, Artzi et al [157] developed a gradient boosting model that used a nongovernmental, nationwide health care database. The predictors consisted of maternal demographics, medical histories, obstetric histories, clinical or obstetric examinations, routine laboratory tests, and medications. These predictors were obtained before pregnancy and up to 22 weeks of gestation to predict gestational diabetes diagnosed at 24 to 28 weeks of gestation. The LR counterparts used maternal demographics or lifestyle [33,100,104], medical histories [33], obstetric histories [104], family histories [33,45], clinical examinations [33], obstetric examinations [33], routine laboratory tests [33,45,94,100,104], medications, and biomarkers [33,45]. The predictor timing was 6 to 14 weeks of gestation [33,45,94,100,104] and >14 to 22 weeks of gestation [45,100,104]. Meanwhile, the outcome timing was 24 to 28 weeks of gestation [33,45,94,100,104].

For vaginal birth after a CS, Macones et al [165] developed an artificial neural network model that used a medical records database. The predictors used maternal characteristics, medical histories, obstetric histories, obstetric examinations, and labor procedures. These were obtained before pregnancy, during pregnancy, and at labor to predict successful vaginal birth after a CS. The LR counterparts used maternal characteristics [36,47,64,83,97], medical histories [57], obstetric histories

[36,47,57,64,83,97], obstetric examinations [97], and labor procedures [97]. These were obtained before pregnancy [36,47,57,64,83,97], during pregnancy [97], and at labor [97]. The models predicted vaginal birth after a CS with the same definition as those of non-LR studies [36,47,57,64,83].

Discussion

Summary of Evidence

Of the 2093 records from 4 literature databases using 144 keywords, we found 142 eligible studies, among which 24 had a low ROB. These eligible studies developed prediction models for outcome categories of premature birth, *in vitro* fertilization, obstetric labor, pregnancy-induced hypertension, fetal distress, gestational diabetes, CS, fetal development, small-for-gestational-age infants, and others.

There were 4 models with non-LR algorithms from low ROB studies that had significantly higher differences in logit AUROCs than those with LR algorithms. The models used random forest algorithms to predict preterm delivery (2.51, 95% CI 1.49-3.53), gradient boosting algorithms to predict CS (2.26, 95% CI 1.39-3.13), random forest algorithms to predict pre-eclampsia (1.2, 95% CI 0.72-1.67), and gradient boosting algorithms to predict gestational diabetes (1.03, 95% CI 0.69-1.37). The first model that applied a random forest used only EHG records to predict preterm delivery. The second random forest model used only maternal demographics and medical histories but excluded obstetric ones for pre-eclampsia prediction. Meanwhile, the first model that applied a gradient boosting algorithm used only CTG records to predict CSs. The last model was developed by applying a gradient boosting algorithm for gestational diabetes. This model used maternal demographics, medical histories, obstetric histories, clinical or obstetric examinations, routine laboratory tests, and medications.

Comparisons With Prior Work

We compared our systematic review and meta-analysis with prior works related to either machine learning algorithms or pregnancy outcomes similar to those in our study. A recent paper described applications of artificial intelligence in obstetrics and gynecology [176]. That paper was a narrative instead of a scoping or systematic review. Our systematic review and meta-analysis covered all pregnancy outcomes in obstetrics, as described in that paper. These were described as fetal heart monitoring and pregnancy surveillance, gestational diabetes mellitus, preterm labor, parturition, and *in vitro* fertilization.

Nevertheless, the predicted outcomes by non-LR models in our review were still insufficient. Diseases that cause maternal deaths should receive higher priority than those causing neonatal deaths. The risks were higher for pregnant women with antepartum hemorrhage (incidence rate ratio [IRR]=3.5, 95% CI 2.0-6.1) or hypertension (IRR=1.5, 95% CI 1.1-2.2) compared with those without these diseases [177]. Maternal sepsis was also associated with fetal or neonatal deaths (odds ratio [OR] 5.78, 95% CI 2.89-11.21) [178]. Accordingly, the impact of the prediction models may be insufficient to reduce both maternal and neonatal deaths.

LR was found in our study to be the most often used algorithm to develop a prediction model in pregnancy care, including predicted outcomes that caused the most maternal deaths, followed by artificial (shallow) neural networks, support vector machines, and deep neural networks. These corresponded to a systematic review and meta-analysis [13] that showed a similar majority of machine learning algorithms in medicine, except that the study reported classification and regression trees to be the second most often used algorithms (30/71, 42%). All models within eligible studies in that review were included instead of only choosing the best one within each study. Using the same summary measures as we did, the aforementioned review demonstrated that non-LR models from low ROB studies did not outperform LR models. A decision tree showed a difference of logit AUROCs of -0.34 (95% CI -0.65 to -0.04 ; $k=16$) compared with an LR. The review selected 125 eligible studies of 927 candidates from one database. Between-study heterogeneity was not described in that review.

Similar to a previous study [13], a systematic review and meta-analysis did not consider LR as a machine learning algorithm and only compared the predictive performances of non-LR algorithms [179]. This study compared machine learning models to predict any outcomes using routinely collected intensive care unit data. Most of the algorithms were artificial neural networks (72/169, 42.6%), support vector machines (40/169, 23.7%), and decision trees (35/169, 20.7%). However, since 2015, most of the algorithms were support vector machines (37/125, 29.6%) and random forests (72/169, 42.6%). These corresponded to the majority of machine learning algorithms for pregnancy care in our systematic review.

We hold a particular assumption to determine whether interaction of predictors and outcome may be best predicted by a prediction algorithm. If the same predictors and outcomes were used by the best prediction algorithm applied in either non-LR or LR models but not used by the other outcomes in this meta-analysis, then the prediction algorithm may be the best for the pregnancy outcome using those predictors. To predict preterm delivery with predictors that included EHG in either non-LR or LR models [115,169], the random forest outperformed the LR algorithm. Similar to this model in terms of using biomedical signals, gradient boosting also outperformed LR using CTG [142], but none of the LR counterparts used the same predictor. Other predictors were used across outcomes and algorithms (LR or non-LR). These included maternal demographics, lifestyle, medical or obstetric histories, clinical examinations, ultrasound measures, routine laboratory tests, biomarkers, and medication or procedures. Family histories were used in the LR models to predict gestational diabetes in this meta-analysis but were not used by the gradient boosting model (the non-LR counterpart). Therefore, we could not find a convincing pattern of predictors with respect to the best algorithms for each of the other pregnancy outcomes beyond preterm delivery.

Interestingly, the random forest significantly outperformed the LR for almost all of the pregnancy outcomes included in the meta-analysis. Although the gradient boosting algorithm significantly outperformed the LR for CS and gestational diabetes instead of the random forest, gradient boosting also

uses multiple decision trees as in the random forest. For ongoing pregnancy predictions in *in vitro* fertilization, a random forest model from low ROB studies also showed the largest difference in logit AUROCs outperforming LR (1.22, 95% CI -0.03 to 2.48) compared with other non-LR algorithms. For predicting vaginal delivery after a CS, a non-LR algorithm, particularly an artificial neural network in our meta-analysis, did not significantly outperform LR.

Comparing differences in AUROCs and focusing on multiple prediction algorithms, a study with individual participant data also compared LR and non-LR algorithms, particularly Poisson regression, random forest, gradient boosting, and an ensemble of a random forest with either LR or support vector machine [180]. Several models were developed to predict all-cause readmissions in patients with heart failure within 30 and 180 days. The random forest significantly outperformed the LR (0.601, 95% CI 0.594-0.607 vs 0.533, 95% CI 0.527-0.538) for 30-day readmissions. Similar to the random forest, the gradient boosting algorithm (0.613, 95% CI 0.607-0.618) also significantly outperformed the LR. The predictors consisted of medical histories and routine laboratory tests.

Massive evaluation of 179 algorithms from 17 machine learning families was conducted using 121 data sets [181]. The best results were achieved using random forests. In our review, there were 13 studies in which the best models applied either a random forest [106,108,115,134,144,147,155,158] or gradient boosting [127,140,142,157,160]. Random forests used multiple subsets of all samples and predictors randomly with replacement to grow multiple parallel decision trees [182]. Although gradient boosting also uses multiple decision trees, the advantages of random forest over gradient boosting are robust to noise and overfitting [183]. Meanwhile, gradient boosting randomly uses multiple subsets of all samples without replacement to sequentially construct additive regression models [184]. The advantages of gradient boosting over random forests are state-of-the-art predictive performance on tabular data and the customizability of loss of function [181,185]. Hence, several gradient boosting algorithms were developed, and some studies in our review applied these algorithms. To predict gestational diabetes, Artzi et al [157] applied LightGBM, a scalable gradient boosting machine. This algorithm was optimized to speed up the training process by up to 20-fold with the same accuracy [186]. Another gradient boosting system (ie, XGBoost) [187] was applied in a study by Qiu et al [140] to predict live births after *in vitro* fertilization. This study was not included in our meta-analysis because there was an insufficient number of LR [61,69] and gradient boosting [140] algorithms for predicting live births.

Of the pregnancy outcomes predicted by non-LR algorithms in this review, most outcomes were *in vitro* fertilization, premature birth, and fetal distress, possibly because of several reasons. Using keywords of “machine learning IVF” in MEDLINE, we found a review paper from 2011 call for a need for artificial intelligence in *in vitro* fertilization [188]. Only one machine learning study for *in vitro* fertilization was found before that study [189]. All machine learning studies for *in vitro* fertilization were published after the review paper, and most studies were identified within 2093 records in our review

[110,140,150,153,158,190-193]. As prediction for *in vitro* fertilization had already begun by 1989 [194], the machine learning prediction (non-LR) possibly arose because of the 2011 review. Meanwhile, for machine learning predictions of premature birth, fetal distress, and CS, many data sets (25/43, 58%) were secondary instead of primarily collected data. The secondary data sets consisted of predictors and outcomes of EHG and preterm delivery [174] (7/25, 28%), CTG, and acidotic blood pH of the umbilical artery [175] (4/25, 16%), CTG and CS [175] (2/25, 8%), CTG and acidotic blood pH of the umbilical artery [195] (3/25, 12%), EHG and preterm delivery [196] (2/25, 8%), and others (7/25, 28%). This implied that shared data sets drive more machine learning predictions compared with self-collected data sets. This indicates that the increase in publicly available data has driven progress in machine learning applications in health care [197].

For non-LR algorithms, the lack of shared data sets may have been the reason for few prediction studies for maternal outcomes compared with those for neonatal outcomes in this systematic review. Meanwhile, pregnancy-induced hypertension was found in pregnant women of newborns who were born prematurely [198]. Prematurity was also associated with maternal sepsis (OR 2.81, 95% CI 1.99-3.96), including antenatal cases [178]. Therefore, more shared data sets for maternal outcomes are needed. Future studies using machine learning algorithms should develop more prediction models for maternal outcomes in pregnancy care.

In addition, sample sizes of data sets for model development may contribute to bias in predictive performance. For example, in our meta-analysis, prediction models of ongoing pregnancy in *in vitro* fertilization had point estimates of AUROCs ranging from 0.575 to 0.982. These were developed using a support vector machine [110], artificial neural networks [132,136,170], random forests [134,158], deep neural networks [148,153], naïve Bayes algorithms [126,135,150], and LRs [73,78,99,158,170]. Compared with a recent systematic review focusing on prediction for *in vitro* fertilization [143,194], the range of AUROCs was wider than that of the previous review. The AUROCs ranged from 0.59 to 0.775 without non-LR machine learning predictions. A previous review also reported that the sample sizes ranged from 110 to 288,161 instances, whereas our review found that studies that applied non-LR algorithms alone or combined with LR had sample sizes ranging from only 46 [158] to 8836 [148] instances. Meanwhile, non-LR machine learning algorithms require larger sample sizes relative to the number of candidate predictors [199].

A meta-analysis of multivariable LR was also previously conducted for premature birth from 4 studies [200]. In a previous systematic review, the 2 highest AUROCs were 0.67 (95% CI 0.62-0.72; low ROB) and 0.64 (95% CI 0.60-0.68; high ROB). Non-LR models of premature birth in our systematic review showed AUROCs of 0.75 (95% CI 0.67-0.82) [121] and 0.911 (95% CI 0.862-0.96) [130], but these models were developed from high ROB studies. The other models only reported point estimates of the AUROC, which were a minimum of 0.6 by a decision tree [156] and a maximum of 0.991 by a support vector machine [143].

Minimizing the bias of model performance is the first thing to consider when developing a clinical prediction model. Several concerns need to be addressed when developing prognostic machine learning predictions of pregnancy care. In our review, most studies had problems of insufficient EPV (either LR and non-LR studies), single imputation (mostly LR studies), and no assessment of calibration (mostly non-LR studies). This may expose the studies to high ROBs [21]. The overestimation of the predictive performance is larger, with fewer participants with events relative to the number of predictor candidates, as described in the PROBAST guidelines. Most ROBs in our review were contributed by the domain of analysis, and answers to which the EPV signaling question mostly led studies to high ROB assessment results. Insufficient EPV mean that the study developed a model using a data set with a sample size that was less than the minimum requirement for events relative to the number of predictors. LR only requires 20 EPV, whereas non-LR algorithms require 50 to 200 EPV. Meanwhile, single imputation means that missing values are imputed by any random value, mean, median, mode, or one-time regression. Multiple imputations are more recommended than single imputations, in which the preferred method is multiple equations by chained equations. For the assessment of calibration, a study should show the incidence of events (true probability) for each subset of samples that belongs to the same range of predicted probability by the model. We recommend these based on PROBAST guidelines and other guidelines for machine learning prognostic predictions in pregnancy care [15,21].

Strengths and Limitations

Our systematic review and meta-analysis will allow investigators or clinicians in pregnancy care to consider whether trying multiple machine learning models provides benefit to their studies. If more prediction models are needed for the outcomes with more specific problems or subpopulations, then predictive modeling may consider comparisons of LR and non-LR algorithms for specific outcomes that were compared in our meta-analysis. We also reported heterogeneity measures to interpret the predictive performances of algorithms across studies.

However, the diverse populations and hyperparameters caused substantial heterogeneity of predictive performance in our meta-analysis. Future meta-analyses will be needed if more machine learning models are developed for the same outcome using the same algorithm. However, we tried to minimize the heterogeneity by excluding several studies to ensure more homogenous outcome definitions and normally distributed AUROCs. We also applied random effects modeling as recommended [172].

Conclusions

Prediction models using non-LR machine learning algorithms significantly outperformed those using LR for several pregnancy outcomes. These non-LR algorithms were random forests for predicting preterm delivery and pre-eclampsia and gradient boosting for predicting CS and gestational diabetes. In our review, studies that developed models using these algorithms had low ROBs. For predicting ongoing pregnancy in *in vitro* fertilization, non-LR algorithms did not significantly outperform

LR. Prediction models using non-LR algorithms for vaginal birth after a CS significantly underperformed LR, but the study with the non-LR algorithm had a high ROB.

On the basis of our meta-analysis, we recommend comparing multiple machine learning models, which include both LR and non-LR algorithms, to develop a prediction model. In our

systematic review, we also found that many studies had high ROB in the domain of analysis. In this domain, many studies lacked EPV to develop a prediction model. Hence, we also recommend the future development of a prediction model to pursue standard EPV and other standards based on guidelines to minimize ROB.

Acknowledgments

This study was funded by the Ministry of Science and Technology of Taiwan under grant number MOST108-2221-E-038-018 and MOST109-2221-E-038-018 to ES. The sponsor had no role in the research design or contents of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details on forest plots, search filter, eligibility criteria, study selection, list of reviewed studies, risk of bias assessment, signaling questions and the answers, predictive performance and sample size, R code for meta-analysis, and records of studies.

[[DOCX File, 2653 KB-Multimedia Appendix 1](#)]

References

1. Domínguez-Almendros S, Benítez-Parejo N, Gonzalez-Ramirez A. Logistic regression models. *Allergol Immunopathol (Madr)* 2011;39(5):295-305. [doi: [10.1016/j.aller.2011.05.002](https://doi.org/10.1016/j.aller.2011.05.002)] [Medline: [21820234](#)]
2. Deo RC. Machine learning in medicine. *Circulation* 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593)] [Medline: [26572668](#)]
3. Higgins JP. Nonlinear systems in medicine. *Yale J Biol Med* 2002;75(5-6):247-260 [FREE Full text] [Medline: [14580107](#)]
4. The Millennium Development Goals Report. United Nations. 2015. URL: [https://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20\(July%201\).pdf](https://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf) [accessed 2019-07-25]
5. Say L, Chou D, Gemmill A, Tunçalp O, Moller A, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Health* 2014 Jun;2(6):e323-e333 [FREE Full text] [doi: [10.1016/S2214-109X\(14\)70227-X](https://doi.org/10.1016/S2214-109X(14)70227-X)] [Medline: [25103301](#)]
6. Lehtonen L, Gimeno A, Parra-Llorca A, Vento M. Early neonatal death: a challenge worldwide. *Semin Fetal Neonatal Med* 2017 Jun;22(3):153-160. [doi: [10.1016/j.siny.2017.02.006](https://doi.org/10.1016/j.siny.2017.02.006)] [Medline: [28238633](#)]
7. Burlinson CE, Sirounis D, Walley KR, Chau A. Sepsis in pregnancy and the puerperium. *Int J Obstet Anesth* 2018 Nov;36:96-107. [doi: [10.1016/j.ijoa.2018.04.010](https://doi.org/10.1016/j.ijoa.2018.04.010)] [Medline: [29921485](#)]
8. Edwards HM. Aetiology and treatment of severe postpartum haemorrhage. *Dan Med J* 2018 Mar;65(3):- [FREE Full text] [Medline: [29510809](#)]
9. Nair TM. Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia. *Comput Biol Chem* 2018 Aug;75:222-230. [doi: [10.1016/j.compbiolchem.2018.05.011](https://doi.org/10.1016/j.compbiolchem.2018.05.011)] [Medline: [29859381](#)]
10. Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. *Science* 2014 Aug 15;345(6198):760-765 [FREE Full text] [doi: [10.1126/science.1251816](https://doi.org/10.1126/science.1251816)] [Medline: [25124429](#)]
11. Nindrea RD, Aryandono T, Lazuardi L, Dwiprahasto I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. *Asian Pac J Cancer Prev* 2018 Jul 27;19(7):1747-1752 [FREE Full text] [doi: [10.22034/APJCP.2018.19.7.1747](https://doi.org/10.22034/APJCP.2018.19.7.1747)] [Medline: [30049182](#)]
12. Lee Y, Raguett R, Mansur RB, Boutillier JJ, Rosenblat JD, Trevizol A, et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord* 2018 Dec 1;241:519-532. [doi: [10.1016/j.jad.2018.08.073](https://doi.org/10.1016/j.jad.2018.08.073)] [Medline: [30153635](#)]
13. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019 Jun;110:12-22. [doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004)] [Medline: [30763612](#)]
14. Gregório T, Pipa S, Cavaleiro P, Atanásio G, Albuquerque I, Chaves PC, et al. Prognostic models for intracerebral hemorrhage: systematic review and meta-analysis. *BMC Med Res Methodol* 2018 Nov 20;18(1):145 [FREE Full text] [doi: [10.1186/s12874-018-0613-8](https://doi.org/10.1186/s12874-018-0613-8)] [Medline: [30458727](#)]
15. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016 Dec 16;18(12):e323 [FREE Full text] [doi: [10.2196/jmir.5870](https://doi.org/10.2196/jmir.5870)] [Medline: [27986644](#)]

16. Nguyen AV, Blears EE, Ross E, Lall RR, Ortega-Barnett J. Machine learning applications for the differentiation of primary central nervous system lymphoma from glioblastoma on imaging: a systematic review and meta-analysis. *Neurosurg Focus* 2018 Nov 1;45(5):E5. [doi: [10.3171/2018.8.FOCUS18325](https://doi.org/10.3171/2018.8.FOCUS18325)] [Medline: [30453459](https://pubmed.ncbi.nlm.nih.gov/30453459/)]
17. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019 Jan 1;170(1):51-58 [FREE Full text] [doi: [10.7326/M18-1376](https://doi.org/10.7326/M18-1376)] [Medline: [30596875](https://pubmed.ncbi.nlm.nih.gov/30596875/)]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
19. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014 Oct;11(10):e1001744 [FREE Full text] [doi: [10.1371/journal.pmed.1001744](https://doi.org/10.1371/journal.pmed.1001744)] [Medline: [25314315](https://pubmed.ncbi.nlm.nih.gov/25314315/)]
20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015 Jan 6;162(1):55-63. [doi: [10.7326/M14-0697](https://doi.org/10.7326/M14-0697)] [Medline: [25560714](https://pubmed.ncbi.nlm.nih.gov/25560714/)]
21. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019 Jan 1;170(1):W1-33 [FREE Full text] [doi: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377)] [Medline: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)]
22. Mitchell TM. *Machine Learning*. New York, NY: McGraw-Hill Inc; 1997.
23. James SB, Bardenet R, Bengio Y, Balázs K. Algorithms for Hyper-Parameter Optimization. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*. 2011 Presented at: NIPS'11; December 21-24, 2011; Granada, Spain URL: <http://papers.nips.cc/paper/4443-algorithms-for-hyper-parameter-optimization.pdf> [doi: [10.5555/2986459.2986743](https://doi.org/10.5555/2986459.2986743)]
24. Snell KI, Hua H, Debray TP, Ensor J, Look MP, Moons KG, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* 2016 Jan;69:40-50 [FREE Full text] [doi: [10.1016/j.jclinepi.2015.05.009](https://doi.org/10.1016/j.jclinepi.2015.05.009)] [Medline: [26142114](https://pubmed.ncbi.nlm.nih.gov/26142114/)]
25. Cheung M, Ho R, Lim Y, Mak A. Conducting a meta-analysis: basics and good practices. *Int J Rheum Dis* 2012 Apr;15(2):129-135. [doi: [10.1111/j.1756-185X.2012.01712.x](https://doi.org/10.1111/j.1756-185X.2012.01712.x)] [Medline: [22462415](https://pubmed.ncbi.nlm.nih.gov/22462415/)]
26. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *Br Med J* 2011 Feb 10;342:d549. [doi: [10.1136/bmj.d549](https://doi.org/10.1136/bmj.d549)] [Medline: [21310794](https://pubmed.ncbi.nlm.nih.gov/21310794/)]
27. Viechtbauer W. Conducting meta-analyses in with the package. *J Stat Soft* 2010;36(3):- [doi: [10.18637/jss.v036.i03](https://doi.org/10.18637/jss.v036.i03)]
28. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016 Mar;7(1):55-79 [FREE Full text] [doi: [10.1002/jrsm.1164](https://doi.org/10.1002/jrsm.1164)] [Medline: [26332144](https://pubmed.ncbi.nlm.nih.gov/26332144/)]
29. Allouche M, Huissoud C, Guyard-Boileau B, Rouzier R, Parant O. Development and validation of nomograms for predicting preterm delivery. *Am J Obstet Gynecol* 2011 Mar;204(3):242.e1-242.e8. [doi: [10.1016/j.ajog.2010.09.030](https://doi.org/10.1016/j.ajog.2010.09.030)] [Medline: [21093847](https://pubmed.ncbi.nlm.nih.gov/21093847/)]
30. Almeida S, Katz L, Coutinho I, Amorim M. Validation of fullPIERS model for prediction of adverse outcomes among women with severe pre-eclampsia. *Int J Gynaecol Obstet* 2017 Aug;138(2):142-147. [doi: [10.1002/ijgo.12197](https://doi.org/10.1002/ijgo.12197)] [Medline: [28475234](https://pubmed.ncbi.nlm.nih.gov/28475234/)]
31. Al-Rubaie ZT, Hudson HM, Jenkins G, Mahmoud I, Ray JG, Askie LM, et al. Prediction of pre-eclampsia in nulliparous women using routinely collected maternal characteristics: a model development and validation study. *BMC Pregnancy Childbirth* 2020 Jan 6;20(1):23 [FREE Full text] [doi: [10.1186/s12884-019-2712-x](https://doi.org/10.1186/s12884-019-2712-x)] [Medline: [31906891](https://pubmed.ncbi.nlm.nih.gov/31906891/)]
32. Bastek JA, Sammel MD, Srinivas SK, McShea MA, Foreman MN, Elovitz MA, et al. Clinical prediction rules for preterm birth in patients presenting with preterm labor. *Obstet Gynecol* 2012 Jun;119(6):1119-1128. [doi: [10.1097/AOG.0b013e31825503e5](https://doi.org/10.1097/AOG.0b013e31825503e5)] [Medline: [22617575](https://pubmed.ncbi.nlm.nih.gov/22617575/)]
33. Benhalima K, van Crombrugge P, Moyson C, Verhaeghe J, Vandeginste S, Verlaenen H, et al. Estimating the risk of gestational diabetes mellitus based on the 2013 WHO criteria: a prediction model based on clinical and biochemical variables in early pregnancy. *Acta Diabetol* 2020 Jun;57(6):661-671. [doi: [10.1007/s00592-019-01469-5](https://doi.org/10.1007/s00592-019-01469-5)] [Medline: [31915927](https://pubmed.ncbi.nlm.nih.gov/31915927/)]
34. Berntorp K, Anderberg E, Claesson R, Ignell C, Källén K. The relative importance of maternal body mass index and glucose levels for prediction of large-for-gestational-age births. *BMC Pregnancy Childbirth* 2015 Oct 29;15:280 [FREE Full text] [doi: [10.1186/s12884-015-0722-x](https://doi.org/10.1186/s12884-015-0722-x)] [Medline: [26514116](https://pubmed.ncbi.nlm.nih.gov/26514116/)]
35. Broekmans FJ, Verweij PJ, Eijkemans MJ, Mannaerts BM, Witjes H. Prognostic models for high and low ovarian responses in controlled ovarian stimulation using a GnRH antagonist protocol. *Hum Reprod* 2014 Aug;29(8):1688-1697 [FREE Full text] [doi: [10.1093/humrep/deu090](https://doi.org/10.1093/humrep/deu090)] [Medline: [24903202](https://pubmed.ncbi.nlm.nih.gov/24903202/)]
36. Fagerberg MC, Källén K. Third-trimester prediction of successful vaginal birth after one cesarean delivery-a Swedish model. *Acta Obstet Gynecol Scand* 2020 May;99(5):660-668. [doi: [10.1111/aogs.13783](https://doi.org/10.1111/aogs.13783)] [Medline: [31788783](https://pubmed.ncbi.nlm.nih.gov/31788783/)]

37. Casikar I, Lu C, Reid S, Condous G. Prediction of successful expectant management of first trimester miscarriage: development and validation of a new mathematical model. *Aust N Z J Obstet Gynaecol* 2013 Feb;53(1):58-63. [doi: [10.1111/ajo.12053](https://doi.org/10.1111/ajo.12053)] [Medline: [23405997](https://pubmed.ncbi.nlm.nih.gov/23405997/)]
38. Cerqueira FR, Ferreira TG, de Paiva Oliveira A, Augusto DA, Krempser E, Corrêa Barbosa HJ, et al. NICeSim: an open-source simulator based on machine learning techniques to support medical research on prenatal and perinatal care decision making. *Artif Intell Med* 2014 Nov;62(3):193-201. [doi: [10.1016/j.artmed.2014.10.001](https://doi.org/10.1016/j.artmed.2014.10.001)] [Medline: [25457563](https://pubmed.ncbi.nlm.nih.gov/25457563/)]
39. Chandrasekaran S, Bastek JA, Turitz AL, Durnwald CP. A prediction score to assess the risk of delivering a large for gestational age infant among obese women. *J Matern Fetal Neonatal Med* 2016;29(1):22-26. [doi: [10.3109/14767058.2014.991709](https://doi.org/10.3109/14767058.2014.991709)] [Medline: [25428834](https://pubmed.ncbi.nlm.nih.gov/25428834/)]
40. Chen L, Luo D, Yu X, Jin M, Cai W. Predicting stress urinary incontinence during pregnancy: combination of pelvic floor ultrasound parameters and clinical factors. *Acta Obstet Gynecol Scand* 2018 Aug;97(8):966-975 [FREE Full text] [doi: [10.1111/aogs.13368](https://doi.org/10.1111/aogs.13368)] [Medline: [29754393](https://pubmed.ncbi.nlm.nih.gov/29754393/)]
41. Ciobanu A, Rouvali A, Syngelaki A, Akolekar R, Nicolaides KH. Prediction of small for gestational age neonates: screening by maternal factors, fetal biometry, and biomarkers at 35-37 weeks' gestation. *Am J Obstet Gynecol* 2019 May;220(5):486.e1-486.11. [doi: [10.1016/j.ajog.2019.01.227](https://doi.org/10.1016/j.ajog.2019.01.227)] [Medline: [30707967](https://pubmed.ncbi.nlm.nih.gov/30707967/)]
42. Cortet M, Maucourt-Boulch D, Deneux-Tharoux C, Dupont C, Rudigoz R, Roy P, et al. Severity of post-partum hemorrhage after vaginal delivery is not predictable from clinical variables available at the time post-partum hemorrhage is diagnosed. *J Obstet Gynaecol Res* 2015 Feb;41(2):199-206. [doi: [10.1111/jog.12528](https://doi.org/10.1111/jog.12528)] [Medline: [25303234](https://pubmed.ncbi.nlm.nih.gov/25303234/)]
43. Crovetto F, Figueras F, Triunfo S, Crispi F, Rodriguez-Sureda V, Dominguez C, et al. First trimester screening for early and late preeclampsia based on maternal characteristics, biophysical parameters, and angiogenic factors. *Prenat Diagn* 2015 Feb;35(2):183-191. [doi: [10.1002/pd.4519](https://doi.org/10.1002/pd.4519)] [Medline: [25346181](https://pubmed.ncbi.nlm.nih.gov/25346181/)]
44. de Oliveira RV, Martins MD, Rios LT, Araujo Júnior E, Simões VM, Nardoza LM, et al. Predictive model for spontaneous preterm labor among pregnant women with contractions and intact amniotic membranes. *Arch Gynecol Obstet* 2012 Oct;286(4):893-900. [doi: [10.1007/s00404-012-2397-0](https://doi.org/10.1007/s00404-012-2397-0)] [Medline: [22674420](https://pubmed.ncbi.nlm.nih.gov/22674420/)]
45. de Wilde MA, Veltman-Verhulst SM, Goverde AJ, Lambalk CB, Laven JS, Franx A, et al. Preconception predictors of gestational diabetes: a multicentre prospective cohort study on the predominant complication of pregnancy in polycystic ovary syndrome. *Hum Reprod* 2014 Jun;29(6):1327-1336. [doi: [10.1093/humrep/deu077](https://doi.org/10.1093/humrep/deu077)] [Medline: [24777850](https://pubmed.ncbi.nlm.nih.gov/24777850/)]
46. Eggebø TM, Wilhelm-Benartzi C, Hassan WA, Usman S, Salvesen KA, Lees CC. A model to predict vaginal delivery in nulliparous women based on maternal characteristics and intrapartum ultrasound. *Am J Obstet Gynecol* 2015 Sep;213(3):362.e1-362.e6. [doi: [10.1016/j.ajog.2015.05.044](https://doi.org/10.1016/j.ajog.2015.05.044)] [Medline: [26008180](https://pubmed.ncbi.nlm.nih.gov/26008180/)]
47. Fagerberg MC, Maršál K, Källén K. Predicting the chance of vaginal delivery after one cesarean section: validation and elaboration of a published prediction model. *Eur J Obstet Gynecol Reprod Biol* 2015 May;188:88-94. [doi: [10.1016/j.ejogrb.2015.02.031](https://doi.org/10.1016/j.ejogrb.2015.02.031)] [Medline: [25801723](https://pubmed.ncbi.nlm.nih.gov/25801723/)]
48. Guo Z, Yang F, Zhang J, Zhang Z, Li K, Tian Q, et al. Whole-genome promoter profiling of plasma DNA exhibits diagnostic value for placenta-origin pregnancy complications. *Adv Sci (Weinh)* 2020 Apr;7(7):1901819 [FREE Full text] [doi: [10.1002/advs.201901819](https://doi.org/10.1002/advs.201901819)] [Medline: [32274292](https://pubmed.ncbi.nlm.nih.gov/32274292/)]
49. Harper LM, Glover AV, Biggio JR, Tita A. Predicting failure of glyburide therapy in gestational diabetes. *J Perinatol* 2016 May;36(5):347-351 [FREE Full text] [doi: [10.1038/jp.2015.216](https://doi.org/10.1038/jp.2015.216)] [Medline: [26796130](https://pubmed.ncbi.nlm.nih.gov/26796130/)]
50. Isono W, Nagamatsu T, Uemura Y, Fujii T, Hyodo H, Yamashita T, et al. Prediction model for the incidence of emergent cesarean section during induction of labor specialized in nulliparous low-risk women. *J Obstet Gynaecol Res* 2011 Dec;37(12):1784-1791. [doi: [10.1111/j.1447-0756.2011.01607.x](https://doi.org/10.1111/j.1447-0756.2011.01607.x)] [Medline: [21793999](https://pubmed.ncbi.nlm.nih.gov/21793999/)]
51. Kang J, Kim HS, Lee EB, Uh Y, Han K, Park EY, et al. Prediction model for massive transfusion in placenta previa during cesarean section. *Yonsei Med J* 2020 Feb;61(2):154-160 [FREE Full text] [doi: [10.3349/ymj.2020.61.2.154](https://doi.org/10.3349/ymj.2020.61.2.154)] [Medline: [31997624](https://pubmed.ncbi.nlm.nih.gov/31997624/)]
52. Kawakita T, Mokhtari N, Huang JC, Landy HJ. Evaluation of risk-assessment tools for severe postpartum hemorrhage in women undergoing cesarean delivery. *Obstet Gynecol* 2019 Dec;134(6):1308-1316. [doi: [10.1097/AOG.0000000000003574](https://doi.org/10.1097/AOG.0000000000003574)] [Medline: [31764744](https://pubmed.ncbi.nlm.nih.gov/31764744/)]
53. Khan N, Ciobanu A, Karampitsakos T, Akolekar R, Nicolaides KH. Prediction of large-for-gestational-age neonate by routine third-trimester ultrasound. *Ultrasound Obstet Gynecol* 2019 Sep;54(3):326-333. [doi: [10.1002/uog.20377](https://doi.org/10.1002/uog.20377)] [Medline: [31236963](https://pubmed.ncbi.nlm.nih.gov/31236963/)]
54. Kok M, van der Steeg J, van der Post J, Mol B. Prediction of success of external cephalic version after 36 weeks. *Am J Perinatol* 2011 Feb;28(2):103-110. [doi: [10.1055/s-0030-1262909](https://doi.org/10.1055/s-0030-1262909)] [Medline: [20661845](https://pubmed.ncbi.nlm.nih.gov/20661845/)]
55. Lafalla O, Esteban LM, Lou AC, Cornudella R, Domínguez M, Sanz G, et al. Clinical utility of thrombophilia, anticoagulant treatment, and maternal variables as predictors of placenta-mediated pregnancy complications: an extensive analysis. *J Matern Fetal Neonatal Med* 2019 May 9:1-11. [doi: [10.1080/14767058.2019.1611764](https://doi.org/10.1080/14767058.2019.1611764)] [Medline: [31018724](https://pubmed.ncbi.nlm.nih.gov/31018724/)]
56. Lee JS, Sultana R, Han NL, Sia AT, Sng BL. Development and validation of a predictive risk factor model for epidural re-siting in women undergoing labour epidural analgesia: a retrospective cohort study. *BMC Anesthesiol* 2018 Nov 29;18(1):176 [FREE Full text] [doi: [10.1186/s12871-018-0638-x](https://doi.org/10.1186/s12871-018-0638-x)] [Medline: [30497401](https://pubmed.ncbi.nlm.nih.gov/30497401/)]

57. Mardy AH, Ananth CV, Grobman WA, Gyamfi-Bannerman C. A prediction model of vaginal birth after cesarean in the preterm period. *Am J Obstet Gynecol* 2016 Oct;215(4):513.e1-513.e7. [doi: [10.1016/j.ajog.2016.05.039](https://doi.org/10.1016/j.ajog.2016.05.039)] [Medline: [27262971](https://pubmed.ncbi.nlm.nih.gov/27262971/)]
58. McCowan LM, Thompson JM, Taylor RS, Baker PN, North RA, Poston L, SCOPE consortium. Prediction of small for gestational age infants in healthy nulliparous women using clinical and ultrasound risk factors combined with early pregnancy biomarkers. *PLoS One* 2017;12(1):e0169311 [FREE Full text] [doi: [10.1371/journal.pone.0169311](https://doi.org/10.1371/journal.pone.0169311)] [Medline: [28068394](https://pubmed.ncbi.nlm.nih.gov/28068394/)]
59. McCowan LM, Thompson JM, Taylor RS, North RA, Poston L, Baker PN, SCOPE Consortium. Clinical prediction in early pregnancy of infants small for gestational age by customised birthweight centiles: findings from a healthy nulliparous cohort. *PLoS One* 2013;8(8):e70917 [FREE Full text] [doi: [10.1371/journal.pone.0070917](https://doi.org/10.1371/journal.pone.0070917)] [Medline: [23940665](https://pubmed.ncbi.nlm.nih.gov/23940665/)]
60. Mehta-Lee SS, Palma A, Bernstein PS, Lounsbury D, Schlecht NF. A preconception nomogram to predict preterm delivery. *Matern Child Health J* 2017 Jan;21(1):118-127. [doi: [10.1007/s10995-016-2100-3](https://doi.org/10.1007/s10995-016-2100-3)] [Medline: [27461021](https://pubmed.ncbi.nlm.nih.gov/27461021/)]
61. Meijerink A, Cissen M, Mochtar M, Fleischer K, Thoonen I, de Melker A, et al. Prediction model for live birth in ICSI using testicular extracted sperm. *Hum Reprod* 2016 Sep;31(9):1942-1951. [doi: [10.1093/humrep/dew146](https://doi.org/10.1093/humrep/dew146)] [Medline: [27406949](https://pubmed.ncbi.nlm.nih.gov/27406949/)]
62. Meister M, Cahill A, Conner S, Woolfolk C, Lowder J. Predicting obstetric anal sphincter injuries in a modern obstetric population. *Am J Obstet Gynecol* 2016 Sep;215(3):310.e1-310.e7. [doi: [10.1016/j.ajog.2016.02.041](https://doi.org/10.1016/j.ajog.2016.02.041)] [Medline: [26902989](https://pubmed.ncbi.nlm.nih.gov/26902989/)]
63. Menon R, Bhat G, Saade GR, Spratt H. Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. *Acta Obstet Gynecol Scand* 2014 Apr;93(4):382-391. [doi: [10.1111/aogs.12344](https://doi.org/10.1111/aogs.12344)] [Medline: [24461165](https://pubmed.ncbi.nlm.nih.gov/24461165/)]
64. Metz TD, Stoddard GJ, Henry E, Jackson M, Holmgren C, Esplin S. Simple, validated vaginal birth after cesarean delivery prediction model for use at the time of admission. *Obstet Gynecol* 2013 Sep;122(3):571-578 [FREE Full text] [doi: [10.1097/AOG.0b013e31829f8ced](https://doi.org/10.1097/AOG.0b013e31829f8ced)] [Medline: [23921867](https://pubmed.ncbi.nlm.nih.gov/23921867/)]
65. Murtoniemi K, Villa PM, Matomäki J, Keikkala E, Vuorela P, Hämäläinen E, et al. Prediction of pre-eclampsia and its subtypes in high-risk cohort: hyperglycosylated human chorionic gonadotropin in multivariate models. *BMC Pregnancy Childbirth* 2018 Jul 3;18(1):279 [FREE Full text] [doi: [10.1186/s12884-018-1908-9](https://doi.org/10.1186/s12884-018-1908-9)] [Medline: [29970026](https://pubmed.ncbi.nlm.nih.gov/29970026/)]
66. Myers J, Kenny L, McCowan L, Chan E, Dekker G, Poston L, SCOPE consortium. Angiogenic factors combined with clinical risk factors to predict preterm pre-eclampsia in nulliparous women: a predictive test accuracy study. *BJOG* 2013 Sep;120(10):1215-1223. [doi: [10.1111/1471-0528.12195](https://doi.org/10.1111/1471-0528.12195)] [Medline: [23906160](https://pubmed.ncbi.nlm.nih.gov/23906160/)]
67. Oates J, Casikar I, Campaign A, Müller S, Yang J, Reid S, et al. A prediction model for viability at the end of the first trimester after a single early pregnancy evaluation. *Aust N Z J Obstet Gynaecol* 2013 Feb;53(1):51-57. [doi: [10.1111/ajo.12046](https://doi.org/10.1111/ajo.12046)] [Medline: [23405996](https://pubmed.ncbi.nlm.nih.gov/23405996/)]
68. Payne BA, Groen H, Ukah UV, Ansermino JM, Bhutta Z, Grobman W, miniPIERS working group. Development and internal validation of a multivariable model to predict perinatal death in pregnancy hypertension. *Pregnancy Hypertens* 2015 Oct;5(4):315-321 [FREE Full text] [doi: [10.1016/j.preghy.2015.08.006](https://doi.org/10.1016/j.preghy.2015.08.006)] [Medline: [26597747](https://pubmed.ncbi.nlm.nih.gov/26597747/)]
69. Pettersson G, Andersen AN, Broberg P, Arce J. Pre-stimulation parameters predicting live birth after IVF in the long GnRH agonist protocol. *Reprod Biomed Online* 2010 May;20(5):572-581. [doi: [10.1016/j.rbmo.2010.02.014](https://doi.org/10.1016/j.rbmo.2010.02.014)] [Medline: [20236862](https://pubmed.ncbi.nlm.nih.gov/20236862/)]
70. Pettersson K, Yousaf K, Ranstam J, Westgren M, Ajne G. Predictive value of traction force measurement in vacuum extraction: development of a multivariate prognostic model. *PLoS One* 2017;12(3):e0171938 [FREE Full text] [doi: [10.1371/journal.pone.0171938](https://doi.org/10.1371/journal.pone.0171938)] [Medline: [28257459](https://pubmed.ncbi.nlm.nih.gov/28257459/)]
71. Ramanah R, Omar S, Guillien A, Pugin A, Martin A, Riethmuller D, et al. Predicting umbilical artery pH during labour: development and validation of a nomogram using fetal heart rate patterns. *Eur J Obstet Gynecol Reprod Biol* 2018 Jun;225:166-171. [doi: [10.1016/j.ejogrb.2018.04.008](https://doi.org/10.1016/j.ejogrb.2018.04.008)] [Medline: [29727787](https://pubmed.ncbi.nlm.nih.gov/29727787/)]
72. Reid S, Lu C, Condous G. Can we improve the prediction of pouch of Douglas obliteration in women with suspected endometriosis using ultrasound-based models? A multicenter prospective observational study. *Acta Obstet Gynecol Scand* 2015 Dec;94(12):1297-1306. [doi: [10.1111/aogs.12779](https://doi.org/10.1111/aogs.12779)] [Medline: [26399692](https://pubmed.ncbi.nlm.nih.gov/26399692/)]
73. Rinaudo P, Shen S, Hua J, Qian S, Prabhu U, Garcia E, et al. (1)H NMR based profiling of spent culture media cannot predict success of implantation for day 3 human embryos. *J Assist Reprod Genet* 2012 Dec;29(12):1435-1442 [FREE Full text] [doi: [10.1007/s10815-012-9877-9](https://doi.org/10.1007/s10815-012-9877-9)] [Medline: [23090745](https://pubmed.ncbi.nlm.nih.gov/23090745/)]
74. Ryu A, Cho NJ, Kim YS, Lee EY. Predictive value of serum uric acid levels for adverse perinatal outcomes in preeclampsia. *Medicine (Baltimore)* 2019 May;98(18):e15462. [doi: [10.1097/MD.00000000000015462](https://doi.org/10.1097/MD.00000000000015462)] [Medline: [31045822](https://pubmed.ncbi.nlm.nih.gov/31045822/)]
75. Sananes N, Meyer N, Gaudineau A, Aissi G, Boudier E, Fritz G, et al. Prediction of spontaneous preterm delivery in the first trimester of pregnancy. *Eur J Obstet Gynecol Reprod Biol* 2013 Nov;171(1):18-22. [doi: [10.1016/j.ejogrb.2013.07.042](https://doi.org/10.1016/j.ejogrb.2013.07.042)] [Medline: [24012451](https://pubmed.ncbi.nlm.nih.gov/24012451/)]
76. Sandström A, Snowden JM, Höijer J, Bottai M, Wikström AK. Clinical risk assessment in early pregnancy for preeclampsia in nulliparous women: a population based cohort study. *PLoS One* 2019;14(11):e0225716 [FREE Full text] [doi: [10.1371/journal.pone.0225716](https://doi.org/10.1371/journal.pone.0225716)] [Medline: [31774875](https://pubmed.ncbi.nlm.nih.gov/31774875/)]
77. Scheinhardt M, Lerman T, König IR, Griesinger G. Performance of prognostic modelling of high and low ovarian response to ovarian stimulation for IVF. *Hum Reprod* 2018 Aug 01;33(8):1499-1505. [doi: [10.1093/humrep/dey236](https://doi.org/10.1093/humrep/dey236)] [Medline: [30007353](https://pubmed.ncbi.nlm.nih.gov/30007353/)]

78. Shi W, Zhang S, Zhao W, Xia X, Wang M, Wang H, et al. Factors related to clinical pregnancy after vitrified-warmed embryo transfer: a retrospective and multivariate logistic regression analysis of 2313 transfer cycles. *Hum Reprod* 2013 Jul;28(7):1768-1775. [doi: [10.1093/humrep/det094](https://doi.org/10.1093/humrep/det094)] [Medline: [23599130](https://pubmed.ncbi.nlm.nih.gov/23599130/)]
79. Sovio U, Smith GC. Blinded ultrasound fetal biometry at 36 weeks and risk of emergency Cesarean delivery in a prospective cohort study of low-risk nulliparous women. *Ultrasound Obstet Gynecol* 2018 Jul;52(1):78-86. [doi: [10.1002/uog.17513](https://doi.org/10.1002/uog.17513)] [Medline: [28452133](https://pubmed.ncbi.nlm.nih.gov/28452133/)]
80. Stamatopoulos N, Lu C, Casikar I, Reid S, Mongelli M, Hardy N, et al. Prediction of subsequent miscarriage risk in women who present with a viable pregnancy at the first early pregnancy scan. *Aust N Z J Obstet Gynaecol* 2015 Oct;55(5):464-472. [doi: [10.1111/ajo.12395](https://doi.org/10.1111/ajo.12395)] [Medline: [26294017](https://pubmed.ncbi.nlm.nih.gov/26294017/)]
81. Stott D, Bolten M, Salman M, Paraschiv D, Douiri A, Kametas NA. A prediction model for the response to oral labetalol for the treatment of antenatal hypertension. *J Hum Hypertens* 2017 Feb;31(2):126-131. [doi: [10.1038/jhh.2016.50](https://doi.org/10.1038/jhh.2016.50)] [Medline: [27465979](https://pubmed.ncbi.nlm.nih.gov/27465979/)]
82. Stroux L, Redman CW, Georgieva A, Payne SJ, Clifford GD. Doppler-based fetal heart rate analysis markers for the detection of early intrauterine growth restriction. *Acta Obstet Gynecol Scand* 2017 Nov;96(11):1322-1329. [doi: [10.1111/aogs.13228](https://doi.org/10.1111/aogs.13228)] [Medline: [28862738](https://pubmed.ncbi.nlm.nih.gov/28862738/)]
83. Tessmer-Tuck JA, El-Nashar SA, Racek AR, Lohse CM, Famuyide AO, Wick MJ. Predicting vaginal birth after cesarean section: a cohort study. *Gynecol Obstet Invest* 2014;77(2):121-126 [FREE Full text] [doi: [10.1159/000357757](https://doi.org/10.1159/000357757)] [Medline: [24525697](https://pubmed.ncbi.nlm.nih.gov/24525697/)]
84. Thériault S, Giguère Y, Massé J, Girouard J, Forest J. Early prediction of gestational diabetes: a practical model combining clinical and biochemical markers. *Clin Chem Lab Med* 2016 Mar;54(3):509-518. [doi: [10.1515/ccim-2015-0537](https://doi.org/10.1515/ccim-2015-0537)] [Medline: [26351946](https://pubmed.ncbi.nlm.nih.gov/26351946/)]
85. Timmerman E, Oude Rengerink K, Pajkrt E, Opmeer BC, van der Post JA, Bilardo CM. Ductus venosus pulsatility index measurement reduces the false-positive rate in first-trimester screening. *Ultrasound Obstet Gynecol* 2010 Dec;36(6):661-667. [doi: [10.1002/uog.7706](https://doi.org/10.1002/uog.7706)] [Medline: [20521242](https://pubmed.ncbi.nlm.nih.gov/20521242/)]
86. Tsur A, Batsry L, Toussia-Cohen S, Rosenstein MG, Barak O, Brezinov Y, et al. Development and validation of a machine-learning model for prediction of shoulder dystocia. *Ultrasound Obstet Gynecol* 2020 Oct;56(4):588-596. [doi: [10.1002/uog.21878](https://doi.org/10.1002/uog.21878)] [Medline: [31587401](https://pubmed.ncbi.nlm.nih.gov/31587401/)]
87. van Baaren GJ, Bruijn MM, Vis JY, Wilms FF, Oudijk MA, Kwee A, et al. Risk factors for preterm delivery: do they add to fetal fibronectin testing and cervical length measurement in the prediction of preterm delivery in symptomatic women? *Eur J Obstet Gynecol Reprod Biol* 2015 Sep;192:79-85. [doi: [10.1016/j.ejogrb.2015.05.004](https://doi.org/10.1016/j.ejogrb.2015.05.004)] [Medline: [26182836](https://pubmed.ncbi.nlm.nih.gov/26182836/)]
88. Van Calster B, Condous G, Kirk E, Bourne T, Timmerman D, Van Huffel S. An application of methods for the probabilistic three-class classification of pregnancies of unknown location. *Artif Intell Med* 2009 Jun;46(2):139-154. [doi: [10.1016/j.artmed.2008.12.003](https://doi.org/10.1016/j.artmed.2008.12.003)] [Medline: [19157812](https://pubmed.ncbi.nlm.nih.gov/19157812/)]
89. van der Ham DP, van Kuijk S, Opmeer BC, Willekes C, van Beek JJ, Mulder AL, PPROMEXIL trial group. Can neonatal sepsis be predicted in late preterm premature rupture of membranes? Development of a prediction model. *Eur J Obstet Gynecol Reprod Biol* 2014 May;176:90-95. [doi: [10.1016/j.ejogrb.2014.02.003](https://doi.org/10.1016/j.ejogrb.2014.02.003)] [Medline: [24630296](https://pubmed.ncbi.nlm.nih.gov/24630296/)]
90. van der Tuuk K, van Pampus MG, Koopmans C, Aarnoudse J, van den Berg PP, van Beek JJ, HYPITAT study group. Prediction of cesarean section risk in women with gestational hypertension or mild preeclampsia at term. *Eur J Obstet Gynecol Reprod Biol* 2015 Aug;191:23-27. [doi: [10.1016/j.ejogrb.2015.05.009](https://doi.org/10.1016/j.ejogrb.2015.05.009)] [Medline: [26070123](https://pubmed.ncbi.nlm.nih.gov/26070123/)]
91. Verhoeven CJ, Nuij C, Janssen-Rolf CR, Schuit E, Bais JM, Oei SG, et al. Predictors for failure of vacuum-assisted vaginal delivery: a case-control study. *Eur J Obstet Gynecol Reprod Biol* 2016 May;200:29-34. [doi: [10.1016/j.ejogrb.2016.02.008](https://doi.org/10.1016/j.ejogrb.2016.02.008)] [Medline: [26967343](https://pubmed.ncbi.nlm.nih.gov/26967343/)]
92. Vieira MC, White SL, Patel N, Seed PT, Briley AL, Sandall J, UPBEAT Consortium. Prediction of uncomplicated pregnancies in obese women: a prospective multicentre study. *BMC Med* 2017 Nov 3;15(1):194 [FREE Full text] [doi: [10.1186/s12916-017-0956-8](https://doi.org/10.1186/s12916-017-0956-8)] [Medline: [29096631](https://pubmed.ncbi.nlm.nih.gov/29096631/)]
93. Visentin S, Londero AP, Camerin M, Grisan E, Cosmi E. A possible new approach in the prediction of late gestational hypertension: the role of the fetal aortic intima-media thickness. *Medicine (Baltimore)* 2017 Jan;96(2):e5515. [doi: [10.1097/MD.0000000000005515](https://doi.org/10.1097/MD.0000000000005515)] [Medline: [28079791](https://pubmed.ncbi.nlm.nih.gov/28079791/)]
94. Wang C, Zhu W, Wei Y, Su R, Feng H, Lin L, et al. The predictive effects of early pregnancy lipid profiles and fasting glucose on the risk of gestational diabetes mellitus stratified by body mass index. *J Diabetes Res* 2016;2016:3013567. [doi: [10.1155/2016/3013567](https://doi.org/10.1155/2016/3013567)] [Medline: [26981541](https://pubmed.ncbi.nlm.nih.gov/26981541/)]
95. Wang L, Matsunaga S, Mikami Y, Takai Y, Terui K, Seki H. Pre-delivery fibrinogen predicts adverse maternal or neonatal outcomes in patients with placental abruption. *J Obstet Gynaecol Res* 2016 Jul;42(7):796-802. [doi: [10.1111/jog.12988](https://doi.org/10.1111/jog.12988)] [Medline: [27075198](https://pubmed.ncbi.nlm.nih.gov/27075198/)]
96. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. *Ann Epidemiol* 2018 Nov;28(11):783-9.e1. [doi: [10.1016/j.annepidem.2018.08.008](https://doi.org/10.1016/j.annepidem.2018.08.008)] [Medline: [30236415](https://pubmed.ncbi.nlm.nih.gov/30236415/)]
97. Xing Y, Qi X, Wang X, Yang F. Development of a modified score system as prediction model for successful vaginal birth after cesarean delivery. *Clin Transl Sci* 2019 Jan;12(1):53-57 [FREE Full text] [doi: [10.1111/cts.12603](https://doi.org/10.1111/cts.12603)] [Medline: [30548202](https://pubmed.ncbi.nlm.nih.gov/30548202/)]

98. Xu H, Feng G, Wei Y, Feng Y, Yang R, Wang L, et al. Predicting ectopic pregnancy using human chorionic gonadotropin (HCG) levels and main cause of infertility in women undergoing assisted reproductive treatment: retrospective observational cohort study. *JMIR Med Inform* 2020 Apr 16;8(4):e17366 [FREE Full text] [doi: [10.2196/17366](https://doi.org/10.2196/17366)] [Medline: [32297865](https://pubmed.ncbi.nlm.nih.gov/32297865/)]
99. Xu H, Wei Y, Yang R, Feng G, Tang W, Zhang H, et al. Prospective observational cohort study: computational models for early prediction of ongoing pregnancy in fresh IVF/ICSI-ET protocols. *Life Sci* 2019 Apr 1;222:221-227. [doi: [10.1016/j.lfs.2019.03.012](https://doi.org/10.1016/j.lfs.2019.03.012)] [Medline: [30858125](https://pubmed.ncbi.nlm.nih.gov/30858125/)]
100. Yang H, Zhu C, Ma Q, Long Y, Cheng Z. Variations of blood cells in prediction of gestational diabetes mellitus. *J Perinat Med* 2015 Jan;43(1):89-93. [doi: [10.1515/jpm-2014-0007](https://doi.org/10.1515/jpm-2014-0007)] [Medline: [24897392](https://pubmed.ncbi.nlm.nih.gov/24897392/)]
101. Yang T, Li N, Qiao C, Liu C. Development of a novel nomogram for predicting placenta accreta in patients with scarred uterus: a retrospective cohort study. *Front Med (Lausanne)* 2019;6:289. [doi: [10.3389/fmed.2019.00289](https://doi.org/10.3389/fmed.2019.00289)] [Medline: [31921868](https://pubmed.ncbi.nlm.nih.gov/31921868/)]
102. Yu C, Zhang R, Li J. A predictive model for high-quality blastocyst based on blastomere number, fragmentation, and symmetry. *J Assist Reprod Genet* 2018 May;35(5):809-816 [FREE Full text] [doi: [10.1007/s10815-018-1132-6](https://doi.org/10.1007/s10815-018-1132-6)] [Medline: [29502189](https://pubmed.ncbi.nlm.nih.gov/29502189/)]
103. Zhao R, Zhang W, Zhou L, Chen Y. Building a predictive model for successful vaginal delivery in nulliparas with term cephalic singleton pregnancies using decision tree analysis. *J Obstet Gynaecol Res* 2019 Aug;45(8):1536-1544. [doi: [10.1111/jog.14011](https://doi.org/10.1111/jog.14011)] [Medline: [31161703](https://pubmed.ncbi.nlm.nih.gov/31161703/)]
104. Zheng T, Ye W, Wang X, Li X, Zhang J, Little J, et al. A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women. *BMC Pregnancy Childbirth* 2019 Jul 19;19(1):252 [FREE Full text] [doi: [10.1186/s12884-019-2374-8](https://doi.org/10.1186/s12884-019-2374-8)] [Medline: [31324151](https://pubmed.ncbi.nlm.nih.gov/31324151/)]
105. Zwertbroek E, Broekhuijsen K, Langenveld J, van Baaren G, van den Berg P, Bremer H, HYPITAT-II Study Group. Prediction of progression to severe disease in women with late preterm hypertensive disorders of pregnancy. *Acta Obstet Gynecol Scand* 2017 Jan;96(1):96-105. [doi: [10.1111/aogs.13051](https://doi.org/10.1111/aogs.13051)] [Medline: [27792243](https://pubmed.ncbi.nlm.nih.gov/27792243/)]
106. Abbas SA, Riaz R, Kazmi SZ, Rizvi SS, Kwon SJ. Cause analysis of caesarian sections and application of machine learning methods for classification of birth data. *IEEE Access* 2018;6(5):67555-67561. [doi: [10.1109/ACCESS.2018.2879115](https://doi.org/10.1109/ACCESS.2018.2879115)]
107. Alberola-Rubio J, Garcia-Casado J, Prats-Boluda G, Ye-Lin Y, Desantes D, Valero J, et al. Prediction of labor onset type: spontaneous vs induced; role of electrohysterography? *Comput Methods Programs Biomed* 2017 Jun;144:127-133. [doi: [10.1016/j.cmpb.2017.03.018](https://doi.org/10.1016/j.cmpb.2017.03.018)] [Medline: [28494996](https://pubmed.ncbi.nlm.nih.gov/28494996/)]
108. Balani J, Hyer S, Shehata H, Mohareb F. Visceral fat mass as a novel risk factor for predicting gestational diabetes in obese pregnant women. *Obstet Med* 2018 Sep;11(3):121-125 [FREE Full text] [doi: [10.1177/1753495X17754149](https://doi.org/10.1177/1753495X17754149)] [Medline: [30214477](https://pubmed.ncbi.nlm.nih.gov/30214477/)]
109. Benalcazar-Parra C, Ye-Lin Y, Garcia-Casado J, Monfort-Ortiz R, Alberola-Rubio J, Perales A, et al. Prediction of labor induction success from the uterine electrohysterogram. *J Sensors* 2019 Nov 15;2019:1-12. [doi: [10.1155/2019/6916251](https://doi.org/10.1155/2019/6916251)]
110. Borup R, Thuesen L, Andersen C, Nyboe-Andersen A, Ziebe S, Winther O, et al. Competence classification of cumulus and granulosa cell transcriptome in embryos matched by morphology and female age. *PLoS One* 2016;11(4):e0153562 [FREE Full text] [doi: [10.1371/journal.pone.0153562](https://doi.org/10.1371/journal.pone.0153562)] [Medline: [27128483](https://pubmed.ncbi.nlm.nih.gov/27128483/)]
111. Chen L, Hao Y. Feature extraction and classification of EHG between pregnancy and labour group using Hilbert-Huang transform and extreme learning machine. *Comput Math Methods Med* 2017;2017:7949507. [doi: [10.1155/2017/7949507](https://doi.org/10.1155/2017/7949507)] [Medline: [28316639](https://pubmed.ncbi.nlm.nih.gov/28316639/)]
112. Chen L, Hao Y, Hu X. Detection of preterm birth in electrohysterogram signals based on wavelet transform and stacked sparse autoencoder. *PLoS One* 2019;14(4):e0214712 [FREE Full text] [doi: [10.1371/journal.pone.0214712](https://doi.org/10.1371/journal.pone.0214712)] [Medline: [30990810](https://pubmed.ncbi.nlm.nih.gov/30990810/)]
113. Cömert Z, Kocamaz AF, Subha V. Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Comput Biol Med* 2018 Aug 1;99:85-97. [doi: [10.1016/j.compbimed.2018.06.003](https://doi.org/10.1016/j.compbimed.2018.06.003)] [Medline: [29894897](https://pubmed.ncbi.nlm.nih.gov/29894897/)]
114. Coppedè F, Grossi E, Migheli F, Migliore L. Polymorphisms in folate-metabolizing genes, chromosome damage, and risk of Down syndrome in Italian women: identification of key factors using artificial neural networks. *BMC Med Genomics* 2010 Sep 24;3:42 [FREE Full text] [doi: [10.1186/1755-8794-3-42](https://doi.org/10.1186/1755-8794-3-42)] [Medline: [20868477](https://pubmed.ncbi.nlm.nih.gov/20868477/)]
115. Despotovic D, Zec A, Mladenovic K, Radin N, Turukalo T. A machine learning approach for an early prediction of preterm delivery. In: 16th International Symposium on Intelligent Systems and Informatics. 2018 Presented at: SISY'18; September 13-15, 2018; Subotica, Serbia. [doi: [10.1109/SISY.2018.8524818](https://doi.org/10.1109/SISY.2018.8524818)]
116. Elaveyini U, Devi SP, Rao KS. Neural networks prediction of preterm delivery with first trimester bleeding. *Arch Gynecol Obstet* 2011 May;283(5):971-979. [doi: [10.1007/s00404-010-1469-2](https://doi.org/10.1007/s00404-010-1469-2)] [Medline: [20449599](https://pubmed.ncbi.nlm.nih.gov/20449599/)]
117. Fergus P, Hussain A, Al-Jumeily D, Huang D, Bouguila N. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. *Biomed Eng Online* 2017 Jul 6;16(1):89 [FREE Full text] [doi: [10.1186/s12938-017-0378-z](https://doi.org/10.1186/s12938-017-0378-z)] [Medline: [28679415](https://pubmed.ncbi.nlm.nih.gov/28679415/)]
118. Fergus P, Idowu I, Hussain A, Dobbins C. Advanced artificial neural network classification for detecting preterm births using EHG records. *Neurocomputing* 2016 May;188:42-49. [doi: [10.1016/j.neucom.2015.01.107](https://doi.org/10.1016/j.neucom.2015.01.107)]

119. Fergus P, Montanez A, Abdulaimma B, Lisboa P, Chalmers C, Pineles B. Utilizing deep learning and genome wide association studies for epistatic-driven preterm birth classification in African-American women. *IEEE/ACM Trans Comput Biol Bioinform* 2020;17(2):668-678. [doi: [10.1109/TCBB.2018.2868667](https://doi.org/10.1109/TCBB.2018.2868667)] [Medline: [30183645](https://pubmed.ncbi.nlm.nih.gov/30183645/)]
120. Figueras F, Savchev S, Triunfo S, Crovetto F, Gratacos E. An integrated model with classification criteria to predict small-for-gestational-age fetuses at risk of adverse perinatal outcome. *Ultrasound Obstet Gynecol* 2015 Mar;45(3):279-285. [doi: [10.1002/uog.14714](https://doi.org/10.1002/uog.14714)] [Medline: [25358519](https://pubmed.ncbi.nlm.nih.gov/25358519/)]
121. Fiset S, Martel A, Glanc P, Barrett J, Melamed N. Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images. *Univ Tor Med J* 2019;96(1):6-9 [FREE Full text]
122. Gao C, Osmundson S, Velez Edwards DR, Jackson G, Malin B, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. *J Biomed Inform* 2019 Dec;100:103334. [doi: [10.1016/j.jbi.2019.103334](https://doi.org/10.1016/j.jbi.2019.103334)] [Medline: [31678588](https://pubmed.ncbi.nlm.nih.gov/31678588/)]
123. Garcés MF, Sanchez E, Cardona LF, Simanca EL, González I, Leal LG, et al. Maternal serum meteorin levels and the risk of preeclampsia. *PLoS One* 2015;10(6):e0131013 [FREE Full text] [doi: [10.1371/journal.pone.0131013](https://doi.org/10.1371/journal.pone.0131013)] [Medline: [26121675](https://pubmed.ncbi.nlm.nih.gov/26121675/)]
124. Georgoulas G, Karvelis P, Spilka J, Chudáček V, Stylios CD, Lhotská L. Investigating pH based evaluation of fetal heart rate (FHR) recordings. *Health Technol (Berl)* 2017;7(2):241-254 [FREE Full text] [doi: [10.1007/s12553-017-0201-7](https://doi.org/10.1007/s12553-017-0201-7)] [Medline: [29201590](https://pubmed.ncbi.nlm.nih.gov/29201590/)]
125. Hamdi M, Limem M, Maaref M. Detection and classification of nonstationary signals: application to uterine EMG for prognostication of premature delivery. *Neurophysiology* 2019 Dec 4;51(4):272-280. [doi: [10.1007/s11062-019-09821-9](https://doi.org/10.1007/s11062-019-09821-9)]
126. Hernández-González J, Inza I, Crisol-Ortiz L, Guembe M, Iñarra MJ, Lozano J. Fitting the data from embryo implantation prediction: learning from label proportions. *Stat Methods Med Res* 2018 Apr;27(4):1056-1066. [doi: [10.1177/0962280216651098](https://doi.org/10.1177/0962280216651098)] [Medline: [27242336](https://pubmed.ncbi.nlm.nih.gov/27242336/)]
127. Jhee JH, Lee S, Park Y, Lee SE, Kim YA, Kang S, et al. Prediction model development of late-onset preeclampsia using machine learning-based methods. *PLoS One* 2019;14(8):e0221202 [FREE Full text] [doi: [10.1371/journal.pone.0221202](https://doi.org/10.1371/journal.pone.0221202)] [Medline: [31442238](https://pubmed.ncbi.nlm.nih.gov/31442238/)]
128. Leonarduzzi R, Spilka J, Frecon J, Wendt H, Pustelnik N, Jaffard S, et al. P-leader multifractal analysis and sparse SVM for intrapartum fetal acidosis detection. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:1971-1974. [doi: [10.1109/EMBC.2015.7318771](https://doi.org/10.1109/EMBC.2015.7318771)] [Medline: [26736671](https://pubmed.ncbi.nlm.nih.gov/26736671/)]
129. Li H, Luo M, Zheng J, Luo J, Zeng R, Feng N, et al. An artificial neural network prediction model of congenital heart disease based on risk factors: a hospital-based case-control study. *Medicine (Baltimore)* 2017 Feb;96(6):e6090. [doi: [10.1097/MD.0000000000006090](https://doi.org/10.1097/MD.0000000000006090)] [Medline: [28178169](https://pubmed.ncbi.nlm.nih.gov/28178169/)]
130. Mas-Cabo J, Prats-Boluda G, Garcia-Casado J, Alberola-Rubio J, Perales A, Ye-Lin Y. Design and assessment of a robust and generalizable ANN-based classifier for the prediction of premature birth by means of multichannel electrohysterographic records. *J Sensors* 2019 Nov 25;2019:1-13. [doi: [10.1155/2019/5373810](https://doi.org/10.1155/2019/5373810)]
131. Mello G, Parretti E, Ognibene A, Mecacci F, Cioni R, Scarselli G, et al. Prediction of the development of pregnancy-induced hypertensive disorders in high-risk pregnant women by artificial neural networks. *Clin Chem Lab Med* 2001 Sep;39(9):801-805. [doi: [10.1515/CCLM.2001.132](https://doi.org/10.1515/CCLM.2001.132)] [Medline: [11601676](https://pubmed.ncbi.nlm.nih.gov/11601676/)]
132. Milewski R, Kuczyńska A, Stankiewicz B, Kuczyński W. How much information about embryo implantation potential is included in morphokinetic data? A prediction model based on artificial neural networks and principal component analysis. *Adv Med Sci* 2017 Mar;62(1):202-206. [doi: [10.1016/j.advms.2017.02.001](https://doi.org/10.1016/j.advms.2017.02.001)] [Medline: [28384614](https://pubmed.ncbi.nlm.nih.gov/28384614/)]
133. Milewski R, Milewska AJ, Więsak T, Morgan A. Comparison of artificial neural networks and logistic regression analysis in pregnancy prediction using the in vitro fertilization treatment. *Stud Log Gramm Rhetor* 2013;35((1)):39-48. [doi: [10.2478/slgr-2013-0033](https://doi.org/10.2478/slgr-2013-0033)]
134. Mirroshandel S, Ghasemian F, Monji-Azad S. Applying data mining techniques for increasing implantation rate by selecting best sperms for intra-cytoplasmic sperm injection treatment. *Comput Methods Programs Biomed* 2016 Dec;137:215-229. [doi: [10.1016/j.cmpb.2016.09.013](https://doi.org/10.1016/j.cmpb.2016.09.013)] [Medline: [28110726](https://pubmed.ncbi.nlm.nih.gov/28110726/)]
135. Morales DA, Bengoetxea E, Larrañaga P, García M, Franco Y, Fresnada M, et al. Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. *Comput Methods Programs Biomed* 2008 May;90(2):104-116. [doi: [10.1016/j.cmpb.2007.11.018](https://doi.org/10.1016/j.cmpb.2007.11.018)] [Medline: [18190996](https://pubmed.ncbi.nlm.nih.gov/18190996/)]
136. Paydar K, Niakan Kalhori SR, Akbarian M, Sheikhtaheri A. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. *Int J Med Inform* 2017 Jan;97:239-246. [doi: [10.1016/j.ijmedinf.2016.10.018](https://doi.org/10.1016/j.ijmedinf.2016.10.018)] [Medline: [27919382](https://pubmed.ncbi.nlm.nih.gov/27919382/)]
137. Petrozziello A, Jordanov I, Aris Papageorghiou T, Christopher Redman WG, Georgieva A. Deep learning for continuous electronic fetal monitoring in labor. *Annu Int Conf IEEE Eng Med Biol Soc* 2018 Jul;2018:5866-5869. [doi: [10.1109/EMBC.2018.8513625](https://doi.org/10.1109/EMBC.2018.8513625)] [Medline: [30441670](https://pubmed.ncbi.nlm.nih.gov/30441670/)]
138. Petrozziello A, Redman C, Papageorghiou A, Jordanov I, Georgieva A. Multimodal convolutional neural networks to detect fetal compromise during labor and delivery. *IEEE Access* 2019;7:112026-112036. [doi: [10.1109/access.2019.2933368](https://doi.org/10.1109/access.2019.2933368)]
139. Qiu H, Yu H, Wang L, Yao Q, Wu S, Yin C, et al. Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. *Sci Rep* 2017 Nov 27;7(1):16417. [doi: [10.1038/s41598-017-16665-y](https://doi.org/10.1038/s41598-017-16665-y)] [Medline: [29180800](https://pubmed.ncbi.nlm.nih.gov/29180800/)]

140. Qiu J, Li P, Dong M, Xin X, Tan J. Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. *J Transl Med* 2019 Sep 23;17(1):317 [FREE Full text] [doi: [10.1186/s12967-019-2062-5](https://doi.org/10.1186/s12967-019-2062-5)] [Medline: [31547822](https://pubmed.ncbi.nlm.nih.gov/31547822/)]
141. Sadi-Ahmed N, Kacha B, Taleb H, Kedir-Talha M. Relevant features selection for automatic prediction of preterm deliveries from pregnancy electrohysterographic (EHG) records. *J Med Syst* 2017 Nov 11;41(12):204. [doi: [10.1007/s10916-017-0847-8](https://doi.org/10.1007/s10916-017-0847-8)] [Medline: [29128973](https://pubmed.ncbi.nlm.nih.gov/29128973/)]
142. Saleem S, Naqvi S, Manzoor T, Saeed A, Ur Rehman N, Mirza J. A strategy for classification of 'vaginal vs cesarean section' delivery: bivariate empirical mode decomposition of cardiocographic recordings. *Front Physiol* 2019;10:246. [doi: [10.3389/fphys.2019.00246](https://doi.org/10.3389/fphys.2019.00246)] [Medline: [30941054](https://pubmed.ncbi.nlm.nih.gov/30941054/)]
143. Shahbakhti M, Beiramvand M, Bavi M, Mohammadi Far S. A new efficient algorithm for prediction of preterm labor. *Annu Int Conf IEEE Eng Med Biol Soc* 2019 Jul;2019:4669-4672. [doi: [10.1109/EMBC.2019.8857837](https://doi.org/10.1109/EMBC.2019.8857837)] [Medline: [31946904](https://pubmed.ncbi.nlm.nih.gov/31946904/)]
144. Signorini MG, Pini N, Malovini A, Bellazzi R, Magenes G. Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Comput Methods Programs Biomed* 2020 Mar;185:105015. [doi: [10.1016/j.cmpb.2019.105015](https://doi.org/10.1016/j.cmpb.2019.105015)] [Medline: [31678794](https://pubmed.ncbi.nlm.nih.gov/31678794/)]
145. Spilka J, Frecon J, Leonarduzzi R, Pustelnik N, Abry P, Doret M. Intrapartum fetal heart rate classification from trajectory in Sparse SVM feature space. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:2335-2338. [doi: [10.1109/EMBC.2015.7318861](https://doi.org/10.1109/EMBC.2015.7318861)] [Medline: [26736761](https://pubmed.ncbi.nlm.nih.gov/26736761/)]
146. Spilka J, Frecon J, Leonarduzzi R, Pustelnik N, Abry P, Doret M. Sparse support vector machine for intrapartum fetal heart rate classification. *IEEE J Biomed Health Inform* 2017 May;21(3):664-671. [doi: [10.1109/JBHI.2016.2546312](https://doi.org/10.1109/JBHI.2016.2546312)] [Medline: [27046884](https://pubmed.ncbi.nlm.nih.gov/27046884/)]
147. Sufriyana H, Wu Y, Su EC. Artificial intelligence-assisted prediction of preeclampsia: development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. *EBioMedicine* 2020 Apr;54:102710 [FREE Full text] [doi: [10.1016/j.ebiom.2020.102710](https://doi.org/10.1016/j.ebiom.2020.102710)] [Medline: [32283530](https://pubmed.ncbi.nlm.nih.gov/32283530/)]
148. Tran D, Cooke S, Illingworth P, Gardner D. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod* 2019 Jun 4;34(6):1011-1018 [FREE Full text] [doi: [10.1093/humrep/dez064](https://doi.org/10.1093/humrep/dez064)] [Medline: [31111884](https://pubmed.ncbi.nlm.nih.gov/31111884/)]
149. Troisi J, Landolfi A, Sarno L, Richards S, Symes S, Adair D, et al. A metabolomics-based approach for non-invasive screening of fetal central nervous system anomalies. *Metabolomics* 2018 May 25;14(6):77. [doi: [10.1007/s11306-018-1370-8](https://doi.org/10.1007/s11306-018-1370-8)] [Medline: [30830338](https://pubmed.ncbi.nlm.nih.gov/30830338/)]
150. Uyar A, Bener A, Ciray HN. Predictive modeling of implantation outcome in an in vitro fertilization setting: an application of machine learning methods. *Med Decis Making* 2015 Aug;35(6):714-725. [doi: [10.1177/0272989X14535984](https://doi.org/10.1177/0272989X14535984)] [Medline: [24842951](https://pubmed.ncbi.nlm.nih.gov/24842951/)]
151. Uyar A, Bener A, Ciray H. ROC Based Evaluation and Comparison of Classifiers for IVF Implantation Prediction. In: *International Conference on Electronic Healthcare*. 2009 Presented at: eHealth'09; September 23-25, 2009; Istanbul, Turkey. [doi: [10.1007/978-3-642-11745-9_17](https://doi.org/10.1007/978-3-642-11745-9_17)]
152. Valensise H, Facchinetti F, Vasapollo B, Giannini F, Monte ID, Arduini D. The computerized fetal heart rate analysis in post-term pregnancy identifies patients at risk for fetal distress in labour. *Eur J Obstet Gynecol Reprod Biol* 2006 Apr 1;125(2):185-192. [doi: [10.1016/j.ejogrb.2005.06.034](https://doi.org/10.1016/j.ejogrb.2005.06.034)] [Medline: [16459010](https://pubmed.ncbi.nlm.nih.gov/16459010/)]
153. VerMilyea M, Hall J, Diakiw S, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod* 2020 Apr 28;35(4):770-784 [FREE Full text] [doi: [10.1093/humrep/deaa013](https://doi.org/10.1093/humrep/deaa013)] [Medline: [32240301](https://pubmed.ncbi.nlm.nih.gov/32240301/)]
154. Vogiatzi P, Poulidakis A, Siristatidis C. An artificial neural network for the prediction of assisted reproduction outcome. *J Assist Reprod Genet* 2019 Jul;36(7):1441-1448 [FREE Full text] [doi: [10.1007/s10815-019-01498-7](https://doi.org/10.1007/s10815-019-01498-7)] [Medline: [31218565](https://pubmed.ncbi.nlm.nih.gov/31218565/)]
155. Xu L, Georgieva A, Redman C, Payne S. Feature selection for computerized fetal heart rate analysis using genetic algorithms. *Annu Int Conf IEEE Eng Med Biol Soc* 2013;2013:445-448. [doi: [10.1109/EMBC.2013.6609532](https://doi.org/10.1109/EMBC.2013.6609532)] [Medline: [24109719](https://pubmed.ncbi.nlm.nih.gov/24109719/)]
156. Amini P, Maroufizadeh S, Samani RO, Hamidi O, Sepidarkish M. Prevalence and determinants of preterm birth in Tehran, Iran: a comparison between logistic regression and decision tree methods. *Osong Public Health Res Perspect* 2017 Jun;8(3):195-200 [FREE Full text] [doi: [10.24171/j.phrp.2017.8.3.06](https://doi.org/10.24171/j.phrp.2017.8.3.06)] [Medline: [28781942](https://pubmed.ncbi.nlm.nih.gov/28781942/)]
157. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. *Nat Med* 2020 Jan;26(1):71-76. [doi: [10.1038/s41591-019-0724-8](https://doi.org/10.1038/s41591-019-0724-8)] [Medline: [31932807](https://pubmed.ncbi.nlm.nih.gov/31932807/)]
158. Blank C, Wildeboer R, DeCroc I, Tilleman K, Weyers B, de Sutter P, et al. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. *Fertil Steril* 2019 Feb;111(2):318-326. [doi: [10.1016/j.fertnstert.2018.10.030](https://doi.org/10.1016/j.fertnstert.2018.10.030)] [Medline: [30611557](https://pubmed.ncbi.nlm.nih.gov/30611557/)]
159. Isakov O, Reicher L, Lavie A, Yogev Y, Maslovitz S. Prediction of success in external cephalic version for breech presentation at term. *Obstet Gynecol* 2019 May;133(5):857-866. [doi: [10.1097/AOG.0000000000003196](https://doi.org/10.1097/AOG.0000000000003196)] [Medline: [30969207](https://pubmed.ncbi.nlm.nih.gov/30969207/)]
160. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inf Sci Syst* 2020 Dec;8(1):14 [FREE Full text] [doi: [10.1007/s13755-020-00105-9](https://doi.org/10.1007/s13755-020-00105-9)] [Medline: [32226625](https://pubmed.ncbi.nlm.nih.gov/32226625/)]

161. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen AC, Joseph KS, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth* 2018 Aug 15;18(1):333 [FREE Full text] [doi: [10.1186/s12884-018-1971-2](https://doi.org/10.1186/s12884-018-1971-2)] [Medline: [30111303](https://pubmed.ncbi.nlm.nih.gov/30111303/)]
162. Kumar SN, Saxena P, Patel R, Sharma A, Pradhan D, Singh H, et al. Predicting risk of low birth weight offspring from maternal features and blood polycyclic aromatic hydrocarbon concentration. *Reprod Toxicol* 2020 Jun;94:92-100. [doi: [10.1016/j.reprotox.2020.03.009](https://doi.org/10.1016/j.reprotox.2020.03.009)] [Medline: [32283251](https://pubmed.ncbi.nlm.nih.gov/32283251/)]
163. Lee K, Ahn KH. Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants. *J Korean Med Sci* 2019 Apr 29;34(16):e128 [FREE Full text] [doi: [10.3346/jkms.2019.34.e128](https://doi.org/10.3346/jkms.2019.34.e128)] [Medline: [31020816](https://pubmed.ncbi.nlm.nih.gov/31020816/)]
164. Liu B, Shi S, Wu Y, Thomas D, Symul L, Pierson E, et al. Predicting pregnancy using large-scale data from a women's health tracking mobile application. *Proc Int World Wide Web Conf 2019 May;2019:2999-3005* [FREE Full text] [doi: [10.1145/3308558.3313512](https://doi.org/10.1145/3308558.3313512)] [Medline: [31538145](https://pubmed.ncbi.nlm.nih.gov/31538145/)]
165. Macones GA, Hausman N, Edelstein R, Stamilio DM, Marder SJ. Predicting outcomes of trials of labor in women attempting vaginal birth after cesarean delivery: a comparison of multivariate methods with neural networks. *Am J Obstet Gynecol* 2001 Feb;184(3):409-413. [doi: [10.1067/mob.2001.109386](https://doi.org/10.1067/mob.2001.109386)] [Medline: [11228495](https://pubmed.ncbi.nlm.nih.gov/11228495/)]
166. Maroufizadeh S, Amini P, Hosseini M, Almasi-Hashiani A, Mohammadi M, Navid B, et al. Determinants of cesarean section among primiparas: a comparison of classification methods. *Iran J Public Health* 2018 Dec;47(12):1913-1922 [FREE Full text] [Medline: [30788307](https://pubmed.ncbi.nlm.nih.gov/30788307/)]
167. Sims CJ, Meyn L, Caruana R, Rao R, Mitchell T, Krohn M. Predicting cesarean delivery with decision tree models. *Am J Obstet Gynecol* 2000 Nov;183(5):1198-1206. [doi: [10.1067/mob.2000.108891](https://doi.org/10.1067/mob.2000.108891)] [Medline: [11084566](https://pubmed.ncbi.nlm.nih.gov/11084566/)]
168. Agopian A, Lupo P, Tinker S, Canfield M, Mitchell L, National Birth Defects Prevention Study. Working towards a risk prediction model for neural tube defects. *Birth Defects Res A Clin Mol Teratol* 2012 Mar;94(3):141-146 [FREE Full text] [doi: [10.1002/bdra.22883](https://doi.org/10.1002/bdra.22883)] [Medline: [22253139](https://pubmed.ncbi.nlm.nih.gov/22253139/)]
169. Fergus P, Cheung P, Hussain A, Al-Jumeily D, Dobbins C, Iram S. Prediction of preterm deliveries from EHG signals using machine learning. *PLoS One* 2013;8(10):e77154 [FREE Full text] [doi: [10.1371/journal.pone.0077154](https://doi.org/10.1371/journal.pone.0077154)] [Medline: [24204760](https://pubmed.ncbi.nlm.nih.gov/24204760/)]
170. Wald M, Sparks A, Sandlow J, Van-Voorhis B, Syrop C, Niederberger C. Computational models for prediction of IVF/ICSI outcomes with surgically retrieved spermatozoa. *Reprod Biomed Online* 2005 Sep;11(3):325-331. [doi: [10.1016/s1472-6483\(10\)60840-1](https://doi.org/10.1016/s1472-6483(10)60840-1)] [Medline: [16176672](https://pubmed.ncbi.nlm.nih.gov/16176672/)]
171. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006 Jun;11(2):193-206. [doi: [10.1037/1082-989X.11.2.193](https://doi.org/10.1037/1082-989X.11.2.193)] [Medline: [16784338](https://pubmed.ncbi.nlm.nih.gov/16784338/)]
172. Deeks J, Higgins J, Altman D. Analysing data undertaking meta-analyses. In: Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. New Jersey, USA: Wiley-Blackwell; 2008:243-296.
173. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods* 2017 Mar;8(1):5-18. [doi: [10.1002/jrsm.1230](https://doi.org/10.1002/jrsm.1230)] [Medline: [28058794](https://pubmed.ncbi.nlm.nih.gov/28058794/)]
174. Fele-Zorz G, Kavsek G, Novak-Antolic Z, Jager F. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Med Biol Eng Comput* 2008 Sep;46(9):911-922. [doi: [10.1007/s11517-008-0350-y](https://doi.org/10.1007/s11517-008-0350-y)] [Medline: [18437439](https://pubmed.ncbi.nlm.nih.gov/18437439/)]
175. Chudáček V, Spilka J, Burša M, Janků P, Hruban L, Huptych M, et al. Open access intrapartum CTG database. *BMC Pregnancy Childbirth* 2014 Jan 13;14:16 [FREE Full text] [doi: [10.1186/1471-2393-14-16](https://doi.org/10.1186/1471-2393-14-16)] [Medline: [24418387](https://pubmed.ncbi.nlm.nih.gov/24418387/)]
176. Iftikhar P, Kuijpers M, Khayyat A, Iftikhar A, DeGouvía de Sa M. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Cureus* 2020 Feb 28;12(2):e7124 [FREE Full text] [doi: [10.7759/cureus.7124](https://doi.org/10.7759/cureus.7124)] [Medline: [32257670](https://pubmed.ncbi.nlm.nih.gov/32257670/)]
177. Khanam R, Ahmed S, Creanga AA, Begum N, Koffi AK, Mahmud A, Projahnmo Study Group in Bangladesh. Antepartum complications and perinatal mortality in rural Bangladesh. *BMC Pregnancy Childbirth* 2017 Mar 7;17(1):81 [FREE Full text] [doi: [10.1186/s12884-017-1264-1](https://doi.org/10.1186/s12884-017-1264-1)] [Medline: [28270117](https://pubmed.ncbi.nlm.nih.gov/28270117/)]
178. Knowles SJ, O'Sullivan NP, Meenan AM, Hanniffy R, Robson M. Maternal sepsis incidence, aetiology and outcome for mother and fetus: a prospective study. *BJOG* 2015 Apr;122(5):663-671. [doi: [10.1111/1471-0528.12892](https://doi.org/10.1111/1471-0528.12892)] [Medline: [24862293](https://pubmed.ncbi.nlm.nih.gov/24862293/)]
179. Shillan D, Sterne JA, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care* 2019 Aug 22;23(1):284 [FREE Full text] [doi: [10.1186/s13054-019-2564-9](https://doi.org/10.1186/s13054-019-2564-9)] [Medline: [31439010](https://pubmed.ncbi.nlm.nih.gov/31439010/)]
180. Mortazavi BJ, Downing NS, Buchholz EM, Dharmarajan K, Manhapra A, Li S, et al. Analysis of machine learning techniques for heart failure readmissions. *Circ Cardiovasc Qual Outcomes* 2016 Nov;9(6):629-640 [FREE Full text] [doi: [10.1161/CIRCOUTCOMES.116.003039](https://doi.org/10.1161/CIRCOUTCOMES.116.003039)] [Medline: [28263938](https://pubmed.ncbi.nlm.nih.gov/28263938/)]
181. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 2014;15(90):3133-3181 [FREE Full text]
182. Breiman L. Random Forests. *Mach Learn* 2001 Oct;45(1):5-32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
183. Fawagreh K, Gaber M, Elyan E. Random forests: from early developments to recent advancements. *Syst Sci Control Engin* 2014 Oct 06;2(1):602-609. [doi: [10.1080/21642583.2014.956265](https://doi.org/10.1080/21642583.2014.956265)]

184. Friedman J. Stochastic gradient boosting. *Comput Stat Data Anal* 2002 Feb;38(4):367-378. [doi: [10.1016/s0167-9473\(01\)00065-2](https://doi.org/10.1016/s0167-9473(01)00065-2)]
185. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21. [doi: [10.3389/fnbot.2013.00021](https://doi.org/10.3389/fnbot.2013.00021)] [Medline: [24409142](https://pubmed.ncbi.nlm.nih.gov/24409142/)]
186. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017 Presented at: NIPS'17; December 1-7, 2017; Long Beach URL: <http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf>
187. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016 Presented at: ACM'16; August 11-15, 2016; San Francisco, California, USA. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
188. Siristatidis C, Poulidakis A, Chrelias C, Kassanos D. Artificial intelligence in IVF: a need. *Syst Biol Reprod Med* 2011 Aug;57(4):179-185. [doi: [10.3109/19396368.2011.558607](https://doi.org/10.3109/19396368.2011.558607)] [Medline: [21375363](https://pubmed.ncbi.nlm.nih.gov/21375363/)]
189. Haake KW, List P, Baier D, Zimmermann G, Pretzsch G, Alexander H. [Risk assessment in ovarian hyperstimulation syndrome (OHS) using the machine learning system (Decision Master) in 155 in-vitro fertilisations and embryo-transfer (IVF/ET) cycles with a long stimulation protocol]. *Zentralbl Gynakol* 1997;119(Suppl 1):23-27. [Medline: [9245120](https://pubmed.ncbi.nlm.nih.gov/9245120/)]
190. Manna C, Nanni L, Lumini A, Pappalardo S. Artificial intelligence techniques for embryo and oocyte classification. *Reprod Biomed Online* 2013 Jan;26(1):42-49. [doi: [10.1016/j.rbmo.2012.09.015](https://doi.org/10.1016/j.rbmo.2012.09.015)] [Medline: [23177416](https://pubmed.ncbi.nlm.nih.gov/23177416/)]
191. Santos Filho E, Noble J, Poli M, Griffiths T, Emerson G, Wells D. A method for semi-automatic grading of human blastocyst microscope images. *Hum Reprod* 2012 Sep;27(9):2641-2648. [doi: [10.1093/humrep/des219](https://doi.org/10.1093/humrep/des219)] [Medline: [22736327](https://pubmed.ncbi.nlm.nih.gov/22736327/)]
192. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit Med* 2019;2:21. [doi: [10.1038/s41746-019-0096-y](https://doi.org/10.1038/s41746-019-0096-y)] [Medline: [31304368](https://pubmed.ncbi.nlm.nih.gov/31304368/)]
193. Liang B, Gao Y, Xu J, Song Y, Xuan L, Shi T, et al. Raman profiling of embryo culture medium to identify aneuploid and euploid embryos. *Fertil Steril* 2019 Apr;111(4):753-62.e1. [doi: [10.1016/j.fertnstert.2018.11.036](https://doi.org/10.1016/j.fertnstert.2018.11.036)] [Medline: [30683589](https://pubmed.ncbi.nlm.nih.gov/30683589/)]
194. Ratna M, Bhattacharya S, Abdulrahim B, McLernon D. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod* 2020 Jan 1;35(1):100-116. [doi: [10.1093/humrep/dez258](https://doi.org/10.1093/humrep/dez258)] [Medline: [31960915](https://pubmed.ncbi.nlm.nih.gov/31960915/)]
195. Doret M, Massoud M, Constans A, Gaucherand P. Use of peripartum ST analysis of fetal electrocardiogram without blood sampling: a large prospective cohort study. *Eur J Obstet Gynecol Reprod Biol* 2011 May;156(1):35-40. [doi: [10.1016/j.ejogrb.2010.12.042](https://doi.org/10.1016/j.ejogrb.2010.12.042)] [Medline: [21257256](https://pubmed.ncbi.nlm.nih.gov/21257256/)]
196. Alexandersson A, Steingrimsdottir T, Terrien J, Marque C, Karlsson B. The Icelandic 16-electrode electrohysterogram database. *Sci Data* 2015;2:150017. [doi: [10.1038/sdata.2015.17](https://doi.org/10.1038/sdata.2015.17)] [Medline: [25984349](https://pubmed.ncbi.nlm.nih.gov/25984349/)]
197. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015 Jul 17;349(6245):255-260. [doi: [10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415)] [Medline: [26185243](https://pubmed.ncbi.nlm.nih.gov/26185243/)]
198. Kennady G, Kottarathara MJ, Kottarathara AJ, Ajith R, Anandakesavan TM, Ambujam K. Maternal and neonatal outcomes in pregnancy induced hypertension: an observational study. *Clin Exp Obstet Gynecol* 2017;44(1):110-112. [Medline: [29714877](https://pubmed.ncbi.nlm.nih.gov/29714877/)]
199. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014 Dec 22;14:137 [FREE Full text] [doi: [10.1186/1471-2288-14-137](https://doi.org/10.1186/1471-2288-14-137)] [Medline: [25532820](https://pubmed.ncbi.nlm.nih.gov/25532820/)]
200. Meertens LJ, van Montfort P, Scheepers HC, van Kuijk SM, Aardenburg R, Langenveld J, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. *Acta Obstet Gynecol Scand* 2018 Aug;97(8):907-920. [doi: [10.1111/aogs.13358](https://doi.org/10.1111/aogs.13358)] [Medline: [29663314](https://pubmed.ncbi.nlm.nih.gov/29663314/)]

Abbreviations

AUROC: area under the receiver operating characteristic curve

CHARMS: checklist for critical appraisal and data extraction for systematic reviews of prediction modeling studies

CS: cesarean section

CTG: cardiotocogram

EHG: electrohysterogram

EPV: events per variable

IRR: incidence rate ratio

LR: logistic regression

MeSH: Medical Subject Heading

MLP-BIOM: guidelines for developing and reporting machine learning predictive models in biomedical research

OR: odds ratio

PI: prediction interval

PICOTS: population, index, comparator, outcomes, timing, and setting

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROBAST: prediction model risk of bias assessment tool

ROB: risk of bias

ROC: receiver operating characteristic

TRIPOD: transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

Edited by G Eysenbach; submitted 07.10.19; peer-reviewed by E Christodoulou, R Ho, WD Dotson, L Kriston; comments to author 13.03.20; revised version received 22.06.20; accepted 24.10.20; published 17.11.20

Please cite as:

Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, Wu YW, Su ECY

Comparison of Multivariable Logistic Regression and Other Machine Learning Algorithms for Prognostic Prediction Studies in Pregnancy Care: Systematic Review and Meta-Analysis

JMIR Med Inform 2020;8(11):e16503

URL: <http://medinform.jmir.org/2020/11/e16503/>

doi: [10.2196/16503](https://doi.org/10.2196/16503)

PMID: [33200995](https://pubmed.ncbi.nlm.nih.gov/33200995/)

©Herdiantri Sufriyana, Atina Husnayain, Ya-Lin Chen, Chao-Yang Kuo, Onkar Singh, Tso-Yang Yeh, Yu-Wei Wu, Emily Chia-Yu Su. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 17.11.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.