Artificial intelligence in obstetrics

Ki Hoon Ahn, MD, PhD[1], Kwang-Sig Lee, PhD[2]

[1]Department of Obstetrics and Gynecology, Korea University Anam Hospital, [2]AI Center, Korea University Anam Hospital, Seoul, Korea

This study reviews recent advances on the application of artificial intelligence for the early diagnosis of various maternal-fetal conditions such as preterm birth and abnormal fetal growth. It is found in this study that various machine learning methods have been successfully employed for different kinds of data capture with regard to early diagnosis of maternal-fetal conditions. With the more popular use of artificial intelligence, ethical issues should also be considered accordingly.

Keywords: Mother; Fetus; Disease; Diagnosis; Artificial intelligence

Introduction

1. Maternal-fetal medicine

Maternal fetal medicine is traditionally a litigation area that accounts for a large portion of compensation payments. The characteristic of the lawsuit is that the cost per case is not small compared to the number of claims. According to a report in South Korea, the number of medical disputes resolved between 2015 and 2019 was between KRW 61 million and the cost was KRW 91 million per case. Medical disputes rank 4th to 6th in 21 departments and 10th in terms of amount. Likewise, the trend was notable in the UK and other countries between 2017 and 2018, with the National Health Service reporting a total of £4.5 billion in compensation, with 48% for obstetric malpractice claims. Interestingly, obstetrics and gynecology accounted for only 15 per cent of all claims. In addition to the costs, abnormal medical outcomes can be fatal socioeconomically given a long life-span with disability. A common complication in obstetric care is hypoxia-induced encephalopathy, with a 5.14 incidence rate for every 1,000 births as reported in 2015. This can sometimes be caused by the misinterpretation of fetal monitoring during labor, and we generally agree that 50% of these cases are preventable [1]. The fundamental reason for this social cost expenditure on obstetric diseases is that it is difficult to predict and diagnose maternal and fetal diseases early, and therefore, even if the disease is found, the effectiveness of treatment is low. The area under the receiver operating characteristic curve (AUC) for diagnosis of conditions, such as preterm birth, preeclampsia, fetal growth restriction, and stillbirth, which are relatively common but have a significant impact on maternal and fetal health, has been reported to be 0.51-0.85, 0.61-0.99, 0.54-0.98, and 0.58-0.76, respectively [2-10]. Evidence-based traditional research methodologies have been the main tools used to solve problems related to obstetric healthcare. These include classical research methodologies such as meta-analysis, clinical trials, cohort studies, cross-sectional studies, case-control studies, and case reporting. However, the limitations of the current prediction model are revealed in most borderline domains, and these have not

yet been solved by existing research. In addition, traditional research methodologies have the disadvantage of requiring a high expenditure and long time for high-quality results. This is the basis for the utilization of artificial intelligence to replace or help existing methodologies in solving such problems.

2. Artificial intelligence

The popularity of the terms "deep learning," "machine learning," and "artificial intelligence" recently have registered a rapid growth worldwide. For instance, their Google trends show a great rise from 10 in 2013 to 100 in 2018. The definition of artificial intelligence provided in the Merriam-Webster dictionary is "the capability of a machine to imitate intelligent human behavior." Machine learning is a division of artificial intelligence and it is defined as "extract knowledge from large amounts of data" [11-14]. There are six major machine learning approaches. (1) The decision tree has internal nodes (the tests on independent variables), branches (the outcomes of the tests), and terminal nodes (the values of the dependent variable). (2) The naïve Bayesian uses Bayes' theorem for classification. (3) The random forest is a group of many decision trees with a majority vote on the dependent variable: in a random forest with 100 decision trees, 100 training sets are sampled with replacements, 100 decision trees are trained with the 100 training sets, and the 100 decision trees take a majority vote on the dependent variable. (4) The support vector machine finds a hyperplane, which is a space dividing data with the maximal space among different groups. (5) The artificial neural network is a group of neurons, which are input/output units linked through weights. Normally, it has 3-5 layers, that is, 1 input layer, 1-3 hidden layers, and 1 output layer. Here, the weights serve as numerical values for the effects of neurons in a previous layer on neurons in the next layer. And (6) deep learning is a special artificial neural network with more hidden layers than five [14]. The unprecedented popularity of the artificial neural network all over the world can be attributed to the great development of computing power in recent decades.

3. Aims of study

Traditional research considered a limited scope of predictors for maternal-fetal conditions and was based on hypothesis with an unrealistic assumption (the variables are acting under a certain distribution). In this context, emerging studies began to use artificial intelligence for a variety of purposes, e.g., classification and pattern recognition for financial and medical sectors [15-21]. They did not depend on unrealistic presumptions of ceteris paribus, and they could examine which factors were more important for the prediction of the dependent variable. The purpose of this study was to review recent advances on the application of artificial intelligence for the early diagnosis of various maternal-fetal conditions such as preterm birth and abnormal fetal growth.

4. Methods of study

Thirteen articles were chosen out of 278 articles in Google and PubMed, with the search words "artificial intelligence," "machine learning," "fetal growth," "maternal-fetal conditions," and "preterm birth." The following eligibility criteria were employed: (1) the intervention(s) of a new deep learning approach or several machine learning methods introduced above; (2) the outcome(s) of accuracy, the AUC, and/or the root mean squared error for the prediction of maternal-fetal conditions (i.e., preterm birth and abnormal fetal growth); (3) the year of publication being 2000 or later; and (4) the language of publication being English. The following summary indicators were used: artificial intelligence approach, sample size, data type, performance indicators, and important factors (independent variables).

Application of machine learning in early diagnosis of maternal-fetal conditions

1. Preterm birth

The performance measures of the artificial neural network and the random forest are reported to be similar to or better than those of traditional approaches such as logistic regression regarding the early diagnosis of preterm birth [17-21]. For instance, recent research employed data with 19,970 participants at the Duke University Medical Center from January 1, 1988 to June 1, 1997 [17-19]. The class (or the dependent variable) was preterm birth, and the attributes (or the independent variables) were 1,622 demographic, socioeconomic, obstetric variables including age, education, insurance, marital status, race, region, and religion. The artificial neural network (0.68) performed better than did logistic regression (0.66) and the decision tree (0.65) in terms of the AUC.

Other recent study made a rare comparison of popular machine learning approaches for the early diagnosis of preterm birth [20]. This study included 596 obstetric patients admitted in Anam Hospital (Seoul, Korea) from 27 March 2014 to 21 August 2018. Variable importance was used to find major predictors of preterm birth. The artificial neural network (0.9115) was similar to logistic regression (0.9180) and the random forest (0.8918) in terms of accuracy. The variable importance results of the artificial neural network put more focus on hypertension, diabetes mellitus, and prior cone biopsy, while their random forest counterparts placed more emphasis on cervical length, age, and prior preterm birth. The former outcomes agreed with those of previous research [22-33], requesting the change of focus from direct to indirect factors of preterm birth.

More recent research employed similar methods for examining preterm birth, gastroesophageal reflux disease, and periodontitis [21]. Previous research made independent hypotheses on a positive linkage between gastroesophageal reflux disease and periodontitis and on a positive relationship between periodontitis and preterm birth [34-40]. In a similar context, one would hypothesize a positive association between gastroesophageal reflux disease and preterm birth. However, no study has been conducted in this regard. The source of data for the study mentioned above was obtained from Anam Hospital in Seoul, Korea, with 731 obstetric patients admitted from January 5, 1995 to August 28, 2018. Variable importance was employed for finding major factors of preterm birth. The random forest (0.8681) was comparable to logistic regression (0.8736) in terms of accuracy. Based on the findings of this study, (1) the importance of gastroesophageal reflux disease (13th) is greater than that of periodontitis (22nd) for preventing preterm birth and (2) for the prevention of preterm birth, preventive approaches for hypertension, gastroesophageal reflux disease, and diabetes mellitus would be essential, together with effective body mass index management and appropriate progesterone and calcium-channel-blocker medications. Other notable attempts in this direction include the employment of machine learning for predicting early stillbirth, late stillbirth, and preterm birth [41] and the use of uterine electrohysterogram data [42,43] (see Table 1, summary of this Study, for these studies).

Finally, the employment of deep learning for predicting preterm birth is addressed here. Different kinds of deep learning have been utilized for different types of data, i.e., convolutional vs. recurrent network for image vs. sequence data. The convolutional neural network uses convolutional, rectified-linear-unit, pooling, and fully-connected layers to acquire the global context of input information [44]. In the recurrent neural network, current output depends on current input and the memory of the network in a "recurrent" pattern [45]. Recent research employed a recurrent neural network ensemble for the early diagnosis of extreme preterm birth (birth before the 28th week of gestational age) [46]. The data source of this study was electronic health records on 25,689 deliveries at the Vanderbilt University Medical Center. The recurrent neural network ensemble (0.965, 0.827) performed better than did logistical regression (0.819, 0.749) and the support vector machine (0.660, 0.728) in terms of sensitivity and the AUC, respectively. In a similar context, other recent research employed a 3-dimensional convolutional neural network for the early diagnosis of preterm birth [47]. Notably, a relevance score was assigned to each input voxel in a backpropagation algorithm (layer-wise relevance propagation). Data came from 157 magnetic resonance imaging (MRI) scans of newborns with 23-42 weeks of gestational age. The performance indicators of the convolutional neural network were very good in terms of accuracy (94%), the true positive rate (100%) and the true negative rate (86%).

2. Abnormal fetal growth

A recent study used machine learning and ultrasound measures for the prediction of estimated fetal weight [48]. Data of 24,910 obstetric patients and their newborns enrolled in the Magee Obstetric Medical and Infant database from the University of Pittsburgh Medical Center was evaluated. Estimated fetal weight was the dependent variable, and five independent variables were accordingly considered, i.e., gestational age at delivery, parity, 1-minute/5-minute Apgar scores, and newborn's gender. The random forest (96.2) was better than the decision tree (104.9) and linear regression (127.5) in terms of the root mean squared error. This is one of the first machine-learning studies with clinical and sonographic markers for the prediction of newborns' estimated fetal weight.

The aim of another recent study was to employ linear regression and ultrasound measures for the prediction of gestational weeks [49]. The source of data was 4,299 mother-newborn pairs enrolled in a multi-center study (International

Table 1. Summary of this study

Study	Method	Sample size	Data type	Performance	Important predictors
Goodwin et al. [18] (2001)	ANN[a] DT LR	19,970	Numeric	AUC: 0.65-0.68	For the prediction of preterm birth Age, race, region, religion, education, insurance, marriage
Lee and Ahn [20] (2019)	ANN[a] DT LR[a] NB RF[a] SVM[a]	596	Numeric	Accuracy: 0.89-0.92 AUC: 0.62-0.64	For the prediction of preterm birth - Body mass index - Hypertension - Diabetes mellitus - Prior cone biopsy - Parity - Cervical length age - Prior preterm birth - Myomas & adenomyosis[b]
Lee et al. [21] (2020)	ANN DT LR[a] NB RF[a] SVM	731	Numeric	Accuracy: 0.79-0.87 AUC: 0.54-0.76	For the prediction of preterm birth Top 01-10 - Body mass index delivery/pregestational - Age - Parity - Blood pressure predelivery systolic/diastolic - Twin - Education (below high school graduation) - Newborn sex - Prior preterm birth Top 11-17 - Medication history - Progesterone - Upper gastrointestinal tract symptom - Gastroesophageal reflux disease - Helicobacter pylori - Region (urban) - Medication history - Calcium-channel-blocker - Gestational diabetes mellitus[b]
Koivu and Sairanen [41] (2020)	ANN[a] DT LR	16,340,661	Numeric	Sensitivity: 0.22-0.24 AUC: 0.62-0.64	For the prediction of preterm birth Demographic - Age - Body mass index - Education - Marital status - Pre-pregnancy smoking - Previous terminations - Race - Special Nutritional program - Weight Obstetric - Diabetes pre-pregnancy/gestational - Hypertension pre-pregnancy/gestational - Hypertension eclampsia - Infertility medication/treatment - Parity - Previous cesarean section - Previous preterm birth - Assisted reproductive technology - Chlamydia - Gonorrhea - Hepatitis C - Syphilis

Table 1. Continued

Study	Method	Sample size	Data type	Performance	Important predictors
Fergus et al. [42] (2013)	DT LR SVM[a]	300	Electrohysterogram	Specificity: 0.86-1.00 AUC: 0.60-0.61	For the prediction of preterm birth Uterine electrical signals (root mean squares, peak frequency, median frequency, sample entropy)
Gao et al. [46] (2019)	LR RNN[a] SVM	25,689	Text (5,602,792 medical concepts)	Sensitivity: 0.66-0.97 AUC: 0.73-0.83	For the prediction of preterm birth Twin pregnancy, short cervical length, hypertensive disorder, systemic lupus erythematosus, hydroxychloroquine sulfate
Grigorescu et al. [47] (2019)	CNN	157	Magnetic resonance imaging	Accuracy: 0.94	For the prediction of preterm birth Increased cerebrospinal fluid and reduced cortical folding due to impaired brain growth
Naimi et al. [48] (2018)	DT LR RF[a]	24,910	Numeric	RMSE: 96.2-127.5	For the prediction of estimated fetal weight Gestational age at delivery, parity, 1-minute/5-minute Apgar scores, newborn's gender
Fung et al. [49] (2020)	LR	4,299	Numeric	Accuracy: 0.99	For the prediction of gestational weeks Abdominal circumference, femur length, head circumference
Signorini et al. [50] (2020)	DT LR NB RF[a] SVM	120	Numeric	Accuracy: 0.85-0.91 Sensitivity: 0.85-0.93 Specificity: 0.83-0.95	For the prediction of intrauterine growth restriction Maternal age, gestational age at cardiotocography, newborn's gender, newborn's weight, 1-minute/5-minute Apgar scores, delivery mode
Pini et al. [51] (2021)	SVM	262	Numeric	Accuracy: 0.93 Sensitivity: 0.93 Specificity: 0.84	For the prediction of intrauterine growth restriction Maternal age, gestational age at cardiotocography, newborn's gender, newborn's weight, 1-minute/5-minute Apgar scores, delivery mode
Lee et al. [52] (2021)	ANN LR[a] RF	3,159	Numeric	RMSE: 1.44-12.44	For the prediction of newborn's body mass index The first abdominal circumference and estimated fetal weight in the week 36 or later, gestational age at delivery, the first abdominal circumference during the week 21 - the week 35, maternal body mass index at delivery, maternal weight at delivery and the first biparietal diameter in the week 36 or later[b]
Sridar et al. [53] (2019)	CNN	4,074	Red-green-blue imaging	Accuracy: 0.97	For the prediction of fetal structures 14 fetal structures including abdomen, arm, blood vessels, cord insertion, face, femur, foot, genitals, hand, head, heart, kidney, leg and spine

ANN, artificial neural network; DT, decision tree; LR, linear regression (fetal growth)/logistic regression (preterm birth); AUC, area under the receiver operating characteristic curve; NB, naïve bayes; RF, random forest; SVM, support vector machine; RNN, recurrent neural network; CNN, (3-dimensional) convolutional neural network; RMSE, root mean squared error.
[a]Method with the best performance; [b]Predictors listed on the basis of the variable importance ranking of the random forest.

Fetal and Newborn Growth Consortium for the 21st Century Study). The number of gestational weeks was the dependent variable, and three independent variables were included accordingly, that is, abdominal circumference, femur length, and head circumference. The accuracy of linear regression was defined as the percentage of correct predictions within

1 week, and it was above 0.99. This is one of the earliest machine-learning studies with clinical and sonographic markers for the prediction of gestational weeks.

Two recent studies involved the combination of machine learning and ultrasound measures to predict intrauterine growth restriction [50,51]. Data were obtained from 120 and

262 mother-newborn pairs in Italy's obstetrics and gynecology clinics, respectively. The dependent variable was intrauterine growth restriction, and seven independent variables were considered accordingly, i.e., maternal age, gestational age at cardiotocography, newborn's gender, newborn's weight, 1-minute/5-minute Apgar scores, and delivery mode. In the first study, the random forest was the best model in terms of accuracy (0.91). In the second study, the support vector machine was the only model used, and its accuracy was slightly higher (0.93) than its random-forest counterpart in the first study with a smaller sample size. The ranges of sensitivity and specificity were 0.85-0.93 and 0.83-0.95, respectively, in these studies. These studies showed that machine learning can be an effective non-invasive tool, together with clinical and sonographic markers, to predict intrauterine growth restriction.

The most recent was the first machine-learning study for the prediction of newborn's body mass index from 64 ultrasound-maternal-delivery variables [52]. This study included 3,159 mother-newborn pairs enrolled in the Korean Society of Ultrasound in Obstetrics and Gynecology Research Group Study (a multi-center retrospective study). Variable importance was employed to find important predictors of newborn's body mass index among ultrasound-maternal-delivery information. For the prediction of newborn's body mass index, linear regression (2.0744) and the random forest (2.1610) performed better than did artificial neural networks with one, two, and three hidden layers (150.7100, 154.7198, and 152.5843, respectively) in terms of the mean squared error. Based on the findings of this study, (1) the week 36 or later is the best period for taking the ultrasound measures and (2) abdominal circumference and estimated fetal weight are the best predictors of newborns' body mass index, together with gestational age and maternal body mass index at delivery.

Finally, one recent study applied the convolutional neural network (Alex) for ultrasound image data to predict each of fetal structures [53]. The fetal structures were related to abnormal fetal growth; hence, this study could be included in this review. Data was obtained from 4,074 ultrasound images (red-green-blue format) in the University of Sydney Nepean Center for Perinatal Care. The dependent variable was binary (no vs. yes) and 14 fetal structures were considered in the study (abdomen, arm, blood vessels, cord insertion, face, femur/humerus, foot, genitals, hand, head, heart, kidney, leg, and spine). The accuracy of Alex varied from 0.94 (face)

to 0.99 (cord insertion, femur, genitals, head). Its mean accuracy was 0.97. The results of this study are expected to aid in the derivation of new indicators of fetal structures (and abnormal fetal growth in a broad context).

Summary of this study

The summary of our study (review) is shown in Table 1. The table has five summary indicators, including artificial intelligence approach, sample size, data type, performance indicators, and important factors (independent variables). Based on the results of this study, different machine learning methods would be appropriate for different kinds of data regarding the early diagnosis of maternal-fetal conditions: logistic regression (numeric), the random forest (numeric), the support vector machine (electrohysterogram), the artificial neural network (numeric), the convolutional neural network (image), and the recurrent neural network (text). The performance indicators registered variations with the following ranges: 0.79-0.99 (accuracy); 0.54-0.83 (the AUC); 0.22-0.97 (sensitivity); and 0.86-1.00 (specificity). Artificial intelligence is a data-driven method; hence, further research needs to be performed for more external validation.

Recent expansion of artificial intelligence in maternal-fetal medicine

Recent research into predicting postpartum hemorrhage is very interesting. Venkatesh et al. [54] defined postpartum hemorrhage as a minimum of 1,000 mL of blood loss and considered 55 risk factors. Among the 152,279 births, 7,279 (4.8% and 95% confidence interval [CI], 4.7-4.9) suffered postpartum hemorrhage. The extreme gradient boosting model had the highest performance in predicting postpartum hemorrhage (C statistic, 0.93; 95% CI, 0.92-0.93) followed by random forests (C statistic, 0.92; 95% CI, 0.91-0.92). Lasso regression models (C statistics, 0.87; 95% CI, 0.86-0.88) and logistic regression (C statistics, 0.87; 95% CI, 0.86-0.87) were shown to have low but good predictive capabilities. The above results retained their predictive power regardless of time and place. Although decision curve analysis provided excellent net benefits for all models when the threshold was between 0% and 80%, the extreme gradient

boosting model showed the greatest net benefits [54].

There has also been a recent study for applying machine learning models to clinical and laboratory features of women with intrahepatic cholestasis of pregnancy to create algorithms for identifying patients without bile acid measurements. The study included 336 pregnant women suffering from rash-free pruritus during the second and third trimesters, using demographic, obstetrical, clinical and laboratory data to predict bile acid level ≥10 mol/L. As a result, the XGB Classifier model was the most accurate (AUC, 0.9), followed by the K-neighbors model (AUC, 0.86), and the support vector classification model (AUC, 0.82). The model with the lowest predictive capability was logistic regression (AUC, 0.72). The sensitivity of the XGB model was 86%, and its specificity was 75% [55].

Postpartum complications are difficult to predict and, if they actually occur, these place a heavy medical and socioeconomic burden on mothers and their families. From January 2009 to October 2015, machine learning analysis of predictions of postpartum complications using data from all live births (n=422,509) in Queensland, Australia showed hypertension disorder (AUC, 0.879; 95% CI, 0.846-0.912) and postpartum wound infections (AUC, 0.856; 95% CI, 0.838-0.873). The authors argued that usage of such methods could play an important role in determining the risk of postpartum complications in advance [56].

Shoulder dystocia must be the least accessible clinical situation for obstetricians. A recent study presents a good model for predicting shoulder dystocia using machine learning. Following the application of the inclusion and exclusion criteria, the derivation cohort included 686 singleton vaginal deliveries, of which 131 cases had shoulder dystocia. In the validation cohort, there were 2,584 deliveries, of which 31 cases had shoulder dystocia. In the derivation cohort, machine learning models provided significantly better predictive power than current clinical guidance based on fetal weight and gestational diabetes (AUC, 0.793±0.041 vs. 0.745±0.044; P=1e-16). Furthermore, the model showed better predictive power than adjusted estimated fetal weight among 273 female subgroups, with a fetus of EFW≥4,000 g in the validation cohort (AUC, 0.775 vs. 0.548, P=0.0002) [57].

Poor neonatal prognosis caused by labor is also a very important issue for obstetricians. The machine learning model for 1,346 cases (1.8%) with severe adverse neonatal outcomes (SANO) among 73,868 single-child pregnancies showed an AUC of 0.761 (95% CI, 0.748-0.774). A third (33.5%; n=24,721) of cohorts were assigned to high-risk groups for SANOs, and SANOs in these high-risk groups were found to occur more frequently than in low-risk groups (odds ratio, 5.3; 95% CI, 4.7-6.0; high risk vs. low risk) [58].

The applicability of machine learning models also appears to be high in predicting successful vaginal delivery. Using the model, we calculated an AUC of 0.817 (95% CI, 0.811-0.823). Models containing cervical exam data showed that their AUC increased to 0.917 (95% CI, 0.913-0.921) at the end of stage 1 of labor [59]. The trial of labor after cesarean (TOLAC) selection is very important for the success of the virtual birth after cesarean delivery. Machine learning algorithm for TOLAC analyses with 989 cases demonstrated more accurate predictive power than conventional maternal–fetal medicine units calculator (0.351 vs. 0.325, P=0.0002) [60].

Continuous fetal heart rate monitoring is widely used to monitor fetal heart rate and uterine contraction as a means of detecting fetal distress. Fetal heart rate monitoring leads to a decrease in neonatal seizures; however, it is associated with an increase in C-sections and instrumental vaginal birth rates. A recent study has shown that artificial intelligence algorithms based on fully convolutional networks can predict non reassuring fetal heart rate patterns with an accuracy of AUC 0.892 [61]. There are also reports of successful abnormal fetal heart rate pattern and fetal electrocardiogram prediction using deep learning [62].

Artificial intelligence shows many possibilities in maternal-fetal imaging techniques. In particular, it is expected to play its role in screening and diagnosis of abnormal findings. In particular, the application of ultrasound artificial intelligence may be useful in probe guidance, fetal biometrics plane finder, anomaly scan completeness, anomaly highlighting, and cyst classification [63]. One example is the effort to detect changes in fetal heart structures or facial expressions using artificial intelligence technologies [64,65].

Efforts to apply artificial intelligence algorithms to prenatal genetic diagnoses have long been made and have shown good results [66]. Recently, there have been many advances in research, including predicting pregnancy rates with fetal MRI using deep learning, discriminating maternal and fetal anatomy with MRI and ultrasound, obtaining diagnostic help using photographs of the placenta, and detecting fetal heart abnormalities [67-70].

Finally, there is an emerging body of literature on the ap-

plication of machine learning with gestational diabetes or hypertension as the dependent variable [71-74]. The independent variables of machine-learning studies on gestational diabetes [71,72] include age, blood pressure, body mass index, diabetes pedigree function, education (elementary, junior high school, senior high school, and university), fasting plasma glucose, gestational diabetes history, pregnancies (number), serum insulin, skin-fold thickness, and smoking. The AUCs of these models vary among naïve Bayes (0.892), logistic regression (0.919), random forest (0.939), adaptive boosting (0.971), gradient boosting (0.973), and XGBoosting (0.993). Likewise, the independent variables of machine-learning research on gestational hypertension [73,74] include age, ethnicity, pre-pregnancy and pregnancy height, weight, and body mass index, chronic hypertension, chemical and hemodynamic factors, abortion, gravidity, parity, cesarean delivery, and gestational age. The accuracy values of XG-Boosting [73] and the support vector machine [74] belong to the range of 0.81-0.93. These studies demonstrate that machine learning can be an effective tool for the diagnosis of gestational diabetes or hypertension.

Current limitations and future perspectives

Existing research on artificial intelligence for the early diagnosis of maternal-fetal conditions has several limitations. First, a cross-sectional design is still common in these studies, and data improvement with a longitudinal design would hence strengthen the performance of artificial intelligence in this area. Second, these studies had not analyzed the possible mediating effects among predictors for the early diagnosis of maternal-fetal conditions. Third, usage of big data (e.g., national health insurance claims data) would be a good strategy for future studies. Fourth, whether binary categories (no, yes) are common for maternal-fetal conditions and analyzing various factors of maternal-fetal conditions based on more refined categories will be an interesting issue for future research. Fifth, the literature is lacking and more examination is needed on possible pathways between gastroesophageal reflux disease and maternal-fetal conditions [21]. Sixth, uniting different kinds of deep learning methods for different types of maternal-fetal data would bring forth more profound clinical implications. Seventh, little research has been conducted

on ethical issues for the application of artificial intelligence in maternal-fetal conditions. It is not surprising given that the application of artificial intelligence in this area has started to expand very recently. Nevertheless, if the current trend continues, this situation is expected to change, and professionals in maternal-fetal conditions would have to devote more attention to these issues and discuss them thoroughly. Finally, no basic or translational research has been performed on the basis of artificial intelligence regarding maternal-fetal conditions.

This study reviewed recent advances on the application of artificial intelligence for the early diagnosis of various maternal-fetal conditions such as preterm birth and abnormal fetal growth. Based on the findings of this study, it is observed that artificial intelligence provides a potential for an effective non-invasive decision support system for the early diagnosis of maternal-fetal conditions.

Conclusion

As in other medical or academic fields, the application of artificial intelligence technologies in maternal-fetal medicine has seen an exponential increase in the number recently and it is getting increasingly utilized in diagnostic and therapeutic medical decisions. Although classical statistical techniques so far have made significant progress in disease causes, pathophysiology, diagnosis, treatment, and prognosis prediction, artificial intelligence technologies appear to be the basis and the next standard methodology for the huge wave of data-based Fourth Industrial Revolution [75]. Since maternal-fetal medicine deals with two lives (mother and fetus), stricter ethical standards will be needed for the application of artificial intelligence technologies; hence, more attention should be paid to the amount, quality, and accuracy of data and research inputs in this direction.

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Obstetrics & Gynecology Science

Ki Hoon Ahn, et al. Artificial intelligence in obstetrics

Ethical approval

Not applicable.

Patient consent

Not applicable.

References

1. National Health Service (NHS). Annual report and accounts 2017/18. Leeds: NHS; 2018.
2. Warrick PA, Hamilton EF, Precup D, Kearney RE. Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. IEEE Trans Biomed Eng 2010;57:771-9.
3. Aung MT, Yu Y, Ferguson KK, Cantonwine DE, Zeng L, McElrath TF, et al. Prediction and associations of preterm birth and its subtypes with eicosanoid enzymatic pathways and inflammatory markers. Sci Rep 2019;9:17049.
4. Burgos-Artizzu XP, Baños N, Coronado-Gutiérrez D, Ponce J, Valenzuela-Alcaraz B, Moreno-Espinosa AL, et al. Mid-trimester prediction of spontaneous preterm birth with automated cervical quantitative ultrasound texture analysis and cervical length: a prospective study. Sci Rep 2021;11:7469.
5. Al-Rubaie ZTA, 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;20:23.
6. Antwi E, Amoakoh-Coleman M, Vieira DL, Madhavaram S, Koram KA, Grobbee DE, et al. Systematic review of prediction models for gestational hypertension and pre-eclampsia. PLoS One 2020;15:e0230955.
7. Leite DFB, Cecatti JG. Fetal growth restriction prediction: how to move beyond. Scientific World Journal 2019;2019:1519048.
8. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Health Inf Sci Syst 2020;8:14.
9. Åmark H, Westgren M, Persson M. Prediction of stillbirth in women with overweight or obesity-a register-based cohort study. PLoS One 2018;13:e0206940.
10. Yerlikaya G, Akolekar R, McPherson K, Syngelaki A, Nicolaides KH. Prediction of stillbirth from maternal demographic and pregnancy characteristics. Ultrasound Obstet Gynecol 2016;48:607-12.
11. Berry MJA, Linoff GS. Data mining techniques. 2nd ed. Indianapolis (IN): Wiley; 2004.
12. Han J, Micheline K. Data Mining: Concepts and Techniques. 2nd ed. San Francisco (CA): Elsevier; 2006.
13. Han J, Micheline K, Pei J. Data Mining: Concepts and Techniques. 3rd ed. San Francisco (CA): Elsevier; 2012.
14. Tan PN, Steinbach M, Karpatne A, Kumar V. Introduction to data mining. 2nd ed. London (UK): Pearson; 2018.
15. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: a survey. Heliyon 2018;4:e00938.
16. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform 2004;107(Pt 1):736-40.
17. Goodwin LK, Maher S. Data mining for preterm birth prediction. Como 2000;1:46-51.
18. Goodwin LK, Iannacchione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data mining methods find demographic predictors of preterm birth. Nurs Res 2001;50:340-5.
19. Goodwin LK, Iannacchione MA. Data mining methods for improving birth outcomes prediction. Outcomes Manag 2002;6:80-5.
20. Lee KS, Ahn KH. Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants. J Korean Med Sci 2019;34:e128.
21. Lee KS, Song IS, Kim ES, Ahn KH. Determinants of spontaneous preterm labor and birth including gastroesophageal reflux disease and periodontitis. J Korean Med Sci 2020;35:e105.
22. Parker MG, Ouyang F, Pearson C, Gillman MW, Belfort MB, Hong X, et al. Prepregnancy body mass index and risk of preterm birth: association heterogeneity by preterm subgroups. BMC Pregnancy Childbirth

2014;14:153.

23. Heude B, Thiébaugeorges O, Goua V, Forhan A, Kaminski M, Foliguet B, et al. Pre-pregnancy body mass index and weight gain during pregnancy: relations with gestational diabetes and hypertension, and birth outcomes. Matern Child Health J 2012;16:355-63.

24. Shin D, Song WO. Prepregnancy body mass index is an independent risk factor for gestational hypertension, gestational diabetes, preterm labor, and small- and large-for-gestational-age infants. J Matern Fetal Neonatal Med 2015;28:1679-86.

25. Sibai BM, Caritis SN, Hauth JC, MacPherson C, VanDorsten JP, Klebanoff M, et al. Preterm delivery in women with pregestational diabetes mellitus or chronic hypertension relative to women with uncomplicated pregnancies. The National institute of Child health and Human Development Maternal- Fetal Medicine Units Network. Am J Obstet Gynecol 2000;183:1520-4.

26. Hedderson MM, Ferrara A, Sacks DA. Gestational diabetes mellitus and lesser degrees of pregnancy hyperglycemia: association with increased risk of spontaneous preterm birth. Obstet Gynecol 2003;102:850-6.

27. Zhang J, Villar J, Sun W, Merialdi M, Abdel-Aleem H, Mathai M, et al. Blood pressure dynamics during pregnancy and spontaneous preterm birth. Am J Obstet Gynecol 2007;197:162.e1-6.

28. O'Hara S, Zelesco M, Sun Z. Cervical length for predicting preterm birth and a comparison of ultrasonic measurement techniques. Australas J Ultrasound Med 2013;16:124-34.

29. Society for Maternal-Fetal Medicine (SMFM), McIntosh J, Feltovich H, Berghella V, Manuck T. The role of routine cervical length screening in selected high- and low-risk women for preterm birth prevention. Am J Obstet Gynecol 2016;215:B2-7.

30. Berghella V, Pereira L, Gariepy A, Simonazzi G. Prior cone biopsy: prediction of preterm birth by cervical ultrasound. Am J Obstet Gynecol 2004;191:1393-7.

31. Bevis KS, Biggio JR. Cervical conization and the risk of preterm delivery. Am J Obstet Gynecol 2011;205:19-27.

32. Pinborg A, Ortoft G, Loft A, Rasmussen SC, Ingerslev HJ. Cervical conization doubles the risk of preterm and very preterm birth in assisted reproductive technology twin pregnancies. Hum Reprod 2015;30:197-204.

33. Cho SH, Park KH, Jung EY, Joo JK, Jang JA, Yoo HN. Maternal characteristics, short mid-trimester cervical length, and preterm delivery. J Korean Med Sci 2017;32:488-94.

34. Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Borgnakke WS, et al. Update on prevalence of periodontitis in adults in the united states: NHANES 2009 to 2012. J Periodontol 2015;86:611-22.

35. Puertas A, Magan-Fernandez A, Blanc V, Revelles L, O'Valle F, Pozo E, et al. Association of periodontitis with preterm birth and low birth weight: a comprehensive review. J Matern Fetal Neonatal Med 2018;31:597-602.

36. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R; Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101:1900-20; quiz 1943.

37. Patrick L. Gastroesophageal reflux disease (GERD): a review of conventional and alternative treatments. Altern Med Rev 2011;16:116-33.

38. Vinesh E, Masthan K, Kumar MS, Jeyapriya SM, Babu A, Thinakaran M. A clinicopathologic study of oral changes in gastroesophageal reflux disease, gastritis, and ulcerative colitis. J Contemp Dent Pract 2016;17:943-7.

39. Deppe H, Mücke T, Wagenpfeil S, Kesting M, Rozej A, Bajbouj M, et al. Erosive esophageal reflux vs. non erosive esophageal reflux: oral findings in 71 patients. BMC Oral Health 2015;15:84.

40. Ali RA, Egan LJ. Gastroesophageal reflux disease in pregnancy. Best Pract Res Clin Gastroenterol 2007;21:793-806.

41. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Health Inf Sci Syst 2020;8:14.

42. 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:e77154.

43. Sadi-Ahmed N, Kacha B, Taleb H, Kedir-Talha M. Relevant features selection for automatic prediction of preterm deliveries from pregnancy ElectroHysterograhic (EHG) records. J Med Syst 2017;41:204.

44. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Siem Reap 2018;1:1097-105.

45. Lee KS, Park KW. Social determinants of the association

Obstetrics & Gynecology Science

Ki Hoon Ahn, et al. Artificial intelligence in obstetrics

among cerebrovascular disease, hearing loss and cognitive impairment in a middle-aged or older population: recurrent neural network analysis of the Korean Longitudinal Study of Aging (2014-2016). Geriatr Gerontol Int 2019;19:711-6.

46. Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. J Biomed Inform 2019;100:103334.

47. Grigorescu I, Cordero-Grande L, Edwards AD, Hajnal J, Modat M, Deprez M. Interpretable convolutional neural networks for preterm birth classification [Internet]. arXiv.org; c2019 [cited 2021 Jun 15]. Available from: https://arxiv.org/abs/1910.00071.

48. Naimi AI, Platt RW, Larkin JC. Machine learning for fetal growth prediction. Epidemiology 2018;29:290-8.

49. Fung R, Villar J, Dashti A, Ismail LC, Staines-Urias E, Ohuma EO, et al. Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. Lancet Digit Health 2020;2:e368-75.

50. 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;185:105015.

51. Pini N, Lucchini M, Esposito G, Tagliaferri S, Campanile M, Magenes G, et al. A machine learning approach to monitor the emergence of late intrauterine growth restriction. Front Artif Intell 2021;4:622616.

52. Lee KS, Kim HY, Lee SJ, Kwon SO, Na S, Hwang HS, et al. Prediction of newborn's body mass index using nationwide multicenter ultrasound data: a machine-learning study. BMC Pregnancy Childbirth 2021;21:172.

53. Sridar P, Kumar A, Quinton A, Nanan R, Kim J, Krishnakumar R. Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks. Ultrasound Med Biol 2019;45:1259-73.

54. Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JSA, et al. Machine learning and statistical models to predict postpartum hemorrhage. Obstet Gynecol 2020;135:935-44.

55. Asali A, Ravid D, Shalev H, David L, Yogev E, Yogev SS, et al. Intrahepatic cholestasis of pregnancy: machine-learning algorithm to predict elevated bile acid based on clinical and laboratory data. Arch Gynecol Obstet 2021;304:641-7.

56. Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019;126:702-9.

57. 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;56:588-96.

58. Guedalia J, Sompolinsky Y, Novoselsky Persky M, Cohen SM, Kabiri D, Yagel S, et al. Prediction of severe adverse neonatal outcomes at the second stage of labour using machine learning: a retrospective cohort study. BJOG 2021;128:1824-32.

59. Guedalia J, Lipschuetz M, Novoselsky-Persky M, Cohen SM, Rottenstreich A, Levin G, et al. Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries. Am J Obstet Gynecol 2020;223:437.e1-437.e15.

60. Meyer R, Hendin N, Zamir M, Mor N, Levin G, Sivan E, et al. Implementation of machine learning models for the prediction of vaginal birth after cesarean delivery. J Matern Fetal Neonatal Med 2020 Oct 25 [Epub]. https://doi.org/10.1080/14767058.2020.1837769.

61. Liu LC, Tsai YH, Chou YC, Jheng YC, Lin CK, Lyu NY, et al. Concordance analysis of intrapartum cardiotocography between physicians and artificial intelligence-based technique using modified one-dimensional fully convolutional networks. J Chin Med Assoc 2021;84:158-64.

62. Zhong W, Liao L, Guo X, Wang G. A deep learning approach for fetal QRS complex detection. Physiol Meas 2018;39:045004.

63. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound Obstet Gynecol 2020;56:498-505.

64. Garcia-Canadilla P, Sanchez-Martinez S, Crispi F, Bijnens B. Machine learning in fetal cardiology: what to expect. Fetal Diagn Ther 2020;47:363-72.

65. Miyagi Y, Hata T, Bouno S, Koyanagi A, Miyake T. Recognition of facial expression of fetuses by artificial intelligence (AI). J Perinat Med 2021;49:596-603.

66. Beksaç MS, Durak B, Ozkan O, Cakar AN, Balci S, Karakaş U, et al. An artificial intelligent diagnostic sys-

tem with neural networks to determine genetical disorders and fetal health by using maternal serum markers. Eur J Obstet Gynecol Reprod Biol 1995;59:131-6.

67. Kojita Y, Matsuo H, Kanda T, Nishio M, Sofue K, Nogami M, et al. Deep learning model for predicting gestational age after the first trimester using fetal MRI. Eur Radiol 2021;31:3775-82.

68. Torrents-Barrena J, Monill N, Piella G, Gratacós E, Eixarch E, Ceresa M, et al. Assessment of radiomics and deep learning for the segmentation of fetal and maternal anatomy in magnetic resonance imaging and ultrasound. Acad Radiol 2021;28:173-88.

69. Looney P, Stevenson GN, Nicolaides KH, Plasencia W, Molloholli M, Natsis S, et al. Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning. JCI insight 2018;3:e120178.

70. Morris SA, Lopez KN. Deep learning for detecting congenital heart disease in the fetus. Nat Med 2021;27:764-5.

71. Gnanadass I. Prediction of gestational diabetes by machine learning algorithms. IEEE Potentials 2020;39:32-7.

72. Wu YT, Zhang CJ, Mol BW, Kawai A, Li C, Chen L, et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. J Clin Endocrinol Metab 2021;106:e1191-205.

73. Hoffman MK, Ma N, Roberts A. A machine learning algorithm for predicting maternal readmission for hypertensive disorders of pregnancy. Am J Obstet Gynecol MFM 2021;3:100250.

74. Yang L, Sun G, Wang A, Jiang H, Zhang S, Yang Y, et al. Predictive models of hypertensive disorders in pregnancy based on support vector machine algorithm. Technol Health Care 2020;28(S1):181-6.

75. Akazawa M, Hashimoto K, Noda K, Yoshida K. The application of machine learning for predicting recurrence in patients with early-stage endometrial cancer: a pilot study. Obstet Gynecol Sci 2021;64:266-73.