Machine learning and disease prediction in obstetrics

Zara Arain [a], Stamatina Iliodromiti [b], Gregory Slabaugh [c], Anna L. David [d], Tina T. Chowdhury [a],*

[a] *Centre for Bioengineering, School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, London, E1 4NS, UK*
[b] *Women's Health Research Unit, Wolfson Institute of Population Health, Queen Mary University of London, 58 Turner Street, London, E1 2AB, UK*
[c] *Digital Environment Research Institute, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 1HH, UK*
[d] *Elizabeth Garrett Anderson Institute for Women's Health, University College London, Medical School Building, Huntley Street, London, WC1E 6AU, UK*

ARTICLE INFO

ABSTRACT

Machine learning technologies and translation of artificial intelligence tools to enhance the patient experience are changing obstetric and maternity care. An increasing number of predictive tools have been developed with data sourced from electronic health records, diagnostic imaging and digital devices. In this review, we explore the latest tools of machine learning, the algorithms to establish prediction models and the challenges to assess fetal well-being, predict and diagnose obstetric diseases such as gestational diabetes, pre-eclampsia, preterm birth and fetal growth restriction. We discuss the rapid growth of machine learning approaches and intelligent tools for automated diagnostic imaging of fetal anomalies and to asses fetoplacental and cervix function using ultrasound and magnetic resonance imaging. In prenatal diagnosis, we discuss intelligent tools for magnetic resonance imaging sequencing of the fetus, placenta and cervix to reduce the risk of preterm birth. Finally, the use of machine learning to improve safety standards in intrapartum care and early detection of complications will be discussed. The demand for technologies to enhance diagnosis and treatment in obstetrics and maternity should improve frameworks for patient safety and enhance clinical practice.

1. Introduction

The digitisation of health records along with advancements in biomedical imaging and medical devices has led to an increase in clinical, biological and imaging data. Within these vast data sets of growing complexity and scale, lies the opportunity to understand challenging disease processes in obstetrics and maternity care where timely interventions can change the outcome for both mothers and babies. New approaches in computer science and statistics are required to identify actionable insights within these clinical conditions (de Marvao et al., 2020). Machine learning is a branch of artificial intelligence that uses computer algorithms to identify patterns within large raw datasets, acquire knowledge and apply this to different tasks (Ahn and Lee, 2022; Dhombres, 2022). A single machine learning model could analyse more data than a clinician would encounter over the duration of the individual's career. The multi-disciplinary field of computer science, engineering and maternal-fetal medicine is increasingly enabling researchers to apply machine learning tools that are transforming obstetrics and maternity care.

1.1. Differences between machine learning and deep learning platforms

Artificial intelligence refers to computer systems which perform tasks that typically require human intelligence. These systems mimic human behaviour and can be programmed to complete automated processes (Fig. 1). One example of artificial intelligence which is prevalent in maternity care is the Dawes Redman computerised Cardiotocograph (CTG) analysis system, that can recognise and diagnose pathological features of electronic fetal heart rate traces before labour. The numerical CTG analysis was developed from a database of over 73,500 CTGs in the largest study of its kind (Pardey et al., 2002). This has subsequently been developed into an automated CTG that is used to determine fetal wellbeing and hypoxia in the antenatal period (Alfirevic et al., 2017). This form of artificial intelligence follows a sequence of programmed commands to formulate an analysis and diagnosis. In contrast, machine learning enables systems to learn from experience and improve performance without specific programming. Compared to traditional statistical methods which classify or regress data, machine learning algorithms can handle multiple dimensional datasets with large numbers of variables (Averbuch et al., 2022). As they do not require interactions to be pre-specified, they are able to detect novel

---

* Corresponding author.
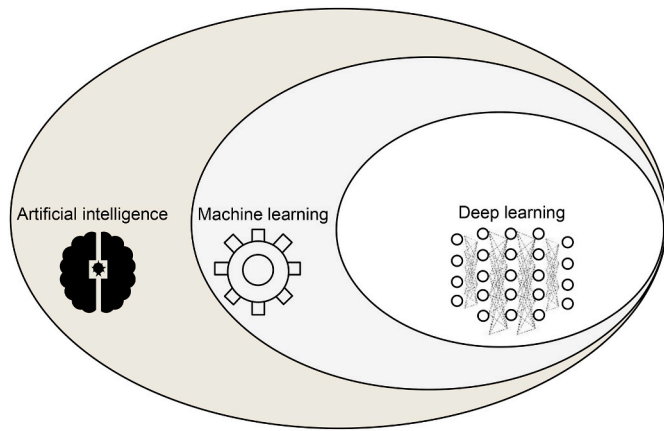 *E-mail address:* t.t.chowdhury@qmul.ac.uk (T.T. Chowdhury).

Fig. 1. Schematic illustrating the relationship between machine learning tools. Artificial intelligence is the programming of machines to mimic human behaviour. Machine learning is a subset of artificial intelligence and refers to computer algorithms which can learn from experience and improve performance without specific programming. Deep learning is a subset of machine learning, uses multi-layered neural networks to progressively learn more complex relationships as data passes through each layer.

relationships. These models are also dynamic and can be continuously updated with training data to reflect temporal changes.

Machine learning algorithms are broadly divided into supervised and unsupervised categories (Fig. 2). Supervised algorithms use labelled data to learn interactions between predictor variables as input and target variables as outcomes. For example in obstetrics, the outputs could be 'preterm birth' and 'term birth." The model adjusts the weighting of predictor variables as more data becomes available and inputted to the system. This approach trains the model to predict outcomes such as preterm birth in unseen datasets. Unsupervised algorithms learn interactions within unlabelled datasets. These systems identify clusters and associations between variables which may not have been evident to the clinician (Li et al., 2021; Averbuch et al., 2022). This can lead to the discovery of new patterns such as a disease mechanism. Additionally, deep learning progressively learns more complex relationships as data passes through each layer within a multilayer neural network and are useful in analysing large data sets such as electronic health records. Convolutional neural networks (CNN) are used more specifically in medical image recognition and can abstract spatial features from input data (de Marvao et al., 2020).

2. Aims of study

The present study reviewed the latest advances on using machine learning tools for use in obstetrics, maternal-fetal medicine and prenatal diagnosis. This includes 1. Risk stratification models for the early diagnosis of gestational diabetes and pre-eclampsia 2. diagnostic imaging tools which automate detection of anatomical structures 3. ensemble models to predict preterm birth, image segmentation models for 3D image reconstruction in fetal MRI 4. antepartum and intrapartum
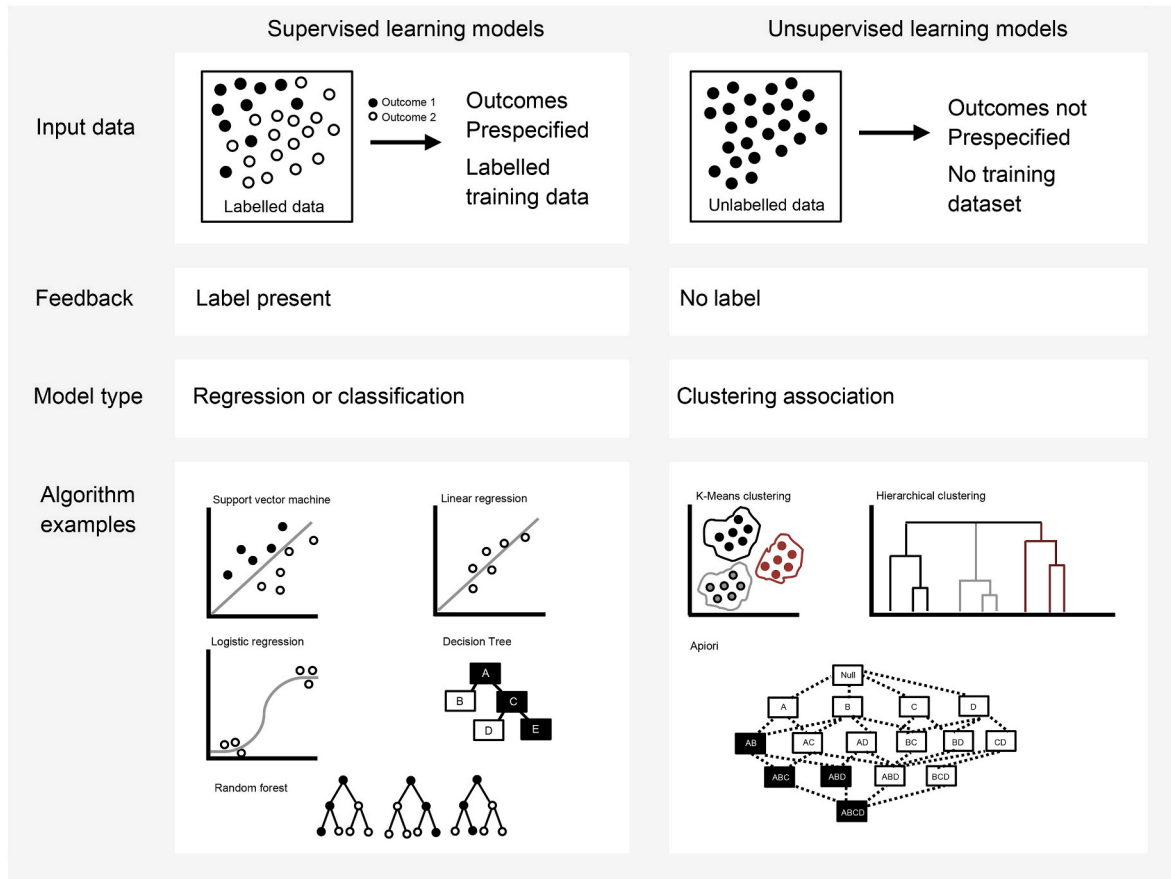

Fig. 2. Differences between supervised and unsupervised machine learning algorithms. Supervised algorithms use labelled data to learn interactions between variables and outcomes using regression and classification models designed to predict outcomes. Some examples of supervised machine learning models include decision tree, random forest, linear regression, logistic regression and supervised vector machine. Unsupervised algorithms learn novel interactions within unlabelled datasets and the models are designed to analyse datasets by identifying clusters and association between variables. Examples of unsupervised machine learning models include K means, hierarchical and apiori.

CTG classification models 5. fetal ECG classification models and 6. decision-assist tools for intrapartum care.

3. Methods

In devising our search strategy, we incorporated the principles from the guidelines for systematic reviews and meta-analysis (PRIMSA), but specific protocol registration was not performed (Page et al., 2021). We searched across Web of Science and PubMed databases using a combination of the following search terms: machine learning, deep learning, obstetrics, maternal medicine, gestational diabetes, pre-eclampsia, preterm birth, fetal growth restriction, prenatal ultrasound, ultrasonography, fetal MRI, CTG classification, intrapartum care, fetal ECG, tocography, electrohysterography and fetal movements monitoring. A total of 60 peer reviewed papers dating from January 1, 2018 onwards were included. We excluded case reports, conference proceedings and opinion pieces. References from review papers were examined manually to identify any additional work. Studies using all machine learning models were included to demonstrate the breadth of current work. Quality assessments for risk of bias were not performed, although evidence quality was considered in the descriptive analysis of included works. Whilst is it not practical to list all the papers that fall in the scope of the selection criteria, we have included representative papers for the purpose of this review and the search list as supplementary information. A quantitative analysis of results did not appear appropriate for this type of scoping review, and this was established prior to commencing database searches.

4. Results

4.1. Risk stratification tools for disease prediction in maternal-fetal medicine

Early screening of conditions in pregnancy enable early intervention strategies that could reduce the risk of disease burden (Bertini et al., 2022). For many conditions, there is a lack of consensus in approaches for antenatal care screening world-wide (Bulletins-Obstetrics, 2018). Current scoring systems are based on conventional methods to classify and regress data using statistical models. Many of these assessment tools have had limited updates over time to account for temporal changes in the population and up to date research (Mateen et al., 2020). In recent years, approaches to stratify patient risk using machine learning algorithms has been explored using electronic health records and more recently data from digital recording devices and are summarised in Table 1.

4.1.1. Gestational diabetes

Gestational diabetes is commonly diagnosed in the second and third trimester of pregnancy with earlier screening offered to women with selected risk factors. Predicting gestational diabetes in the first trimester could allow lifestyle intervention strategies to be implemented at a time of greater impact (Mateen et al., 2020). Table 1 summarises the key studies with values for performance indicators. A recent meta-analysis compared 30 machine learning algorithms to predict gestational diabetes using electronic health records and other datasets (X. Zhang et al., 2022). Of these, 16 models for first trimester gestational diabetes screening performed well in predicting the disease compared to current clinical assessment tools from the National Institute of Health (area under receive operating curve, 0.86 and 0.67, respectively, Table 1)

Table 1
Machine learning model with performance data to predict gestational diabetes and pre-eclampsia.

Study	Machine learning model	Prediction	Performance	Key findings
Early prediction of gestational diabetes				
Artzi et al., (2020)	Gradient boosted model, 2355 features. Electronic data for 588,622 pregnancies, 0–20 weeks gestational age	Early prediction (<20 weeks gestational age) of gestational diabetes	auROC = 0.85 Simpler model based on 9 feature selection auROC = 0.80	Large, granular dataset, model out-performed current baseline risk scores (auROC = 0.68), retrospective data, no external validation, model bias not assessed
Araya et al., 2021	Primary component analysis, 29 features. Electronic data for 39 pregnancies, 12–28 weeks gestational age	Maternal thyroid profiles classification with gestational diabetes	Thyroid profiles separated distribution of patients	Differences in gestational diabetes and normal glucose tolerance. Small dataset, no external validation, limited performance assessment
Yang et al., (2022)	3 models: multilinear regression, random forest and XGBoost. 14 features. Data from EHR and bluetooth monitoring devices for 1857 pregnancies	Identify patients with gestational diabetes at risk of high blood glucose levels	XGBoost MSE 0.020 (0.020–0.021) R2 0.519 (0.505–0.530) MAE 0.108 (0.107–0.110)	Multicentre design, external validation performed. comprehensive performance assessment. Improved accuracy required prior to transition to clinical practice
Early prediction of pre-eclampsia				
Marić et al., (2020)	2 models: elastic net, gradient boosting algorithm. 67 features. Electronic data for 16370 births	Develop an early prediction model for pre-eclampsia	Elastic net Prediction of early onset pre-eclampsia auROC 0.89 (0.84–0.95) TPR 72.3%, FPR 8.8%	New insights, appropriate feature selection. Retrospective, single centre study, no external validation, model bias not assessed
Lafuente-Ganuza et al., 2020	Decision tree model. Electronic data for 309 pregnancies	Identify and validate cut off values screening tests for early onset pre-eclampsia	Combined sFlt-1/PlGF ratio >45 and NT-proBNP value > 174: PPV 86% (95% CI: 79.2–92.6)	Suitable controls, new insights provided, cost-effective clinical application risk of model bias not assessed
Zhang et al., (2022)	Light gradient boosting machine, 43 features. Electronic data for 248 pregnancies	Correlation between severe pre-eclampsia and blood data characteristics	Binary classification using AST, direct bilirubin and APTT: sens. 88.37%, spec. 77.27%, AUC 89.74% PPV 65.96%.	Clear inclusion and exclusion criteria for participants. Risk of bias (selection and model) not assessed. Single centre no external validation

All studies sourced data from electronic health records, paper clinical records or patient monitoring devices. Sample size ranged 39 pregnancies to 588,622 pregnancies. Several studies used gradient boosting machines. Features refer to the selected variables inputted into the predictive model. Performance parameters include area under the receiver operating curve (auROC), mean squared error (MSE), R-squared (R2), mean absolute error (MAE), sensitivity (sens.), specificity (spec.), true positive rate (TPR), false positive rate (FPR), negative predictive value (NPV), positive predictive value (PPV).

(Mateen et al., 2020). Though many of the predictive variables identified amongst early screening models are present in current assessment tools, some new insights were provided for variables such as age and fasting glucose. One prominent study included in this meta-analysis was conducted by Artzi and colleagues who developed a gradient boosted machine learning algorithm from 2355 features using retrospective data sourced from electronic health records for 588,622 pregnancies (Artzi et al., 2020). The model's accuracy in predicting gestational diabetes (auROC 0.85) outperformed current baseline risk scores (auROC 0.68), even with a simplified model using 8 features (auROC 0.80). However, external validation of performance in a more ethnically diverse prospective population is required. Another study considered the potential for thyroid function tests to be used as a secondary tool for diagnosing GDM in borderline cases. Using a principal component analysis in a small population they determined that thyroid patterns correlated with the development of gestational diabetes (Araya et al., 2021). Though this provided novel insights, an objective assessment of the model performance and risk of bias was not performed and results will need to be replicated in a larger, more diverse dataset. More recently, a digital monitoring system was used to identify women at highest risk of developing gestational hypergylcaemia (Yang et al., 2022). In this study, data from Bluetooth-enabled digital blood glucose management system was used to develop a machine learning based regression model (Table 1). This is one of the few externally validated models where model performance was assessed but there was a need for improved predictive accuracy prior to a transition to clinical practice. In summary, machine learning algorithms with data from digital devices could be used to streamline clinical workflows and provide a more cost-effective patient-centred care.

4.1.2. Pre-eclampsia

Machine learning algorithms are being developed to formulate early prediction of pre-eclampsia, a condition characterised by raised blood pressure and proteinuria in the last trimester of pregnancy. The models can consider more variables than current risk assessment tools, with high performance, using data sourced from electronic patient health records (Marić et al., 2020). They can also optimise the use of predictive biomarkers, such as soluble FMS-like tyrosine kinase 1 (sFlt1), placental growth factor (PIGF) and B-type natriuretic peptide (BMP) used in early screening tests to improve test accuracy (Lafuente-Ganuza et al., 2020). Algorithms can identify novel biomarkers for integration into existing clinical-based patient screening methods. In one study, liver function tests and coagulation tests appeared useful in screening for early onset pre-eclampsia (Z. Zhang et al., 2022). These studies highlight the potential to develop algorithms for early disease prediction and to optimise approaches in patient screening that in the future could also provide mechanistic insights. Unsupervised models which cluster data could be useful in understanding the disease aetiology although this area is less explored. Commonly, current models lack external validation and assessments of bias within the model which could inflate estimates of predictive accuracy. In addition, a multicentric approach using datasets that correlate findings in a larger and more diverse population could be integrated into screening methods and would have a broader clinical application.

4.2. Applications of machine learning technologies for diagnostic imaging

4.2.1. Fetal ultrasonography

Table 2 summarises the machine learning models to classify and diagnose abnormalities seen on obstetric ultrasound (Chen et al., 2021). 3D convolutional neural networks have been applied to automate the

Table 2
Deep learning model with performance data to predict fetal abnormalities by ultrasonography.

Study	Deep learning model	Datasets	Prediction aims	Performance
Yang et al., 2018	3D Fully Convolutional network and recurrent neural networks	104 prenatal ultrasound volumes from 104 women 10–14 weeks gestation	Fully automated segmentation of anatomical structures in first trimester ultrasound	Correlation with experts: test set 0.938 validation set 0.922
Arnaout et al., (2021)	Ensemble neural network Aggregated residual visual block net (ARVBNet)	107,823 images sourced from 1326 echocardiograms, 18–24 weeks gestational age	Identify recommended cardiac views and distinguish between normal hearts and complex CHD	AUC 0.99 Sensitivity 95% CI 84–99% Specificity 96% CI 95–97% NPV 100%
Dong et al., (2019)	Convolutional neural Network	2032 positive samples [cardiac four-chamber planes, 5000 negative samples (non-CFPs)	Automate quality control of fetal ultrasound cardiac four-chamber planes	highest mean average precision (mAP) 93.52% at 101 frames per second (FPS) PASCAL VOC dataset: highest mAP 81.2%
Miyagi and Miyake, 2020	Neural network	Japan Society of Ultrasonics in Medicine dataset (2003)	Accurate estimation of fetal weight based on bi-parietal diameter, abdominal circumference, and femur length.	Difference between observed and predicted value Test data set 12.5% All datasets 20%
Fung et al., (2020)	Geometric Machine Learning	INTERGROWTH-21st dataset (n = 4607) INTERBIO-21st Fetal Study (n = 3076)	To improve estimates of fetal gestational age and provide personalised predictions of future growth.	Gestational age could be estimated for pregnancies between 20 and 30 weeks gestation to within 3 days
Naimi et al., (2018)	Quantile regression, random forest (RF), Bayesian additive regression trees (BART), generalised boosted models (GBM)	Magee-Womens obstetric maternal and infant data (18,517 pregnancies) high risk pregnancies dataset (n = 240)	Predict fetal weight over the course of gestation using ex-utero information. Assess relationship between smoking and fetal weight	Pearson correlation coefficient (actual and predicted weight) RF 0.985, Generalised linear 0.977, BART 0.986, quantile regression 0.976, GBM 0.986
Lu et al., (2020)	Ensemble random forest (RF), XGBoost, light GBM	4212 clinical electronic health records	Predict fetal weight at different gestational ages in the absence of ultrasound examination	Mean Relative Error 6% accuracy 64.3%
Baños et al., 2018	Regression Primary Component Analysis (PCA)	Single centre prospective cohort of 677 women 19–24 + 6 gestation	Identify women at risk of spontaneous preterm birth	discriminative performance of cervical texture based score sens.70.4%, spec.77.4% +LR 3.1 (2.2–4.3), -LR 0.4 (0.2–0.7)

Several studies used neural networks which appeared high performing. Publicly available datasets were used in several studies. Performance was assessed using correlation coefficients, sensitivity (sens.), specificity (spec.), positive predictive value (PPV), negative predictive value (NPV), area under receiver operating curve (auROC), mean average precision (mAP), mean relative error (MRE), positive likelihood ratio (+LR) and negative likelihood ratio (-LR).

identification of structures such as the fetal pole, gestational sac and placenta seen on first trimester volumetric ultrasound (Yang et al., 2018). Neural networks have also been applied to improve the diagnosis of congenital heart defects on fetal echocardiography and show similar predictive performance to expert clinicians (Arnaout et al., 2021). Ensemble models can automate quality control in fetal echocardiography. One study used convolutional neural networks and aggregated residual visual block net (ARVBNet) to recognise internal structures of the fetal heart with high precision (93.52%) and scored whether this recognition was true (Dong et al., 2019). An example of how the process of assessing fetal echocardiography using convolutional neural networks for predicting fetal heart structures is shown in Fig. 3. A four chamber view of the fetal heart is inputted to the model and undergoes a complex series of steps including feature extraction, convolution, pooling, flattening, classification and probabilistic distribution to generate an output which predicts the fetal structure of interest.

Furthermore, estimated fetal weight has typically been investigated using applied regression models. However, significant discrepancies persist between fetal biometry and actual birth weights. Deep learning has been applied to improve the accuracy of estimated fetal weight using ultrasound biometry (Miyagi and Miyake, 2020). Though superior to regression models, this method could not reliably estimate the weight of non-standard fetuses such as those in multiple pregnancy or with congenital anomalies. Other studies have used large datasets sourced from electronic health records, INTERGROWTH-21st and the Magee Obstetric Maternal and Infant (MOMI) database to forecast fetal growth trajectories with varying performance (Lu et al., 2020), Fung et al. (2020), Naimi et al. (2018); Lu et al. (2020). Cervical insufficiency is an important risk factor leading to recurrent pregnancy loss and spontaneous preterm birth. Cervical length and funnelling can be assessed with transvaginal ultrasound. Researchers performed a quantitative analysis of tissue texture in 700 ultrasound images of the cervix to identify women at risk of preterm birth (Baños et al., 2018). This broader evaluation of multiple features appears superior in specificity for preterm birth than traditional measurements of cervical length for the assessment of cervical function. Another study combined sonographic features for asymptomatic women with cervical shortening, amniotic fluid metabolomics, proteomics, clinical and demographic factors with good predictions for preterm birth (auROC 0.883) using deep learning model (Bahado-Singh et al., 2019). It is possible that more machine learning approaches will develop tools in diagnostic imaging with pooled datasets which can circumvent uncertainties in current predictive methods.

4.2.2. Fetal magnetic resonance imaging

MRI is being increasingly used to image the fetus and placenta in prenatal diagnosis. It provides a more detailed view of soft tissues which can be difficult to characterise by ultrasound (Ronneberger et al., 2015). Fetal and maternal motion can corrupt 2D image slices making it difficult to reconstruct images. Studies applying machine learning in fetal MRI are summarised in Table 3. Several studies have developed automated pipelines for fetal brain reconstruction using ensemble approaches (Ebner et al., 2020; Salehi et al., 2018; Ronneberger et al., 2015). The model constructed by Ebner and colleagues appeared to have fewer imbalances and improved efficiency working on the whole image when assessed using Dice coefficient scores (Ebner et al., 2020). Another study developed a deep learning model to predict neurodevelopmental outcomes in very preterm infants using clinical and image data (He et al., 2020). The early identification of neurodevelopmental deficits in cognition, language and motor skills could allow early targeted interventions to improve clinical outcomes.

Accurate reconstructions of placental images are important when planning fetal surgery in conditions such as twin to twin transfusion syndrome (Wang et al., 2018). Variation in the placental orientation and location can make it challenging to automate image segmentation. A deep learning framework provided accurate image segmentation in 2D and 3D, preserving image resolution with good performance (Wang et al., 2018). Placental imaging is also important in the diagnosis of Placenta Accreta Spectrum, a condition characterised by abnormal trophoblast invasion of part or all the placenta into the myometrium of the uterine wall which is associated with high maternal mortality (7%). Improving detection of this disease has been explored with ensemble deep learning models (Ye et al., 2022) and MRI derived-texture analysis (Romeo et al., 2019). Many of the studies within this area use advanced machine learning models for image segmentation but few studies combine MRI and large clinical datasets which reflect a diverse patient population from multiple centres.

4.3. Decision-assist tools for intrapartum management

4.3.1. Prediction and classification models for fetal monitoring

Cardiotocography is the most prevalent diagnostic tool used for continuous fetal monitoring and assessments of fetal distress (Cömert et al., 2019). They display biophysical signals for the fetal heart rate and uterine contractions (Hoodbhoy et al., 2019; Chen and Yin, 2022). Using current standards for interpretation there is a high degree of inter and intra-observer variation. To reduce the disparities in cardiotocography interpretation, computational methods are being developed. Researchers used a combination of machine learning algorithms to assess 30 diagnostic features in intrapartum cardiotocography monitoring using advanced signal processing. 12 features were identified as the most relevant in predicting fetal hypoxia with 88.6% accuracy and 94% specificity (Hoodbhoy et al., 2019). Many more recent studies have
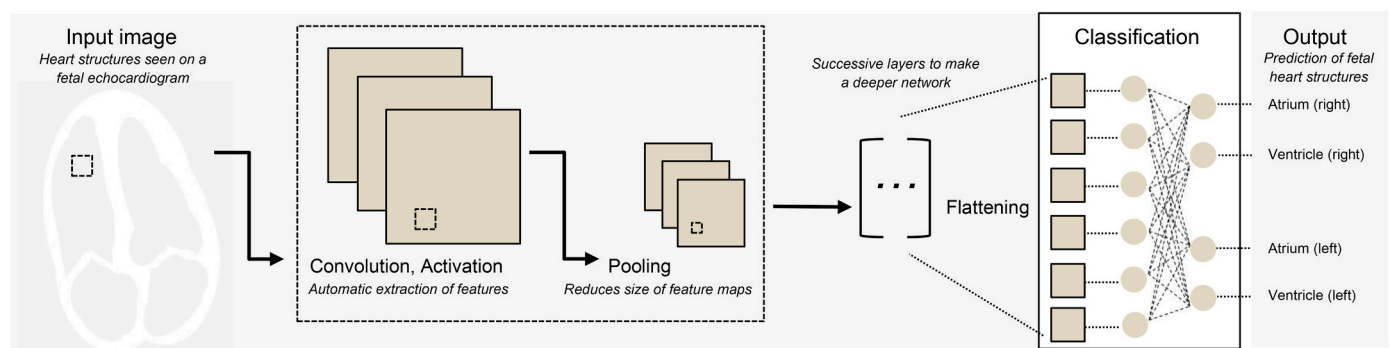

Fig. 3. Application of convolutional neural networks in prenatal ultrasound. Anatomical structures were identified to differentiate between normal and abnormal findings of the fetal heart. This was achieved by inputting images of the four chamber view of the fetal heart as seen on a fetal echocardiogram. After extraction of features, data was inputted to the model and underwent a complex series of steps involving convolution, pooling, flattening, classification and probabilistic distribution to generate an output. The output is a prediction of the anatomical structures in the input image and in this example is labelled as the right ventricle (indicated by black dashed box).

Table 3
Deep learning models to predict fetal structures using magnetic resonance imaging.

Study	Deep learning model	Datasets	Prediction	Performance
Ebner et al., (2020)	Convolutional neural networks (P-net, U net) Support Vector Regression Super Resolution Reconstruction	Single centre, group A: 134 stacks (37 fetuses), group B1: 167 stacks (32 fetuses), group B2 105 stacks (15 fetuses)	Automate localisation, segmentation and super-resolution reconstruction of fetal brain MRI	P-net accuracy Group A 86.54%, Group B1 84.74% Group B2 83.67% P-Net + ML DICE Group A 93.21%, Group B1 93.87% Group B2 92.94%
Salehi et al., (2018)	Fully convolutional network (U-net, voxelwise pp, voxelwise). Machine learning model (SIFT features, random forest, conditional random field)	Single centre 285 stacks from 33 fetal MRIs at 22–38 weeks gestation	Automatic segmentation method which independently segments sections of the fetal brain in 2D fetal MRI slices in real-time.	Highest performing model using challenging test set U net DICE 78.83% sens. 71.97% spec.99.82%.
He et al., (2020)	Deep transfer learning neural network	ABIDE-I repository + prospective recruitment (2 centres). 884 children/adults 291 neonates, 33 preterm infants	Early prediction of neurodevelopment in very preterm infant	Cognition model: Accuracy 81.5% (SD 3.2%), sens. 74% (SD 4.9%), spec. 88.9% (SD 3.1%) LR+ 6.6 (SD 1.9) FPR (11.1% (SD 3.1%) AUC 0.86 (SD 0.05)
Wang et al., (2018)	Convolutional neural networks (P- net, R-netCRF-net)	T2-weighted MRI scans of 25 pregnant women (trimester 2):925 slices. Brain Tumor Segmentation Challenge (BraTS) dataset: 274 cases	Accurate medical image segmentation method compared to automatic segmentation by convolution neural network	Placenta segmentation DICE 89.31% (SD 5.33%) ASSD(pixels) 1.22 (SD 0.55) Brain tumor segmentation DICE 89.93% (SD 6.49%) ASSD(pixels) 1.43 (SD 1.16)
Ye et al., (2022)	Logistic regression support vector machine primary component analysis	Retrospective recruitment from 2 centres: 407 pregnant women undergoing preoperative MRI	Model for the prediction of Placenta Accreta Spectrum using data from clinical records and MRI.	Combined learning model evaluation following external testing AUC 0.857 (0.808–0.894) Accuracy 0.852 Sensitivity 0.904 Specificity 0.769
Romeo et al., (2019)	Random Forest K-nearest neighbour (k-NN) Naïve Bayes (NB) Multilayer perceptron (MLP)	Placenta praevia and suspicious Placenta Accreta Spectrum (n = 64). Positive MRI (n = 20) 12 accreta, 7 increta 1 percreta. Negative MRI (n = 44)	Assess the presence of placenta accreta spectrum in patients with placenta praevia	Highest performing model *k-NN* Accuracy 98.1 Precision, 98.7, Sensitivity 97.5, Specificity 98.7, N° Features 26

Image segmentation models typically followed steps of localisation, segmentation, intensity correction, motion correction, and space alignment. Most datasets were sourced from single centres. Performance measures included DICE coefficient scores (DICE), accuracy, sensitivity (sens.), specificity (spec.), false positive rate (FPR), area under the receiver operating curve (AUC), average symmetric surface distance (ASSD) and precision.

achieved higher levels of accuracy. One example is a deep learning classification model for cardiotocography using a combination of Alex-Net architecture and support vector machines (Mehbodniya et al., 2022). Here 23 attributes were dynamically programmed and fed through the algorithm with accuracy (99.7%). In the last year, many promising studies have been published within this area of research (Fei et al., 2022; Muhammad Hussain et al., 2022; Mehbodniya et al., 2022; Park et al., 2022; Cheng et al., 2022).

Non-invasive fetal electrocardiography (ECG), an alternative to cardiotocography, monitors fetal wellbeing using abdominal ECG signals acquired from electrodes placed on the mother's abdomen (Zhang and Yu, 2020). The main challenge faced when using this technology is signal interference, mainly attributed to the dominant maternal ECG which overlaps in temporal and frequency domains. Machine learning and deep learning approaches are being developed to reduce noise interference in signal processing (Abel et al., 2022). Early approaches subtract 'repetitive noise' attributed to the maternal ECG from the abdominal ECG signal (Ungureanu et al., 2009). More recently, researchers have used deep learning approaches to detect fetal QRS complexes from raw signals and recorded improved performance (PPV 92.25% F1-score 94.1%) after extraction of a fetal ECG signal from a single channel abdominal ECG (Zhong et al., 2019). Removal of maternal ECG from the abdominal ECG signal with short-time Fourier transform and a convolutional auto-encoder appeared effective (Zhong et al., 2020). In another study, clustering and primary component analysis to extract fetal ECG had high performance (PPV of 95.35%, F1-measure of 95.78%.) (Zhong et al., 2020) and a hidden Markov model (HMM)-based supervised algorithm demonstrated high accuracy and sensitivity (97.1% and 100%, respectively) (Huque et al., 2019).

External tocography is used to monitor uterine activity during labour and in women presenting with threatened preterm labour. Periods of poor signal quality are common when using this technology, and human interpretation is subjective. An adaboost algorithm was developed to approximate human interpretation with 93.8% classification accuracy in women >35 weeks gestation (Reynolds et al., 2020). This appeared more robust than earlier methods to automate uterine contraction detection which showed falls in sensitivity and PPV on low quality recordings (Horoba et al., 2016). Recently, algorithms from electrohysterography signals, acquired by electrodes placed on the mother's abdomen are being developed to predict embryo implantation for IVF patients (Sammali et al., 2021) and preterm birth (Cheng et al., 2022; Prats-Boluda et al., 2021, Nieto-del-Amor et al., 2022). One method using novel multichannel entropy features appeared accurate in predicting preterm birth (90.5%) (Cheng et al., 2022). Fetal movements are an important indicator of fetal wellbeing and are presently monitored by maternal perception. This method is prone to inaccuracy due to variation in an individual's perception of vibration, the buffering effect of amniotic fluid which impairs detection of subtle movements such as hand activity, and the false perception of intestinal peristalsis as fetal movements (Hijazi et al., 2010). Wearable devices for out of hospital monitoring of fetal movements are now being developed using accelerometers based on machine learning (Zhang et al., 2022; Xu et al., 2022). An Extra Trees Classifier model showed good accuracy (86.6%) and precision (86.1%) in detecting fetal movements in 20 women, comparable to ultrasound. The use of collaborative datasets replicative in a larger pregnant population and external, real-time validation may accelerate the transition of predictive algorithms predictive algorithms models for fetal monitoring to clinical practice.

4.3.2. Predictive models to optimise intrapartum care

A number of preliminary studies have explored applications of machine learning in intrapartum care. Pain management is an important aspect of intrapartum care. Epidural analgesia is the current gold standard for labour pain relief. However, it is estimated that 0.9%–25% of patients experience breakthrough pain (Tan et al., 2021). One study developed two machine learning models (Random Forest, XGBoost) that identify patients who are most likely to experience pain and compared this to a Logistic regression model (Tan et al., 2021). The authors identified 13 features which were predictive across the three models, with similar performance. However, the omission of predictive parameters from the algorithms may have impacted performance. Malposition of the fetal head in labour is associated with increased obstetric intervention and failed operative vaginal delivery (Ghi et al., 2022). A study developed a feed-forward neural network model (88.7% accuracy, 85.4% precision), using transperineal ultrasound images from 15 maternity units to correctly identify fetal head position during the second stage of labour. It was concluded that automating the assessment of fetal head position could assist in making clinical decisions around the management of labour and mode of delivery. Other studies developed algorithms to predict complications in labour. A supervised machine learning model to predict shoulder dystocia, an obstetric emergency, showed superior accuracy (AUC 0.866) compared to estimates using fetal weight alone (AUC 0.772) or in combination with maternal diabetes (AUC 0.784) (Tsur et al., 2020). Machine learning and regression models to predict post-partum haemorrhage were developed from 152, 279 deliveries, showing good to excellent discrimination (Venkatesh et al., 2020). A gradient boosted model was developed to stratify patients as high or low risk of obstetric anal sphincter injury (AUC 0.756, 95% CI 0.732–0.780) (Chill et al., 2021). Such models could support decisions for the management of labour. Interestingly, a random forest model (eCART) predicted the risk of patient death or transfer to intensive care (AUC 0.86) with higher accuracy than current early warning scores (Arnolds et al., 2022). This study proposed that early warning algorithms could be optimised with machine learning to enable the appropriate implementation of early interventions. In summary, machine learning algorithms for intrapartum care show great variation in model design, dataset and feature selection and not ready for a clinical setting. The role of quality standards will be important when developing reputable models for a multi-centre framework in clinical practice. Whilst intrapartum care is a high-risk area of obstetrics, intelligent tools could improve the delivery of safe intrapartum care for early prediction and detection of complications without clinical deterioration.

5. Future strategies of machine learning

The breadth of machine learning applications in obstetrics indicates the potential of algorithms to improve early prediction, optimise and standardise diagnostic imaging and improve patient care. However, many studies require further development before transition to a clinical setting. Whilst increasing numbers of machine learning algorithms are being developed using electronic healthcare records, it is important that quality standards are consistent. CODE-EHR was recently developed to improve the design and reporting of research studies using electronic healthcare data (Kotecha et al., 2022). The framework specifies minimum standards for dataset construction, quality assessment of data, definition of outcomes, data analysis that follow ethics and governance. Adhering to these standards will improve comparability between models and their impact. In predictive risk stratification models, there are significant disparities in the selection of features and sample sizes, making comparisons between works more challenging. Many models aimed to improve diagnostic accuracy, sensitivity and specificity through supervised approaches rather than uncovering new patterns between variables to understand disease processes. Many researchers model identical data sets using multiple methods to identify which tool will perform best. Some models incorporated large numbers of features to improve

diagnostic accuracy. These types of models may be impractical within a clinical setting, narrowing the scope of their application (Chen et al., 2021). Validating models with an external dataset is an important step which has not been performed consistently. This process is key when assessing the performance of a model, particularly when the training dataset has been developed from a small or less diverse population. It is also important to consider bias within the model which may lead to the overestimation of performance. The prediction model risk of bias assessment tool (PROBAST) is a quality assessment tool for artificial intelligence which could identify this (Collins et al., 2021). Smart devices are being developed for blood pressure monitoring and it is possible that future risk stratification models for pre-eclampsia will utilise this data (Yang et al., 2022). In diagnostic imaging particularly fetal MRI, the field of computer vision is advancing rapidly enabling methods that are transferable to medical imaging. Though some larger collaborative studies have been completed in prenatal ultrasonography (Fung et al., 2020, Kim et al., 2022), the approaches have not yet been observed in fetal MRI where many works use datasets from single centres and with smaller numbers. In intrapartum management, studies lacked external validation or quality assessment (Tan et al., 2021; Ghi et al., 2022; Tsur et al., 2020; Venkatesh et al., 2020; Chill et al., 2021) but are a first step in developing clinical decision assist tools to improve safety standards. Developing multicentric databases with public availability and having multi-disciplinary and international collaborations will therefore accelerate research and improve the diversity of datasets to better represent minority groups (Meshaka et al., 2022, Malani et al., 2023).

6. Conclusion

Machine learning applications in obstetrics and maternity care are rapidly evolving. Increasingly, predictive algorithms incorporate data from electronic health records and frameworks are now being developed to standardise the quality and design of studies using large datasets, particularly in diagnostic imaging. Presently many models lack external validation, a necessary step before algorithms can be transitioned into clinical practice. Quality assessment for bias within models using tools such as PROBAST may reduce the inflation of performance measures. Decision-assistive tools for the delivery of intrapartum care are in preliminary stages and so far, pilot studies have identified areas of interest. As smart technology advances, it is possible that more algorithms using datasets from these recording devices will be created. Developing multicentric databases, increasing multi-disciplinary working and promoting international collaboration will accelerate the advancement of this emerging area in research.

CRediT authorship contribution statement

Zara Arain: Writing – original draft, Conceptualization, Visualization, Writing – review & editing. **Stamatina Iliodromiti:** Supervision, Writing – review & editing. **Gregory Slabaugh:** Supervision, Writing – review & editing. **Anna L. David:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Tina T. Chowdhury:** Conceptualization, Visualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

References

Abel, J.D.K., Dhanalakshmi, S., Kumar, R., 2022. A comprehensive survey on signal processing and machine learning techniques for non-invasive fetal ECG extraction. Multimedia Tools Appl. 1–28.

Ahn, K.H., Lee, K.S., 2022. Artificial intelligence in obstetrics. Obstetr. Gynecol. Sci. 65 (2), 113–124.

Alfirevic, Z., Gyte, G.M., Cuthbert, A., Devane, D., 2017. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst. Rev. (2).

Araya, J., Rodriguez, A., Lagos-SanMartin, K., Mennickent, D., Gutiérrez-Vega, S., Ortega-Contreras, B., Valderrama-Gutiérrez, B., Gonzalez, M., Farías-Jofré, M., Guzmán-Gutiérrez, E., 2021. Maternal thyroid profile in first and second trimester of pregnancy is correlated with gestational diabetes mellitus through machine learning. Placenta 103, 82–85.

Arnaout, R., Curran, L., Zhao, Y., Levine, J.C., Chinn, E., Moon-Grady, A.J., 2021. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nat. Med. 27 (5), 882–891.

Arnolds, D.E., Carey, K.A., Braginsky, L., Holt, R., Edelson, D.P., Scavone, B.M., Churpek, M., 2022. Comparison of early warning scores for predicting clinical deterioration and infection in obstetric patients. BMC Pregnancy Childbirth 22 (1), 1–9.

Artzi, N.S., Shilo, S., Hadar, E., Rossman, H., Barbash-Hazan, S., Ben-Haroush, A., Balicer, R.D., Feldman, B., Wiznitzer, A., Segal, E., 2020. Prediction of gestational diabetes based on nationwide electronic health records. Nat. Med. 26 (1), 71–76.

Averbuch, T., Sullivan, K., Sauer, A., Mamas, M.A., Voors, A.A., Gale, C.P., Metra, M., Ravindra, N., Van Spall, H.G.C., 2022. Applications of artificial intelligence and machine learning in heart failure. Eur. Heart J. Digit. Health 3, 311–322.

Bahado-Singh, R.O., Sonek, J., McKenna, D., Cool, D., Aydas, B., Turkoglu, O., Bjorndahl, T., Mandal, R., Wishart, D., Friedman, P., Graham, S.F., 2019. Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. Ultrasound Obstet. Gynecol. 54 (1), 110–118.

Baños, N., Perez-Moreno, A., Julià, C., Murillo-Bravo, C., Coronado, D., Gratacos, E., Deprest, J., Palacio, M., 2018. Quantitative analysis of cervical texture by ultrasound in mid-pregnancy and association with spontaneous preterm birth. Ultrasound Obstet. Gynecol. 51 (5), 637–643.

Bertini, A., Salas, R., Chabert, S., Sobrevia, L., Pardo, F., 2022. Using Machine learning to predict complications in pregnancy: a systematic review. Front. Bioeng. Biotechnol. 9, 1385.

Bulletins-Obstetrics, C., 2018. ACOG practice bulletin no. 190: gestational diabetes mellitus. Obstet. Gynecol. 131 (2), e49–e64.

Chen, M., Yin, Z., 2022. Classification of Cardiotocography based on Apriori algorithm and multi-model ensemble classifier. Front. Cell Dev. Biol. 844.

Chen, Z., Liu, Z., Du, M., Wang, Z., 2021. Artificial intelligence in obstetric ultrasound: an update and future applications. Front. Med. 1431.

Cheng, A., Yao, Y., Jin, Y., Chen, C., Vullings, R., Xu, L., Mischi, M., 2022. Novel multichannel entropy features and machine learning for early assessment of pregnancy progression using electrohysterography. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 69 (12), 3728–3738.

Chill, H.H., Guedalia, J., Lipschuetz, M., Shimonovitz, T., Unger, R., Shveiky, D., Karavani, G., 2021. Prediction model for obstetric anal sphincter injury using machine learning. Int. Urogynecol. J. 32, 2393–2399.

Collins, G.S., Dhiman, P., Navarro, C.L.A., Ma, J., Hooft, L., Reitsma, J.B., Logullo, P., Beam, A.L., Peng, L., Van Calster, B., van Smeden, M., 2021. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 11 (7), e048008.

Cömert, Z., Şengür, A., Budak, Ü., Kocamaz, A.F., 2019. Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models. Health Inf. Sci. Syst. 7, 1–9.

de Marvao, A., Dawes, T.J., Howard, J.P., O'Regan, D.P., 2020. Artificial intelligence and the cardiologist: what you need to know for 2020. Heart 106 (5), 399–400.

Dhombres, et al., 2022. Contributions of artificial intelligence reported in obstetrics and gynecology journals: systematic review. J. Med. Internet Research 24 (4), p.e35465.

Dong, J., Liu, S., Liao, Y., Wen, H., Lei, B., Li, S., Wang, T., 2019. A generic quality control framework for fetal ultrasound cardiac four-chamber planes. IEEE J. Biomed. Health Inf. 24 (4), 931–942.

Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P.A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A.L., 2020. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. Neuroimage 206, 116324.

Fei, Y., Chen, F., He, L., Chen, J., Hao, Y., Li, X., Liu, G., Chen, Q., Li, L., Wei, H., 2022. Intelligent classification of antenatal cardiotocography signals via multimodal bidirectional gated recurrent units. Biomed. Signal Process Control 78, 104008.

Fung, R., Villar, J., Dashti, A., Ismail, L.C., Staines-Urias, E., Ohuma, E.O., Salomon, L.J., Victora, C.G., Barros, F.C., Lambert, A., Carvalho, M., 2020. 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 2 (7), e368–e375.

Ghi, T., Conversano, F., Ramirez Zegarra, R., Pisani, P., Dall'Asta, A., Lanzone, A., Lau, W., Vimercati, A., Iliescu, D.G., Mappa, I., Rizzo, G., 2022. Novel artificial intelligence approach for automatic differentiation of fetal occiput anterior and non-occiput anterior positions during labor. Ultrasound Obstet. Gynecol. 59 (1), 93–99.

He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J.R., Parikh, N.A., 2020. A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. Sci. Rep. 10 (1), 1–13.

Hijazi, Z.R., Callan, S.E., East, C.E., 2010. Maternal perception of foetal movement compared with movement detected by real-time ultrasound: an exploratory study. Aust. N. Z. J. Obstet. Gynaecol. 50 (2), 144–147.

Hoodbhoy, Z., Noman, M., Shafique, A., Nasim, A., Chowdhury, D., Hasan, B., 2019. Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. Int. J. Appl. Basic Med. Res. 9 (4), 226.

Horoba, K., Wrobel, J., Jezewski, J., Kupka, T., Roj, D., Jezewski, M., 2016. Automated detection of uterine contractions in tocography signals–Comparison of algorithms. Biocybern. Biomed. Eng. 36 (4), 610–618.

Huque, A.S.A., Ahmed, K.I., Mukit, M.A., Mostafa, R., 2019. HMM-based supervised machine learning framework for the detection of fECG RR peak locations. IRBM 40 (3), 157–166.

Kim, H.Y., Cho, G.J., Kwon, H.S., 2022. Applications of artificial intelligence in obstetrics. Ultrasonography 42 (1), 2–9.

Kotecha, D., Asselbergs, F.W., Achenbach, S., Anker, S.D., Atar, D., Baigent, C., Banerjee, A., Beger, B., Brobert, G., Casadei, B., Ceccarelli, C., 2022. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. Eur. Heart J. 43 (37), 3578–3588.

Lafuente-Ganuza, P., Lequerica-Fernandez, P., Carretero, F., Escudero, A.I., Martinez-Morillo, E., Sabria, E., Herraiz, I., Galindo, A., Lopez, A., Martinez-Triguero, M.L., Alvarez, F.V., 2020. A more accurate prediction to rule in and rule out pre-eclampsia using the sFlt-1/PlGF ratio and NT-proBNP as biomarkers. Clin. Chem. Lab. Med. 58 (3), 399–407.

Li, Y., Rao, S., Hassaine, A., Ramakrishnan, R., Canoy, D., Salimi-Khorshidi, G., Mamouei, M., Lukasiewicz, T., Rahimi, K., 2021. Deep Bayesian Gaussian processes for uncertainty estimation in electronic health records. Sci. Rep. 11 (1), 1–13.

Lu, Y., Fu, X., Chen, F., Wong, K.K., 2020. Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning. Artif. Intell. Med. 102, 101748.

Malani IV, S.N., Shrivastava, D., Raka, M.S., Raka IV, M.S., 2023. A comprehensive review of the role of artificial intelligence in obstetrics and gynecology. Cureus 15 (2).

Marić, I., Tsur, A., Aghaeepour, N., Montanari, A., Stevenson, D.K., Shaw, G.M., Winn, V. D., 2020. Early prediction of preeclampsia via machine learning. Am. J. Obstetr. Gynecol. MFM 2 (2), 100100.

Mateen, B.A., David, A.L., Denaxas, S., 2020. Electronic health records to predict gestational diabetes risk. Trends Pharmacol. Sci. 41 (5), 301–304.

Mehbodniya, A., Lazar, A.J.P., Webber, J., Sharma, D.K., Jayagopalan, S., Singh, P., Rajan, R., Pandya, S., Sengan, S., 2022. Fetal health classification from cardiotocographic data using machine learning. Expet Syst. 39 (6), e12899.

Meshaka, R., Gaunt, T., Shelmerdine, S.C., 2022. Artificial intelligence applied to fetal MRI: a scoping review of current research. Br. J. Radiol. 95, 20211205.

Miyagi, Y., Miyake, T., 2020. Potential of artificial intelligence for estimating Japanese fetal weights. Acta Med. Okayama 74 (6), 483–493.

Muhammad Hussain, N., Rehman, A.U., Othman, M.T.B., Zafar, J., Zafar, H., Hamam, H., 2022. Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data. Sensors 22 (14), 5103.

Naimi, A.I., Platt, R.W., Larkin, J.C., 2018. Machine learning for fetal growth prediction. Epidemiology (Cambridge, Mass.) 29 (2), 290.

Nieto-del-Amor, F., Prats-Boluda, G., Garcia-Casado, J., Diaz-Martinez, A., Diago-Almela, V.J., Monfort-Ortiz, R., Hao, D., Ye-Lin, Y., 2022. Combination of feature selection and resampling methods to predict preterm birth based on electrohysterographic signals from imbalance data. Sensors 22 (14), 5098.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.

Pardey, J., Moulden, M., Redman, C.W., 2002. A computer system for the numerical analysis of nonstress tests. Am. J. Obstet. Gynecol. 186 (5), 1095–1103.

Park, T.J., Chang, H.J., Choi, B.J., Jung, J.A., Kang, S., Yoon, S., Kim, M., Yoon, D., 2022. Machine learning model for classifying the results of fetal cardiotocography conducted in high-risk pregnancies. Yonsei Med. J. 63 (7), 692.

Prats-Boluda, G., Pastor-Tronch, J., Garcia-Casado, J., Monfort-Ortíz, R., Perales Marín, A., Diago, V., Roca Prats, A., Ye-Lin, Y., 2021. Optimization of imminent labor prediction systems in women with threatened preterm labor based on electrohysterography. Sensors 21 (7), 2496.

Reynolds, A.J., Waldron, O.M., Halpern, E.M., McGarvey, C.M., Murray, M.L., Ater, S.B., Geary, M.P., Hayes, B.C., 2020. A wavelet-based algorithm for automated analysis of external tocography: how does it compare to human interpretation? Comput. Biol. Med. 122, 103814.

Romeo, V., Ricciardi, C., Cuocolo, R., Stanzione, A., Verde, F., Sarno, L., Improta, G., Mainenti, P.P., D'Armiento, M., Brunetti, A., Maurea, S., 2019. Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa. Magn. Reson. Imag. 64, 71–76.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference. Springer International Publishing, Munich, Germany, pp. 234–241. October 5-9, 2015, Proceedings, Part III 18.

Salehi, S.S.M., Hashemi, S.R., Velasco-Annis, C., Ouaalam, A., Estroff, J.A., Erdogmus, D., Warfield, S.K., Gholipour, A., 2018. Real-time automatic fetal brain extraction in fetal MRI by deep learning. April. In: 2018 IEEE 15th International *Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, pp. 720–724.

Sammali, F., Blank, C., Bakkes, T.G., Huang, Y., Rabotti, C., Schoot, B.C., Mischi, M., 2021. Multi-modal uterine-activity measurements for prediction of embryo implantation by machine learning. IEEE Access 9, 47096–47111.

Tan, H.S., Liu, N., Sultana, R., Han, N.R., Tan, C.W., Zhang, J., Sia, A.T.H., Sng, B.L., 2021. Prediction of breakthrough pain during labour neuraxial analgesia: comparison of machine learning and multivariable regression approaches. Int. J. Obstet. Anesth. 45, 99–110.

Tsur, A., Batsry, L., Toussia-Cohen, S., Rosenstein, M.G., Barak, O., Brezinov, Y., Yoeli-Ullman, R., Sivan, E., Sirota, M., Druzin, M.L., Stevenson, D.K., 2020. Development and validation of a machine-learning model for prediction of shoulder dystocia. Ultrasound Obstet. Gynecol. 56 (4), 588–596.

Ungureanu, G.M., Bergmans, J.W., Oei, S.G., Ungureanu, A., Wolf, W., 2009. The event synchronous canceller algorithm removes maternal ECG from abdominal signals without affecting the fetal ECG. Comput. Biol. Med. 39 (6), 562–567.

Venkatesh, K.K., Strauss, R.A., Grotegut, C., Heine, R.P., Chescheir, N.C., Stringer, J.S., Stamilio, D.M., Menard, M.K., Jelovsek, J.E., 2020. Machine learning and statistical models to predict postpartum hemorrhage. Obstet. Gynecol. 135 (4), 935.

Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T., 2018. DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (7), 1559–1572.

Xu, J., Zhao, C., Ding, B., Gu, X., Zeng, W., Qiu, L., Yu, H., Shen, Y., Liu, H., 2022. Fetal movement detection by wearable accelerometer duo based on machine learning. IEEE Sensor. J. 22 (12), 11526–11534.

Yang, X., Yu, L., Li, S., Wen, H., Luo, D., Bian, C., Qin, J., Ni, D., Heng, P.A., 2018. Towards automated semantic segmentation in prenatal volumetric ultrasound. IEEE Trans. Med. Imag. 38 (1), 180–193.

Yang, J., Clifton, D., Hirst, J.E., Kavvoura, F.K., Farah, G., Mackillop, L., Lu, H., 2022. Machine learning-based risk stratification for gestational diabetes management. Sensors 22 (13), 4805.

Ye, Z., Xuan, R., Ouyang, M., Wang, Y., Xu, J., Jin, W., 2022. Prediction of placenta accreta spectrum by combining deep learning and radiomics using T2WI: a multicenter study. Abdom. Radiol. 47 (12), 4205–4218.

Zhang, Y., Yu, S., 2020. Single-lead noninvasive fetal ECG extraction by means of combining clustering and principal components analysis. Med. Biol. Eng. Comput. 58, 419–432.

Zhang, X., Chen, Y., Salerno, S., Li, Y., Zhou, L., Zeng, X., Li, H., 2022. Prediction of severe preeclampsia in machine learning. Med. Novel Technol. Dev. 15, 100158.

Zhang, Z., Yang, L., Han, W., Wu, Y., Zhang, L., Gao, C., Jiang, K., Liu, Y., Wu, H., 2022. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J. Med. Internet Res. 24 (3), e26634.

Zhong, W., Liao, L., Guo, X., Wang, G., 2019. Fetal electrocardiography extraction with residual convolutional encoder–decoder networks. Australas. Phys. Eng. Sci. Med. 42, 1081–1089.

Zhong, W., Guo, X., Wang, G., 2020. Maternal ECG removal using short time Fourier transform and convolutional auto-encoder. Int. J. Data Min. Bioinf. 23 (2), 160–175.