THE JOURNAL OF MATERNAL-FETAL & NEONATAL MEDICINE
2025, VOL. 38, NO. 1, 2532099
https://doi.org/10.1080/14767058.2025.2532099

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Perinatal artificial intelligence in ultrasound (PAIR) study: predicting delivery timing

Neil Patel[a], John O'Brien[b], Robert Bunn[c], Brandon Schanbacher[d], John Bauer[d] and Garrett K. Lam[b]

[a]Department of Maternal Fetal Medicine, Ascension Sacred Heart, Pensacola, FL, USA; [b]Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, University of Kentucky, Lexington, KY, USA; [c]Ultrasound AI, CO, USA; [d]Department of Pediatrics, University of Kentucky, Lexington, KY, USA

ABSTRACT

OBJECTIVE: To evaluate the ability of a proprietary artificial intelligence (AI) model to predict the number of days until delivery using ultrasound images alone and to assess the continuous improvement of prediction accuracy, particularly for preterm births, through model retraining.

METHODS: An AI software was developed and trained using de-identified ultrasound images from a cohort of women who delivered at the University of Kentucky from 2017 to 2021. Initially, 5,714 pregnant women, with 19,940 unique ultrasound exams and 877,141 total ultrasound images were utilized from this timeframe. Images from 79% of this cohort (4,505 patients) trained the AI to estimate the number of days until delivery and secondarily optimize predictions related to preterm birth (<37 weeks gestational age). The output consisted of days until delivery which was subsequently categorized as either preterm or term birth.

The remaining 21% of the cohort (1,209 patients) was reserved for derivation and validation of test characteristics. Delivery outcomes for this subgroup were blinded from the AI by an independent third-party data monitor. Unique predictions were made for each patient after each ultrasound exam, and the AI's performance was evaluated against the actual delivery date using metrics such as R^2 values and mean absolute error (MAE) compared to actual days until delivery. After initial testing, the AI was retrained x3 more using the same data (Version 2, V2) and later with an additional 1,165,618 images obtained by extension of the study to include data from our center until 2023 (Version 3 (V3), Version 4 (V4)- consisted of retraining on V3)

RESULTS: Preterm birth rates were similar between the training (18.4%) and validation (18.6%) sets in the initial study set. The initial AI model exhibited a sensitivity of 39% and specificity of 93% for preterm birth prediction, with an AUC of 0.757. The AI's predictions of days to delivery versus actual in the validation set yielded R^2 of 0.90 for term births, 0.88 for spontaneous preterm birth plus term births, and 0.48 for spontaneous preterm birth alone. The MAE in predicting the number of days until delivery showed similar accuracy across all trimesters that were assessed by image analysis. Finally, retraining with improvements in AI architecture and training methodology using additional images provided improved preterm birth prediction, with R^2 values for all births increasing from 0.85 (V1) to 0.88 (V3) to 0.92 (V4). For spontaneous PTB, MAE was 19.99 days in V4.

CONCLUSIONS: AI can predict timing until delivery from ultrasound data alone. This technology can also predict preterm delivery with limited sensitivity. Retraining the AI with supervised and unsupervised learning has the potential to further improve performance.

ARTICLE HISTORY
Received 16 January 2025
Revised 4 June 2025
Accepted 6 July 2025

KEYWORDS
Preterm birth AI; delivery timing AI; ultrasound AI; predicting delivery timing; predicting delivery

Introduction

Preterm birth (PTB) is the leading cause of neonatal mortality worldwide, but global rates remained unchanged from 2010 to 2020 [1]. Current knowledge for PTB prevention consists of a limited

CONTACT Neil Patel drneilbpatel@gmail.com 5153 North 9th Ave #201, Pensacola, Fl32504
Supplemental data for this article can be accessed online at https://doi.org/10.1080/14767058.2025.2532099.

number of associations relating ongoing pathophysiology with outcomes [2]. Cervical length represents a well-studied ultrasonographic biomarker for PTB [3]. One strategy to improve the utility of sonography is to employ serial cervical length surveillance, particularly in high-risk patients, which has been shown to increase the positive predictive value (PPV) substantially compared to a single examination [4]. However, the diagnostic accuracy of cervical length and other modalities may vary appreciably given inter-observer variability [5].

Artificial intelligence (AI) techniques, particularly deep learning algorithms, have been increasingly utilized in a variety of medical applications, especially medical imaging [6]. A distinct advantage of AI its ability to continually refine predictions. The integration of AI into sonographic imaging, particularly in image analysis for diagnostics, can reduce inter-observer variability [7].

Several studies have explored the potential application of AI in predicting PTB based on risk factors alone, [8–10] including a study with a high AUC of 0.93 [11]. Júnior et al. constructed an AI with 524 patients to predict PTB with risk factors and cervical length measurements with AUC ranging from 0.318 to 0.808 [12]. Bahado-Sing et al. constructed an AI trained on 26 patients with asymptomatic short cervix utilizing the presence of sonographic markers like amniotic sludge with PTB risk factors and amniotic fluid proteomics and metabolomics with an AUC of 0.890 [13]. These studies show that AI's multimodal capabilities may enhance PTB prediction.

In this study, we evaluated the ability of a proprietary AI software to predict delivery timing based on ultrasound image characteristics alone, using only a large volume of imaging data for both training and validation of its capabilities. Secondarily, we sought to identify performance characteristics of the AI in predicting PTB and to demonstrate that retraining of the AI improved its predictive ability (i.e. self-learning led to improved performance).

Materials and methods

Study design and participants

The PAIR (Perinatal Artificial Intelligence in ultRasound) study is a retrospective cohort study conducted at the University of Kentucky (UK), which adhered to SPIRIT-AI extension guidelines and Strengthening the reporting of observational studies in epidemiology (STROBE) checklist [14,15]. The study population included pregnant patients who underwent ultrasound examinations at UK central campus or satellite locations between 2017 to 2021. In addition, delivery data was required for inclusion. Exclusion criteria included participants lacking recorded gestational age (GA) at delivery. Medically indicated deliveries adhered to ACOG guidelines for delivery timing [16]. The study was approved by the UK Institutional Review Board (Protocol #75514) and was not supported by external funding.

Data collection and preparation

Ultrasound examinations followed the ultrasound guidelines by the American Institute of Ultrasound in Medicine. These standards enhanced consistency and reduced hardware-related bias across sites. The majority of ultrasound images (>95%) were acquired using General Electric machines with <5% of the total images obtained from Philips machines. A sensitivity analysis by the manufacturer was not performed. Scans included all trimesters. First trimester ultrasounds, usually 11 to 13 + 6 weeks, included nuchal translucency, CRL, adnexa, placenta, and maternal uterine artery Doppler waveforms. Anatomic ultrasounds, usually 18 to 22 weeks, included adnexa and cervical assessment with transvaginal imaging upon patient approval. Third trimester ultrasounds consisted of follow up growth scans, biophysical profiles and umbilical artery Doppler assessment [17].

Still ultrasound images (excluding video) were downloaded directly from image storage. Python code was used to de-identify ultrasound images. Each patient, ultrasound study, and ultrasound image were given a unique ID, which was associated to the GA at delivery to ensure confidentiality. A computer-generated random allocation was used to assign patients to the training set or the validation set. The GA at birth for each patient was obtained from a delivery database maintained at

United Kingdom. The de-identification process, ID assignment, sequestration of the validation data from the training set, and usage of GA at birth data were performed independently and verified by a third-party data monitor.

AI model development

A proprietary deep learning algorithm was built to analyze deidentified ultrasound images and predict PTB with the first data tranche (images from 2017 to 2021), labeled "V1". The AI was first trained on a dataset consisting of 79% of the patient exams (across all trimesters) collected in the first tranche. The data input consisted of all de-identified ultrasound images from individual ultrasound exams along with the corresponding number of days until delivery from the date the study was performed. No designations of specific structures were given to the AI. Deidentified images were broken down by the AI into digital signals, which allowed the AI to learn and identify patterns that were consistent with time estimates (days) until delivery. It then filtered and pretrained those images to optimize delivery timing prediction. Supervised and unsupervised machine learning techniques were utilized for image training. Within each exam, features from every image were aggregated with conventional machine-learning classifiers to yield one study-level prediction before any patient-level analysis. To avoid overfitting, training logic involved image subsets and a proprietary block box algorithm to generate new training images. Label smoothing functionality addressed incorrect labels and prevented overconfidence. Label Smoothing Cross Entropy was used to improve prediction while minimizing overfitting. A Frequentist approach was applied to estimate the AI's model parameters and validate its predictions.

A logical extension of prediction of delivery timing allowed for a focus on optimizing prediction for PTB (<37 weeks GA). Neural networks were used to recognize features and patterns within the digital signals of the images. Two types of neural networks were used for training: convolutional neural networks (CNNs) and transformers. Both looked at the images in different ways. CNNs were used for the original work including PTB analysis and days until delivery. During the training process, in a select number of cases, class-activation maps created by the programmer and were used by the AI as part of its deep learning process to indicate the anatomic regions of the images that were most discriminative in the AI's Decision making, thus allowing it to refine its prediction for both PTB and time until delivery. Transformers were used for the retraining predictions.

Validation of the AI model

After the AI was trained on how to identify digital signals and learned the patterns consistent within the digital signals to predict both time-to-delivery and PTB, its performance was evaluated using the sequestered validation dataset. The input to the AI for the validation set was only de-identified ultrasound images of individual scans, while the output was the AI's determination of time to delivery and PTB. There was no input of risk factors or actual measurements of structures. The AI did not have access to the actual GA at the time of delivery for any of the images in the validation set. Only two people in the study had the GA at delivery for the validation set: the primary author, and the independent third-party data monitor. Once the AI's analysis of PTB was performed, the AI's estimation of PTB was then compared to the patient's actual GA at date of delivery. The AI developer and AI were blinded from the GA of patients in validation set. The integrity of the validation process, harms and auditing were ensured through independent third-party data monitoring.

Redevelopment for delivery timing prediction

Following initial development, the AI underwent further refinement through a combination of supervised and unsupervised learning techniques. Adjustments were made to the neural networks to include

transformers instead of CNNs to enhance their predictive capabilities. Using the same dataset of deidentified ultrasound images as the input, the fine-tuning shifted from predicting PTB to predicting the number of days until delivery in both V1 and V2. Estimation of continuous days-until-delivery was not part of the original study protocol and was explored after the preterm-versus-term classifier proved feasible. Number of days until delivery became an outcome variable in V2. The training dataset was expanded with 1,165,618 additional images including exams through 2023. AI development version history is below:

Version	Data Included
V1	(original dataset) 877,141
V2	(7 months of retraining V1) 877,141
V3	(V2 + 1,165,618 images) 2,042,759
V4	(8 months of retraining V3) 2,042,759

Accuracy per trimester analysis

The validation set was stratified into trimester-based subgroups, comparing confidence intervals within each trimester. Potential confounders were assessed per trimester against a reference subgroup.

Retraining for PTB prediction

The original PTB AI underwent 7 months of retraining on the same dataset to further refine its PTB prediction,"V2". This retraining process improved the AI's predictive capabilities based on the initial dataset, while attempting to enhance PTB accuracy. The AI retraining involved adding an unsupervised pretraining step that created a base AI model which was then fine-tuned with supervised and unsupervised learning.

Additional images for training set

1,165,618 additional ultrasound images were incorporated into the training set labeled "V3" by including additional patients. Both supervised and unsupervised learning techniques were employed to maximize the AI's ability to extract meaningful features from these images. Synthetic data generation was utilized to augment the training process, improving the AI's robustness in predicting delivery timing and PTB risks. The V3 AI underwent refinement for 8 months with additional CNNs labeled "V4".

Test characteristics assessment

The AI's predictions for delivery timing were compared to the actual GA at delivery. Test characteristics for the PTB AI performance included: sensitivity, specificity, PPV, negative predictive value (NPV), and area under the receiver operating characteristic curve. The MAE and R^2 statistics assessed the AI's accuracy in predicting the number of days until delivery.

Analysis

All statistical analyses were conducted using Python. Calibration metrics were not calculated for this analysis; and will be reported in a follow-up study. Descriptive statistics summarized the baseline data of the study population. These statistics provided a clear overview of the data, aiding in understanding our sample's composition. Comparisons between the training and validation sets used the following statistical tests: independent samples t-tests for continuous variables, chi-square tests for categorical variables, and Wilcoxon rank-sum tests for non-normally distributed data. Pearson correlation coefficients to explore relationships between variables of interest. Statistical significance was defined as a two-sided p-value < 0.05.

Results

Population and demographics

The initial version of the AI (V1) included 5714 patients with 19,940 unique ultrasound examinations and 877,141 total ultrasound images. This patient cohort was divided into a training set, comprising 4505 patients (79%), and a separate, independently held validation set, consisting of 1209 patients (21%). The baseline data of the patients in both sets were comparable, Table 1. The overall PTB rates in the training set (18.4%) and validation set (18.6%) were similar. Stillbirth outcomes were not available in the source database.

AI PTB prediction performance

Evaluation of the AI's V1 performance in predicting PTB revealed a sensitivity of 39% and specificity of 93%. The PPV was 56%, and the NPV was 87%, Table 2. The area under the curve (AUC) for predicting PTB was 0.757 compared to the AUC of 0.842 for PTB for scans performed < 30 days until delivery, Figure 1. The AI demonstrated high specificity, its initial sensitivity in predicting PTBs based on ultrasound images alone was modest. The validation set accuracy was 81.3% for V1 and 82.7% for V2.

Table 1. Patient characteristics compared in training set versus validation set.

	Training Set ($n = 4505$)	Validation Set ($n = 1209$)
Maternal age¥	29.2 (± 5.9)	29.1 (± 5.8)
Race	221 (4.9%)	55 (4.5%)
Asian	546 (12.1%)	153 (12.7%)
Black	3513 (78.0%)	935 (77.3%)
White	35 (0.8%)	9 (0.7%)
Other	190 (4.2%)	57 (4.7%)
Unknown		
Hispanic	551 (12.2%)	150 (12.4%)
BMI¥	29.2 (± 10.6)	28.9 (± 7.8)
<20	215 (4.8%)	77 (6.4%)
20-30	2578 (57.2%)	677 (56.0%)
>30	1505 (33.4%)	398 (32.9%)
Unknown	207 (4.6%)	57 (4.7%)
Gravida¥	2.7 (± 1.8)	2.7 (± 1.9)
Parity¥	1.2 (± 1.4)	1.2 (± 1.4)
History of preterm birth	545 (12.1%)	152 (12.6%)
Number of preterm births	3959 (87.9%)	1056 (87.4%)
0	449 (10.0%)	120 (9.9%)
1	73 (1.6%)	21 (1.7%)
2	23 (0.5%)	11 (0.9%)
3+		
Prelabor rupture of membranes (PROM)	131 (2.9%)	38 (3.1%)
Tobacco use	693 (15.4%)	193 (15.4%)
Singleton	4342 (96.4%)	1169 (96.4%)
Gestational Age at birth¥	37.8 (± 3.5)	37.81 (± 3.6)
Cesarean Section	1627 (36.1%)	457 (37.8%)
NICU Admission	987 (21.9%)	268 (22.2%)

No statistical difference between both groups as defined by P value less than 0.05.
Data reported as the number of patients and percentage of patients within dataset, unless otherwise specified.
¥Data reported as mean ± standard deviation.

Table 2. AI predictions for preterm birth in the validation set with initial AI (V1) vs retrained AI (V2).

	Initial AI (V1)	Retrained AI (V2)	Initial AI (V1)	Retrained AI (V2)
Test characteristics	<37w GA	<37w GA	<37w GA and <30 Days until delivery	<37w GA and <30 Days until delivery
PPV	56%	65%	86%	90%
NPV	87%	87%	74%	78%
Sensitivity	39%	40%	51%	59%
Specificity	93%	95%	94%	96%
Area Under Curve	0.757	0.825	0.842	0.900

37w = 37 weeks, GA = Gestational Age, PPV = Positive Predictive Value, NPV = Negative Predictive Value.

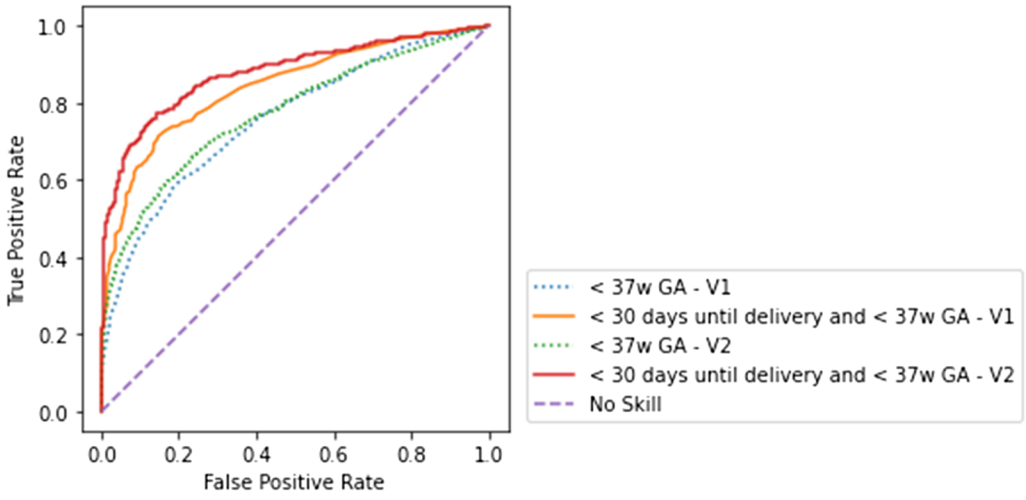

Figure 1. Receiver operating characteristic curve for AI predictions of preterm birth (V1) and after 7 months of retraining (V2).

Table 3. AI's Prediction vs actual days until delivery in validation set per individual study: V1 (877,141 images) vs V3 (2,042,759 images) vs V4 (2,042,759 images retrained).

	Correlation Coefficient R^2		
	V1	V3	V4
Term + all PTB	0.85	0.88	0.92
Term Births Alone	0.90	0.91	0.95
Term + Spontaneous PTB	0.88	0.90	0.94
Spontaneous PTB	0.48	0.64	0.72
Iatrogenic PTB	0.52	0.63	0.75

*Predictions made independent of each study, not per patient.
V1 = original AI with 877,141 US images.
V3 = V1 original AI with 1,165,618 additional US images in training set.
V4 = V3 AI with over 8 months of retraining.

Accuracy of delivery timing prediction AI

In assessing the accuracy of the AI for predicting delivery timing, the R^2 values of the AI's predictions of time until delivery versus the actual days until delivery in the validation set were as follows: the overall set of patients (term plus all PTB), R^2 = 0.85; term births alone, R^2 = 0.90; spontaneous PTB plus term births, R^2 = 0.88; spontaneous PTB alone, R^2 = 0.48; and iatrogenic PTB alone, R^2 = 0.52, Table 3.

Figure 2 illustrates the comparison between the AI's 50% quantile prediction of the number of days until delivery and the actual days until delivery. The standard deviation was 18 days. 80% of these predictions were within 1 standard deviation.

Correlation analysis stratified by factors

The correlation analysis, based on 1,714 patient exams in the validation set, assessed the AI's prediction accuracy for delivery timing. Across trimesters, the MAE was consistent: 15.06 days (95% CI 12.13–17.98) in the first trimester, 14.25 days (95% CI 12.59–15.91) in the second, and 12.16 days (95% CI 10.79–13.53) in the third. Overlapping confidence intervals suggest the AI's predictive performance was preserved across trimesters, Table 4.

Additional images for training set

The addition of 1,165,618 images to the training set (V3) improved performance metrics compared to V1. R^2 correlation values for prediction time to delivery for the whole dataset (term plus PTB) improved from V1 = 0.85 to V3 = 0.88. Spontaneous PTB values improved from V1 = 0.48 to V3 = 0.64, Table 3.

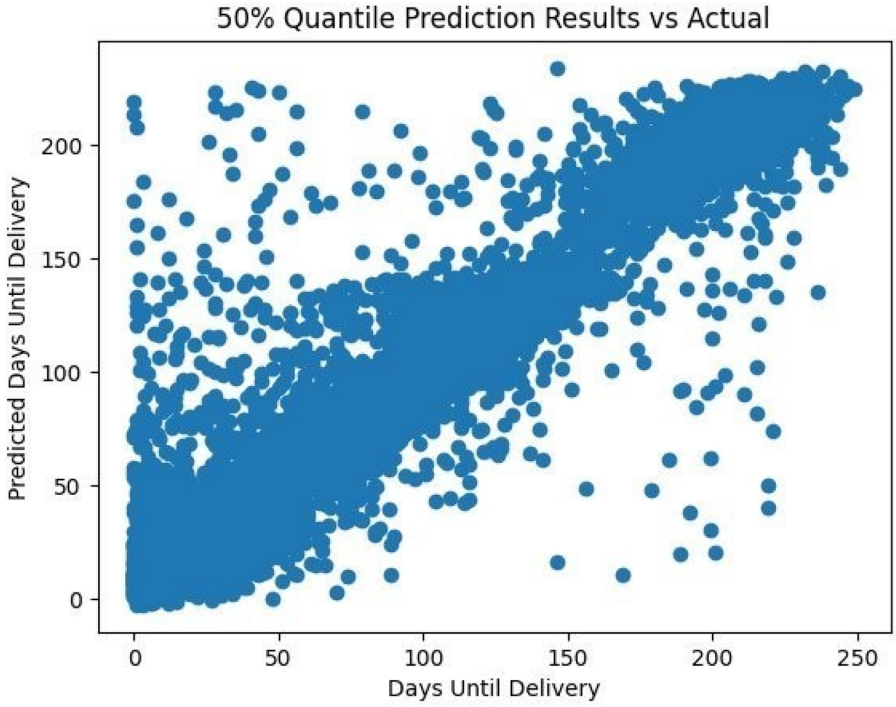

Figure 2. Prediction of the number of days until delivery vs. actual days until delivery.

Table 4. Correlation of predicted versus actual gestational age at delivery per trimester stratified by factors that may impact AI performance.

	1st Trimester (11w0d-13w6d) $n=327$			2nd Trimester (18w6d-22w6d) $n=791$			3rd Trimester (28w0d-32w6d) $n=596$		
	Frequency (%)	Prediction Mean Absolute Error ± SD	P Value	Frequency (%)	Prediction Mean Absolute Error ± SD	P Value	Frequency (%)	Prediction Mean Absolute Error ± SD	P Value
Age	$n=327$			$n=791$			$n=596$		
<20 years	15 (4.6%)	29.78 ± 55.27	0.66	43 (5.5%)	14.46 ± 17.97	0.824	41 (6.9%)	10.30 ± 14.23	0.490
20-30 years	140 (42.8%)	14.49 ± 26.63	REF	350 (44.2%)	15.32 ± 24.50	REF	249 (41.8%)	12.26 ± 17.20	REF
>30	172 (52.6%)	14.24 ± 22.87	0.929	398 (50.3%)	13.30 ± 23.74	0.253	306 (51.3%)	12.33 ± 17.39	0.962
BMI	$n=324$			$n=775$			$n=561$		
<20	13 (3.8%)	19.98 ± 28.37	0.236	54 (7.0%)	16.77 ± 28.39	0.467	31(5.5%)	11.03 ± 14.17	0.986
20-30	196 (60.5%)	12.73 ± 20.77	REF	449 (57.9%)	13.33 ± 22.61	REF	285 (50.8%)	11.08 ± 15.17	REF
30-35	51 (15.7%)	15.20 ± 26.96	0.479	128 (16.5%)	14.64 ± 24.23	0.570	101 (18.0%)	13.22 ± 19.48	0.261
>35	64 (19.8%)	20.18 ± 39.44	0.053	144 (18.6%)	14.78 ± 24.63	0.513	144 (25.7%)	12.54 ± 17.37	0.486
History of PTB	$n=327$			$n=791$			$n=595$		
No	289 (88.4%)	14.88 ± 26.81	REF	675 (85.3%)	13.53 ± 23.15	REF	486 (81.7%)	12.18 ± 17.26	REF
Yes	38 (11.6%)	16.42 ± 22.72	0.735	116 (14.7%)	18.47 ± 26.01	0.038	109 (18.3%)	12.12 ± 16.42	0.974
Race/Ethnicity	$n=327$			$n=791$			$n=596$		
Asian	24 (7.3%)	12.53 ± 21.54	0.638	37 (4.7%)	8.08 ± 10.46	0.121	16 (2.7%)	10.08 ± 9.47	0.654
Black	45 (13.8%)	20.91 ± 38.27	0.113	101 (12.8%)	17.90 ± 29.69	0.154	69 (11.6%)	13.81 ± 19.04	0.421
Hispanic	9 (2.8%)	14.29 ± 24.79	0.961	24 (3.0%)	13.95 ± 23.47	0.969	17 (2.9%)	12.00 ± 17.01	0.998
White	247 (75.5%)	14.64 ± 25.26	REF	617 (78%)	14.14 ± 23.58	REF	482 (80.9%)	12.01 ± 17.10	REF

*REF: Reference subgroup was aged 20-30 yrs with BMI 20-30, no history of Preterm Birth and White.
**Races less than 2% of population were excluded from analysis.
***Patients had multiple US exams in pregnancy, but predictions were made on single US exams irrespective of other US exams.

Retraining analysis

Retraining of the AI model enhanced its predictive performance. In V2, retraining on the original dataset raised sensitivity from 39% to 40%, and specificity from 93% to 95%, Table 2. The AUC for PTB prediction increased from 0.757 to 0.825, and delivery prediction within 30 days showed improvement from 0.842 to 0.900 (Figure 1). Retraining of the AI model with additional training set images (V4) enhanced R^2 correlation values for prediction of term and PTB (V1 = 0.85 to V4 = 0.92), Table 3. MAE values improved from AI retraining V3 to V4 in all categories, Table 5.

Table 5. AI's Prediction vs actual days until delivery in validation set per individual study: V3 (2,042,759 images) vs V4 (2,042,759 images retrained).

	Mean Absolute Error (days)	
	V3	V4 (Retrained V3)
Term + all PTB	14.30	12.90
Term Births Alone	11.71	10.76
Term + Spontaneous PTB	12.67	11.62
Spontaneous PTB	22.08	19.99
Iatrogenic PTB	22.44	19.33

*Predictions made independent of each study, not per patient.
V3 & V4 = updated AI with 2,042,759 US images.

Discussion

Our study demonstrates the AI's capability to predict delivery timing and PTB using ultrasound images alone. Notable improvements in performance were seen across successive AI model versions thus demonstrating that AI can improve its performance with more data and time to learn. The initial V1 model, trained on 877,141 ultrasound images, provided a foundational level of predictive performance with preterm-specific limitations. Enhancements in the V3 model were achieved by incorporating 1,165,618 additional images, raising the R^2 value for combined term and PTBs to 0.88 and for spontaneous PTBs to 0.64, which demonstrated a beneficial impact of dataset expansion on predictive accuracy.

Further refinement of the V4 model through retraining V3 on the same expanded dataset (2,042,759 images) yielded the highest performance metrics. R^2 improved to 0.92 for all births and 0.72 for spontaneous PTBs, highlighting gains from iterative learning. These results highlight the critical roles of data quantity, diversity, and targeted retraining in enhancing the precision and reliability of AI predictions.

Currently, assessment for PTB risk includes assessment of historical risk factors, cervical length measurement, serum biomarkers and improved ultrasound performance [18]. Our AI demonstrates a novel ability to analyze ultrasound images independent of known risk factors or sonographer measurements to predict PTB and time-to-delivery. This method is agnostic to risk factors and methodology of risk assessment, and may be improved if our image analysis is combined with biomarkers [8–12]. Our initial AI performance assessments revealed accurate predictive ability for term and overall births. Continuous retraining with additional training images have suggested that enhancement of the AI's performance and accuracy is feasible. Of particular interest is the algorithm written which highlights which regions were most influential in the AI's class activation mapping. This feature of the AI will help clinicians validate the significance of hot spot structures known to be influential in PTB prediction (i.e. cervix) and to potentially discover other structures not known or thought to be germane to PTB (i.e. ovary, endometrium). Interpretability remains limited because only coarse class-activation maps are available; the proprietary network architecture and weightings are not directly inspectable.

We emphasize that AI achieves its performance by analyzing digital signals inherent to the ultrasound images alone, and does not incorporate structural measurements produced by the sonographer/operator on the machine. Thus instantaneous prediction is feasible once images are available and are unbiased by operator variability. We foresee versatility with the use of this AI, as it can be deployed in both resource-rich and resource-limited settings, allowing a reliable assessment of risk from anywhere at any time. Furthermore, the AI, as shown, will only improve as it continues to learn with more utilization and more data input for analysis.

Strengths and limitations

Strengths of our study include the novel use of an AI algorithm to predict delivery timing. We showed the capabilities of the AI to learn and improve its performance through continuous training to enhance predictive accuracy over time. The validation cohort was kept separately and independently from the training set and thus the evaluation of the AI performance is non-biased. The data input of images came from a variety of sonographers across a heterogeneous group of sites within the same state. Video clips were excluded owing to higher computational cost and acquisition variability. There were

several limitations. The dataset in this study was obtained from a single institution, which may raise concerns regarding generalizability. No formal power calculation was performed, as total image numbers exceeded typical sample thresholds for detecting AUC ≥ 0.75. The analysis was limited to single exams within a patient's pregnancy course. Individual scans were purposefully used to limit potential bias for optimizing test characteristics. We foresee the probability of enhanced predictive abilities by combining images from serial trimester exams. Integration of multiple image sets for individual patients may offer a particular value for this technology. Future work will prioritize (i) integrating biochemical biomarkers (e.g. fetal-fibronectin, pro-inflammatory cytokines), (ii) extracting Doppler-flow and fetal-movement features from video, and (iii) leveraging serial ultrasound exams to capture longitudinal change. Finally, our study focused on ultrasound images alone, without incorporating other potentially relevant patient data, such as PTB history or biochemical testing. Multiple data points (including history and structural measurements) could be incorporated rapidly which may aid in more accurate predictions. Such markers may lead to personalized treatment modalities such as progesterone therapy while enhancing patient and clinician awareness of risk.

Conclusions

Our study advances the application of AI in obstetrics, showing promise in predicting delivery timing and identifying individuals at risk of PTB. Sensitivity for PTB prediction remains constrained because spontaneous PTB is multifactorial. Further research is warranted to refine AI models, validate findings in diverse populations, and assess the integration of AI predictions into clinical practice to improve maternal and neonatal outcomes.

During the preparation of this work the author(s) used ChatGPT in order to write python code for deidentification of ultrasound images, grammatical errors, and cohesiveness in the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgments

University of Kentucky (UK) ultrasonographers.

Author contributions

CRediT: **Neil Patel**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing; **John O'Brien**: Investigation, Project administration, Writing – original draft, Writing – review & editing; **Robert Bunn**: Conceptualization, Formal analysis, Methodology, Software, Writing – review & editing; **Brandon Schanbacher**: Validation; **John Bauer**: Resources, Writing – review & editing; **Garrett K. Lam**: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Disclosure statement

Dr. Lam has a stock interest in Ultrasound AI. Robert Bunn is the President and Founder of Ultrasound AI. The other authors report there are no competing interests to declare.

Funding

This was completed with no external funding.

Presentations

Poster at SMFM Annual meeting in San Francisco, CA: February 2023. Oral presentation at AIUM Annual Meeting in Orlando, FL: March 2023. Poster at SMFM Annual meeting in National Harbor, MD: February 2024. Oral presentation at AIUM Annual Meeting in Austin, TX: April 2024. Poster at SMFM Annual meeting in Denver, CO: February 2025.

Précis

This cohort study evaluated a novel artificial intelligence to predict delivery timing using ultrasound images alone.

Data availability statement

Due to the nature of the research, due to ethical and privacy concerns, supporting data is not available.

References

[1] Ohuma EO, Moller A-B, Bradley E, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. Lancet. 2023;402(10409):1261–1271. doi:10.1016/S0140-6736(23)00878-4.
[2] Goldenberg RL, Culhane JF, Iams JD, et al. Epidemiology and causes of preterm birth. Lancet. 2008;371(9606):75–84. doi:10.1016/S0140-6736(08)60074-4.
[3] Gudicha DW, Romero R, Kabiri D, et al. Personalized assessment of cervical length improves prediction of spontaneous preterm birth: a standard and a percentile calculator. Am J Obstet Gynecol. 2021;224(3):288.e1–288.e17. e1. Epub 2020 Sep 9. PMID: 32918893; PMCID: PMC7914140. doi:10.1016/j.ajog.2020.09.002.
[4] Owen J, Yost N, Berghella V, National Institute of Child Health and Human Development, Maternal-Fetal Medicine Units Network., et al. Mid-trimester endovaginal sonography in women at high risk for spontaneous preterm birth. JAMA. 2001;286(11):1340–1348. PMID: 11560539. doi:10.1001/jama.286.11.1340.
[5] Dubbins PA. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 2017;8(1):171–182.
[6] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. doi:10.1038/nature14539.
[7] Byra M, Styczynski G, Szmigielski C, et al. Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. Sci Rep. 2018;8(1):15919.
[8] Akazawa M, Hashimoto K. Prediction of preterm birth using artificial intelligence: a systematic review. J Obstet Gynaecol. 2022;42(6):1662–1668. Epub 2022 Jun 1. PMID: 35642608. doi:10.1080/01443615.2022.2056828.
[9] Hong YM, Lee J, Cho DH, et al. Predicting preterm birth using machine learning techniques in oral microbiome. Sci Rep. 2023;13(1):21105–21130. Nov. doi:10.1038/s41598-023-48466-x.
[10] Chakoory O, Barra V, Rochette E, et al. DeepMPTB: a vaginal microbiome-based deep neural network as artificial intelligence strategy for efficient preterm birth prediction. Biomark Res. 2024;12(1):14–25. Feb. doi:10.1186/s40364-024-00557-1.
[11] Zhang Y, Du S, Hu T, et al. Establishment of a model for predicting preterm birth based on the machine learning algorithm. BMC Pregnancy Childbirth. 2023;23(1):710–779. Nov. doi:10.1186/s12884-023-06058-7.
[12] Andrade Júnior VL, França MS, Santos RAF, et al. A new model based on artificial intelligence to screening preterm birth. J Matern Fetal Neonatal Med. 2023;36(2):2241100. doi:10.1080/14767058.2023.2241100.
[13] Bahado-Singh RO, Sonek J, McKenna D, et al. Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. Ultrasound Obstet Gynecol. 2019;54(1):110–118. doi:10.1002/uog.20168.
[14] Cruz Rivera S, Liu X, Chan AW, SPIRIT-AI and CONSORT-AI Working Group., et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;2(10):e549–e560. doi:10.1016/S2589-7500(20)30219-3.
[15] Liu X, Cruz Rivera S, Moher D, SPIRIT-AI and CONSORT-AI Working Group., et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364–1374. Epub 2020 Sep 9. PMID: 32908283; PMCID: PMC7598943. doi:10.1038/s41591-020-1034-x.
[16] American College of Obstetricians and Gynecologists' Committee on Obstetric Practice, Society for Maternal-Fetal Medicine. Medically Indicated Late-Preterm and Early-Term Deliveries: ACOG Committee Opinion, Number 831. Obstet Gynecol. 2021;138(1):e35–e39. doi:10.1097/AOG.0000000000004447.
[17] AIUM Practice Parameter for the Performance of Standard Diagnostic Obstetric Ultrasound. Journal of Ultrasound in Medicine: official Journal of the American Institute of Ultrasound in Medicine. 2024;43:E20–E32. 6. doi:10.1002/jum.16406.
[18] Celik E, To M, Gajewska K, Fetal Medicine Foundation Second Trimester Screening Group., et al. Cervical length and obstetric history predict spontaneous preterm birth: development and validation of a model to provide individualized risk assessment. Ultrasound Obstet Gynecol. 2008;31(5):549–554. PMID: 18432605. doi:10.1002/uog.5333.