Artificial Intelligence in Obstetrics

Current Applications, Opportunities, and Clinical Implementation Challenges

Eileen Deuster, MD, MSc, and Asma Khalil, MD, PhD†‡§*

Abstract: Artificial intelligence is transforming obstetric practice through applications in diagnostic imaging, risk prediction, and clinical decision-making. Deep learning algorithms have achieved diagnostic accuracy comparable to that of experienced clinicians. However, gaps persist between algorithmic capability and clinical implementation. Critical challenges include limited external validation and algorithmic bias. This review examines current AI applications in obstetrics across multiple clinical domains: automated fetal biometry, structural anomaly detection, prediction of pregnancy complications, and intrapartum fetal surveillance. It highlights persistent technical, ethical, and implementation barriers. Key recommendations include multicenter validation across diverse populations, explainable AI approaches, and creating strong regulatory frameworks.

Key Words: artificial intelligence, machine learning, deep learning, fetal biometry, anomaly detection, preeclampsia prediction

(*Clin Obstet Gynecol* 2026;69:54–61)

Artificial intelligence (AI), particularly machine learning and deep learning technologies, represents one of the most transformative advances in modern obstetrics. Deep learning algorithms, predominantly utilizing convolutional neural network architectures, have achieved diagnostic accuracy comparable to or exceeding that of experienced clinicians across multiple domains.[1,2] The exponential growth of AI applications in women's health reflects both technological maturation and the recognition that pregnancy care generates vast quantities of complex, multimodal data, which are ideally suited to computational analysis. From 2016 to 2024, FDA authorizations for AI-enabled medical devices increased from 2 to 69 annually, with obstetrics and gynecology emerging as key application domains.[3,4]

The appeal of AI in obstetrics stems from several factors. First, pregnancy monitoring relies heavily on imaging modalities—ultrasound, MRI, and fetal echocardiography—where pattern recognition algorithms excel. Second, obstetric care demands risk stratification to identify high-risk pregnancies requiring intensified surveillance or early intervention. Third, significant inter-observer variability exists in the interpretation of cardiotocography, ultrasound biometry, and cervical assessment, areas where AI can provide objective, standardized measurements.[5] Finally, the global shortage of trained sonographers and maternal-fetal medicine specialists, particularly in low- and middle-income countries, creates urgent demand for AI-assisted diagnostic tools that can democratize access to quality prenatal care.

However, critical gaps persist in external validation, prospective clinical trials demonstrating improved outcomes, algorithmic bias mitigation, and generalizability across diverse populations and health care settings.[6] Analysis of neural network architecture selection, training dataset requirements, and multicenter validation strategies reveals that model performance is highly dependent on data

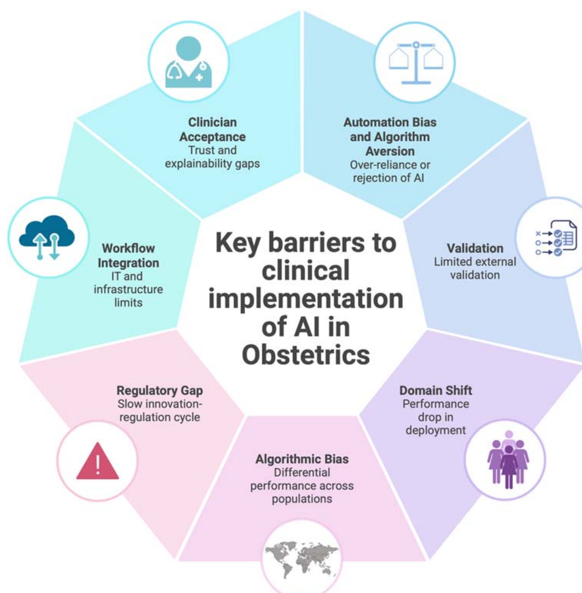

FIGURE 1. Challenges and barriers to clinical implementation of AI in obstetrics. This figure synthesizes the major barriers preventing translation of AI algorithms from research to clinical practice in obstetrics. full color online

Received for publication November 2, 2025; accepted November 12, 2025.

From the *Department of Obstetrics and Perinatal Medicine, Goethe University Frankfurt and University Hospital Frankfurt am Main, Germany; †Vascular Biology Research Centre, Molecular and Clinical Sciences Research Institute, City St George's University of London, London, United Kingdom; ‡Fetal Medicine Unit, St George's University Hospital, London, United Kingdom; and §Fetal Medicine Unit, Liverpool Women's Hospital, Liverpool, United Kingdom.

The authors declare that they have nothing to disclose.

Correspondence: Asma Khalil, MD, PhD, Department of Obstetrics and Gynaecology, University of London, St. George's University Hospital, Blackshaw Road, Tooting, London SW17 0QT, United Kingdom. E-mails: akhalil@sgul.ac.uk; asmakhalil79@googlemail.com

DOI: 10.1097/GRF.0000000000000980

TABLE 1. Summary of Artificial Intelligence Applications in Obstetrics

Domain	Task/application	Algorithm(s)	Performance metric(s)	Key result(s)	Ref
Fetal biometry	BPD, HC, AC, FL measurement	CNN	ICC, Reproducibility Time	ICC >0.95; lower variability; 30%-50% time reduction	9–11
CNS anomaly detection	Brain abnormality classification	Hybrid U-Net and VGG architectures	Accuracy	91.5%-92.9% accuracy	12,13
Cardiac anomaly detection	Congenital heart disease	Deep learning models	Sens, Spec	Pooled Sens 0.89, Spec 0.91	2
Fetal MRI anomaly	Organ classification	FOAC-Net	Accuracy, Precision, Recall	85.3% overall; varied by organ	14
Preeclampsia prediction	First-trimester risk	Neural networks, Logistic regression	AUC, Sens, Spec	Sens 58%-64%; AUC 0.75	15,16
Preterm birth	Risk prediction	Machine learning models (multifactorial)	AUC	0.63-0.75	17
Twin growth discordance	Trajectory analysis	K-means	Cluster patterns	5 trajectories; $>25\%$ discordance = adverse outcomes	18
CTG interpretation	FHR analysis	Deep learning and rule-based techniques	Precision	High precision for pattern identification	19,20

All findings are derived from peer-reviewed literature cited in the main text and reference list of this review.

AC indicates abdominal circumference; AUC, area under the receiver operating characteristic curve; BPD, biparietal diameter; CNN, convolutional neural network; CNS, central nervous system; CTG, cardiotocography; FHR, fetal heart rate; FL, femur length; FOAC-Net, Fetal Organ Anomaly Classification Network; HC, head circumference; ICC, intraclass correlation coefficient; Sens, sensitivity; Spec, specificity; U-Net, encoder-decoder neural network architecture; VGG, Visual Geometry Group network.

quality, sample size adequacy, and population representativeness[7] (Fig. 1).

This review provides a comprehensive examination of AI applications across obstetrics, organized by clinical domain, followed by a critical analysis of implementation challenges, ethical considerations, and future directions.

AI IN OBSTETRIC IMAGING AND FETAL BIOMETRY

Automated fetal biometry represents one of the most mature applications of AI in obstetrics, with multiple systems demonstrating accuracy comparable to experienced sonographers.[8] Deep learning models can identify standard imaging planes and measure biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL) with minimal operator intervention, potentially reducing examination time by 30% to 50% while improving measurement consistency[9] (Table 1).

A 2025 study by Venturini et al[10] introduced a paradigm shift in biometric assessment by aggregating measurements from every frame across an entire ultrasound scan rather than relying on single operator-selected images. Using convolutional neural networks to classify each frame of ultrasound video recordings, the system measured fetal biometrics in every frame where appropriate anatomy was visible and used Bayesian estimation to determine the true biometric value from thousands of measurements while probabilistically rejecting outliers. Evaluated retrospectively on 1457 recordings comprising 48 million frames from 20-week scans, this whole-examination approach achieved human-level performance in estimating fetal biometrics with well-calibrated credible intervals.[10]

Recent work has specifically addressed repeatability and reproducibility—essential prerequisites for clinical adoption. A 2024 study examining AI-assisted 3-dimensional ultrasound for fetal head biometry found excellent intra-observer agreement (intraclass correlation coefficient >0.95) and inter-observer reliability, with automated measurements showing lower variability than manual techniques.[11] However, the study found that image quality has a significant impact on AI performance, with suboptimal image acquisition leading to measurement failures or inaccuracies.[11]

A fundamental challenge in AI-enabled biometry involves generalizability across different populations, ultrasound equipment, and clinical settings. Sendra-Balcells et al[21] conducted an investigation of fetal ultrasound deep learning model transferability from high-income to low-resource settings. A classifier trained on 1792 patients from Spain was evaluated on 1008 patients in Denmark under optimal conditions, achieving an average AUC $>95\%$ when training samples exceeded 500.[21] However, when applied to 5 African centers with 25 patients each, performance was inconsistent and substantially lower. Implementing transfer learning domain adaptation by fine-tuning the Spain-trained model using small local datasets enabled performance recovery. This finding demonstrates that AI models can be adapted to new populations and imaging conditions but require the acquisition of local validation data and model recalibration.

Portable ultrasound devices combined with AI show particular promise for expanding access in low-resource settings. A 2024 validation study of Butterfly iQ and Clarius C3 portable ultrasound machines for obstetric biometry in

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

the United States and Zambia found high accuracy compared with standard ultrasound equipment, and limits of agreement within clinically acceptable ranges.[22] This supports the feasibility of AI-enhanced point-of-care ultrasound for settings where access to conventional imaging is limited.

Fetal Anomaly Detection: Deep Learning Performance

Artificial intelligence has substantially advanced the detection of structural fetal anomalies, particularly in the central nervous system and congenital cardiac disease[2,11,19,23] (Table 1).

Ultrasound-based anomaly detection has progressed markedly with systems such as Fetal-Net, which integrates multiscale convolutional neural networks with transformer layers trained on over 12,000 expert-annotated images.[24] This system achieved 97.5% accuracy, 96.5% precision, and 97.8% recall for anomaly detection across varied anatomic planes and ultrasound equipment.[24] U-Net architectures combined with VGG networks achieved 91.5% to 92.9% accuracy for fetal brain anomaly classification in multiple studies.[12,13] Real-time clinical systems such as PAICS have translated these advances into practical tools, achieving diagnostic performance comparable to expert sonologists while reducing interpretation time.[12,13]

Fetal MRI has benefited from the Fetal Organ Anomaly Classification Network (FOAC-Net), which achieved 85.3% accuracy in organ-level anomaly identification, substantially exceeding conventional architectures.[14] Performance varied by system: cardiac anomalies demonstrated 82.2% precision, spinal cord defects 85.8% precision with 89.3% recall, while brain abnormalities showed 85% precision with lower recall due to heterogeneous pathology.[14]

Congenital heart disease, affecting 1.5 per 1000 live births, is a high-priority AI application.[2] Deep learning models for fetal echocardiography achieve cardiac segmentation with mean Dice coefficients of 0.890 and sensitivity exceeding 90%.[2] A 2024 validation study demonstrated that higher image quality significantly improved agreement between AI and expert assessment ($P<0.001$).[25] AI-enabled standard plane detection, capable of automatically identifying and retrieving clinically relevant cardiac views from continuous ultrasound video streams, substantially reduces sonographers cognitive load by automating the mechanics of image acquisition, allowing attention to focus on anatomic interpretation and anomaly recognition.[25] A recent meta-analysis on the diagnostic performance of AI-enabled prenatal cardiac ultrasound included fifteen studies and found that AI models performed better than operators with lower expertise and were nearly comparable to experts (pooled sensitivity of 0.89 and specificity of 0.91).[2]

A particularly promising frontier is the application of AI-enhanced CHD screening in resource-limited settings, where access to pediatric cardiologists and specialized fetal echocardiography expertise remains severely constrained.[26,27] The combination of portable, battery-operated ultrasound systems with integrated AI interpretation offers the potential to enable frontline health care workers—including midwives and community health extension workers—to conduct standardized cardiac screening with automated referral protocols flagging suspected anomalies for specialist evaluation.[26,27]

AI for Prediction of Preeclampsia

Early, personalized risk prediction for preeclampsia, preterm birth, and fetal growth restriction represents a critical need in contemporary obstetrics.[3,28,29] Machine learning models that integrate electronic health records, maternal characteristics, and increasingly omics and wearable sensor data now provide a foundation for multivariate risk assessment across the pregnancy continuum (Table 1).

Preeclampsia prediction exemplifies artificial intelligence's potential for early risk stratification enabling timely preventive interventions. The Fetal Medicine Foundation competing-risks algorithm, integrating maternal characteristics, mean arterial pressure, uterine artery pulsatility index, and placental biomarkers (PAPP-A, PlGF), is promising, but poorly implemented.[30,31] Recent work has explored whether artificial neural networks can enhance predictive performance by leveraging both cell-free DNA biomarkers and continuous physiological data from wearable sensors.[15]

A recent study analyzed the prospective multicenter SMART study (17,520 participants), comparing neural networks with logistic regression for preeclampsia prediction using thirteen maternal variables and 2 cfDNA parameters.[15] Results demonstrated phenotype-specific advantage. For preterm preeclampsia, neural networks demonstrated clinically meaningful superiority with 58% sensitivity at 15% screen-positive rate. For term preeclampsia, both models showed limited utility (AUC ~0.71).[15] These findings underscore that the artificial intelligence advantage depends on the disease mechanism and does not represent universal improvement.

A 2025 US-based validation study of the FMF algorithm in over 8000 nulliparous women from the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be cohort reported AUC 0.75 for preterm preeclampsia, with sensitivity 64.0% and specificity 72.5% at optimal cutoff.[16] Calibration analysis revealed the model underestimated risk, though recalibration improved alignment.[16] This underscores the critical importance of population-specific validation and recalibration when applying algorithms developed in one population to another.

Machine learning approaches now integrate real-time wearable data to augment preeclampsia detection. A 2025 Scripps Research study published in *Lancet eBioMedicine* analyzed smartwatch and fitness tracker data (Apple Watch, Garmin, Fitbit) from over 5000 participants, with a subcohort of individuals providing detailed longitudinal data from before pregnancy through 6 months postpartum across 42 US states.[32] Researchers identified physiological patterns that aligned with the fluctuation of key pregnancy hormones, including estrogen, progesterone, and human chorionic gonadotropin.[32] Heart rate data showed characteristic patterns throughout pregnancy, with exploratory analysis revealing that pregnancies ending in adverse outcomes, such as miscarriage or stillbirth, demonstrated different heart rate patterns compared with healthy pregnancies.[32] The integration of continuous wearable monitoring with traditional biomarkers and ultrasound parameters offers potential for enhanced risk identification. However, systematic reviews identify several barriers, including sensor calibration variability, data privacy and security concerns, user compliance challenges, and limited real-world validation in free-living conditions.[33]

For preeclampsia specifically, continuous blood pressure monitoring via wearables offers theoretical value in

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

resource-limited settings where limited antenatal visits make early detection challenging. Most pregnant mothers in low- and middle-income countries do not have personal blood pressure devices and depend on readings during infrequent antenatal visits, where early preeclampsia detection is often missed. However, rigorous validation in these populations and real-world deployment studies remain limited.

AI for Prediction of Preterm Birth

Machine learning models for preterm birth prediction achieve moderate discriminative performance (AUC 0.63 to 0.75), limited by multifactorial etiology, with prior preterm birth, multiple gestation, and body mass index emerging as consistent predictors.[17]

An innovative application of machine learning to twin pregnancy growth discordance was conducted by Prasad et al,[18] employing unsupervised algorithms to identify distinct trajectories of intertwin fetal growth discordance. Using k-means clustering, 5 distinct discordance patterns were identified: low-stable, moderate-stable, high-stable, early-increasing, and late-increasing trajectories. Consistent high discordance (> 25%) was associated with significantly increased rates of adverse perinatal outcomes, including small for gestational age, neonatal intensive care admission, and perinatal mortality[18] (Table 1).

Complementary work by Lopian et al[34] developed machine learning models for the accurate prediction of growth-restricted neonates at term. The study integrated multimodal third-trimester data, including biometry, Doppler indices, and maternal factors, to predict small-for-gestational-age and growth restriction, with associated adverse perinatal outcomes.[34] This exemplifies how machine learning can extract clinically meaningful patterns from longitudinal surveillance data, enabling more nuanced risk stratification than traditional static assessments.

AI in Fetal Monitoring and Cardiotocography

Cardiotocography (CTG), the primary intrapartum fetal surveillance method, suffers from substantial inter- and intra-observer variability, with interpretation heavily dependent on clinician experience.[35,36] AI-assisted CTG interpretation aims to address this limitation through objective, standardized analysis (Table 1).

Recent studies demonstrate that machine learning models can predict fetal well-being from CTG signals with high accuracy.[20] Google Research developed end-to-end neural network models predicting both objective measures (fetal arterial cord blood pH, fetal acidosis) and subjective measures (Apgar scores), evaluating performance with varying inputs, including fetal heart rate (FHR) only, FHR plus uterine contractions (UC), and FHR plus UC plus metadata.[20] A novel 2025 AI algorithm for interpreting FHR and uterine activity achieved high precision, in identifying accelerations, decelerations, and contractions using deep learning and rule-based techniques.[19]

Despite promising performance metrics, clinical implementation of AI-CTG faces challenges. A 2024 meta-analysis noted that while AI excels in technical performance, gaps remain in external validation and prospective clinical trials demonstrating improved maternal-fetal outcomes.[3] In addition, clinician acceptance requires trust in AI recommendations, necessitating explainable AI approaches that provide rationale rather than simple outputs.

Challenges in External Validation and Generalizability

Despite the exponential proliferation of AI studies in obstetrics, rigorous external validation remains largely absent from the literature. A 2024 systematic review examining machine learning studies for fetal growth restriction and small-for-gestational-age prediction identified critical methodological deficiencies: fewer than one-third of studies adhered to established Delphi consensus definitions for outcomes, only 25% met recommended sample size requirements, and adherence to TRIPOD+AI reporting guidelines demonstrated consistent shortcomings in addressing model fairness, population heterogeneity, and calibration assessment.[37] This pattern may reflect a broader phenomenon whereby researchers prioritize algorithm development over the laborious, resource-intensive work of external validation, thereby creating a gap between proof-of-concept demonstrations and clinically deployable tools.

Regulatory analysis reveals the magnitude of this challenge. Approximately 43% of FDA-approved AI medical devices lack published clinical validation data, with many relying on phantom images or computer-generated data rather than genuine patient populations.[4] Notably, obstetrics and gynecology represent only 1.2% of FDA device approvals by specialty classification, with merely 3 devices focusing specifically on fetal or obstetric ultrasound applications.[38] This marked underrepresentation, despite obstetrics accounting for substantial health care utilization and generating rich imaging datasets, highlights critical regulatory and developmental gaps in translating AI research into approved medical devices.

When external validation is attempted, performance degradation is the norm rather than the exception. Prospective studies consistently demonstrate that algorithms achieving AUC values exceeding 0.90 in development cohorts frequently decline to 0.70 to 0.80 when applied to new institutions, populations, or imaging equipment.[21,39] This phenomenon, termed "domain shift," arises from differences in patient demographics, disease prevalence, imaging protocols, equipment manufacturers, and clinical workflows between development and deployment settings.[40] The critical clinical implication is that high internal validation performance provides insufficient evidence for real-world implementation—external validation across multiple independent cohorts representing diverse populations is a prerequisite for responsible clinical adoption.

Ethical, Legal, and Regulatory Considerations

The rapid deployment of AI in clinical obstetrics raises profound questions regarding patient autonomy, informed consent, data governance, and algorithmic fairness. These considerations are not peripheral to implementation; they are central to the responsible translation of research into practice.

A fundamental ethical imperative involves ensuring AI benefits are equitably distributed across all pregnant populations. Training datasets predominantly derived from high-income countries, tertiary centers, and specific demographic groups risk amplifying existing health care disparities.[41] When algorithms trained predominantly on European populations are deployed to underrepresented groups, performance degradation is well-documented, with sensitivity differences exceeding 10 to 15 percentage points in some studies.[42] This phenomenon manifests across multiple dimensions: racial and ethnic minorities,

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

differences of body mass index, and low-resource versus high-resource health care systems all experience differential algorithm performance.[42]

Addressing algorithmic bias requires intentional, equity-centered development practices.[43] These include: deliberate oversampling of underrepresented populations during training, routine, disaggregated reporting of performance metrics across demographic subgroups, and prospective validation in diverse settings, including community hospitals and low-resource contexts.[43] Importantly, bias mitigation is not a one-time intervention but requires continuous monitoring and recalibration as populations, clinical practices, and disease patterns evolve.

Data Governance, Privacy, and Consent

Obstetric datasets encompass exceptionally sensitive information, including medical histories, genetic data, imaging, biomarkers, and pregnancy outcomes, which carry implications not only for the pregnant individual but also for their offspring and family members. Robust data governance frameworks must strike a balance between the societal benefits of AI development and individual privacy rights and autonomy. Contemporary frameworks, including the General Data Protection Regulation in Europe and evolving guidance from the FDA in the United States, emphasize that patients possess fundamental rights: to understand when algorithms influence clinical decisions, to know how their data contributes to model training, and to receive explanations for algorithmic recommendations.[44,45]

Regulatory pathways for AI medical devices remain in evolution, with frameworks struggling to keep pace with technological advancement. The FDA distinguishes between "locked" algorithms (fixed after training) and "continuously learning" systems that adapt based on new data.[45] Most current obstetric device approvals utilize locked algorithms through the 510(k) premarket notification pathway, which requires demonstrating substantial equivalence to existing predicate devices.[45] This approach, however, proves inadequate for genuinely novel AI applications lacking clear precedents.

Critical legal questions regarding new AI technologies remain unresolved, such as: Who bears liability when algorithmic recommendations contribute to adverse outcomes? What documentation standards ensure transparency and accountability? These ambiguities create legal and professional uncertainties that substantially inhibit clinical adoption, even when algorithms demonstrate strong performance in validation studies.

Professional societies, including ISUOG have emphasized through educational initiatives and research publications that AI should augment rather than replace clinical expertise, that clinicians retain ultimate responsibility for patient care, and that algorithms must provide explainable outputs enabling professional verification.[29] These principles establish important boundaries, positioning AI as a decision support system rather than an autonomous decision-maker, thereby making a critical distinction for maintaining the physician-patient relationship and ensuring professional accountability.

Implementation Challenges and Workflow Integration

The path from validated algorithm to routine clinical use involves substantial technical, organizational, and human factors challenges. Even well-validated AI tools frequently fail during implementation, not due to technical deficiencies but because of workflow incompatibility, inadequate infrastructure, or insufficient stakeholder engagement.[46]

Successful implementation requires seamless integration with existing ultrasound equipment, picture archiving and communication systems, and electronic health records.[46] Many health care institutions operate heterogeneous IT systems with legacy systems lacking standardized data formats or application programming interfaces necessary for AI integration.[46] The result is solutions that require substantial investment in technical infrastructure, which is frequently unavailable, particularly in community hospitals and low-resource settings.

Surveys consistently reveal mixed attitudes among clinicians towards AI in obstetrics, with concerns about over-reliance undermining clinical skills, liability uncertainties, and distrust of algorithmic recommendations.[28,47] These concerns are not unfounded resistance to innovation but legitimate questions about how AI changes the professional practice. Addressing these issues requires transparent performance reporting, the meaningful involvement of clinicians in algorithm development, and education initiatives that foster AI literacy without requiring technical expertise.

A particular challenge involves "automation bias," defined as the tendency to over-rely on automated systems even when outputs conflict with clinical judgment.[48] This phenomenon, well-documented in aviation and other domains, poses risks in obstetrics, where algorithms may fail in edge cases or unanticipated scenarios. Conversely, "algorithm aversion" (rejecting algorithmic recommendations even when superior to human judgment) also occurs, particularly after observing algorithmic errors.[49] Optimal implementation requires cultivating appropriate trust: neither blind acceptance nor reflexive rejection, but rather critical engagement and an understanding of the algorithm's strengths and limitations.

Future Directions

The trajectory of AI in obstetrics points toward increasingly sophisticated, multimodal systems that integrate diverse data streams for comprehensive, personalized pregnancy care.

Federated learning (an approach where algorithms train across multiple institutions without centralizing sensitive patient data) represents a transformative approach for developing globally representative models while preserving privacy.[50,51] Distributed learning enables institutions to collaboratively develop and validate algorithms without the logistical complexities of data sharing.[51] Early applications in medical imaging demonstrate feasibility.[52] For obstetrics, federated approaches could enable rare disease research where no single center possesses sufficient cases, global validation incorporating diverse populations and practice settings, and continuous learning as algorithms improve with expanding real-world experience.

Next-generation obstetric AI will synthesize imaging, continuous physiological monitoring via wearables, genomics, proteomics, metabolomics, electronic health records, and social determinants of health.[53,54] This multimodal integration promises precision medicine approaches where interventions are tailored to individual risk profiles. For example, combining ultrasound-derived placental volumetry, maternal genomic variants associated with preeclampsia susceptibility, continuous blood pressure monitoring from wearables, and biomarker trajectories could enable individualized aspirin prophylaxis, prescribing preventive therapy specifically to those most likely to benefit while avoiding unnecessary treatment and potential side effects in low-risk individuals.

Machine learning excels at identifying complex, nonlinear interactions across heterogeneous data types, relationships that are difficult or impossible to detect through traditional statistical approaches. However, realizing multimodal AI's potential requires substantial infrastructure investment, standardized data formats enabling interoperability, and carefully designed studies validating that algorithmic predictions translate to improved clinical decision-making and patient outcomes.[54]

AI for Global Health and Low-resource Settings

The global burden of maternal and perinatal mortality falls disproportionately on low- and middle-income countries.[55] AI-enabled point-of-care ultrasound, combined with telemedicine linking frontline workers to specialist expertise, offers transformative potential for expanding access. Recent studies have demonstrated that minimally trained operators using AI-assisted portable ultrasound can achieve gestational age estimation and fetal presentation assessment with accuracy comparable to that of expert sonographers.[56] Scaling these innovations requires addressing infrastructure challenges, including connectivity, power availability, and device durability, as well as local capacity building, ensuring communities can operate, interpret, and maintain AI-enabled tools.

Critically, global health applications must prioritize equity-focused development: algorithms trained on data from target populations, validated in real-world deployment conditions, and involving participatory design with local health care workers and communities. Without careful attention to specific contextual needs and limitations, AI risks becoming another technology that widens rather than reduces global health disparities.

CONCLUSION

Artificial intelligence has transitioned from experimental technology to a clinically viable tool across multiple obstetric domains. Deep learning algorithms now achieve expert-level diagnostic performance in automated biometry, anomaly detection, and risk stratification for major pregnancy complications. Rigorous multicenter research has established consensus definitions, validated explainable AI systems prospectively, and demonstrated the feasibility of federated learning approaches. These efforts represent the methodological rigor and clinical pragmatism necessary for responsible translation from research to practice.

Yet critical gaps persist that demand urgent attention. The majority of published obstetric AI studies lack external validation, employ inadequate sample sizes, and fail to address algorithmic bias systematically. Fewer than 5% of FDA-approved AI medical devices serve the field of obstetrics, despite a substantial clinical need and rich data availability. Most importantly, prospective randomized trials demonstrating that AI improves maternal and perinatal outcomes, rather than just improving prediction metrics, remain rare.

This evidence gap reveals that technical capability has substantially outpaced clinical validation infrastructure, regulatory maturity, and equity considerations.

Realizing AI's transformative potential requires 3 fundamental commitments. First, rigorous validation methodology: multicenter, multinational, prospective studies with transparent performance reporting stratified across diverse populations and health care settings. Second, explainability should be the standard, not the exception: algorithms must provide clinically meaningful explanations that enable verification of reasoning while maintaining professional oversight and accountability. Third, equity-centered development prioritizing underrepresented populations, continuous fairness monitoring, and real-world validation in resource-limited settings—the essential prerequisite for ensuring AI bridges rather than widens existing health disparities.

The central concern is not whether AI will change obstetric practice—since change is already underway—but how we ensure this change promotes health equity, maintains clinical safety, preserves professional expertise, and improves outcomes for all pregnant individuals, regardless of their location, socioeconomic background, or demographic characteristics. This demands unprecedented collaboration among clinicians, data scientists, ethicists, regulators, patients, and communities. Guided by a commitment to transparency, accountability, and equity, artificial intelligence can help achieve a more accessible and truly personalized obstetric care for every pregnancy worldwide.

REFERENCES

1. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271–e297.
2. D'Alberti E, Patey O, Smith C, et al. Artificial intelligence-enabled prenatal ultrasound for the detection of fetal cardiac abnormalities: a systematic review and meta-analysis. *EClinicalMedicine*. 2025;84:103250.
3. Lin X, Liang C, Liu J, et al. Artificial Intelligence–augmented clinical decision support systems for pregnancy care: systematic review. *J Med Internet Res*. 2024;26:e54737.
4. Muralidharan V, Adewale BA, Huang CJ, et al. A scoping review of reporting gaps in FDA-approved AI medical devices. *NPJ Digit Med*. 2024;7:273.
5. Sarris I, Ioannou C, Chamberlain P, et al. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol*. 2012;39:266–273.
6. Della Ripa S, Santos N, Walker D. AI-enabled obstetric point-of-care ultrasound as an emerging technology in low- and middle-income countries: provider and health system perspectives. *BMC Pregnancy Childbirth*. 2025;25:729.
7. Riley RD, Ensor J, Snell KIE, et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. *Lancet Digit Health*. 2025;7:100857.
8. Lee C, Willis A, Chen C, et al. Development of a machine learning model for sonographic assessment of gestational age. *JAMA Netw Open*. 2023;6:2248685.
9. Ambroise Grandjean G, Hossu G, Bertholdt C, et al. Artificial intelligence assistance for fetal head biometry: Assessment of automated measurement software. *Diagn Interv Imaging*. 2018; 99:709–716.
10. Venturini L, Budd S, Farruggia A, et al. Whole examination AI estimation of fetal biometrics from 20-week ultrasound scans. *npj Digit Med*. 2025;8:1.
11. Mlodawski J, Zmelonek-Znamirowska A, Mlodawska M, et al. Repeatability and reproducibility of artificial intelligence-

acquired fetal brain measurements (SonoCNS) in the second and third trimesters of pregnancy. *Sci Rep*. 2024;14:1.
12. Lin M, Zhou Q, Lei T, et al. Deep learning system improved detection efficacy of fetal intracranial malformations in a randomized controlled trial. *npj Digit Med*. 2023;6:1.
13. Lin M, He X, Guo H, et al. Use of real-time artificial intelligence in detection of abnormal image patterns in standard sonographic reference planes in screening for fetal intracranial malformations. *Ultrasound Obstet Gynecol*. 2022;59:304–316.
14. Lo J, Lim A, Wagner MW, et al. Fetal Organ Anomaly Classification Network for Identifying Organ Anomalies in Fetal MRI. *Front Artif Intell*. 2022;5:832485.
15. Khalil A, Bellesia G, Norton ME, , et al The role of cell-free DNA biomarkers and patient data in the early prediction of preeclampsia: an artificial intelligence model. Accessed November 1, 2025. Available from: https://doi.org/10.1016/j.ajog.2024.02.299
16. Martins JG, Miller E, Aboukhater D, et al. Performance of a first-trimester combined screening for preterm preeclampsia in the United States population using the fetal medicine foundation competing risks model. *Am J Obstet Gynecol MFM*. 2025;7:101803.
17. Bitar G, Liu W, Tunguhan J, et al. A machine learning algorithm using clinical and demographic data for all-cause preterm birth prediction. *Am J Perinatol*. 2024;41(S 01):E3115–E3123.
18. Prasad S, Ayhan I, Mohammed D, et al. Longitudinal twin growth discordance patterns and adverse perinatal outcomes. *Am J Obstet Gynecol*. 2025;233:73.e1–73.e14.
19. Pardasani R, Vitullo R, Harris S, et al. Development of a novel artificial intelligence algorithm for interpreting fetal heart rate and uterine activity data in cardiotocography. *Front Digit Health*. 2025;7:1638424.
20. Chiou N, Young-Lin N, Kelly C, et al. Development and evaluation of deep learning models for cardiotocography interpretation. *npj Women's Health*. 2025;3:1.
21. Sendra-Balcells C, Campello VM, Torrents-Barrena J, et al. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. *Sci Rep*. 2023;13:1.
22. Rittenhouse KJ, Vwalika B, Sebastião Y, et al. Accuracy of portable ultrasound for obstetric biometry. *Ultrasound Obstet Gynecol*. 2024;63:772.
23. Day TG, Matthew J, Budd SF, , et al Artificial Intelligence to Assist in the Screening Fetal Anomaly Ultrasound Scan (PROMETHEUS): A Randomised Controlled Trial. Accessed Jun 18, 2025. Available from: https://doi.org/10.1101/2024.05.23.24307329
24. Islam U, Ali YA, Al-Razgan M, et al. Fetal-Net: enhancing maternal-fetal ultrasound interpretation through multi-scale convolutional neural networks and transformers. *Sci Rep*. 2025;15. Accessed November 1, 2025. https://pubmed.ncbi.nlm.nih.gov/40665017/
25. Taksoee-Vester CA, Mikolaj K, Bashir Z, et al. AI supported fetal echocardiography with quality assessment. *Sci Rep*. 2024;14:1.
26. Gomes RG, Bellington Vwalika B, Lee C, et al. AI system for fetal ultrasound in low-resource settings. 2022. Accessed November 1, 2025. https://arxiv.org/abs/2203.10139v1
27. Valderrama CE, Ketabi N, Marzbanrad F, et al. A review of fetal cardiac monitoring, with a focus on low- and middle-income countries. *Physiol Meas*. 2020;41:11TR01.
28. Fischer A, Rietveld A, Teunissen P, et al. What is the future of artificial intelligence in obstetrics? A qualitative study among healthcare professionals. *BMJ Open*. 2023;13:e076017.
29. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol*. 2020;56:498–505.
30. O'gorman N, Wright D, Syngelaki A, et al Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation. 2015. Accessed November 1, 2025. http://dx.doi.org/10.1016/j.ajog.2015.08.034
31. Mazer Zumaeta A, Wright A, Syngelaki A, et al. Screening for pre-eclampsia at 11-13 weeks' gestation: use of pregnancy-associated plasma protein-A, placental growth factor or both. *Ultrasound Obstet Gynecol*. 2020;56:400–407.
32. Milan G, Ariniello L, Baca-Motes K, et al. Association between wearable sensor signals and expected hormonal changes in pregnancy. *EBioMedicine*. 2025;119. Accessed November 2, 2025. https://pubmed.ncbi.nlm.nih.gov/40876228/
33. Alim A, Imtiaz MH. Wearable sensors for the monitoring of maternal health—a systematic review. *Sensors (Basel)*. 2023;23:2411.
34. Lopian M, Ulusoy CO, Prasad S, et al. Accurate prediction of growth-restricted neonates at term using machine learning. *Am J Obstet Gynecol*. 2025;232:e170–e173.
35. Schiermeier S, Westhof G, Leven A, et al. Intra- and interobserver variability of intrapartum cardiotocography: a multicenter study comparing the FIGO Classification with Computer Analysis Software. *Gynecol Obstet Invest*. 2011;72:169–173.
36. Alfirevic Z, Devane D, Gyte GML, et al. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst Rev*. 2017;2017. Accessed November 1, 2025. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD006066.pub3/full
37. Zamagni G, Stampalija T. Machine learning methods to predict foetal growth restriction and small for gestational age: a systematic and critical review. *Italian J Gynaecol Obstetr*. 2024;36(supplement 03):3.
38. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Health*. 2023;5:e618–e626.
39. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *npj Digit Med*. 2022;5:1.
40. Roland T, Böck C, Tschoellitsch T, et al. Domain shifts in machine learning based Covid-19 diagnosis from blood tests. *J Med Syst*. 2022;46:23.
41. Hasanzadeh F, Josephson CB, Waters G, et al. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *NPJ Digit Med*. 2025;8. Accessed November 1, 2025. https://pubmed.ncbi.nlm.nih.gov/40069303/
42. Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLoS Digital Health [Internet]*. 2022;1:e0000022.
43. Laguitan RVM, Bernardino GD. Can AI bridge or widen maternal health inequities? *Public Health Challenges* . 2025;4:e70119.
44. Van Kolfschooten HB. A health-conformant reading of the GDPR's right not to be subject to automated decision-making. *Med Law Rev*. 2024;32:373–391.
45. Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles | FDA [Internet]. Accessed November 1, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles
46. Ye J, Woods D, Jordan N, et al. The role of artificial intelligence for the application of integrating electronic health records and patient-generated data in clinical decision support. *AMIA Jt Summits Transl Sci Proc*. 2024;2024:459.
47. Day TG, Matthew J, Budd S, et al. Sonographer interaction with artificial intelligence: collaboration or conflict? *Ultrasound Obstet Gynecol*. 2023;62:167–174.
48. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2011;19:121.
49. Filiz I, Judek JR, Lorenz M, et al. The extent of algorithm aversion in decision-making situations with varying gravity. *PLoS One*. 2023;18:e0278751.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

50. Pati S, Kumar S, Varma A, et al. Privacy preservation for federated learning in health care. *Patterns*. 2024;5:100974.
51. Sarma KV, Harmon S, Sanford T, et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J Am Med Inform Assoc*. 2021;28:1259–1264.
52. Adnan M, Kalra S, Cresswell JC, et al. Federated learning and differential privacy for medical image analysis. *Sci Rep*. 2022;12:1.
53. Ravindra NG, Espinosa C, Berson E, et al. Deep representation learning identifies associations between physical activity and sleep patterns during pregnancy and prematurity. *NPJ Digit Med*. 2023;6. Accessed November 1, 2025. https://pubmed.ncbi.nlm.nih.gov/37770643/
54. Kiang MV, Victor A. The role of artificial intelligence in maternal and child health: Progress, controversies, and future directions. *PLoS Digital Health*. 2025;4:e0000938.
55. Shartyanie NP, Hanifa IN, Khan N. Digital health interventions in emergency obstetric and newborn care services in low- and middle-income countries: scoping review. *J Med Internet Res*. 2025;27. Accessed November 2, 2025. https://pubmed.ncbi.nlm.nih.gov/41150886/
56. Stringer JSA, Pokaprakarn T, Prieto JC, et al. Diagnostic accuracy of an integrated AI tool to estimate gestational age from blind ultrasound sweeps. *JAMA*. 2024;332:649–657.