Central Medical College Journal
Vol 9 No 2 July 2025
DOI: https://doi.org/10.3329/cemecj.v9i2.90819

Review Article:

Artificial Intelligence in Obstetrics and Gynecology: Current Applications and Future Prospects

Aosaf S[1], Nahar S[2]

Abstract

Artificial intelligence is moving from demonstration projects to clinically relevant tools across obstetrics and gynecology. The specialty is especially well suited to algorithmic support because it combines image-heavy workflows, longitudinal electronic records, time-series monitoring, pathology, operative video, and urgent decisions that are often made under uncertainty. This review synthesizes current applications of artificial intelligence in fetal ultrasound, pregnancy risk prediction, intrapartum monitoring, cervical cancer screening, gynecologic oncology, benign gynecologic imaging, assisted reproductive technology, and large language model-enabled education and communication. Recent evidence indicates that the most mature use cases are image-based and tightly bounded tasks such as gestational-age estimation from ultrasound sweeps, cytology triage, and structured embryo assessment. Risk-prediction models for preeclampsia, postpartum hemorrhage, and fetal acidemia are increasingly sophisticated, but their translational value depends less on headline area-under-the-curve results than on calibration, external validation, workflow integration, and the quality of the actions triggered by the prediction. In gynecology, artificial intelligence is strengthening screening and subspecialty diagnostics rather than replacing expert judgment. In reproductive medicine, algorithmic embryo scoring is approaching routine adjunctive use, yet outcome prediction remains constrained by the fact that implantation and live birth depend on much more than embryo morphology alone. Across domains, the central implementation barriers are dataset shift, limited representativeness, labeling quality, automation bias, regulatory uncertainty, and uneven digital infrastructure. The near future of artificial intelligence in obstetrics and gynecology will likely be defined by multimodal models, human-centered interfaces, prospective impact trials, privacy-preserving multicenter collaboration, and equity-focused development. The strongest conclusion from the current literature is that artificial intelligence is most valuable when it functions as a carefully monitored clinical partner rather than an autonomous substitute for obstetric and gynecologic expertise.

Keywords: *artificial intelligence; obstetrics; gynecology; fetal ultrasound; cervical cancer screening; machine learning; reproductive medicine; IVF; digital health*

1. Introduction: Obstetrics and gynecology sits at the intersection of routine population-scale care and high-acuity emergencies. A normal antenatal visit, a labor ward tracing, a colposcopy image, an IVF embryo video, and a gynecologic pathology slide all contain information that can alter management, yet the expertise needed to interpret those signals is not evenly distributed. This imbalance is one reason artificial intelligence has attracted unusual attention in women's health. The field generates rich digital data, but it also faces workforce shortages, interobserver variability, limited access to subspecialists, and large global disparities in outcomes[1-3].

The public-health case for improvement is strong. The World Health Organization (WHO) reported that in 2023 more than 700 women died every day from preventable causes related to pregnancy and childbirth, and that over 90% of maternal deaths occurred in low- and lower-middle-income countries [1]. Cervical cancer remains another stark example of preventable harm: WHO estimated roughly 660,000 new cases and 350,000 deaths in 2022, with the heaviest burden falling on settings that lack robust vaccination, screening, and treatment systems[2,3]. In both obstetrics and gynecology, therefore, the attraction of AI is not simply computational novelty. It is the possibility of making skilled interpretation more available, more consistent, and sometimes earlier.

1. **Sabab Aosaf,** *MSc student, Dept. of Computer Science Engineering (CSE), Bangladesh University of Engineering & Technology (BUET), Dhaka*
2. **Prof. Dr. Shamsun nahar,** *Professor & Head, Dept. of Obstetrics & Gynecology, Central Medical College, Cumilla*

Correspondence: Prof. Dr. Shamsun nahar *Mobile: +88 01716551941, Email: drsn101@gmail.com*

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

In this review, the term artificial intelligence includes machine learning, deep learning, computer vision, natural language processing, and newer generative systems such as large language models. In practical clinical terms, these tools are being used in four main ways: first, to automate measurements or classifications that experts already perform; second, to estimate future risk from combinations of weak signals; third, to support workflow through triage and prioritization; and fourth, to generate or organize language for education and documentation. Each use case raises different expectations. An algorithm that marks fetal biometric landmarks is being asked to solve a very different problem from a model that predicts severe preeclampsia or drafts a patient information sheet.

The aim of this paper is to provide a substantially expanded, critical narrative review of current AI applications in obstetrics and gynecology and to assess the field's future prospects. Rather than treating every reported model as equally mature, the discussion focuses on where evidence is strongest, where implementation barriers are most serious, and why several highly publicized systems still remain better described as research assets than clinical products. Throughout the paper, the central argument is that the decisive questions are no longer whether algorithms can perform well on retrospective datasets, but whether they can be trusted, monitored, and integrated in ways that improve real care without worsening inequity.

2. Review Approach and Scope

This paper is a narrative review rather than a formal systematic review. It synthesizes peer-reviewed studies, systematic reviews, meta-analyses, prospective validation studies, and official guidance from WHO, the U.S. Food and Drug Administration (FDA), and related regulatory bodies published mainly between 2020 and early 2026, while also incorporating selected earlier landmark papers when they remain highly relevant [1-16]. Priority was given to work with one or more of the following features: clinically interpretable endpoints, external validation, prospective assessment, multicenter design, or explicit regulatory significance.

That scope has consequences. A narrative review allows wider clinical interpretation, but it is also more selective than a systematic review and cannot claim exhaustive coverage of every technical paper. In AI-enabled medicine, this matters because publication volume is high and terminology is inconsistent across computer science and clinical journals. Furthermore, performance claims often depend on data curation decisions that are difficult to compare across studies. For this reason, the present review emphasizes clinical maturity over benchmark novelty and gives particular attention to reporting quality, calibration, representativeness, and deployment readiness.

Several broader reviews already indicate why this distinction matters. Recent obstetric and gynecologic overviews describe rapid growth in AI publications but also emphasize that comparatively few models have reached robust real-world use, especially when judged against standards for external validation, bias assessment, and lifecycle monitoring[12-16]. The present review builds on that literature but rewrites the synthesis in original language and with a stronger focus on clinical translation, implementation science, and future design choices.

3. Why Obstetrics and Gynecology is a Distinctive Domain for AI

Few specialties combine as many AI-friendly data types as obstetrics and gynecology. Prenatal care produces ultrasound images and videos, laboratory trends, blood-pressure trajectories, and free-text notes over time. Labor wards generate continuous cardiotocography and electronic fetal monitoring traces. Gynecology adds colposcopy images, cervical cytology slides, pelvic ultrasound, magnetic resonance imaging, histopathology, and increasingly surgical video. Reproductive medicine contributes highly standardized laboratory imaging, including time-lapse embryo videos and structured embryology annotations. Each modality invites a different class of model, but together they create a landscape in which AI can act as a second reader, measurement engine, risk stratifier, triage system, or language assistant.

Yet the same variety that makes the specialty attractive also makes evaluation difficult. A strong classification score on curated cervical cytology images may tell us little about how a system behaves on routine slides with artifacts. A fetal ultrasound model trained on one vendor's machine may falter when probe quality, operator technique, or maternal body habitus changes. A risk model for postpartum hemorrhage can appear impressive on retrospective area-under-the-curve statistics while remaining

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

clinically unhelpful if it overestimates risk at the threshold used to mobilize blood products. In other words, data abundance does not remove the need for careful clinical framing.

The past five years have produced a more coherent methodological response to these issues. SPIRIT-AI and CONSORT-AI extended trial design and reporting guidance for AI interventions, while TRIPOD+AI updated expectations for clinical prediction models that use regression or machine learning[9-11]. These frameworks are especially relevant to obstetrics and gynecology because they push authors beyond narrow accuracy claims and toward details that determine usability: intended population, missing-data handling, model updating, human oversight, failure modes, and outcome relevance. In a field where consequences can include emergency delivery, operative intervention, cancer referral, or embryo selection, that kind of rigor is not a luxury. It is the basis for trust.

Table 1. Representative AI use cases in obstetrics and gynecology

Domain	Representative task	Main data type(s)	Current maturity	Illustrative references
Prenatal ultrasound	Gestational-age estimation, biometry, plane recognition, anomaly support	B-mode images, cineloop sweeps, video	One of the most advanced clinical areas; strongest evidence for bounded tasks	[17-25]
Pregnancy risk prediction	Preeclampsia, postpartum hemorrhage, severe maternal morbidity	EHR variables, vitals, laboratory data, obstetric history	Promising but still limited by calibration and external validation gaps	[26-31]
Intrapartum monitoring	Prediction of fetal acidemia and abnormal labor patterns	Cardiotocography, fetal heart-rate time series	Methodologically active but outcome-impact evidence remains limited	[32-34]
Cervical screening	Cytology triage, HPV-positive triage, colposcopy assistance	Slide images, cervix images, screening metadata	High translational potential with clear workflow fit	[35-38]
Gynecologic oncology	Adnexal mass risk, MRI staging, digital pathology	Ultrasound, MRI, histopathology images	Useful specialist support emerging; strong need for multicenter validation	[39-42]
Reproductive medicine	Embryo assessment and ranking	Time-lapse embryo video, morphology, laboratory metadata	Commercial translation underway, but endpoint complexity limits certainty	[45-47]
Generative AI	Educational support, drafting, patient communication	Text prompts, notes, educational materials	Useful under supervision; not reliable enough for autonomous counseling	[48-50]

4. Current Applications in Obstetrics

Obstetrics contains some of the best-known clinical AI use cases because the specialty blends image interpretation, longitudinal risk assessment, and urgent physiologic monitoring. Even so, the obstetric literature is not uniform in maturity. Ultrasound support is moving into prospective evaluation and even low-resource testing, whereas many prediction models remain retrospective and center-specific. The following subsections separate those domains because they pose different translational problems.

4.1 Fetal Ultrasound, Prenatal Imaging, and Measurement Automation

Fetal ultrasound is an unusually strong match for computer vision because a large portion of routine scanning involves repeated pattern-recognition tasks: finding a standard plane, identifying landmarks, measuring structures, and flagging images that look abnormal. Recent reviews describe rapid progress in plane recognition, segmentation, biometric measurement, and anomaly support[19,20]. Importantly, these tasks are not only technically tractable; they are also clinically meaningful because ultrasound access is often limited by the availability of trained operators rather than by hardware alone.

Among the most consequential recent advances is AI-based gestational-age estimation from blind ultrasound sweeps. In the 2024 JAMA study by Stringer and colleagues, an integrated AI tool estimated gestational age from sweep data with accuracy comparable to standard sonographer biometry in pregnancies between 14 and 27 weeks, and novice users achieved similar performance to trained sonographers[17]. This work built on earlier NEJM Evidence data showing that blind-sweep AI could outperform biometry in low-resource settings, including scans collected by untrained users with low-cost devices[18]. The significance is practical rather than merely statistical. Many patients present late, may not know their last menstrual period with confidence, and receive care where expert sonography is unavailable. A dating tool that works with simplified acquisition could improve triage, referral, and timing of intervention in exactly those settings where conventional ultrasound expertise is thinnest.

AI is also being applied to more specialized prenatal imaging problems. Taksøe-Vester and colleagues demonstrated that automated cardiac biometrics could improve prenatal screening for coarctation of the aorta, a lesion whose subtle sonographic features are often missed on routine examination [21]. Xie and colleagues showed that deep learning can classify fetal brain ultrasound images as normal or

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

abnormal while localizing suspicious regions[22]. These are promising examples because they target areas where expert review is scarce and missed diagnoses are costly. At the same time, most anomaly-oriented models face a tougher generalization problem than dating or biometry. Rare pathologies, class imbalance, operator dependence, and the diversity of fetal position all complicate external performance.

Another underappreciated use of ultrasound AI is education. Lei and colleagues reported that a real-time AI system enhanced trainee performance in obstetric ultrasound, suggesting that algorithms can improve not only final interpretation but also the acquisition process itself[23]. This educational role may be especially valuable in low-volume training environments or where novice clinicians must learn essential obstetric scanning quickly. Recent work on AI-assisted birth-weight estimation at term and broader reviews on fetal growth disorders point in the same direction: the most immediate benefit of ultrasound AI may come from raising the consistency of everyday scanning rather than from replacing expert anomaly diagnosis [24,25].

Taken together, current evidence suggests that prenatal ultrasound is one of the nearest-term translational successes in obstetric AI. The most realistic implementations are bounded systems that assist acquisition, automate measurement, or estimate gestational age under constrained conditions. Fully autonomous anomaly diagnosis remains farther away. It requires much broader validation across devices, populations, and care settings, as well as careful design for communicating uncertainty when image quality is poor.

4.2 Prediction of Preeclampsia and Other Adverse Pregnancy Outcomes

Risk prediction is the largest and perhaps the most conceptually ambitious branch of obstetric AI. Pregnancy complications such as preeclampsia, gestational diabetes, spontaneous preterm birth, fetal growth restriction, and severe maternal morbidity arise from interacting biological and social factors that are difficult to summarize with simple rule-based tools. This complexity has motivated a shift from traditional regression toward machine learning models that can identify nonlinear patterns across maternal characteristics, laboratory values, blood pressure trends, and historical data [12-14].

Preeclampsia has been the leading test case. Recent systematic reviews suggest that machine learning often outperforms classical regression in discrimination for preeclampsia prediction, particularly when multimodal features are used [26,27]. Yet those same reviews repeatedly identify the field's central weakness: many studies are retrospective, single-center, and inadequately reported, with limited external validation and incomplete calibration assessment. That weakness is more than a methodological footnote. Obstetric decisions are threshold-sensitive. A model can achieve an attractive area under the receiver operating characteristic curve yet still mislead clinicians if it systematically overpredicts or underpredicts risk around the threshold that would trigger aspirin prophylaxis, transfer, heightened surveillance, or early delivery.

The broader first-trimester literature reinforces this caution. In a recent systematic review and meta-analysis, van Eekhout and colleagues found only moderate discrimination for first-trimester prediction models targeting preeclampsia, small-for-gestational-age birth, spontaneous preterm birth, and gestational diabetes; most models also carried high risk of bias, and calibration was infrequently reported[28]. This should temper the tendency to equate statistical sophistication with clinical readiness. In pregnancy risk prediction, what matters is not simply whether a model can sort patients broadly into lower and higher risk groups, but whether it does so in a way that remains stable across settings and supports a clinically actionable response.

Some studies nonetheless show what more useful models can look like. The PIERS-ML model for women with established preeclampsia improved stratification of short-term maternal risk, especially at the lowest and highest ends of the risk spectrum [29]. This type of model is potentially valuable because it is tied to a time-sensitive decision environment: who requires transfer, intensive monitoring, or immediate escalation of care within the next forty-eight hours. As a result, it points toward a future in which AI contributes less as a generic forecasting engine and more as a context-specific triage tool embedded within an established clinical pathway.

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

4.3 Postpartum Hemorrhage, Severe Maternal Morbidity, and Obstetric Deterioration

Postpartum hemorrhage illustrates another reason clinicians are interested in AI. Existing risk assessment tools are often static and coarse, classifying patients as low or high risk based on a limited set of factors. In practice, severe bleeding can still arise in patients who appear low risk at admission. Interpretable machine-learning approaches attempt to improve on this by continuously combining signals from the electronic health record, labor characteristics, prior history, and physiologic data.

Lengerich and colleagues developed an interpretable model that predicted postpartum hemorrhage with severe maternal morbidity in a lower-risk laboring population, highlighting clinically plausible contributors rather than only producing a black-box score[30]. That emphasis on interpretability matters in obstetrics, where clinicians need to understand whether a prediction reflects anemia, prolonged labor, uterine overdistension, operative delivery risk, or some combination of factors. Meyer and colleagues further showed that external validation of postpartum hemorrhage models can reveal important variability in transportability across electronic health record datasets[31]. The lesson is clear: local performance checks are not optional. Even apparently universal obstetric endpoints are shaped by institutional documentation, transfusion practice, case mix, and trigger definitions.

These findings suggest a useful design principle for maternal-risk AI. The score itself should not be treated as the clinical product. The product is the workflow it activates: heightened nursing attention, blood-bank preparation, senior review, or more frequent reassessment. Without that linkage, predictive analytics risk becoming another alert layer that adds cognitive load without changing outcomes.

4.4 Intrapartum Monitoring and Cardiotocography

Intrapartum electronic fetal monitoring remains one of the most debated monitoring technologies in obstetrics. Cardiotocography is widespread, but its interpretation is subjective, its false-positive burden is substantial, and its relationship to meaningful neonatal outcomes is imperfect. That combination has made labor ward traces a major target for machine learning[32-34].

Recent reviews show both promise and fragmentation. Francis and colleagues described a heterogeneous literature in which models, endpoints, and datasets vary widely, limiting comparison and clinical translation[33]. Aeberhard and colleagues reached a similar conclusion, noting that many cardiotocography studies remain small or methodologically inconsistent[34]. Against this background, the deep-learning study by McCoy and colleagues is notable because it used intrapartum fetal heart rate monitoring to predict acidemia at birth with good discrimination[32]. Such work suggests that algorithms may eventually standardize part of a monitoring process long criticized for interobserver inconsistency.

Still, intrapartum AI should be approached carefully. Biochemical acidemia is not itself the endpoint patients care about; the meaningful outcomes are neurologic injury, avoidable operative delivery, and neonatal well-being. A model that predicts a surrogate endpoint may still worsen care if it amplifies intervention without improving newborn outcomes. For this reason, intrapartum monitoring will likely require some of the strictest prospective impact testing in all of obstetric AI.

5. Current Applications in Gynecology

In gynecology, AI is progressing fastest in areas where image interpretation is repetitive, specialist capacity is limited, and diagnostic delay carries major consequences. Cervical cancer screening is the clearest example, but the field also now includes adnexal mass assessment, endometrial cancer imaging and pathology, benign gynecologic ultrasound, and emerging surgical applications. As in obstetrics, however, the maturity of these tools varies sharply by task.

5.1 Cervical Cancer Screening, Cytology, and Colposcopy

Cervical cancer prevention is one of the strongest population-health arguments for AI in women's health. WHO's elimination strategy sets explicit 90-70-90 targets for vaccination, screening, and treatment, yet many health systems still struggle with pathologist availability, cytology throughput, and consistent colposcopic assessment [2,3]. Because cervical screening generates large numbers of relatively standardized visual samples, it provides exactly the sort of environment in which algorithmic triage can produce operational benefits.

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

Evidence for cytology assistance is already substantial. Bao and colleagues reported that AI-assisted cytology detected high-grade cervical lesions with high sensitivity in a large multicenter observational study, performing at a level comparable to skilled cytologists and reducing the dependence on less experienced manual screening[35]. In HPV-positive triage, Xue and colleagues found that AI-enabled liquid-based cytology could support population-based screening workflows[36]. More recently, Wang and colleagues described an AI system capable of precise cervical cytology grading and cervical cancer detection, with strong prospective performance and measurable benefit when used to assist pathologists rather than replace them[37].

Colposcopy is following a similar path. In a resource-limited population, Chang and colleagues showed that AI-assisted colposcopy improved effectiveness, with particular gains among postmenopausal women and less experienced examiners[38]. This is clinically important because colposcopic interpretation is notoriously variable, especially when transformation zones are difficult to visualize. The likely near-term role for AI is therefore not full automation of cervical diagnosis, but smarter triage, quality assurance, prioritization of suspicious cases, and support for clinicians whose experience varies.

Cervical screening may become a template for responsible medical AI deployment. The task is high volume, the label structure is relatively established, the disease burden is significant, and the human-AI partnership is intuitive: algorithms help classify or prioritize slides and images, while clinicians retain responsibility for confirmation, sampling, and management. That combination of public-health need and bounded technical scope is one reason cervical AI has moved closer to real-world implementation than many other gynecologic applications.

5.2 Adnexal Masses and Ovarian Tumor Risk Stratification

Ultrasound evaluation of adnexal masses is another clinically attractive AI problem. Experienced sonologists can often distinguish benign from malignant masses using subtle morphologic patterns, but that level of expertise is not universally available. Algorithms therefore offer an appealing second-reader function, especially in settings where referral decisions depend on whether malignancy is suspected. He and colleagues compared machine-learning models for malignancy risk in ovarian tumors and found that although expert subjective assessment remained the best-performing method, deep-learning architectures achieved high diagnostic accuracy and approached expert-level performance[39]. Fanizzi and colleagues added an important dimension by developing an explainable model for solid adnexal masses that paired predictions with interpretable feature contributions[40]. This matters because gynecologists are more likely to trust a risk estimate when the pathway to that estimate is visible and clinically plausible rather than opaque.

The most credible role for these tools is selective enhancement of general practice rather than replacement of expert ultrasound specialists. An explainable second opinion could support referral to gynecologic oncology, reduce unnecessary anxiety and surgery for clearly benign lesions, and improve consistency where advanced expertise is sparse. To reach that point, however, models must prove robustness across scanners, operators, and populations with different baseline malignancy prevalence.

5.3 Endometrial Cancer Imaging and Gynecologic Oncology Pathology

AI is also expanding in gynecologic oncology, particularly in preoperative imaging, digital pathology, and operative planning. Restaino and colleagues reviewed this landscape and described a field moving from diagnostic classification toward broader support across the care continuum, including surgery[16]. Two specific areas illustrate both the promise and the implementation challenges.

First, radiomics-based MRI analysis may improve staging in endometrial cancer. Fang and colleagues found that multiparametric MRI radiomics improved preoperative diagnostic performance for assessing local disease extent, including deep myometrial and cervical stromal invasion[41]. These are clinically consequential endpoints because they influence surgical planning and the anticipated extent of staging. Second, pathology is beginning to benefit from privacy-preserving collaboration. Yeom and colleagues reported that federated learning on endometrial cancer pathology images was feasible and in some respects competitive with centralized approaches[42]. This is important because individual hospitals often lack enough labeled gynecologic pathology images to build robust models on their own, yet regulatory and privacy concerns may limit direct data pooling.

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

Central Medical College Journal
Vol 9 No 2 July 2025
DOI: https://doi.org/10.3329/cemecj.v9i2.90819

Even here, caution is warranted. Gynecologic oncology datasets are frequently small, enriched for tertiary-care pathology, and vulnerable to scanner and staining differences that degrade external performance. Successful deployment will therefore depend on multicenter quality control, domain adaptation, and a willingness to treat model maintenance as an ongoing responsibility rather than a one-time validation exercise.

5.4 Endometriosis and Benign Gynecologic Imaging

Benign gynecology has historically received less AI attention than cancer screening and fertility care, but that is beginning to change. Endometriosis is a particularly important target because diagnosis is often delayed, symptoms are heterogeneous, and high-quality imaging interpretation is concentrated among specialized operators. Recent review work suggests that AI could improve pattern recognition on ultrasound and MRI for endometriosis and related benign disorders, but the evidence remains early and methodologically diverse [43,44].

This relatively immature evidence base should not be mistaken for low importance. Delayed recognition of benign disease can impose years of pain, infertility, repeated consultations, and unnecessary procedures. If future datasets become larger and more carefully labeled, benign gynecologic imaging may turn out to be one of the areas where AI has large downstream quality-of-life benefits. At present, however, the field is best characterized as promising but not yet stable enough for broad deployment claims.

5.5 Gynecologic Surgery

Surgical AI in gynecology is newer and less clinically mature than imaging AI, but it deserves attention because operative video is becoming easier to capture and annotate. Pipes and colleagues describe a landscape that includes phase recognition, workflow analysis, skills assessment, instrument tracking, and decision support during gynecologic surgery[15]. The immediate opportunities are likely to be indirect rather than autonomous: objective feedback for training, semi-automated documentation, post hoc review of adverse events, and intraoperative prompts that enhance situational awareness.

Fully real-time operative assistance is a more demanding frontier. Surgical anatomy varies, visual fields change rapidly, occlusion is common, and the tolerance for error is low. Accordingly, the strongest near-term impact of AI in gynecologic surgery may come from making operative practice more measurable and teachable rather than from giving direct instructions in the moment of dissection.

6. Reproductive Medicine and IVF

Reproductive medicine is one of the most commercially active AI subfields in obstetrics and gynecology because IVF laboratories already operate in highly digitized, image-rich, and tightly protocolized environments. Embryo morphology, time-lapse imaging, semen analysis, and laboratory process data can all be captured in a structured way, making them amenable to algorithmic modeling [45].

Early enthusiasm focused on the idea that AI could reduce subjectivity in embryo grading. Zaninovic and colleagues described how machine learning could standardize parts of the embryology workflow, potentially improving consistency in embryo assessment and laboratory decision-making[45]. A later systematic review by Salih and colleagues found that AI often outperformed embryologists or clinical teams in specific predictive tasks, especially those centered on embryo images and morphology[46]. Those results help explain why fertility care has produced some of the first commercially visible AI products in women's health.

Yet the IVF literature also illustrates a recurring danger in medical AI: overinterpreting a prediction problem that is intrinsically noisy. Implantation and live birth do not depend on embryo appearance alone. Endometrial receptivity, maternal age, hormonal environment, laboratory conditions, and male factors all influence outcome. In a 2025 diagnostic meta-analysis, Mina and colleagues found only modest overall performance for AI in predicting pregnancy outcomes from embryo assessment[47]. This does not negate the value of AI, but it changes the claim. Rather than identifying the single objectively best embryo in a deterministic way, current systems are better understood as ranking embryos probabilistically and helping embryologists make more consistent choices.

Regulatory translation in this area has also begun. The FDA's list of AI-enabled medical devices includes CHLOE BLAST, an obstetrics-and-gynecology device cleared in 2025, indicating that embryo-related AI has started to enter formal device pathways[5]. Even so, routine adoption should proceed

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

with care. Fertility decisions are emotionally and financially intense, and clinicians must guard against automation bias, opaque scoring, and exaggerated claims about what embryo-centered prediction can actually deliver.

7. Large Language Models, Generative AI, Education, and Communication

The most visible recent wave of AI in medicine has come from large language models (LLMs). In obstetrics and gynecology, their relevance extends well beyond image interpretation. LLMs can draft notes, summarize literature, answer educational questions, translate text, generate patient instructions, and simulate exam-style reasoning. These abilities make them attractive to trainees and clinicians under time pressure, but they also introduce risks that differ from those posed by more conventional prediction models.

The educational potential is real. In the PERFORM study, Martinelli and colleagues found that AI systems could outperform human residents on cross-sectional obstetrics-gynecology scenarios, especially under language and time constraints[48]. This suggests a role for LLMs in question generation, formative feedback, structured review, and perhaps rapid evidence orientation. However, exam performance should not be confused with safe clinical reasoning. Specialty-specific assessment remains mixed. In urogynecology, Yadav and colleagues reported that none of the tested LLMs achieved a passing score on a self-assessment examination[49]. That gap between plausible language and reliable domain competence is central to understanding what LLMs can and cannot currently do.

Patient communication presents a similar tension. Daram and colleagues found that AI-generated gynecology education materials often failed to meet ideal readability targets even when the output appeared polished[50]. A superficially coherent patient handout can still be too complex, too generic, or subtly wrong. As a result, the best current use of LLMs in obstetrics and gynecology is supervised drafting. They can accelerate first-pass documentation, simplify repeated writing tasks, and support educational scaffolding, but they should not be treated as autonomous counselors, especially where recommendations may affect medication use, cancer screening, pregnancy management, or informed consent.

Generative systems therefore expand the AI conversation from prediction and classification to communication and judgment. Their future role in OB-GYN will depend less on raw fluency than on grounding, auditability, update mechanisms, and the quality of human review surrounding their outputs.

8. Cross-Cutting Challenges to Clinical Implementation

The literature reviewed above makes one pattern unmistakable: the hardest problems in obstetrics and gynecology AI are no longer purely algorithmic. Many models can achieve impressive retrospective performance. The limiting factors are instead representativeness, calibration, workflow fit, ongoing monitoring, and the ability to define a clinically responsible role for the human-AI team. Recent reviews of the field repeatedly emphasize these translational gaps[12-16].

8.1 External Validity, Calibration, and Clinical Actionability

External validity is particularly important in OB-GYN because local practice patterns strongly shape data. The same label-severe preeclampsia, postpartum hemorrhage, inadequate colposcopy, suspicious adnexal mass-may be operationalized differently across institutions. When models are trained in tertiary referral centers, they often inherit a case mix, imaging quality, and documentation style that is not representative of community practice. A system that looks strong in development can therefore degrade quickly when deployed elsewhere.

Calibration deserves equal emphasis. A probability estimate matters only if it corresponds to observed risk in the target population. This is crucial in pregnancy because many interventions are binary or threshold-based. A slightly miscalibrated model can trigger unnecessary admissions or, worse, false reassurance. TRIPOD+AI was developed partly to address this recurring weakness in predictive model reporting, but the field has not yet fully met that standard[11]. Future obstetric and gynecologic AI studies should treat calibration plots, decision thresholds, and clinically relevant net-benefit analyses as basic requirements rather than optional extras.

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

8.2 Bias, Equity, and Data Representation

Equity is not an optional secondary concern in women's health AI. Maternal morbidity and mortality, access to cervical screening, time to diagnosis of endometriosis, and availability of fertility care all reflect structural inequities. If training data are drawn primarily from high-resource centers, well-imaged patients, or historically advantaged populations, the resulting models may widen the very disparities they are advertised to reduce [1-3,12-16].

Bias can enter at multiple levels: who gets imaged, who gets labeled, which outcomes are recorded, and which patient groups are abundant enough for the model to learn reliably. In fetal ultrasound, image-quality differences across devices and operators may correlate with geography and resource level. In cervical screening, cytology or colposcopy labels may themselves reflect variable access to pathology confirmation. In reproductive medicine, commercial datasets may overrepresent patients treated in high-cost private systems. Because of this, fairness work in OB-GYN must go beyond subgroup accuracy tables. It should include dataset auditing, clinically meaningful error analysis, and explicit design for low-resource and historically excluded populations.

8.3 Human Factors, Trust, and Workflow Design

The best AI model can still fail if it enters the clinic in the wrong way. Labor wards are cognitively busy. Colposcopy visits are brief. IVF laboratories are tightly timed. Ultrasound examinations already require continuous visual, tactile, and communicative attention. An algorithm that interrupts workflow, explains itself poorly, or offers recommendations at the wrong moment may be ignored or, alternatively, trusted too readily. Both failure modes are dangerous.

The transparency principles articulated by FDA, Health Canada, and MHRA are therefore highly relevant to OB-GYN practice. They emphasize clear communication of intended use, performance, limitations, workflow role, and maintenance across the product lifecycle[7]. In practical terms, this means clinicians need to know what data a model expects, when its output is likely to be unreliable, how uncertainty is presented, and whether the system is meant to inform, prioritize, or recommend. It also means that developers should optimize for the performance of the human-AI team, not merely for isolated model accuracy[6,7].

8.4 Privacy, Interoperability, and Continual Learning

Many of the most promising OB-GYN applications are inherently multimodal. A clinically useful maternal-risk system may need ultrasound findings, laboratory data, blood-pressure trajectories, and prior obstetric history. A gynecologic oncology model may need radiology, pathology, operative notes, and biomarker data. Such integration is difficult in fragmented health systems where data are stored across noninteroperable platforms. Even when technical linkage is possible, privacy and governance constraints can limit multicenter model development

Federated learning offers one partial solution by allowing sites to collaborate without moving raw patient data centrally, as shown in recent endometrial pathology work[42]. But federated development does not remove the need for data standards, local validation, and lifecycle oversight. Regulatory thinking is moving in the same direction. FDA's 2025 draft guidance on AI-enabled device software functions frames evidence generation across the total product lifecycle and expects documentation that supports evaluation of safety and effectiveness before and after deployment[4]. Related guidance on predetermined change control plans recognizes that machine-learning devices may evolve over time and therefore require bounded, risk-based, and transparent updating mechanisms[8].

8.5 Evidence Standards, Regulation, and Medicolegal Responsibility

AI in obstetrics and gynecology can no longer be evaluated only as an academic prediction exercise. Once algorithms influence screening, triage, or treatment, they enter the domain of device regulation, liability, and professional accountability. FDA's guidance and public listings of AI-enabled medical devices show that the regulatory environment is maturing, but it remains stricter for a reason: clinical software can alter real decisions about pregnancy continuation, operative timing, cancer workup, and embryo selection [4,5].

Good Machine Learning Practice guidance and related transparency principles make clear that medical AI should be developed with representative data, human-centered design, lifecycle monitoring, and explicit communication of limitations [6,7].

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

Meanwhile, SPIRIT-AI and CONSORT-AI extend familiar clinical-trial expectations into AI research, and TRIPOD+AI strengthens reporting for prediction models[9-11]. Together, these frameworks imply a more demanding standard than is still common in the OB-GYN literature. It is no longer enough to publish a retrospective model with strong discrimination. Developers must show how the model fits practice, how it is maintained, what harms are plausible, and who remains responsible when the system is wrong.

That last question is particularly important. AI does not dissolve professional responsibility. In current practice, accountability still rests with clinicians and institutions, not with the model. This is one reason the most acceptable near-term tools are likely to be narrow, well-bounded adjuncts that support specific decisions rather than general autonomous systems that attempt to replace specialist judgment.

Table 2. Key implementation barriers and practical responses

Barrier	Why it matters in OB-GYN	Typical failure mode	Practical response
Dataset shift	Device type, case mix, and workflow vary sharply across sites	Loss of performance after deployment in a new hospital or region	Require external validation, local recalibration, and post-deployment monitoring
Poor calibration	Many obstetric decisions are threshold-based and time-sensitive	Over-triage, unnecessary intervention, or false reassurance	Report calibration metrics, decision curves, and clinically justified thresholds
Bias and under-representation	Women's health outcomes already reflect structural inequity	Worse performance in low-resource settings or under-represented groups	Audit subgroup error, diversify training data, and design explicitly for equity
Weak workflowfit	Labor wards, ultrasound suites, and IVF labs are fast moving environments	Alert fatigue or blind acceptance of model output	Co-design interfaces with clinicians and optimize the human-AI team
Opaque updating	Learning systems may change over time	Unnoticed performance drift or undocumented model revisions	Use lifecycle governance and bounded change-control plans
Privacy and interoperability limits	High-value models often require multimodal, multicenter data	Inability to train or maintain robust models	Adopt common data standards and consider federated collaboration
Unclear liability	Clinical responsibility cannot be delegated to software	Confusion about accountability after an adverse event	Define intended use narrowly and preserve clinician oversight in policy and training

9. Future Prospects

The next phase of AI in obstetrics and gynecology is likely to be less about single-task models and more about clinically integrated systems. Several trajectories appear especially important.

First, multimodal modeling will probably become the dominant technical direction. Many important outcomes in women's health are not visible in one data stream alone. Combining ultrasound, laboratory data, vital-sign trends, pathology, genomics, and longitudinal record information may improve risk estimation and diagnostic support, especially for conditions such as preeclampsia, fetal growth disorders, ovarian malignancy, and infertility [14,19-25,39-47]. The central challenge will be to make such systems not only accurate but also maintainable and interpretable enough for real clinical use.

Second, low-resource and edge deployment may become one of the field's most important impact pathways. Blind-sweep gestational-age estimation already demonstrates how AI can reduce dependence on highly specialized imaging expertise[17,18]. Similar approaches may eventually support basic anomaly triage, postpartum risk warning, and cervical screening in settings where specialist infrastructure is limited. If this happens thoughtfully, OB-GYN could become a leading example of AI used to narrow global inequity rather than simply optimize already well-resourced care.

Third, prospective implementation studies will matter more than additional retrospective benchmarks. The clinically meaningful question is whether AI changes outcomes, throughput, access, or cost without introducing unacceptable harms. Future studies should therefore ask pragmatic questions: Does AI-assisted cytology increase detection of high-grade lesions at the program level? Does ultrasound guidance improve the performance of newly trained users? Does a hemorrhage alert reduce response time or transfusion delay? Does AI in IVF improve cumulative live birth per transfer, not just embryo ranking? The field will mature only when such outcomes become routine endpoints.

Fourth, model governance will increasingly be treated as part of the intervention itself. As regulators and health systems gain experience, successful products will likely be those that can document their data provenance, communicate uncertainty, support audit trails, and adapt within clearly bounded change-control plans[4-8]. In practice, this means the future of OB-GYN AI is as much about infrastructure and stewardship as it is about neural-network architecture.

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

DOI: https://doi.org/10.3329/cemecj.v9i2.90819

Fifth, human-AI collaboration will likely remain the dominant design philosophy. In cervical screening, the most realistic goal is augmented reading and triage. In ultrasound, it is guided acquisition and measurement assistance. In IVF, it is more consistent embryo ranking under embryologist supervision. In labor wards, it may be real-time pattern support rather than autonomous delivery decisions. This is not a sign of technological weakness. It reflects the reality that obstetrics and gynecology involve judgment, context, ethics, and communication that cannot be reduced to a single model output.

10. Conclusion

Artificial intelligence is already changing obstetrics and gynecology, but not in a uniform way. The strongest current evidence supports bounded applications in fetal ultrasound, cervical screening, selected maternal-risk prediction tasks, embryo assessment, and a growing set of oncologic imaging and pathology tools[17-18,26-32,35-47]. Across these domains, AI is proving most useful when it improves consistency, extends specialist capacity, or highlights risk in data that humans already collect but cannot always synthesize efficiently.

At the same time, the literature makes clear that performance alone is an insufficient endpoint. Many published models still lack external validation, robust calibration, fairness auditing, transparent failure analysis, or convincing workflow integration [11-16]. The future of AI in obstetrics and gynecology will therefore depend less on isolated technical advances than on disciplined clinical evaluation, better data stewardship, and thoughtful design of the human-AI partnership.

In the near term, the field's most responsible vision is neither automation hype nor blanket skepticism. It is careful augmentation: algorithms that help clinicians scan, screen, prioritize, explain, and monitor more effectively while preserving professional oversight and patient-centered care. If developed and governed well, AI could become one of the most important enabling technologies in modern obstetrics and gynecology. If deployed carelessly, it could also entrench bias, add noise, and erode trust. The difference will be determined not only by model quality, but by the quality of the clinical systems into which those models are placed.

References

1. World Health Organization. Maternal mortality. Geneva: WHO; 2025. Available from: https:// www.who.int/news-room/fact-sheets/detail/maternal-mortality
2. World Health Organization. Cervical cancer. Geneva: WHO; 2025. Available from: https:// www.who.int/news-room/fact-sheets/detail/cervical-cancer
3. World Health Organization. Global strategy to accelerate the elimination of cervical cancer as a public health problem. Geneva: WHO; 2020. Available from: https:// www.who. int/ publications/i/item/9789240014107
4. U.S. Food and Drug Administration. Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. Draft guidance for industry and Food and Drug Administration staff. Silver Spring, MD: FDA; 2025. Available from: https:// www. fda.gov/ regulatory-information /search- fda-guidance- documents/ artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing
5. U.S. Food and Drug Administration. Artificial Intelligence-Enabled Medical Devices. Silver Spring, MD: FDA; 2025. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices
6. U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles. 2021; updated 2025. Available from: https:// www. fda.gov/ medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
7. U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Transparency for Machine Learning -Enabled Medical Devices: Guiding Principles. 2024. Available from: https:// www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding- principles

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

8. U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles. 2024. Available from: https:// www.canada. ca/en/health-canada/services/drugs-health-products/medical-devices/good-machine-learning-practice-medical-device-development/predetermined-change-control-plans-machine-learning-enabled-medical-devices.html

9. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, Ashrafian H, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature Medicine. 2020. doi:10. 1038/s 41591 -020 -1037-7

10. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. The Lancet Digital Health. 2020. doi:10. 1016/ S2589 -7500(20)30218-1

11. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024. doi:10.1136 /bmj-2023 -078378

12. Aftab N. Artificial Intelligence in Obstetrics and Gynaecology: Advancing Precision and Personalised Care. Cureus. 2025. doi: 10.7759/ cureus.86929

13. Devoe LD, et al. Current state of artificial intelligence model development in obstetrics. Obstetrics and Gynecology. 2025. doi:10. 1097/ AOG.0000000000005944

14. Mennickent D, Rodriguez A, Opazo MC, Riedel CA, Castro E, Eriz-Salinas A, et al. Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. Frontiers in Endocrinology. 2023. doi:10.3389/fendo.2023.1130139

15. Pipes GM, Hung AJ, Kim KH. Artificial intelligence and gynecologic surgery. Obstetrics and Gynecology. 2025. doi:10.1097/ AOG. 00000 00000006052

16. Restaino S, et al. Artificial intelligence in gynecological oncology from diagnosis to surgery. Cancers. 2025. doi:10.3390/cancers17071060

17. Stringer JSA, Pokaprakarn T, Prieto JC, Vwalika B, Chari SV, Sindano N, et al. Diagnostic accuracy of an integrated AI tool to estimate gestational age from blind ultrasound sweeps. JAMA. 2024. doi:10.1001/jama.2024.10770

18. Pokaprakarn T, Prieto JC, Price JT, Kasaro MP, Sindano N, Shah HR, et al. AI estimation of gestational age from blind ultrasound sweeps in low-resource settings. NEJM Evidence. 2022. doi:10.1056/evidoa2100058

19. Fiorentino MC, et al. A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical Image Analysis. 2023. doi: 10. 1016/ j.media.2022.102629

20. Xiao S, Wang Y, Li Y, et al. Application and progress of artificial intelligence in fetal ultrasound. Journal of Clinical Medicine. 2023. doi:10.3390/jcm12093298

21. Taksoe-Vester CA, et al. Role of artificial -intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta. Ultrasound in Obstetrics and Gynecology. 2024. doi:10.1002/uog.27608

22. Xie HN, Wang N, He M, Zhang L, Cai HM, Xian JB, et al. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound in Obstetrics and Gynecology. 2020. doi:10.1002/uog.21967

23. Lei T, Zheng Q, Feng J, Zhang L, Zhou Q, He M, et al. Enhancing trainee performance in obstetric ultrasound through an artificial intelligence system: randomized controlled trial. Ultrasound in Obstetrics and Gynecology. 2024. doi:10. 1002/ uog.29101

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

24. Horky A, Wasenitz M, Iacovella C, Bahlmann F, Al Naimi A. The performance of sonographic antenatal birth weight assessment assisted with artificial intelligence compared to that of manual examiners at term. Archives of Gynecology and Obstetrics. 2025. doi:10. 1007/ s00404 -025 -08042-2

25. Lugowski F, Babinska J, Stanirowski PJ, et al. Application of artificial intelligence in diagnosis and management of fetal growth disorders: a comprehensive review. Frontiers in Medicine. 2026. doi:10.3389/fmed.2025.1737391

26. Ranjbar A, et al. Machine learning models for predicting preeclampsia: a systematic review. BMC Pregnancy and Childbirth. 2024. doi:10. 1186/s12884-023-06220-1

27. Tiruneh SA, Vu TTT, Rolnik DL, Teede HJ, Enticott J. Machine learning algorithms versus classical regression models in pre-eclampsia prediction: a systematic review. Current Hypertension Reports. 2024. doi:10.1007/s11906-024-01297-1

28. van Eekhout JCA, Becking EC, Scheffer PG, Koutsoliakos I, Bax CJ, Henneman L, et al. First-trimester prediction models based on maternal characteristics for adverse pregnancy outcomes: a systematic review and meta-analysis. BJOG. 2025. doi:10.1111/1471-0528.17983

29. Montgomery-Csoban T, et al. Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study. The Lancet Digital Health. 2024. doi: 10. 1016/ S2589 -7500 (23) 002 67-4

30. Lengerich BJ, Caruana R, Painter I, Weeks WB, Sitcov K, Souter V. Interpretable machine learning predicts postpartum hemorrhage with severe maternal morbidity in a lower-risk laboring obstetric population. American Journal of Obstetrics and Gynecology MFM. 2024. doi:10. 1016/j.ajogmf.2024.101391

31. Meyer SR, et al. External validation of postpartum hemorrhage prediction models using electronic health record data. American Journal of Perinatology. 2024. doi:10.1055/a-1745-1348

32. McCoy JA, et al. Intrapartum electronic fetal heart rate monitoring to predict acidemia at birth with the use of deep learning. American Journal of Obstetrics and Gynecology. 2025. doi: 10. 1016/ j. ajog.2024.04.022

33. Francis F, Luz S, Wu H, Stock SJ, Townsend R. Machine learning on cardiotocography data to classify fetal outcomes: a scoping review. Computers in Biology and Medicine. 2024. doi:10.1016/j.compbiomed.2024.108220

34. Aeberhard JL, et al. Artificial intelligence and machine learning in cardiotocography: a scoping review. European Journal of Obstetrics, Gynecology, and Reproductive Biology. 2023. doi:10.1016/j.ejogrb.2022.12.008

35. Bao H, et al. Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: a multicenter, clinical-based, observational study. Gynecologic Oncology. 2020. doi:10.1016/j.ygyno.2020.07.099

36. Xue P, Xu HM, Tang HP, Wu WQ, Seery S, Han X, et al. Assessing artificial intelligence enabled liquid-based cytology for triaging HPV-positive women: a population-based cross-sectional study. Acta Obstetricia et Gynecologica Scandinavica. 2023. doi:10.1111/aogs.14611

37. Wang J, Yu Y, Tan Y, Wan H, Zheng N, He Z, et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. Nature Communications. 2024. doi: 10.1038/ s41467-024-48705-3

38. Chang Y, et al. Effectiveness of artificial intelligence-assisted colposcopy in a resource-limited population. Obstetrics and Gynecology. 2025. doi:10. 1097/ AOG. 00000000 00006014

39. He X, et al. Machine learning models in evaluating the malignancy risk of ovarian tumors: a comparative study. Journal of Ovarian Research. 2024. doi:10.1186/s13048-024-01544-8

40. Fanizzi A, et al. An explainable machine learning model to solid adnexal masses diagnosis based on clinical data and qualitative ultrasound indicators. Cancer Medicine. 2024. doi:10.1002/cam4.7425

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026

41. Fang R, et al. Multiparametric MRI radiomics improves preoperative diagnostic performance for local staging in patients with endometrial cancer. Abdominal Radiology. 2024. doi:10. 1007/ s00261 -023-04149-9

42. Yeom JC, Kim JH, Kim YJ, Kim J, Kim KG. A comparative study of performance between federated learning and centralized learning using pathological image of endometrial cancer. Journal of Imaging Informatics in Medicine. 2024. doi:10.1007/s10278-024-01020-1

43. Mittal S, Tong A, Young S, Jha P. Artificial intelligence applications in endometriosis imaging. Abdominal Radiology. 2025. doi:10. 1007/ s00261 -025-04897-w

44. Moro F, et al. Application of artificial intelligence to ultrasound imaging for benign gynecological disorders: systematic review. Ultrasound in Obstetrics and Gynecology. 2025. doi:10. 1002/ uog.29171

45. Zaninovic N, et al. Artificial intelligence in human in vitro fertilization and embryology. Fertility and Sterility. 2020. doi:10. 1016/j. fertnstert. 2020. 09. 157

46. Salih M, Austin C, Warty RR, Tiktin C, Rolnik DL, Momeni M, et al. Embryo selection through artificial intelligence versus embryologists: a systematic review. Human Reproduction Open. 2023. doi:10.1093/hropen/hoad031

47. Mina A, Younesi M, Doohandeh T, Darzi S, Ajabi Ardehjani N, Sheibani S, et al. Predicting pregnancy outcomes in IVF cycles: a systematic review and diagnostic meta-analysis of artificial intelligence in embryo assessment. Contraception and Reproductive Medicine. 2025. doi:10. 1186/ s40834-025-00400-4

48. Martinelli C, Giordano A, Carnevale V, Burk SR, Porto L, Vizzielli G, et al. The PERFORM study: artificial intelligence versus human residents in cross-sectional obstetrics-gynecology scenarios across languages and time constraints. Mayo Clinic Proceedings: Digital Health. 2025. doi:10.1016/j. mcpdig.2025.100206

49. Yadav GS, Pandit K, Connell PT, Erfani H, Nager CW. Comparative analysis of performance of large language models in urogynecology. Urogynecology. 2025. doi:10.1097/ SPV.0000 000000001545

50. Daram NR, Maxwell RA, D'Amato J, Massengill JC. Can artificial intelligence improve the readability of patient education information in gynecology? American Journal of Obstetrics and Gynecology. 2025. doi:10.1016/j.ajog.2025.06.047

Received date: 01/02/2026
Accepted reviewed version date: 09/03/2026