GHURKI JOURNAL OF MEDICAL
AND HEALTH SCIENCES ISSN: 3106-7360

GJMHS

Volume 2 Issue 1 (2026)

ARTIFICIAL INTELLIGENCE IN GYNECOLOGY AND WOMEN'S HEALTH: A NARRATIVE REVIEW OF CURRENT APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS

Review Article

Dr Atiqa Ijaz[1]*, Dr Farah Niaz Awan[2]
[1]MBBS, FCPS (PGR), PGR, Diagnostic Radiology, Doctors Hospital and Medical Center, Lahore Pakistan
[2]Anaesthetist, Pain Specialist and Women Medical Officer, THQ Hospital Ferozewala, Punjab, Pakistan

Corresponding Author: Dr Atiqa Ijaz, atiqaijaz572@gmail.com, MBBS, FCPS (PGR), PGR, Diagnostic Radiology, Doctors Hospital and Medical Center, Lahore Pakistan
Acknowledgement: The authors acknowledge the researchers and healthcare professionals whose published work contributed to the development of this narrative review. The authors also appreciate the institutional and academic support provided during the literature review and manuscript preparation.

Conflict of Interest: None Grant Support & Financial Support: None

ABSTRACT

Background: Artificial intelligence (AI) is being used in more and more settings in obstetrics and gynecology, providing novel screening, diagnostic, treatment planning and monitoring solutions in women's health. AI tools can potentially address the gaps in service provision in low- and middle-income countries where medical access to specialist gynecologic and reproductive health care is not always available.

Objectives: To provide an update on the applications, benefits, limitations, and ethical implications of AI in gynecologic oncology, reproductive medicine, obstetric and gynecologic imaging, and gynecologic surgery.

Methods: A narrative literature review was conducted that incorporated structured data from PubMed, Scopus, and Google Scholar, and keywords included "artificial intelligence," "machine learning," "deep learning," "gynecology," "obstetrics," and "women's health," and combinations thereof, with the vast majority of searches occurring between 2020 and 2026. Key metrics were extracted from systematic reviews, meta-analyses and primary studies reporting quantitative diagnostic data (accuracy, sensitivity, specificity or area under the curve [AUC]).

Results: Twenty-one primary sources were included, representing over 150 individual primary sources. AI-enabled cervical cancer screening had an accuracy of up to 95% in 75 combined studies. The machine learning classifiers for PCOS diagnosis showed accuracy ranging from 89-100%, and AUC of 73-100% in 31 studies. The machine-learning model for the diagnosis of endometriosis had an AUC of 0.80 during validation. The pooled AUC of testing for the presence of PM in ovarian cancer was 0.81 for the radiomics-based models and 0.87 for the models that incorporated both clinical and radiomics-based data. Embryo-selection and IVF-outcome prediction models were more consistent than unaided visual grading, but prospective validation is still lacking.

Conclusion: AI has significant potential to enhance diagnostic precision, efficiency, and personalization of women's health care in the field of gynecology; however, its responsible use demands diverse training datasets, prospective validation, clear regulatory oversight, and careful efforts aimed at equity, particularly in low-resource countries like Pakistan.

Keywords: Artificial Intelligence, Machine Learning, Gynecology, Women's Health, Reproductive Medicine, Gynecologic Oncology, Narrative Review.

GHURKI JOURNAL OF MEDICAL AND HEALTH SCIENCES ISSN: 3106-7360

GJMHS

INTRODUCTION

Artificial intelligence (AI), which includes machine learning, deep learning, and natural language processing, is revolutionizing the way healthcare is delivered, with the ability to identify patterns in complex clinical, imaging, and laboratory data that are hard for the human brain to process at scale.1 This change is underway in almost every area of obstetrics and gynecology, including the process of cervical cancer screening, embryo selection in fertility services, fetal ultrasound interpretation, and robotic surgical guidance2,3. In the last ten years what was once a small body of proof-of-concept algorithms has grown into a rapidly growing evidence base: systematic reviews now combine the results of dozens to over a hundred original studies in individual subspecialties, including gynecologic oncology and reproductive endocrinology, and their number has grown with the speed of technical advancement and clinical interest in these tools. This expansion is taking place in the face of significant unmet need in women's health. Despite the fact that many cases of cervical cancer can be prevented by early screening, it is one of the most common causes of cancer death amongst women in low- and middle-income countries (LMICs) due to the fact that polycystic ovary syndrome (PCOS) and endometriosis are often not diagnosed or diagnosed too late because of heterogeneity in presentation, invasive testing methods, and lack of subspecialist evaluation. On the other hand, health institutions that lack trained gynecologists, radiologists, embryologists, and cytopathologists are overburdened by issues of infertility, high-risk pregnancies, and gynecologic malignancy. AI-based decision support has therefore been suggested as being more than just a novelty; it could be a way to bring the skills of the experts to a program or provider that might not otherwise have timely access to it.

Concurrently, women's health is under-represented in biomedical research and the development of clinical algorithms, and concerns are growing that AI systems trained on non-representative data sets could perpetuate or even exacerbate existing sex and gender-based inequities in access to health care, or care4,5,6. The majority of large validation studies published so far come from well-resourced academic centres in high-income countries and may raise questions about the performance of these models in the more diverse imaging equipment and laboratory environments, as well as in more heterogeneous patient groups, common in LMIC countries such as Pakistan. One possibility for making specialist-level decision support available to under-resourced environments is that of community-based screening with the help of AI-assisted image interpretation, but this potential is only partly realised in the current routine practice6. The narrative review aims to integrate the recently published systematic reviews and meta-analyses on the use of AI in gynecologic oncology, reproductive medicine/infertility, obstetric and gynecologic imaging and gynecologic surgery, focusing on the quantitative diagnostic performance reported in these reviews. It also explores the ethical, regulatory and equity issues to be overcome before these tools can be safely and broadly embraced, and looks at the implications for resource limited health systems.

METHODS

The articles were searched in three databases (PubMed, Scopus, Google Scholar) from January 2020 to mid-2026, by doing a narrative literature search with structured data extraction. Supplementary manual screening of the reference list of the articles of the literature search of the identified systematic reviews was conducted. The search terms included the combinations of the terms "artificial intelligence," "machine learning," "deep learning," "gynecology," "obstetrics," "reproductive medicine," and "women's health," as well as application-specific terms like "cervical cancer screening" and "colposcopy," "in vitro fertilization," "embryo selection," "PCOS," "endometriosis," "ovarian cancer," and "ultrasound. The inclusion criteria were: (i) peer reviewed systematic reviews, meta-analysis or primary studies; (ii) explicit use of an AI, machine learning, or deep learning method to a gynecologic, obstetric, or reproductive health question; and (iii) if available, a clearly reported quantitative performance measure (e.g., diagnostic accuracy, sensitivity, specificity, or AUC). Results of studies that did not provide outcome data, conference abstracts that were not available as a full text, and non-peer-reviewed preprints, were not included unless a peer-reviewed alternative was not available for a particular application. The number of primary studies included in each systematic review or meta-analysis, the AI method(s) applied and the reported range or a pooled estimate of diagnostic performance were collated into a structured table for cross domain comparison for each systematic review/meta analysis included.

Given the rapid pace of publication in this area, and the diversity of study design, population and reference standards, this review does not follow a formal systematic review protocol (PRISMA) and does not combine results across domains into an overall summary statistic. Rather, it seeks to combine typical, contemporary, and peer-reviewed evidence and to communicate the taken-out performance data in a way that makes it easy for the reader to make his own decision on whether the information is comparable or not.

RESULTS

Twenty-one primary sources were included in this synthesis – 19 of which are systematic reviews, two are meta-analyses, and two are narrative reviews – as well as primary studies. As a group, they rely on more than 150 different underlying primary studies, from one representative computer-aided diagnosis (CAD) system to systematic reviews combining 75 studies on cervical cancer screening and 45 studies on diagnosing endometriosis with machine learning. The representative AI applications found in each of the four main domains studied are summarized in Table 1, and the quantitative diagnostic performance are extracted from the systematic review and meta-analyses found in the search in Table 2, by domain.

Table 1: Representative applications of artificial intelligence across major domains of gynecology and women's health

Domain	Representative AI applications	Reported performance / status
Gynecologic oncology	AI-assisted Pap smear/colposcopy interpretation, radiomics-based risk stratification for cervical, ovarian, and endometrial cancer, treatment-response prediction	Reported accuracies of up to ~90-95% for cervical screening in validation cohorts
Reproductive medicine / infertility	Embryo grading and selection from time-lapse imaging, ovarian stimulation/dose prediction, PCOS and endometriosis pattern recognition, IVF live-birth prediction models	Improved consistency of embryo selection versus unaided visual grading in several cohorts
Obstetric & gynecologic imaging	Automated fetal biometry and anomaly detection, endometrial receptivity assessment on ultrasound, ovarian mass characterization	Reduced inter-observer variability in several validation studies
Gynecologic surgery	Preoperative planning, intraoperative image-guided navigation, surgical skill assessment and training, analysis of robotic-assisted surgery data	Emerging evidence; largely early-stage / feasibility studies
Maternal-fetal & emergency care	Prediction of pre-eclampsia, postpartum hemorrhage, preterm birth, and gynecologic/obstetric emergencies from clinical and laboratory data	Promising discrimination in retrospective cohorts; limited prospective validation

AI in Gynecologic Oncology

Gynecologic cancers, cervical, ovarian, and endometrial are still present as major causes of morbidity and mortality among women worldwide, and early diagnosis is the most important factor in survival. An analysis of 75 peer-reviewed studies published from 2017 to 2024 showed that AI-based Pap smear analysis and colposcopy showed high diagnostic accuracy in cervical cancer screening, and there was an improvement in ovarian cancer risk stratification, ovarian cancer diagnosis and treatment planning, and endometrial cancer risk stratification and treatment planning2 (Table 2). One CAD system analysed to distinguish grades of CIN on colposcopy was able to detect 75% of lesions and 85% of those that were detected, and the accuracy was 80%18. Radiomics- and deep-learning-based machine learning models have also been developed to predict the response to treatment, and in gynecologic malignancies, to guide radiotherapy and chemoradiotherapy planning7, with imaging features alone, or in combination with clinical variables, predicting peritoneal metastasis in ovarian cancer with a pooled AUC of 0.81 (six studies)17.

Although diagnostic imaging is the primary domain of application of LLMs, they are now also being tested for their use in clinical decision making across various oncology specialties, such as gynecologic oncology, where they can be compared to existing treatment guidelines; LLMs have potential uses for patient education and getting access to information quickly, but at this time, evidence shows that they should be used in addition to, not instead of, guidelines provided by specialists8,9. AI in clinical medicine is also a topic of many publications, and a more comprehensive analysis warns against the potential for overconfidence in AI-generated suggestions and inaccuracies, which sometimes are explained by the Dunning-Kruger framework, especially if these suggestions are followed blindly by non-specialists without proper clinical oversight9. Despite the good results in validation studies in controlled environments, technical challenges such as protocol variations in image acquisition, limited generalizability of models trained on single-center data, and lack of interpretability of "black-box" deep-learning outputs for clinical use remain a major problem in gynecologic imaging.

AI in Reproductive Medicine and Infertility

One of the most active fields in which AI is being utilized to help women is assisted reproduction. New machine learning models have been developed to forecast suboptimal ovarian response and to customize gonadotropin doses in PCOS patients undergoing IVF, to avoid under- and over-stimulation11. The 10-year history of the use of AI in the IVF laboratory has a range of applications from sperm selection to assessment of the quality of the oocytes and embryos based on time-lapse imaging, and predicting the likelihood of implantation and live-birth rates; some deep-learning models for embryo selection have performed more consistently than human embryologists, who are not always reliable12,13.

AI was also directly used in the diagnosis of PCOS. Among the 31 studies analyzed and selected from an initial set of 135 that used standardized PCOS diagnostic criteria14 (Table 2), the AUC ranged from 73% to 100% and diagnostic accuracy from 89% to 100%, reflecting the challenge of diagnosis due to the very variable expression of PCOS and the absence of a single predominant biomarker. A related meta-analysis of machine-learning models for endometriosis diagnosis, that pooled 45 studies, reported a pooled validation-set AUC of 0.796 and a pooled sensitivity of 80% and a pooled specificity of 76% for clinical-feature-based models, and much higher performance (AUC 0.982) for models using genetic-regulator data16. While this is

GHURKI JOURNAL OF MEDICAL AND HEALTH SCIENCES ISSN: 3106-7360

positive progress, patient-facing research reveals mixed attitudes toward the integration of AI in fertility treatments, a survey of patients going for IVF showed overall "acceptance" of embryo selection with AI, but continued trust that the final choice should be made by their fertility specialist15.

AI for OBGYN imaging.AI in OBGYN Imaging.

In obstetrics and gynecology, Ultrasound is the cornerstone technique, and AI image analysis has been used to aid in automation of fetal biometry, the detection of fetal structural anomalies, the evaluation of endometrial receptivity before embryo transfer and the characterization of adnexal and ovarian masses1. An AI review in gynecologic imaging highlighted the potential of AI tools in decision support to minimize inter-observer variation, a known shortcoming of ultrasound interpretation, and the fact that implementation of these tools in gynecologic imaging has been slower to come than in other imaging-intensive specialties10. Another emerging application of particular relevance to time critical and resource-limited settings is the use of AI for obstetric and gynecologic emergencies, such as ectopic pregnancy, ovarian torsion, and postpartum hemorrhage19.

AI in Gynecologic Surgery

In gynecologic surgery, AI application is focused in three areas: preoperative planning with imaging-based risk models, intraoperative assistance with understanding complex surgical anatomy through semantic segmentation during laparoscopy and robot-assisted surgery; and post-surgery analysis of large data sets generated by the robotic surgery platforms to train, credential, and predict outcomes of surgery for surgeons3. While these applications are still in their early stages compared to AI applications in oncologic and reproductive clinical research, the majority of evidence stems from single-center feasibility studies, and not from larger prospective multicenter trials.

Table 2: Quantitative diagnostic performance of AI models by domain, extracted from included systematic reviews and meta-analyses

Domain	AI method(s)	Studies (n)	Reported performance	Ref.
Cervical cancer screening	AI-assisted Pap smear / colposcopy image analysis	75 (2017-2024)	Diagnostic accuracy up to 95%	[2]
Colposcopy CIN differentiation	Computer-aided diagnosis (CAD)	1 representative system	Accuracy 80%; sensitivity 85%; specificity 75%	[18]
PCOS diagnosis / classification	ML classifiers (SVM, KNN, regression, others)	31 (of 135 screened)	AUC 73-100%; accuracy 89-100%	[14]
Endometriosis diagnosis	ML models, clinical-feature based (validation set)	45	AUC 0.796; sensitivity 80%; specificity 76%	[16]
Ovarian cancer peritoneal metastasis	Radiomics-based ML	6 (5 meta-analyzed)	Pooled AUC 0.81 (radiomics); 0.87 (combined with clinical variables)	[17]

GHURKI JOURNAL OF MEDICAL AND HEALTH SCIENCES ISSN: 3106-7360

GJMHS

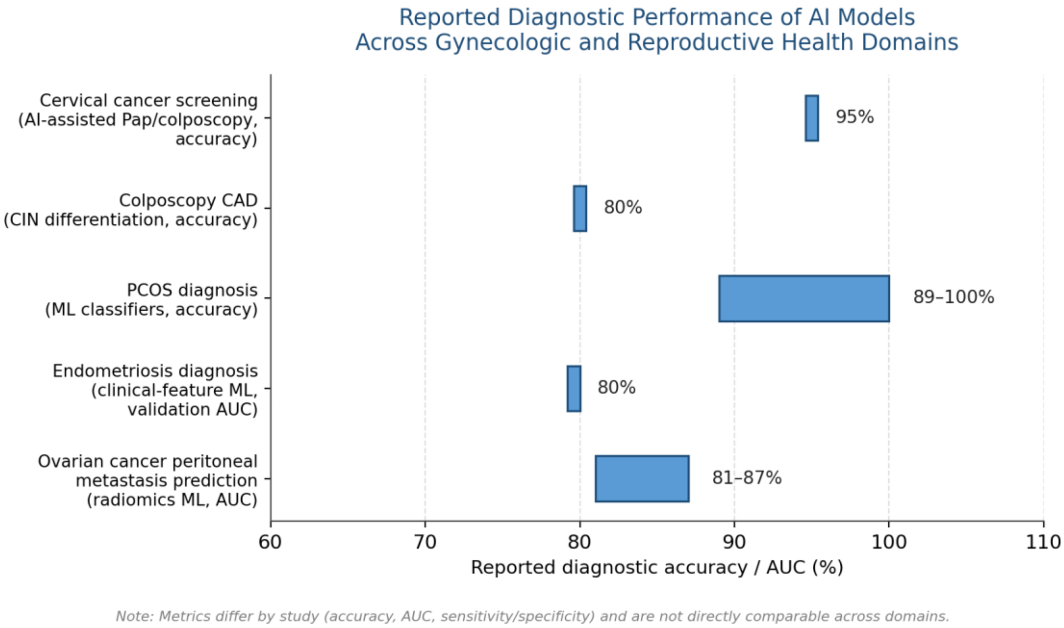

Figure 1: Reported diagnostic accuracy or AUC of AI models across gynecologic and reproductive health domains, extracted from the systematic reviews and meta-analyses in Table 2.

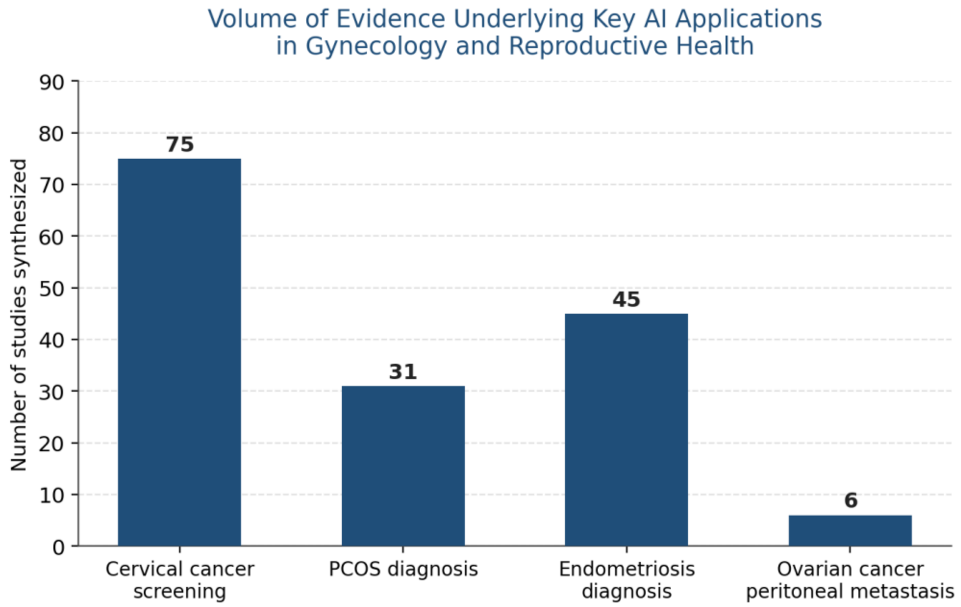

Figure 2: Number of primary studies synthesized within the systematic reviews/meta-analyses underlying four key AI applications in gynecology.

DISCUSSION

The data presented collectively suggest that AI is now much more than a nascent technology but has already proven valuable, albeit not fully confirmed, in clinical applications throughout the field of gynecologic oncology, reproductive medicine, obstetric imaging, and gynecologic surgery. The best evidence currently available is for the use of AI in cervical cancer screening, the diagnosis of PCOS and embryo selection in IVF, which are all supported by dozens of studies with consistently reported performance measures2,12,14 (Table 2, Figure 1). The evidence of AI in how surgery is guided and the real-time triage of obstetric emergencies is relatively limited and is based on a small number of single centers feasibility studies (Figure 2). Caution should be used when interpreting the performance figures obtained in this review. The metrics reported across domains are not directly comparable since some reflect diagnostic accuracy against a defined reference standard, some reflect AUC in an internal or external validation cohort, and reported ranges are often quite broad, as a result of differences in study design, patient population and definition of the outcome. Furthermore, the range of accuracy reported for PCOS classifiers14 (89-100%) represents a wide variety of studies, with different diagnostic criteria, and varying sample sizes, and the drop in the performance of the model from training set to validation set shown in the endometriosis meta-analysis (0.810 to 0.796)16 reminds us that high accuracy in the training set does not necessarily translate to high accuracy in the test set.

Ethical and regulatory protections to ensure the safe and responsible use of AI in women's health have not kept pace with its clinical use. One of the key issues is algorithmic bias: machine-learning models are trained using historical data, which can perpetuate and mirror past sex- and gender-based inequities in medical research and care, as well as in the quality of care received, especially if training data sets do not represent women from low-resource, rural, and minority communities4,5. AI principles and ethics review papers generally cite data privacy, transparency of AI systems and algorithms, informed consent, and indeterminate liability for the actions of AI systems as emerging issues that need to be addressed by regulators and health systems. 20,21. These concerns are further exacerbated in LMICs like Pakistan due to further structural challenges: inadequate digital health infrastructure, weak EMR systems, lack of staff capacity building, and financial constraints of acquiring and sustaining AI based diagnostic tools. If left unaddressed, AI can further contribute to the disparities in access to specialist care for women, especially when it comes to gynecologic and obstetric care, between high-resource and low-resource settings, in fact compounding the disparities6.

There are a number of caveats to both the evidence and this review. First, the vast majority of published AI models have been created and internally tested using data from one center or country, and therefore cannot be assumed to be generalizable across other centers or countries. Second, the number of prospective, randomized, evaluations of AI-assisted clinical decision-making is quite small compared to the number of retrospective and proof-of-concept studies. Third, deep learning models remain opaque and hinder clinicians' confidence in, and accountability for, their use. Fourth, this review is narrative rather than systematic in design, did not apply systematic review risk-of-bias tools to each of the underlying primary studies, and was based on secondary quantitative analysis of primary studies from the various systematic reviews included in Table 2 rather than primary study re-analysis; the figures in Table 2 should therefore be interpreted as an illustrative summary rather than a formal meta-analytic estimate. Future research and implementation suggestions involve prospective validation of current AI tools in multiple centers across varied populations, the creation of explainable AI models that can be easily understood by clinicians and trusted by patients, regulatory pathways tailored for AI-driven medical devices in reproductive and gynecologic education, and a conscious effort to incorporate women from LMIC and under-resourced communities into the datasets used to train and validate AI tools. Context-specific and locally developed AI solutions, like low cost AI based cervical screening tools for community based outreach, could provide the most immediate benefit in certain countries with high burden of late-presenting gynecologic disease, such as Pakistan. The key takeaway for clinicians and policymakers from this literature is that AI tools must be assessed beyond accuracy to their feasibility, affordability, and fairness in practice for their potential application in the real world with limited resources in LMICs.

CONCLUSION

From cervical cancer screening to diagnosis of systemic conditions like PCOS and endometriosis, from IVF and embryo selection to obstetric imaging, and gynecologic surgery, the impact of AI is coming through with promising quantitative data on the effectiveness of AI in improving diagnostic accuracy and efficiency across various areas. The type, quality, and generalizability of the performance metrics reported for each of these domains however, is often very different, and evidence from prospective, real-world, and LMIC-representative studies is comparatively limited. To achieve the full potential of AI responsibly, a variety of diverse and representative training data is required, prospective validation is essential, regulators need to be transparent and oversight needs to be in place, and investment in the digital health infrastructure and clinician training necessary to deploy these tools equitably – especially in low and middle-income countries like Pakistan, where the burden of preventable gynecologic and obstetric disease is highest.

REFERENCES

1. Aftab N. Artificial intelligence in obstetrics and gynaecology: advancing precision and personalised care. Cureus. 2025;17(6):e86929.
2. Paiboonborirak C, et al. Artificial intelligence in the diagnosis and management of gynecologic cancer. Int J Gynaecol Obstet. 2025.
3. Pipes GM, Hung AJ, Kim KH. Artificial intelligence and gynecologic surgery. Obstet Gynecol. 2025;146(5):672-8.
4. Joshi A. Big data and AI for gender equality in health: bias is a big challenge. Front Big Data. 2024;7:1436019.
5. Most WB. The gender bias built into AI—and its threat to women's health [Internet]. 2025.
6. dos Santos GG. Using artificial intelligence as a technological tool in gynecologic and obstetric health: a narrative literature review. Int J Gynaecol Obstet. 2025.
7. Guo L, Zhang S, Chen H, Li Y, Liu Y, Liu W, et al. Application of artificial intelligence in assisting treatment of gynecologic tumors: a systematic review. Vis Comput Ind Biomed Art. 2025.
8. Reicher L, Lutsker G, Michaan N, Grisaru D, Laskov I. Exploring the role of artificial intelligence, large language models: comparing patient-focused information and clinical decision support capabilities to the gynecologic oncology guidelines. Int J Gynaecol Obstet. 2025;168(2):419-27.

9. Pavlik EJ, Land Woodward J, Lawton F, Swiecki-Sikora AL, Ramaiah DD, Rives TA. Artificial intelligence in relation to accurate information and tasks in gynecologic oncology and clinical medicine—Dunning-Kruger effects and ultracrepidarianism. Diagnostics (Basel). 2025;15(6):735.

10. Yu J, Wang Y, Li Z, Shen Y. Advances in artificial intelligence for gynecological imaging: technical bottlenecks and future engineering solutions. Intell Med. 2026;6(3):250-9.

11. Xu X, Jiang Y, Du J, Song C, Hu Y. Development and validation of a prediction model for suboptimal ovarian response in PCOS patients undergoing GnRH-antagonist protocol in IVF/ICSI cycles. J Ovarian Res. 2024;17(1):116.

12. Jiang VS, Bormann CL. Artificial intelligence in the in vitro fertilization laboratory: a review of advancements over the last decade. Fertil Steril. 2023;120(1):17-23.

13. Zhang Q, Liang X, Chen Z. A review of artificial intelligence applications in in vitro fertilization. J Assist Reprod Genet. 2025;42(1):3-14.

14. Barrera FJ, Brown EDL, Rojo A, Obeso J, Plata H, Lincango EP, et al. Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. Front Endocrinol (Lausanne). 2023;14:1106625.

15. Cromack SC, Lew AM, Bazzetta SE, Xu S, Walter JR. The perception of artificial intelligence and infertility care among patients undergoing fertility treatment. J Assist Reprod Genet. 2025;42(3):855-63.

16. Zhang B, Lv X, Li D, Zhang L, Ru Z, Ma Y. Diagnostic accuracy of machine learning for endometriosis: a systematic review and meta-analysis. Front Endocrinol (Lausanne). 2026;17:1735567.

17. Salimi M, Vadipour P, Abdolizadeh A, Fayedeh F, Seifi S. Radiomics-based machine learning in the prediction of peritoneal metastasis in ovarian cancer: a systematic review and meta-analysis. BMC Med Imaging. 2025.

18. Brandão M, Mendes F, Martins M, Cardoso P, Macedo G, Mascarenhas T, et al. Revolutionizing women's health: a comprehensive review of artificial intelligence advancements in gynecology. J Clin Med. 2024;13(4):1061.

19. Elbiss HM, Abu-Zidan FM. Artificial intelligence in gynecologic and obstetric emergencies. Int J Emerg Med. 2025;18:20.

20. Williamson SM, Prybutok V. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. Appl Sci. 2024;14(2):675.

21. Xie Z, Wang H, Dai L, Wang Z, Song H, Qian J. Ethical issues in multi-agent AI systems for healthcare: a narrative review. Front Public Health. 2026;14:1792627.

AUTHOR's CONTRIBUTION:

Author	Contribution
Dr Atiqa Ijaz	Conceptualization, Methodology, Formal Analysis, Writing - Original Draft, Validation, Supervision
Dr Farah Niaz Awan	Methodology, Investigation, Data Curation, Writing - Review & Editing