

Application of artificial intelligence to ultrasound imaging for benign gynecological disorders: systematic review

F. MORO^{1,2} , M. T. GIUDICE¹, M. CIANCIA¹, D. ZACE³, G. BALDASSARI⁴, M. VAGNI⁵, H. E. TRAN⁴, G. SCAMBIA^{1,6} and A. C. TESTA^{1,6}

¹Dipartimento Scienze della Salute della Donna, del Bambino e di Sanità Pubblica, Fondazione Policlinico Universitario Agostino Gemelli, IRCCS, Rome, Italy; ²UniCamillus International Medical University, Rome, Italy; ³Infectious Disease Clinic, Department of Systems Medicine, Tor Vergata University, Rome, Italy; ⁴Radiomics G-STEP Research Core Facility, Fondazione Policlinico Universitario A. Gemelli, IRCCS, Rome, Italy; ⁵Istituto di Radiologia, Università Cattolica del Sacro Cuore, Rome, Italy; ⁶Dipartimento Universitario Scienze della Vita e Sanità Pubblica, Università Cattolica del Sacro Cuore, Rome, Italy

KEYWORDS: artificial intelligence; gynecology; machine learning; ultrasonography

ABSTRACT

Objective Although artificial intelligence (AI) is increasingly being applied to ultrasound imaging in gynecology, efforts to synthesize the available evidence have been inadequate. The aim of this systematic review was to summarize and evaluate the literature on the role of AI applied to ultrasound imaging in benign gynecological disorders.

Methods Web of Science, PubMed and Scopus databases were searched from inception until August 2024. Inclusion criteria were studies applying AI to ultrasound imaging in the diagnosis and management of benign gynecological disorders. Studies retrieved from the literature search were imported into Rayyan software and quality assessment was performed using the Quality Assessment Tool for Artificial Intelligence-Centered Diagnostic Test Accuracy Studies (QUADAS-AI).

Results Of the 59 studies included, 12 were on polycystic ovary syndrome (PCOS), 11 were on infertility and assisted reproductive technology, 11 were on benign ovarian pathology (i.e. ovarian cysts, ovarian torsion, premature ovarian failure), 10 were on endometrial or myometrial pathology, nine were on pelvic floor disorder and six were on endometriosis. China was the most highly represented country (22/59 (37.3%)). According to QUADAS-AI, most studies were at high risk of bias for the subject selection domain (because the sample size, source or scanner model was not specified, data were not derived from open-source datasets and/or imaging preprocessing was not performed) and the index test domain (AI models

were not validated externally), and at low risk of bias for the reference standard domain (the reference standard classified the target condition correctly) and the workflow domain (the time between the index test and the reference standard was reasonable). Most studies (40/59) developed and internally validated AI classification models for distinguishing between normal and pathological cases (i.e. presence vs absence of PCOS, pelvic endometriosis, urinary incontinence, ovarian cyst or ovarian torsion), whereas 19/59 studies aimed to automatically segment or measure ovarian follicles, ovarian volume, endometrial thickness, uterine fibroids or pelvic floor structures.

Conclusion The published literature on AI applied to ultrasound in benign gynecological disorders is focused mainly on creating classification models to distinguish between normal and pathological cases, and on developing models to automatically segment or measure ovarian volume or follicles. © 2025 The Author(s). *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.

INTRODUCTION

Ultrasound examination is the first-line method for the diagnosis and management of many gynecological diseases¹. In recent decades, ultrasound has played a fundamental role in the fields of endometriosis^{2–4}, reproductive medicine^{5–7} and pelvic floor disorders. Advances in the quality and availability of transvaginal

Correspondence: Dr F. Moro, Fondazione Policlinico Universitario Agostino Gemelli, IRCCS, Dipartimento Scienze della Salute della Donna, del Bambino e di Sanità Pubblica, Largo A. Gemelli 8, Rome, 00168, Italy (e-mail: morofrancy@gmail.com)

Accepted: 5 December 2024

ultrasound machines have led to improvements in the diagnosis of deep endometriosis^{8–11} and the monitoring of infertile women undergoing assisted reproductive technology (ART) (e.g. assessment of follicular maturity and endometrial factors affecting embryo implantation)¹². Ultrasound is the imaging method of choice in the perioperative assessment of patients with pelvic floor dysfunction to evaluate the urethra, bladder neck, rectum and anorectal junction as well as the integrity of the pelvic muscles¹³.

The introduction of artificial intelligence (AI) in imaging is revolutionizing the diagnosis and management of gynecological disease. AI includes machine learning (ML) models, which are trained to recognize patterns and relationships from input data without explicit programming¹⁴, and deep learning (DL) models, a subset of ML using artificial neural networks with multiple layers (deep architectures) capable of learning complex representations from data. Convolutional neural networks (CNNs) are a subtype of DL models that can automatically learn spatial hierarchies of characteristics from input images¹⁵. CNNs are composed of multiple layers: convolutional layers for extracting image characteristics; pooling layers for reducing dimensionality; and fully connected layers for classification¹⁶. A detailed description of key concepts in AI is provided in Appendix S1.

ML and DL techniques are often employed in radiomics research to analyze and interpret large quantities of data generated from medical images. Radiomics is a process that extracts features from medical images and provides a quantitative description of the imaging data. The radiomics workflow involves the segmentation of the region of interest (ROI) from the studied image, image processing for subsequent analysis, extraction of radiomics features from the ROI and analysis of the extracted features to identify patterns, correlations and associations with clinical outcomes¹⁷.

There are several applications of AI in ultrasound imaging, including detection (i.e. the automatic identification of organ structures and other objects of interest), classification (i.e. the analysis of ultrasound images to assess disease status or classify pathology into a specific category) and segmentation (i.e. the delineation of precise lesion boundaries, such as those of ovarian follicles or cysts)¹⁸.

Many authors have explored AI in gynecology^{19–87}, but there is a lack of synthesis of the available evidence regarding ultrasound-based AI methods. The aim of this systematic review was to synthesize and evaluate the existing evidence on how AI technologies can enhance the accuracy, efficiency and predictive capability of ultrasound imaging for the diagnosis, management and monitoring of benign gynecological disorders, considering the strengths, limitations and potential gaps that may guide future research in this field.

METHODS

This systematic review was conducted and reported according to the Preferred Reporting Items for Systematic

reviews and Meta-Analyses (PRISMA) statement. The protocol was registered with the PROSPERO database (CRD42023427088).

Search strategy

A literature search was conducted in Web of Science, PubMed and Scopus to retrieve potentially eligible articles, published from inception until 4 August 2024. A search string for PubMed was constructed from medical subject headings (MeSH) terms, keywords and free-text words, such as ‘radiomics’, ‘ultrasound-based radiomics’, ‘artificial intelligence’, ‘machine learning’, ‘deep learning’, ‘ultrasonography’, ‘gynecology’, ‘gynecological diseases’, ‘endometrium’, ‘uterus’, ‘uterine’, ‘ovary’, ‘ovarian’, ‘ovaries’ and ‘fallopian tube’. The search was restricted to studies performed on humans and reported in the English language. No other restrictions were applied. The search string was adapted for use in the two other electronic databases. The full search strategy for all databases is provided in Appendix S2.

Inclusion and exclusion criteria

We included articles reporting the role of AI applied to ultrasound in benign gynecological disorders, focusing specifically on diagnosis, outcome, image acquisition, segmentation, type of AI, model input and model performance. We excluded systematic reviews, non-empirical studies, conference abstracts, editorials, commentaries, book reviews and abstracts not accompanied by a full-text article. Furthermore, animal and modeling studies were excluded.

Study selection

All studies retrieved from the literature search were imported to Rayyan software⁸⁸ and duplicates were removed. Rayyan is a web-based tool designed to facilitate the screening and selection of studies for systematic reviews⁸⁹. We chose to use this tool because it improves efficiency, reduces bias and helps to organize documents and the decision-making process, thus ensuring a more rigorous and transparent review.

Two researchers (F.M., M.T.G.) independently performed the first round of screening based on titles and abstracts. Discrepancies were resolved by discussion. Subsequently, studies with the full text available were read in their entirety by four researchers (F.M., M.T.G., M.C., G.B.) to select the final articles to include in the review and disagreements were resolved by discussion. If more than one study was published for the same cohort with identical endpoints, the report containing the most comprehensive information on the population was included to avoid overlapping populations. The reference lists of the included studies were searched manually for additional studies. When it was not possible to retrieve the full text online, we contacted the corresponding author of the article.

Data extraction

Data extraction was performed by four researchers (F.M., M.T.G., M.V., G.B). A dedicated data extraction form was used to retrieve the following information for each eligible study: (1) study identifiers (first author, title, publication year); (2) study characteristics (study period, country, population, type of gynecological disorder); (3) objective of the study; (4) AI specifics (type of AI being assessed, model input); and (5) main findings and model performance.

Because a quantitative synthesis was not feasible because of the high heterogeneity of the included studies, resulting from differences in study design, population, interventions, outcomes and measurement methods, we produced a qualitative summary of the findings in the form of a narrative synthesis. This approach allowed us to systematically explore and describe these differences, providing a structured and comprehensive summary of the evidence that, despite the diversity of the data, offers valuable insights and a deeper understanding of the topic. The information retrieved from the included articles was categorized according to the type of AI assessed and the gynecological disease studied, and was structured using Excel spreadsheets (Microsoft, Redmond, WA, USA). The findings were summarized in a dedicated table, including the specific AI used, the study setting, the gynecological disease and the objective for each study.

For studies that developed multiple AI models, we reported the results for all models. Performance was reported as the area under the receiver-operating-characteristics curve (AUC) or, if unavailable, as accuracy, for studies aiming to discriminate between categories (e.g. presence *vs* absence of polycystic ovary syndrome (PCOS), pelvic floor disorder or pelvic endometriosis). For studies with the sole aim of automated segmentation (e.g. of the ovary or follicle), the Dice similarity coefficient (DICE) was reported. DICE is a statistical measure used to assess the similarity between two groups, such as the automated segmentation produced by an algorithm and the ground truth segmentation⁹⁰. It is calculated as follows:

$$\text{DICE} = 2 \times \frac{|A \cap B|}{|A| + |B|}$$

where A and B represent the two segmentations being compared. The DICE value ranges from 0 to 1, with 1 indicating perfect agreement and 0 indicating no overlap. We chose to report the DICE value, among other metrics, because it is the measure used most widely to assess the quality of segmentation⁸⁸. There are several other metrics that can be used to assess the performance of models. The structural similarity index is a perceptual metric used to measure the similarity between two given images, taking into account changes in structural information, luminance and contrast, and providing an assessment of image quality as perceived by human vision⁹¹. Its value ranges from -1, meaning that the compared images are very different, to 1, meaning that the compared images are the same.

Missed segmentation ratio quantifies the proportion of pixels that the algorithm fails to detect or correctly segment³². It is computed as the ratio between the number of segmented pixels and the total number of pixels inside the ground truth segmentation. Its value ranges from 0, indicating perfect segmentation, to 1, indicating failed segmentation. The intraclass correlation coefficient is a statistical measure used to assess the consistency of measurements made by different raters; its value ranges from 0, indicating no agreement, to 1, indicating perfect agreement⁹². The F1 score is a metric used to evaluate the performance of a classification model, especially in scenarios in which the balance between precision and recall is crucial; it is given by the harmonic mean of precision and recall⁸⁸. An F1 score of 1 indicates perfect precision and recall, while 0 indicates that either precision or recall is zero. The performance reported herein refers to the external validation set or, when absent, to the internal validation set; if no validation was performed, this was reported.

Quality assessment

The overall quality of the included studies was assessed using the Quality Assessment Tool for Artificial Intelligence-Centered Diagnostic Test Accuracy Studies (QUADAS-AI)⁹³. The specific criteria are listed in Table S1. This tool represents an extension and revision of QUADAS-2⁹⁴ and QUADAS-C⁹⁵ guidelines, and uses four domains (subject selection, index test, reference standard, workflow) to evaluate the risk of bias in AI-centered studies.

RESULTS

Search strategy

Our literature search retrieved a total of 4994 unique articles. After the screening and selection process, 59 studies reporting on benign gynecological conditions were accepted for inclusion in the systematic review^{29–87} (Figure 1). Of those, 12 were on PCOS, 11 were on infertility and ART, 11 were on benign ovarian pathology (i.e. ovarian cysts, ovarian torsion, premature ovarian failure), 10 were on endometrial or myometrial pathology, nine were on pelvic floor disorder and six were on endometriosis.

The results of the quality assessment of the included studies using the QUADAS-AI tool are presented in Table S2. Most studies were at high risk of bias for the subject selection domain (because the sample size or source was not specified, data were not derived from open-source datasets, imaging preprocessing was not performed and/or information on the scanner model used to acquire images was not specified) and the index test domain (the AI model was not tested in an external population in most studies). However, there was generally a low risk of bias for the reference standard domain (in most studies, the reference standard classified the target condition correctly) and the workflow domain (the time elapsed between the index test and the reference standard was generally reasonable).

Characteristics of included studies

The characteristics of the 59 included studies are summarized in Table S3. China was the most highly represented country (22 (37.3%) studies)^{30,33,37,52,55,58–62,65,69,72–78,81,82,87}, followed by India (nine (15.3%) studies)^{31,36,38,46–49,83,84}. The number of patients analyzed ranged between 16²⁹ and 25 535⁴⁴, although in 20 (33.9%) studies^{31,32,34,38,40,41,43,45–48,53,56–58,79,80,83,84,86}, only the number of images was specified.

Most papers (47 (79.7%))^{3,29–31,36,37,39,40,42,45–49,51–55,57–63,65–74,76–86} were published between 2021 and 2024; the oldest studies included in the present review were published in 2007^{43,75}. Most studies (33 (55.9%)) developed DL models, 22 (37.3%) developed ML models, and both ML and DL models were developed in four (6.8%) studies. The AI models were implemented using different programming languages, namely Python ($n=20$), MATLAB ($n=10$) and R ($n=5$), or software

applications, including SPSS ($n=2$) and MedCalc ($n=1$); for the remaining 21 studies, no information was provided. The model code was publicly available for three studies^{68,82,86}, available on request for six and presented in pseudocode in four, while no information was available in the remaining 46 studies.

Regarding manufacturer details, most studies (33 (55.9%)) did not report the vendor of the machine used to obtain the ultrasound images. Of the studies that did report the scanner model, 19/26 (73.1%) used only one scanner while 7/26 (26.9%) used two or more scanners.

Infertility and assisted reproductive technology

Most studies (6/11) on infertile women undergoing ART developed models to automatically segment or measure the ovarian follicle, measure ovarian volume or track follicle growth^{31–34,36,87}. These studies developed either DL models ($n=5$)^{31,33,34,36,87} or ML models ($n=1$)³², using two-dimensional or three-dimensional (3D) ultrasound images as the model input, and showed good-to-high performance in terms of DICE. Two studies developed models to predict hyper-response during ART cycles and included both ultrasound and clinical features as the model input: the ML model in one study had an AUC of 0.91³⁵ and the DL model in the other study had an AUC of 0.88³⁰. One study developed a ML model to differentiate between a favorable and non-favorable uterine cavity for embryo implantation using amplitude- and frequency-based features and reported accuracy of 0.94²⁹. One study developed a ML model to evaluate endometrial elasticity measured by shear-wave elastography in patients with unexplained infertility and demonstrated an AUC of 0.89³⁷. Finally, Boneš *et al.* collected 3D ultrasound images from women in the general population, infertile women and women with recurrent spontaneous miscarriage, and developed a DL model for segmentation and alignment of uterine shape, which achieved high performance in terms of DICE (0.90)⁸⁶.

Polycystic ovary syndrome

Twelve studies developed ML models ($n=7$)^{42–44,46–49}, DL models ($n=3$)^{38,39,41} or both ($n=2$)^{40,45} to discriminate between PCOS and normal ovary/ovarian cysts on ultrasound examination. The models included one or more of the following feature families as the model input: shape, intensity, ultrasound, clinical, morphological, stereological, geometrical, statistical and textural. The model performance was reported in terms of accuracy in 11 studies (ranging from 0.82 to 0.99) and as an F1 score (0.76) in one study. The largest study, performed by Cheng and Mahalingaiah in 2019⁴⁴, including 25 535 patients and 39 093 ultrasound examinations, reported accuracy of 0.98 to discriminate between PCOS and normal ovarian morphology.

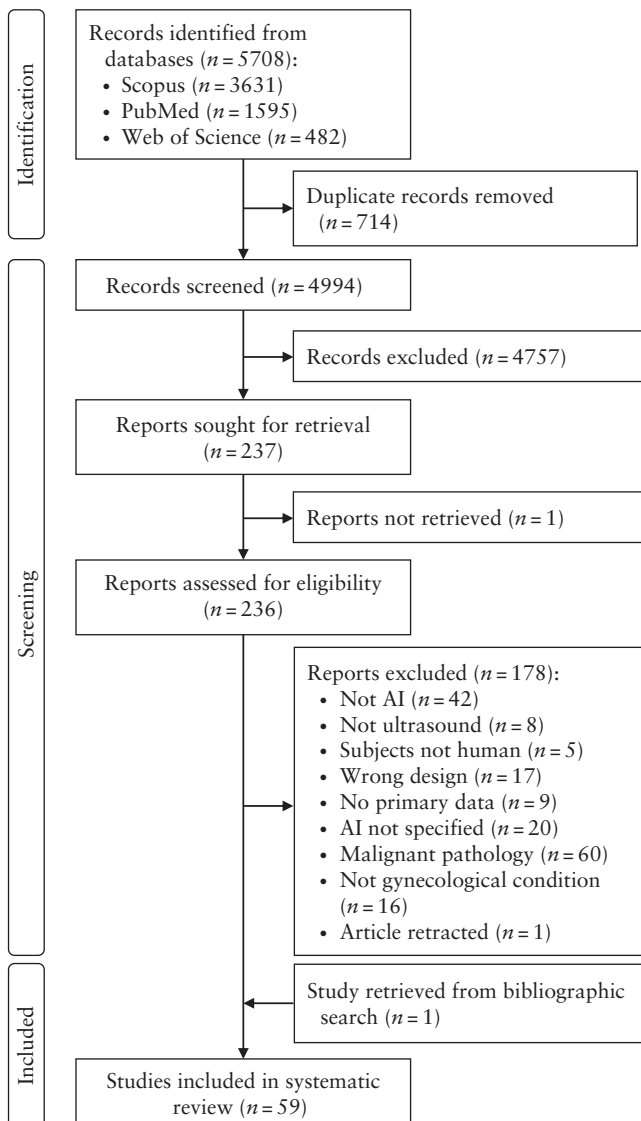


Figure 1 Flowchart summarizing inclusion of articles in systematic review. AI, artificial intelligence.

Endometriosis

Three studies developed DL models to detect deep endometriosis, of which one used ultrasound and clinical data as the model input⁵⁰ and two used only ultrasound images as the model input^{52,53}, with an AUC of 0.82, accuracy of 0.97 and DICE value of 0.82, respectively. One study developed a DL model, using ultrasound video clips as the model input, to detect indirect signs of pelvic endometriosis (i.e. sliding sign), with an AUC of 0.96⁵¹. Finally, two studies developed models to distinguish between endometrioma and other types of adnexal lesion: one used a ML model including textural features and reported an AUC of 1⁵⁴, while the other used a DL model based on ultrasound images and reported an AUC of 0.99⁵⁵.

Ovarian pathology (ovarian cysts, ovarian torsion, premature ovarian failure)

Six studies developed models to discriminate between different types of ovarian cyst^{56–61}. Sohail *et al.*⁵⁶ used an ML model including both clinical and ultrasound features to discriminate between three types of cyst, showing accuracy of 0.88. Narmatha *et al.*⁵⁷ used a DL method to distinguish between seven types of ovarian lesion, reporting accuracy of 0.97. Fan *et al.*⁵⁸ developed a DL model to discriminate between normal pelvic cyst, ovarian cyst and non-pure ovarian cyst, with accuracy of 0.96. Two studies developed models to distinguish endometrioma from other lesions: Liu *et al.*⁶⁰ used a ML model to differentiate endometrioma from ovarian dermoid cyst and reported accuracy of 0.99, whereas Miao *et al.*⁶¹ developed a DL model with an AUC of 0.90 to discriminate between endometrioma and mucinous cystadenoma. Finally, Li *et al.*⁵⁹ focused on cyst segmentation and endometrioma classification using a hybrid DL/ML model: for the first task, the authors developed a model composed of Swin Transformer and UPerNet, which showed a DICE value of 0.85; for the second task, the model comprised ResNet50 and Vision Transformer (ViT) and showed accuracy of 0.91.

Four studies developed ML models including both clinical and ultrasound features to identify adnexal torsion^{63–66}. Otjen *et al.*⁶⁴ analyzed pediatric patients treated surgically for ovarian torsion at a quaternary pediatric hospital over an 11-year period. They built a decision tree to identify torsion based on multiple features from pelvic ultrasound, yielding an AUC of 0.96. Chen *et al.*⁶⁵ analyzed pediatric patients who underwent surgical confirmation of ovarian torsion over a 5-year period. They built a ML model to identify torsion based on clinical and sonographic features, showing an AUC of 0.87. Turki and Raml⁶⁶ built a support vector machine (SVM) with a linear kernel to discriminate between ovarian torsion and acute appendicitis, with an AUC of 0.99. Atia *et al.*⁶³ developed a multivariate logistic regression model to predict adnexal torsion in patients who underwent laparoscopy for suspected torsion. The predictive performance was fair, with an AUC of 0.75.

Finally, Yu and Qing⁶² developed an AI segmentation algorithm using color Doppler ultrasound images (with peak systolic velocity, end-diastolic velocity, resistance index and pulsatility index) that showed clearly the functional status and hemodynamic characteristics of the ovaries, which was used to diagnose idiopathic premature ovarian failure.

Pelvic floor disorder

Six studies developed AI models to identify pelvic floor disorders^{69,70,72–75}. Of those, two studies developed AI models to predict stress urinary incontinence: one used a ML model including both clinical and ultrasound features and reported an AUC of 0.81⁷⁴, and the other used a DL model including only ultrasound features and reported an AUC of 0.94⁷⁵. The remaining four studies^{69,70,72,73} developed AI models including ultrasound images, video footage or clinical and ultrasound features in patients with prolapse. For example, Duan *et al.*⁷³ developed a DL algorithm including 3D ultrasound images as the model input to identify different types of pelvic organ prolapse (anterior *vs* middle *vs* posterior), reporting an AUC of 0.79. Xu *et al.*⁷² built a predictive ML model including clinical and ultrasound features to select the type of pessary (ring *vs* Gellhorn) for women with symptomatic prolapse, showing a AUC of 0.81.

Three studies focused on the identification and analysis of the anatomical structures involved in pelvic floor disorders^{67,68,71}. Of those, two developed DL models, with DICE values of 0.79⁶⁷ and 0.86⁶⁸, respectively. One study⁷¹ developed a ML model with an AUC of 0.91.

Uterine pathology

Two different DL models have been proposed for the assessment of endometrial thickness and endometrial adhesions using 3D ultrasound images^{76,77}. Wu and Zhang⁷⁷ used an extreme learning machine denoising algorithm for the diagnosis of adhesions and found that it performed similarly to hysteroscopy. Wang *et al.*⁷⁶ applied a 3D U-Net model to 3D ultrasound images to perform automatic segmentation of the endometrium and recognition of endometrial adhesions, showing high segmentation accuracy (DICE, 0.91).

Seven studies developed models for the detection of fibroids in the uterus, using a DL approach ($n = 6$)^{78–82,84} or a combined DL/ML method ($n = 1$)⁸³. Dilna *et al.*⁷⁹ used an ultrasound-based DL model to detect the presence of fibroids, with accuracy of 0.95. Huo *et al.*⁷⁸ analyzed 3870 ultrasound images from 667 patients using a deep convolutional neural network (DCNN), which performed well (AUC, 0.95). Shahzad *et al.*⁸⁰ tested a dual-path DCNN architecture on 1057 ultrasound images and reported accuracy of 0.998 in automatic fibroid detection, whereas Yang *et al.*⁸² showed a mean average precision of 0.98 for their DL model in a cohort of 871 patients. Cai *et al.*⁸¹ developed a hybrid DL model composed of MobileNetV2 and a deep convolutional generative

adversarial network for the automatic detection of uterine fibroids, with accuracy of 0.97. The convolutional recurrent neural network of Chinna and Pathrose Mary⁸⁴ had accuracy of 0.998 in detecting fibroids in uterine ultrasound images. Finally, Kaveramma *et al.*⁸³ aimed to classify ultrasound images of the uterus as fibrotic or normal using ML and DL classifiers. Among ML classifiers, SVM produced a classification accuracy of 0.93, whereas that for ViT was 0.98.

Raimondo *et al.*⁸⁵ evaluated the performance of a DL model to diagnose adenomyosis on uterine ultrasonographic images and the accuracy was low (0.51).

DISCUSSION

In the present systematic review, we summarized the findings of studies that applied AI to ultrasound imaging in benign gynecological disorders. Most studies created classification models for distinguishing between normal and pathological cases (e.g. presence *vs* absence of PCOS, endometriosis and stress urinary incontinence). Others developed models to automatically segment or measure ovarian follicles or ovarian volume, with the aim of facilitating improved surveillance of the ovarian cycle, thereby optimizing ART in infertile women. Some limitations were identified in the included studies. For example, most did not perform external validation, which is a critical step in determining the reproducibility and generalizability of a prediction model to new and varied patient settings. In addition, the number and type of variables used for modeling varied widely between studies, making it difficult to compare models and assess model relevance. In addition, most studies were conducted in a single center, resulting in a limited number of cases with poor representation of patient characteristics. In addition, AI models were often trained on specific datasets, which can introduce bias, especially if the training data do not adequately represent diverse patient populations.

Several systematic reviews have been published on the use of AI in gynecological malignancy, presenting the principal findings and most common applications of AI to various imaging modalities^{96–98}. The few systematic reviews that have been published on AI and ultrasound in the context of benign gynecological disorders have focused on a single condition^{99,100}. Avery *et al.*⁹⁹ summarized the role of modern diagnostic techniques (i.e. transvaginal ultrasound examination, combined ultrasound and magnetic resonance imaging, and AI) in endometriosis. Of the 49 studies included, only three were on ultrasound imaging and AI. In their systematic review, Barrera *et al.*¹⁰⁰ included 31 studies assessing the performance of ML algorithms in the diagnosis or classification of PCOS. Fourteen of these studies used ultrasound imaging, of which six were included in the present review; we excluded the remaining eight studies because they were reported in databases not included in our search strategy or the full text was not available.

Our review demonstrates that AI models are useful in several benign gynecological conditions, including endometriosis, endocrine disorders and pelvic floor dysfunction. Endometriosis and endocrine disorders are very common among women of reproductive age. Transvaginal ultrasound examination is usually the first-line approach and is an excellent diagnostic tool when performed by experts. However, its diagnostic accuracy depends on operator experience level, ultrasound equipment and scanning time. AI models proved to be as accurate as an expert operator and therefore can be used in settings where expert operators are not available. We also found some applications of AI in pelvic floor disorders for which diagnosis using ultrasound can be difficult. The potential to predict pelvic floor disorder would be of great value in the preventive care of patients at highest risk, namely menopausal or postpartum women.

To the best of our knowledge, this is the first systematic review dedicated specifically to AI applied to ultrasound across the spectrum of benign gynecological disorders. We conducted an extensive literature search in multiple databases to ensure the rigor of the study. We reported comprehensive data including sample size, number of images, year of publication, geographical distribution and outcomes, as well as the type of AI and the families of variables included. Moreover, we assessed the quality of studies using QUADAS-AI⁹³, which is specifically adapted for AI research. However, some limitations should be acknowledged. First, the significant heterogeneity of the studies, in terms of AI algorithms, ultrasound techniques, patient populations and outcomes, prevented us from performing a meta-analysis, which limited our ability to generalize the findings. In addition, rapid progress in the field of AI may diminish the long-term relevance of these findings. Notable gaps in the literature include the lack of longitudinal studies evaluating the long-term efficacy and safety of AI applications in ultrasound imaging for benign gynecological conditions, as well as a lack of studies addressing the practical challenges, barriers to clinical integration and ethical considerations associated with these technologies.

We believe that the present review can help readers to better understand the role of AI applied to ultrasound imaging. Incorporating AI systems into clinical practice can improve patient management and prognosis, reduce healthcare costs and reduce gynecologists' workload by increasing their efficiency and accuracy. Despite the advantages and developments described herein, the clinical translation of AI in gynecological diagnosis warrants further research and is a long-term process.

In conclusion, the published literature on AI applied to ultrasound in benign gynecological disorders is focused mainly on creating classification models to distinguish between normal and pathological cases, and on developing models to automatically segment or measure ovarian follicles or volume. The present review should help readers to better understand the applications of AI to gynecological ultrasound imaging.

ACKNOWLEDGMENTS

The authors thank Sofia Gambigliani Zoccoli for her contribution to analyzing articles and quality assessment, and Francesco D'Antonio for his supervision. Open access publishing facilitated by Università Cattolica del Sacro Cuore, as part of the Wiley - CRUI-CARE agreement.

REFERENCES

- American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 83: management of adnexal masses. *Obstet Gynecol.* 2007;110(1):201-214.
- Exacoustos C, Manganaro L, Zupi E. Imaging for the evaluation of endometriosis and adenomyosis. *Best Pract Res Clin Obstet Gynaecol.* 2014;28(5):655-681.
- Guerriero S, Ajossa S, Garau N, Alcazar JL, Mais V, Melis GB. Diagnosis of pelvic adhesions in patients with endometrioma: the role of transvaginal ultrasonography. *Fertil Steril.* 2010;94(2):742-746.
- Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. The efficiency of transvaginal ultrasonography in the diagnosis of endometrioma. *Fertil Steril.* 1993;60(5):776-780.
- Kondagari L, Kahn J, Singh M. *Sonography gynecology infertility assessment, protocols, and interpretation.* StatPearls; 2023.
- Klenov VE, Van Voorhis BJ. Ultrasound in infertility treatments. *Clin Obstet Gynecol.* 2017;60(1):108-120.
- Campbell S. Ultrasound evaluation in female infertility. *Obstet Gynecol Clin North Am.* 2019;46(4):683-696.
- Nisenblat V, Bossuyt PM, Farquhar C, Johnson N, Hull ML. Imaging modalities for the non-invasive diagnosis of endometriosis. *Cochrane Database Syst Rev.* 2016;2(2):CD009591.
- Moore J, Copley S, Morris J, Lindsell D, Golding S, Kennedy S. A systematic review of the accuracy of ultrasound in the diagnosis of endometriosis. *Ultrasound Obstet Gynecol.* 2002;20(6):630-634.
- Moro F, Leombroni M, Testa AC. Ultrasound imaging in endometriosis. *Obstet Gynecol Clin North Am.* 2019;46(4):643-659.
- Moro F, Ianieri MM, De Cicco NA, et al. Comparison of clinical and ultrasound examinations in assessing the parametria in patients with deep infiltrating endometriosis: a multicentre prospective study. *Reprod Biomed Online.* 2024;48(4):103733.
- Dishuck CF, Perchik JD, Porter KK, Gunn DD. Advanced imaging in female infertility. *Curr Urol Rep.* 2019;20(11):77.
- Bahrami S, Khatri G, Sheridan AD, et al. Pelvic floor ultrasound: when, why, and how? *Abdom Radiol (NY).* 2021;46(4):1395-1413.
- Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *RadioGraphics.* 2017;37(2):505-515.
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
- Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging.* 2018;9(4):611-629.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278(2):563-577.
- Shen YT, Chen L, Yue WW, Xu HX. Artificial intelligence in ultrasound. *Eur J Radiol.* 2021;139:109717.
- Moro F, Vagni M, Tran HE, et al. Radiomics analysis of ultrasound images to discriminate between benign and malignant adnexal masses with solid ultrasound morphology. *Ultrasound Obstet Gynecol.* 2025;65(3):353-363.
- Moro F, Albanese M, Boldrini L, et al. Developing and validating ultrasound-based radiomics models for predicting high-risk endometrial cancer. *Ultrasound Obstet Gynecol.* 2022;60(2):256-268.
- Van Calster B, Van Hoorde K, Valentin L, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ.* 2014;349:g5920.
- Lucidarne O, Akakpo JP, Granberg S, et al. A new computer-aided diagnostic tool for non-invasive characterisation of malignant ovarian masses: Results of a multicentre validation study. *Eur Radiol.* 2010;20(8):1822-1830.
- Gao Y, Zeng S, Xu X, et al. Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. *Lancet Digit Health.* 2022;4(3):e179-e187.
- Li J, Chen Y, Zhang M, et al. A deep learning model system for diagnosis and management of adnexal masses. *Cancers (Basel).* 2022;14(21):5291.
- Ruan H, Chen S, Li J, et al. Development and validation of a nomogram prediction model for endometrial malignancy in patients with abnormal uterine bleeding. *Yonsei Med J.* 2023;64(3):197-203.
- Zhou H, Wang S, Zhang T, Liu D, Yang K. Ultrasound image analysis technology under deep belief networks in evaluation on the effects of diagnosis and chemotherapy of cervical cancer. *J Supercomput.* 2021;77(4):4151-4171.
- Chiappa V, Bogani G, Interlenghi M, et al. The Adoption of Radiomics and machine learning improves the diagnostic processes of women with Ovarian MAsses (the AROMA pilot study). *J Ultrasound.* 2021;24(4):429-437.
- Fragomeni SM, Moro F, Palluzzi F, et al. Evaluating the risk of inguinal lymph node metastases before surgery using the morphonode predictive model: a prospective diagnostic study in vulvar cancer patients. *Cancers (Basel).* 2023;15(4):1121.
- Sammali F, Blank C, Bakkes TGH, et al. Multi-modal uterine-activity measurements for prediction of embryo implantation by machine learning. *IEEE Access.* 2021;9:47096-47111.
- Liang X, Liang J, Zeng F, et al. Evaluation of oocyte maturity using artificial intelligence quantification of follicle volume biomarker by three-dimensional ultrasound. *Reprod Biomed Online.* 2022;45(6):1197-1206.
- Gupta S, Suryanarayana KV, Kudavelly SR, Gottumukkala Achyuta R. Ovarian assessment using deep learning based 3D ultrasound super resolution. In: Drukker K, Mazurowski MA, eds. *Medical Imaging 2021: Computer-Aided Diagnosis.* SPIE; 2021:17.
- Chen T, Zhang W, Good S, Zhou KS, Comaniciu D. Automatic ovarian follicle quantification from 3D ultrasound data using global/local context with database guided segmentation. 2009 *IEEE 12th International Conference on Computer Vision.* IEEE; 2009:795-802.
- Li H, Fang J, Liu S, et al. CR-Unet: A composite network for ovary and follicle segmentation in ultrasound images. *IEEE J Biomed Health Inform.* 2019;24(4):974-983.
- Carvalho C, Marques S, Peixoto C, et al. Deep learning approaches for gynaecological ultrasound image segmentation: a radio-frequency vs B-mode comparison. *Image Recognition and Analysis.* Springer; 2019:295-306.
- Robertson I, Chmiel FP, Cheong Y. Streamlining follicular monitoring during controlled ovarian stimulation: a data-driven approach to efficient IVF care in the new era of social distancing. *Hum Reprod.* 2021;36(1):99-106.
- Srivastava D, Gupta S, Kudavelly SK, Ga R. Unsupervised deep learning based longitudinal follicular growth tracking during IVF cycle using 3D transvaginal ultrasound in assisted reproduction. 2021 *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* IEEE; 2021:3209-3212.
- Li ZY, Cai L, Zhang ZJ, et al. Evaluation of endometrial receptivity in women with unexplained infertility by shear wave elastography. *Insights Imaging.* 2024;15(1):85.
- Kumar HP, Srinivasan S. Classification of ovary abnormality using the probabilistic neural network (PNN). *Technol Health Care.* 2014;22(6):857-865.
- Alamoudi A, Khan IU, Aslam N, et al. A deep learning fusion approach to diagnosis of the polycystic ovary syndrome (PCOS). *Appl Computational Intell Soft Comput.* 2023;2023:1-15.
- Suha SA, Islam MN. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. *Sci Rep.* 2022;12(1):17123.
- Cahyono B, Adiwijaya MMS, Wisesty UN. An implementation of convolutional neural network on PCO classification based on ultrasound image. 2017 *5th International Conference on Information and Communication Technology (ICICT7).* IEEE; 2017:1-4.
- Madhumitha J, Kalaiyarasi M, Ram SS. Automated polycystic ovarian syndrome identification with follicle recognition. 2021 *3rd International Conference on Signal Processing and Communication (ICSPC).* IEEE; 2021:98-102.
- Lawrence MJ, Eramian MG, Pierson RA, Neufeld E. Computer assisted detection of polycystic ovary morphology in ultrasound images. *Fourth Canadian Conference on Computer and Robot Vision (CRV '07).* IEEE; 2007:105-112.
- Cheng JJ, Mahalingaiah S. Data mining polycystic ovary morphology in electronic medical record ultrasound reports. *Fertil Res Pract.* 2019;5(1):13.
- Sumathi M, Chitra P, Sheela S, Ishwarya C. Study and implementation of automated system for detection of PCOS from ultrasound scan images using artificial intelligence. *Imag Sci J.* 2024;72:828-839.
- Jeevitha S, Priya N. Identifying and classifying an ovarian cyst using SCBOD (Size and Count-Based Ovarian Detection) algorithm in ultrasound image. *Int J Elec Comput Eng Systems.* 2022;13(9):799-806.
- Gopalakrishnan C, Iyapparaja M. Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning. *Int J System Assur Eng Manag.* 2021;23:879.
- Suganya Y, Ganesan S, Valarmathi P. Comparative analysis of ovarian images classification for identification of cyst using ensemble method machine learning approach. *ECS Trans.* 2022;107(1):7407-7415.
- Kiruthika V, Sathya S, Ramya MM, Sakthidasan SK. An intelligent machine learning approach for ovarian detection and classification system using ultrasonogram images. *Eng Sci.* 2023;23:879.
- Guerriero S, Pascual M, Ajossa S, et al. Artificial intelligence (AI) in the detection of rectosigmoid deep endometriosis. *Eur J Obstet Gynecol Reprod Biol.* 2021;261:29-33.
- Maicas G, Leonardi M, Avery J, et al. Deep learning to diagnose pouch of Douglas obliteration with ultrasound sliding sign. *Reprod Fertil.* 2021;2(4):236-243.
- Yang M, Liu M, Chen Y, He S, Lin Y. Diagnostic efficacy of ultrasound combined with magnetic resonance imaging in diagnosis of deep pelvic endometriosis under deep learning. *J Supercomput.* 2021;77(7):7598-7619.
- Podda AS, Balia R, Barra S, et al. Multi-scale deep learning ensemble for segmentation of endometriotic lesions. *Neural Comput Appl.* 2024;36(24):14895-14908.
- Ştefan RA, Ştefan PA, Mişu CM, et al. Ultrasonography in the differentiation of endometriomas from hemorrhagic ovarian cysts: the role of texture analysis. *J Pers Med.* 2021;11(7):611.
- Hu P, Gao Y, Zhang Y, Sun K. Ultrasound image-based deep learning to differentiate tubal-ovarian abscess from ovarian endometriosis cyst. *Front Physiol.* 2023;14:1101810.
- Sohail ASM, Bhattacharya P, Mudur SP, Krishnamurthy S, Gilbert L. Content-based retrieval and classification of ultrasound medical images of ovarian cysts. *Artificial Neural Networks in Pattern Recognition.* Springer; 2010:173-184.
- Narmatha C, Manimegalai P, Krishnadas J, Valsalan P, Manimurugan S, Mustafa M. Ovarian cysts classification using novel deep reinforcement learning with Harris Hawks Optimization method. *J Supercomput.* 2023;79(2):1374-1397.

58. Fan J, Liu J, Chen Q, Wang W, Wu Y. Accurate ovarian cyst classification with a lightweight deep learning model for ultrasound images. *IEEE Access*. 2023;11:110681-110691.
59. Li Y, Zhao B, Wen L, Huang R, Ni D. Multi-purposed diagnostic system for ovarian endometrioma using CNN and transformer networks in ultrasound. *Biomed Signal Process Control*. 2024;91:105923.
60. Liu L, Cai W, Zhou C, et al. Ultrasound radiomics-based artificial intelligence model to assist in the differential diagnosis of ovarian endometrioma and ovarian dermoid cyst. *Front Med (Lausanne)*. 2024;11:1362588.
61. Miao K, Lv Q, Zhang L, Zhao N, Dong X. Discriminative diagnosis of ovarian endometriosis cysts and benign mucinous cystadenomas based on the ConvNeXt algorithm. *Eur J Obstet Gynecol Reprod Biol*. 2024;298:135-139.
62. Yu L, Qing X. Diagnosis of idiopathic premature ovarian failure by color doppler ultrasound under the intelligent segmentation algorithm. *Comput Math Methods Med*. 2022;2022:1-10.
63. Atia O, Hazan E, Rotem R, et al. A scoring system developed by a machine learning algorithm to better predict adnexal torsion. *J Minim Invasive Gynecol*. 2023;30(6):486-493.
64. Otjen JP, Stanescu AL, Alessio AM, Parisi MT. Ovarian torsion: developing a machine-learned algorithm for diagnosis. *Pediatr Radiol*. 2020;50(5):706-714.
65. Chen S, Gao Z, Qian Y, Chen Q. Key clinical predictors in the diagnosis of ovarian torsion in children. *J Pediatr (Rio J)*. 2024;100(4):399-405.
66. Turki A, Raml E. Enhancing pediatric adnexal torsion diagnosis: prediction method utilizing machine learning techniques. *Children*. 2023;10(10):1612.
67. García-Mejido JA, Solís-Martín D, Martín-Morán M, Fernández-Conde C, Fernández-Palacin F, Sainz-Bueno JA. Applicability of deep learning to dynamically identify the different organs of the pelvic floor in the midsagittal plane. *Int Urogynecol J*. 2024;35:2285-2293.
68. Rabbat N, Qureshi A, Hsu KT, et al. Automated segmentation of levator ani muscle from 3d endovaginal ultrasound images. *Bioengineering*. 2023;10(8):894.
69. Zhu Y, Zhang J, Ji Z, et al. Ultrasound evaluation of pelvic floor function after transumbilical laparoscopic single-site total hysterectomy using deep learning algorithm. *Comput Math Methods Med*. 2022;2022:1-10.
70. Van den Noort F, Manzini C, Hofsteenge M, Sirmacek B, Van der Vaart CH, Slump CH. Unsupervised convolutional autoencoders for 4D transperineal ultrasound classification. *J Med Imaging (Bellingham)*. 2023;10(1):014004.
71. Muta M, Takahashi T, Tamai N, et al. Pelvic floor muscle contraction automatic evaluation algorithm for pelvic floor muscle training biofeedback using self-performed ultrasound. *BMC Womens Health*. 2024;24(1):219.
72. Xu H, Wu W, Wang X, Xia Z. A predictive model of choosing pessary type for women with symptomatic pelvic organ prolapse. *Menopause*. 2021;28(11):1279-1286.
73. Duan L, Wang Y, Li J, Zhou N. Exploring the clinical diagnostic value of pelvic floor ultrasound images for pelvic organ prolapses through deep learning. *J Supercomput*. 2021;77(9):10699-10720.
74. Liu W, Qian L. Establishment and validation of a risk prediction model for postpartum stress urinary incontinence based on pelvic floor ultrasound and clinical data. *Int Urogynecol J*. 2022;33(12):3491-3497.
75. Huang YL, Chen HY. Computer-aided diagnosis of urodynamic stress incontinence with vector-based perineal ultrasound using neural networks. *Ultrasound Obstet Gynecol*. 2007;30(7):1002-1006.
76. Wang X, Bao N, Xin X, et al. Automatic evaluation of endometrial receptivity in three-dimensional transvaginal ultrasound images based on 3D U-Net segmentation. *Quant Imaging Med Surg*. 2022;12(8):4095-4108.
77. Wu J, Zhang Z. Extreme learning machine denoising algorithm based analysis of transvaginal 3-dimensional ultrasonic image for the diagnostic effect of intrauterine adhesion. *Sci Program*. 2021;2021:1-7.
78. Huo T, Li L, Chen X, et al. Artificial intelligence-aided method to detect uterine fibroids in ultrasound images: a retrospective study. *Sci Rep*. 2023;13(1):3714.
79. Dilna KT, Anitha J, Angelopoulou A, Kapetanios E, Chaussalet T, Hemanth DJ. Classification of uterine fibroids in ultrasound images using deep learning model. *Computational Science – ICCS 2022 Lecture Notes in Computer Science*. Springer; 2022:50-56.
80. Shahzad A, Mushtaq A, Sabeeh AQ, et al. Automated uterine fibroids detection in ultrasound images using deep convolutional neural networks. *Healthcare*. 2023;11(10):1493.
81. Cai P, Yang T, Xie Q, Liu P, Li P. A lightweight hybrid model for the automatic recognition of uterine fibroid ultrasound images based on deep learning. *J Clin Ultrasound*. 2024;52(6):753-762.
82. Yang T, Yuan L, Li P, Liu P. Real-time automatic assisted detection of uterine fibroid in ultrasound images using a deep learning detector. *Ultrasound Med Biol*. 2023;49(7):1616-1626.
83. Kaveramma UPP, Snehalatha U, Karthik V, Anuradha M. Ultrasound-based machine learning-aided detection of uterine fibroids: integrating vision transformer for improved analysis. *Biomed Eng (Singapore)*. 2023;35(6):2350033.
84. Chinna LM, Pathrose Mary JP. Efficient feature extraction and hybrid deep learning for early identification of uterine fibroids in ultrasound images. *Int J Imaging Syst Technol*. 2024;34(3):1-17.
85. Raimondo D, Raffone A, Aru AC, et al. Application of deep learning model in the sonographic diagnosis of uterine adenomyosis. *Int J Environ Res Public Health*. 2023;20(3):1724.
86. Boneš E, Gergolet M, Bohak C, Lesar Ž, Marolt M. Automatic segmentation and alignment of uterine shapes from 3D ultrasound data. *Comput Biol Med*. 2024;178:108794.
87. Liang X, Fang J, Li H, et al. CR-Unet-Based ultrasonic follicle monitoring to reduce diameter variability and generate area automatically as a novel biomarker for follicular maturity. *Ultrasound Med Biol*. 2020;46(11):3125-3134.
88. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29.
89. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
90. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297-302.
91. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.
92. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
93. Sounderajah V, Ashrafi H, Rose S, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med*. 2021;27(10):1663-1665.
94. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
95. Yang B, Mallett S, Takwoingi Y, et al. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med*. 2021;174(11):1592-1599.
96. Shrestha P, Poudyal B, Yadollahi S, et al. A systematic review on the use of artificial intelligence in gynecologic imaging – Background, state of the art, and future directions. *Gynecol Oncol*. 2022;166(3):596-605.
97. Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: Current status and future challenges – A systematic review. *Artif Intell Med*. 2021;120:102164.
98. Moro F, Ciancia M, Zacc D, et al. Role of artificial intelligence applied to ultrasound in gynecology oncology: A systematic review. *Int J Cancer*. 2024;155(10):1832-1845.
99. Avery JC, Deslandes A, Freger SM, et al. Noninvasive diagnostic imaging for endometriosis part 1: a systematic review of recent developments in ultrasound, combination imaging, and artificial intelligence. *Fertil Steril*. 2024;121(2):164-188.
100. Barrera FJ, Brown EDL, Rojo A, 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.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:



Appendix S1 Key concepts in artificial intelligence

Appendix S2 Search strategy

Table S1 Description of quality assessment based on QUADAS-AI domains

Table S2 Risk-of-bias assessment for 59 studies included in systematic review

Table S3 Characteristics and key findings of 59 studies included in systematic review



Aplicación de la inteligencia artificial a la ecografía para trastornos ginecológicos benignos: revisión sistemática

RESUMEN

Objetivo. Aunque la inteligencia artificial (IA) se aplica cada vez más a la ecografía en ginecología, los esfuerzos para sintetizar la evidencia disponible han sido inadecuados. El objetivo de esta revisión sistemática fue resumir y evaluar la bibliografía sobre el papel de la IA aplicada a la ecografía en los trastornos ginecológicos benignos.

Métodos. Se hicieron búsquedas en las bases de datos bibliográficas Web of Science, PubMed y Scopus desde su inicio hasta agosto de 2024. Los criterios de inclusión fueron los estudios que aplicaban la IA a las imágenes ecográficas en el diagnóstico y tratamiento de los trastornos ginecológicos benignos. Los estudios recuperados de la búsqueda bibliográfica fueron importados al programa informático Rayyan y la evaluación de la calidad se realizó mediante QUADAS-AI, que son las siglas en inglés para la herramienta de evaluación de la calidad para estudios de precisión de pruebas diagnósticas centrados en la inteligencia artificial.

Resultados. De los 59 estudios incluidos, 12 eran sobre síndrome de ovario poliquístico (SOP), 11 sobre infertilidad y tecnología de reproducción asistida, 11 sobre patología ovárica benigna (es decir, quistes ováricos, torsión ovárica, o fallo ovárico prematuro), 10 sobre patología endometrial o miometrial, nueve sobre trastorno del suelo pélvico y seis sobre endometriosis. China fue el país más representado (22/59 (37,3%)). Según QUADAS-AI, la mayoría de los estudios presentaban un alto riesgo de sesgo en el ámbito de la selección de sujetos (porque no se especificó el tamaño de la muestra, la fuente o el modelo de escáner, los datos no procedían de conjuntos de datos de fuentes abiertas y/o no se realizó preprocesamiento de imágenes) y en el ámbito de la prueba índice (los modelos de IA no se validaron externamente), y un bajo riesgo de sesgo en el ámbito del estándar de referencia (el estándar de referencia clasificó correctamente la afección objetivo) y en el ámbito del flujo de trabajo (el tiempo transcurrido entre la prueba índice y el estándar de referencia fue razonable). La mayoría de los estudios (40/59) desarrollaron y validaron internamente modelos de clasificación de IA para distinguir entre casos normales y patológicos (es decir, presencia frente a ausencia de SOP, endometriosis pélvica, incontinencia urinaria, quiste ovárico o torsión ovárica), mientras que 19/59 estudios trataban de segmentar o medir automáticamente los folículos ováricos, el volumen ovárico, el grosor endometrial, los fibromas uterinos o las estructuras del suelo pélvico.

Conclusión. La bibliografía publicada sobre IA aplicada a la ecografía en trastornos ginecológicos benignos se centra principalmente en la creación de modelos de clasificación para distinguir entre casos normales y patológicos, y en el desarrollo de modelos para segmentar o medir automáticamente el volumen ovárico o los folículos.

人工智能在妇科良性疾病超声成像中的应用：系统综述

摘要

目的 虽然人工智能 (AI) 越来越多地应用于妇科超声成像, 但对现有证据进行综合的研究尚不充分。本系统综述旨在总结和评估有关人工智能应用于妇科良性疾病超声成像时的作用的文献。

方法 检索从研究开始到 2024 年 8 月的 Web of Science、PubMed 和 Scopus 数据库。纳入标准是将人工智能应用于妇科良性疾病超声成像诊断和管理的研究。从文献检索中检索到的研究被导入Rayyan软件, 并使用以人工智能为中心的诊断测试准确性研究质量评估工具 (QUADAS-AI) 进行质量评估。

结果 在纳入的 59 项研究中, 12 项涉及多囊性卵巢综合征 (PCOS), 11 项涉及不孕症和辅助生殖技术, 11 项涉及卵巢良性病变 (即卵巢囊肿、卵巢扭转、卵巢早衰), 10 项涉及子宫内膜或子宫肌层病变, 9 项涉及盆底功能障碍, 6 项涉及子宫内膜异位症。中国是参与比例最高的国家 (22/59 (37.3%))。根据 QUADAS-AI, 大多数研究在受试者选择领域 (因为未说明样本量、来源或扫描仪型号, 数据并非来自开源数据集和/或未进行成像预处理) 和指标测试领域 (人工智能模型未经外部验证) 存在高偏倚风险, 而在参考标准领域 (参考标准正确分类了目标病症) 和工作流程领域 (指标测试和参考标准之间的时间间隔合理) 存在低偏倚风险。大多数研究 (40/59) 开发并在内部验证了人工智能分类模型, 用于区分正常和病理病例 (即有无PCOS、盆腔子宫内膜异位症、尿失禁、卵巢囊肿或卵巢扭转), 而 19/59 项研究旨在自动分割或测量卵巢滤泡、卵巢体积、子宫内膜厚度、子宫肌瘤或盆底结构。

结论 已发表的有关将人工智能应用于妇科良性疾病超声检查的文献主要集中在创建分类模型以区分正常和病理病例, 以及开发自动分割或测量卵巢体积或卵巢滤泡的模型。