Artificial Intelligence in Gynecologic Imaging

Morgan Briggs, MD, Ayesha Saif, MD, MBBS,**
Timothy L. Kline, PhD, MS,† Wendaline VanBuren, MD,†
Sarah L. Cohen Rassier, MD, MPH, and Isabel C. Green, MD, MHPE**

Abstract: What was done? A review of artificial intelligence (AI) applications for the imaging of uterine fibroids, endometriosis, and adenomyosis. What was found? AI models can assist with the recognition, segmentation, and localization of uterine fibroids, and the differentiation of benign fibroids and sarcomas. Models can aid in the diagnosis of adenomyosis and endometriosis, and the prediction of the impact of endometriosis on fertility. What the findings mean? Deployed thoughtfully, AI tools could reduce variability, shorten read times, and add objective measurements to routine care. Studies evaluating these models are limited by single-institution designs and continued reliance on expert sonologists and radiologists.

Key Words: artificial intelligence, pelvic imaging, fibroids, endometriosis, machine-learning

(*Clin Obstet Gynecol* 2026;69:26–35)

Artificial Intelligence (AI) is a broad term that encompasses the use of computers to perform complex tasks and solve problems that are typically accomplished by human reasoning. Publications regarding the use of AI in the field of medicine date back to the 1950s, with a precipitous increase in research interest on this topic over the past decade. In 2015, there were just over 6800 citations in PubMed.gov on the topic of "artificial intelligence," increasing to over 22,700 indexed papers by 2020 and over 60,200 to date in 2025. Within the field of Gynecology, there are many potential applications of AI, including diagnostic and outcome predictions, preoperative planning, surgical skill assessment, surgical education, anatomic recognition, and automation.[1] The goal of this review is to establish the clinical application and utility of AI within the realm of gynecologic imaging, focusing specifically on the complex benign disease states of uterine fibroids, endometriosis, and adenomyosis. By harnessing the power of AI and adhering to ethical principles surrounding its use, providers can improve and individualize the care provided to patients.

AI DEFINITIONS AND TOOLS

Table 1 includes the definitions of the key AI tools that will be referenced throughout this review. Machine learning (ML), a subtype of AI, uses algorithms to learn from data and make predictions (Table 1). In imaging, common ML tasks include segmentation, detection, and classification, with models trained and validated on curated imaging data sets. Deep learning (DL), a further subset of ML, relies on multilayer neural networks, a structure of algorithms modeled on the human brain, that learn image features automatically.

Segmentation models produce pixel (2-dimensional)/voxel (3-dimensional pixel) wise masks that precisely outline anatomy or lesions (ie, a tracing of the structures border). For benign uterine fibroids, automatic masks of the uterus and individual fibroids enable objective quantification of fibroid location and total burden, procedure planning, and therapy monitoring. The most common model for segmentation is UNet (Table 1).

Similarly, detection models are types of AI models that are created to identify and locate specific lesions/objects. For endometriosis, a detector can highlight suspected deep infiltrating endometriosis (DIE) nodules across pelvic compartments to guide targeted reading. For example, the YOLO (You Only Look Once) family are fast, one-stage detection models that can rapidly flag potential DIE sites during a read (Table 1).

Convolutional neural networks (CNNs) remain the backbone for image tasks. A CNN is a system of algorithms modeled on the human brain that uses a set of filters to extract spatial features such as organ shape. Two types of classification CNNs are DenseNet and MobileNet (Table 1).

In supervised learning, models learn from labeled examples, for example, a UNet trained with expert annotation masks to segment fibroids on ultrasound. Unsupervised learning discovers structure in unlabeled data and can reveal imaging phenotypes—for instance, clustering patients by radiomic patterns to identify subtypes of fibroids or endometriosis that correlate with symptoms or outcomes. Radiomics converts regions of interest into quantitative descriptors (shape, intensity, and texture) that can be combined with clinical variables (such as age and parity) to predict diagnosis, prognosis, or treatment response—for example, estimating the likelihood of fibroid response to uterine sparing therapy from T2/T1 signal characteristics and texture features.

Deployed thoughtfully, these tools can reduce variability, shorten read times, and add objective measurements to routine care. Automated burden and location mapping supports procedure selection (myomectomy vs. uterine artery embolization), DIE detection can prompt protocol optimization or subspecialty referral, and consistent junctional zone metrics may help track adenomyosis over time. Importantly, model training should reflect diverse scanners and patient populations, with rigorous external validation, transparent reporting, and human oversight to mitigate bias and automation complacency.

From the *Department of Obstetrics and Gynecology, Division of Minimally Invasive Surgery; and †Department of Radiology, Mayo Clinic, Rochester, Minnesota.
The authors declare that they have nothing to disclose.
Correspondence: Isabel C. Green, MD, Rochester, MN. E-mail: green.isabel@mayo.edu

DOI: 10.1097/GRF.0000000000000984

TABLE 1. Core AI Terms for Gynecologic Imaging (Fibroids, Endometriosis, and Adenomyosis)

Term	Plain-English definition	Example
Artificial intelligence (AI)	Computational methods that perform tasks requiring human-like perception, reasoning, or decision-making	End-to-end system that ingests MRI, suggests imaging protocol tweaks, segments uterus/fibroids, and generates a structured report draft for clinician review
Machine learning (ML)	A subset of AI where models learn patterns from data rather than being explicitly programmed	Model predicting the likelihood of adenomyosis from imaging-derived features plus age/parity
Deep learning (DL)	A subset of ML using multilayer neural networks to automatically learn hierarchical features from data	3D CNN or transformer that segments the uterus, junctional zone, and fibroids on pelvic imaging
Segmentation model	Pixel/voxel-wise delineation of anatomy or lesions	Automatic uterine, junctional zone, and fibroid masks to compute burden and location
Detection model	Finds candidate lesions/objects and returns boxes or points	MRI detection of deep-infiltrating endometriosis nodules in compartments
Classification model	Assigns a label or category to an image/case	Classify adenomyosis (present/absent) or fibroid type (submucosal/intramural/subserosal)
Radiomics	Quantitative features (shape/texture/intensity) extracted from regions-of-interest	Predict fibroid response to uterine-sparing therapy from T2/T1 and texture features
Supervised learning	Model learns from labeled examples (inputs + correct outputs)	US U-Net trained with annotation masks to segment fibroids
Unsupervised learning	Model learns patterns/representations from unlabeled data (no ground-truth labels)	Identify fibroid/endometriosis phenotypes by how patients cluster together based on radiomic imaging patterns
Convolutional neural network (CNN)	Neural network using convolutional layers to extract spatial features; excels at image tasks with local patterns	Pelvic MRI pipeline with CNN backbone for feature extraction across segmentation, classification, and detection heads
U-Net (segmentation)	Most common model architecture for segmentation tasks with skip connections, shortcuts, that preserve fine detail while capturing context	Segment the uterus, junctional zone, and fibroids on T2 MRI to compute burden and location
YOLO family (detection)	You Only Look Once (YOLO) one-stage, real-time detectors optimized for speed and deployment	Rapid screening for suspected DIE locations to guide targeted reading
DenseNet (classification)	CNN with dense connections that reuse features previously learned to improve efficiency	Classify fibroid type (submucosal/intramural/subserosal) from MRI stacks
MobileNet (classification)	CNN using depthwise separable convolutions, for mobile/edge devices	Point-of-care ultrasound model to flag suspected adenomyosis for expert review

FIBROIDS

Diagnosis

Fibroids are the most prevalent solid gynecologic benign tumors of the uterus.[2] By the age of 50, the estimated cumulative incidence of fibroids is more than 80% for Black women and nearly 70% for white women.[3] Accurate diagnosis and mapping of uterine fibroids is clinically important because lesion number, size, and location directly determine eligibility and strategy for surgical and interventional treatments. Transvaginal ultrasound (TVUS) is usually the first line diagnostic test for uterine fibroids because it is accessible and inexpensive, and it has a high diagnostic accuracy for the presence of fibroids. Magnetic resonance imaging (MRI) shows similar accuracy for detection, but is superior for mapping, especially when the uterus is large (> 375 mL) or contains 4 or more fibroids.[4] MRI also has the additional benefit of risk stratification for potential malignancy. Number, size, and location directly determine eligibility and strategy for surgical and interventional treatments. Transvaginal ultrasound (TVUS) is usually the first line diagnostic test for uterine fibroids because it is accessible and inexpensive, and it has a high diagnostic accuracy for the presence of fibroids. Magnetic resonance imaging (MRI) shows similar accuracy for detection, but is superior for mapping, especially when the uterus is large (> 375 mL) or contains 4 or more fibroids.[4] MRI also has the additional benefit of risk stratification for potential malignancy.[5]

Even with high diagnostic accuracy, conventional imaging remains time-intensive and labor-intensive, and the degree of detail reported is not uniform. Ultrasound interpretation is highly operator-dependent, and performance declines when fibroids are numerous, or when the uterus is enlarged.[6] MRI interpretation can vary between readers, and discrepancies have been documented in the reporting of qualitative features such as vascularization, degeneration, and restricted diffusion.[7] These features are important not only for diagnosing fibroids by differentiating them from malignancy but also for guiding therapeutic interventions.

Ultrasound

AI use for fibroid diagnosis on ultrasound is still at a foundational stage, with work focused primarily on automated recognition rather than segmentation or mapping of fibroids. Three recent studies have explored the use of deep learning on ultrasound for uterine fibroid diagnosis, but they do so with different goals: one detects and localizes fibroids in real time, while the other 2 classify images as fibroid or nonfibroid without localization.[8–10]

Yang and colleagues developed an object-detection model to identify fibroids directly on 2D ultrasound images. The model generated rectangular bounding boxes over regions it identified as fibroids. This approach attained an F1 score of 95%, an average precision of 98.38%, and a detection speed of 0.28s per image.[8]

TABLE 2. Statistical Terms and Definitions for Evaluation of AI Model Performance

Area under the curve (AUC)	AUC represents the probability that the model will assign a higher score to a randomly selected positive case than to a randomly selected negative case. An AUC of 1.0 indicates perfect discrimination, whereas an AUC of 0.5 reflects performance equivalent to random chance
Accuracy	Accuracy is the proportion of all predictions that are correct. It is the sum of true positives and true negatives divided by the total number of cases
Recall	Recall is the proportion of true positive cases correctly identified by the model. It reflects the sensitivity of the model
Precision	Precision is the proportion of cases predicted as positive that are truly positive. It reflects the positive predictive value (PPV) of the model
F1 score	F1 score is the harmonic mean of precision and recall, providing a single measure of a model's ability to balance false positives and false negatives. A perfect F1 score is 1.0. It is particularly useful in imbalanced data sets, where accuracy might be misleading
Dice score	The Dice score is mathematically similar to the F1 score and is its equivalent in image segmentation tasks. It compares the overlap between the model's segmented region and the ground-truth annotation, with higher values indicating closer agreement

Table 2 defines key statistical terms utilized in the evaluation of AI model performance and referenced through this review.

Xi and Wang used a different approach. They fine-tuned an EfficientNetB0 classifier with an added attention mechanism to focus the model on the most informative regions of the ultrasound image. They worked on a data set of 1990 images (fibroid vs. nonfibroid), later expanded to 10,000 through augmentation, and achieved 0.99 accuracy.[9] Cai and colleagues developed a hybrid model using real clinical ultrasound images and supplemented them with additional synthetic fibroid images generated to increase the size of the training data set. A MobileNetV2 classifier was then trained on this augmented data set. This hybrid method achieved 0.97 accuracy and 0.9741 F1-score, with a real-time classification speed of 40 frames per second.[10]

Both prior studies addressed the small sample size by increasing the number of training images. Xi and Wang augmented real images (rotations, flips, and contrast adjustments), whereas Cai and colleagues generated synthetic fibroid images. The latter approach introduces the possibility that the model may learn image characteristics that are specific to the synthetic, generated data rather than reflecting the real ultrasound images.[9,10]

Despite methodological differences, all 3 studies share important limitations: (1) retrospective, single-center analysis with no external validation, (2) no correlation of AI predictions with histopathology or clinical outcomes, and (3) no account for common confounders such as adenomyosis, poor acoustic windows, or postsurgical anatomy.[8–10]

Clinical application: These models may assist clinicians by flagging likely fibroids and reducing operator dependence, but binary classification does not meet the clinical requirement for anatomic mapping needed for treatment planning. Clinical implementation will depend on prospective validation and further development of models to localize lesions in addition to indicating the presence of a fibroid.

MRI

There is a growing body of work exploring the use of AI for MRI-based diagnosis of uterine fibroids, ranging from models that aim to automate the detection and localization of fibroids to those that focus on improving automated segmentation of fibroids on MRI. To provide a visual example, we have included images from a project by the author (T.K.) using AI to differentiate leiomyoma from leiomyosarcoma. Figure 1 includes images of both segmentation and classification using radiomics. Both AI principles will be described further as we explore recent literature within MRI-based diagnoses of fibroids.

Lin et al[11] trained an AI model on sagittal and axial T2-weighted images with the goal of automating the detection and segmentation of uterine fibroids.[11] Manual segmentation was performed by a junior radiologist and then verified by a senior radiologist. Results were stratified by fibroid size and FIGO category. On internal testing, precision for fibroid detection was 0.989 (sagittal) and 0.979 (axial). Results remained high on external data (0.96 and 0.944). Overall recall declined on the external cohort (0.581 sagittal; 0.547 axial) but remained higher within the clinically significant subgroups: for fibroids ≥4 cm it reached 0.937–0.953 internally and 0.946–0.1 externally, and FIGO 2–5 lesions were detected with 100% recall in both cohorts. The study excluded cases with adenomyosis and other copathology, and although it was tested on an external cohort, generalizability to routine mixed-pathology MRI remains uncertain.[12]

Unlike Liu et al,[13] which evaluated segmentation within a diagnostic framework, studies by Zhang et al[14] and by Liu et al[13] were limited to improving the segmentation step itself. Zhang et al[15] modified an open-source U-Net model by adding modules to allow the model to focus on the region that contains fibroid and to demarcate fibroid boundaries in low-contrast settings, resulting in improved and uniform segmentation. In a cohort of 150 patients, the resulting DARU-Net model outperformed the baseline U-Net and other models, with a Dice score of 0.8066, precision 0.8233, and recall 0.7913, with the greatest difference seen in cases where the fibroids were poorly defined.[14]

Another MRI study that addressed segmentation alone was conducted by Liu and colleagues. In the study, a publicly available V-Net model was adapted with attention gates and deep supervision in a 3-dimensional format to produce more stable fibroid outlines on T2-weighted MRI in the context of planning high-intensity focused ultrasound (HIFU) treatment. In a single-center data set of 245 cases (147 train, 49 validation, and 49 test), the modified model (3D DA-VNet) outperformed standard 2D and 3D U-Net/V-Net baselines, with a Dice score of 0.878, sensitivity 0.879, and precision 0.885, with the largest gains in scans where the fibroid-myometrium boundary was indistinct.[14]

Both studies were confined to single-center T2-weighted MRI data sets without coexisting pathology

Copyright © 2026 Wolters Kluwer Health, Inc. All rights reserved.

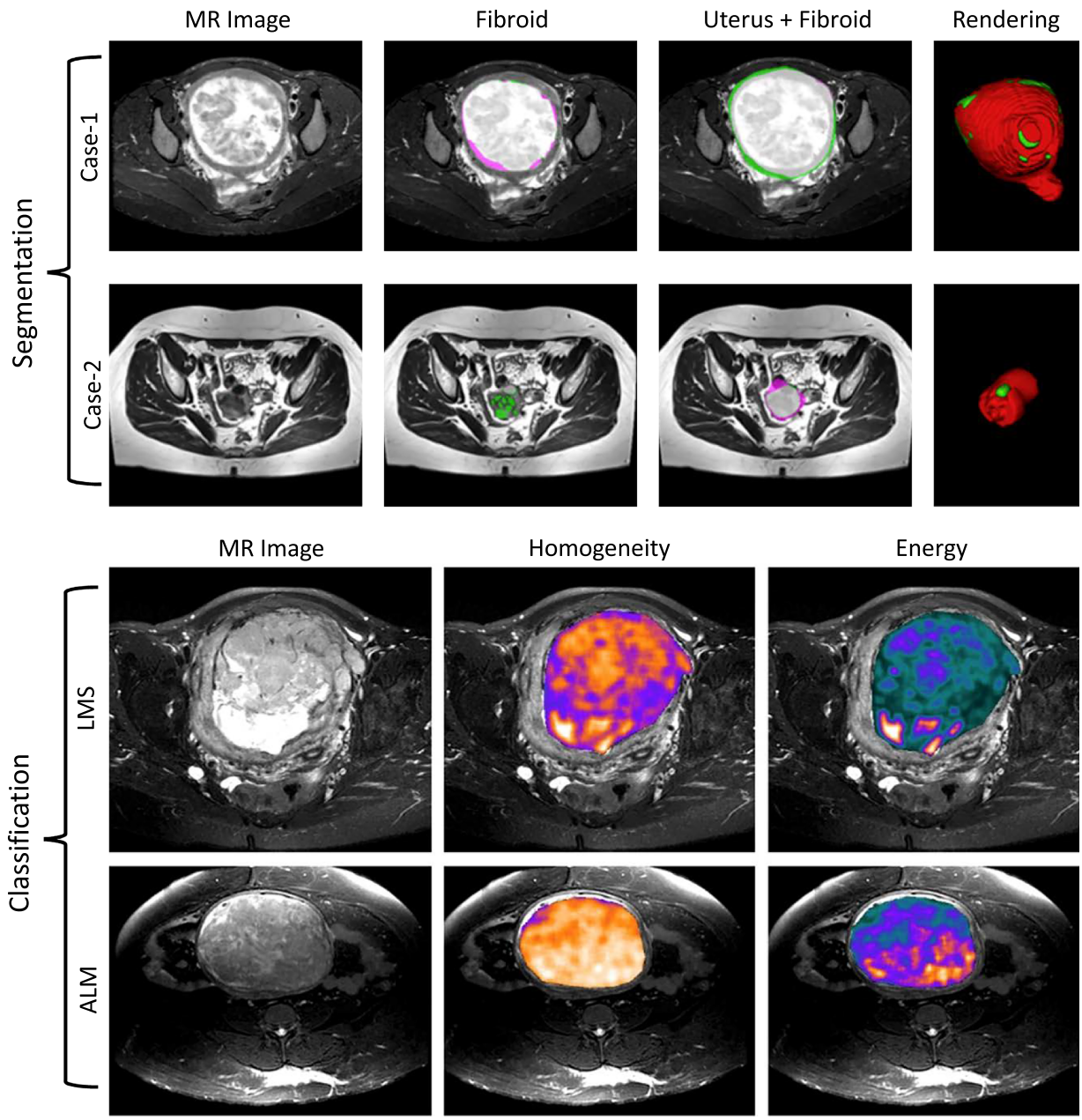

FIGURE 1. Examples of segmentation and classification of fibroids. In the top panel, 2 example cases are shown comparing manual segmentation (ie, the reference standard) in green, and the automated segmentation model results in purple. The model was trained to segment both the uterus and fibroids. Case 2 highlights an example with multiple leiomyomas, including a cellular leiomyoma. On the bottom is an example from a classification model trained to distinguish benign versus malignant and degenerated leiomyoma versus leiomyosarcoma. Radiomic features overlaid on the image show drastic differences in image texture features. The radiomic texture features that are extracted are then used to train machine learning models to distinguish between types of uterine tissue. full color online

such as adenomyosis. Also, the studies did not include external validation or make the trained models publicly available for independent testing. The work in both papers is therefore limited to demonstrating improvements in segmentation performance without evidence of performance in more heterogeneous or real-world clinical settings.[12,13]

Clinical Application: The clinical relevance of this work lies in the possibility that MRI studies could eventually be delivered with fibroids already outlined in a consistent and reproducible way. This would make review faster, reduce variation between readers, and allow objective comparison across time and institutions. Although these models are not yet ready for routine use, they point toward a future in which MRI scans arrive with standardized, preprocessed fibroid information rather than requiring manual interpretation on every case.

Differentiation from Malignancy

When a uterine mass is detected, the clinical priority is to rule out leiomyosarcoma without subjecting patients with benign fibroids to unnecessary radical surgery. Ultrasound is usually the first-line modality, but its diagnostic performance for distinguishing fibroids from sarcoma is limited. In a systematic review of 972 women, ultrasound showed a pooled sensitivity of 0.76 and a specificity of 0.89. The pooled AUC was 0.8925, indicating only moderate overall discriminative ability.[16] Diagnosis is further complicated by the fact that clinical presentation is often indistinguishable, and ultrasound offers only moderate discriminatory performance, leading many cases to remain ambiguous at the point of initial imaging.[17]

In current practice, MRI is considered the most reliable noninvasive tool for distinguishing benign fibroids from uterine leiomyosarcoma. A meta-analysis of 8 studies with 2495 women (2253 with uterine leiomyomas and 179 with uterine sarcomas) reported a pooled sensitivity of 0.90, specificity of 0.96, and AUC of 0.9759 for MRI in distinguishing fibroids from uterine sarcoma.[18] Although MRI is considered the most reliable noninvasive modality for uterine mass characterization, preoperative distinction between fibroids and leiomyosarcoma remains challenging because imaging features frequently overlap, especially when fibroids show a T2 hyperintense signal with enhancement and restricted diffusion, which can make differentiation between hypercellular fibroids, STUMP, and sarcomas challenging. the rarity of leiomyosarcomas and that they may have benign imagining features.[5]

Differentiating Malignancy and Benign Fibroids With AI

Despite the strong performance of conventional imaging, diagnostic uncertainty persists in a nontrivial subset of cases. AI tries to reduce this ambiguity by standardizing feature extraction. In an ultrasound study by Chiappa and colleagues, 70 surgically confirmed cases (50 fibroids and 20 sarcomas) were analyzed using a standardized IBSI-compliant texture extraction framework (TRACE4). Three hundred nineteen radiomics features were extracted and 308 radiomics features were found stable. Different machine learning classifiers were created and the best classifier achieved an AUC of about 0.86 with a negative predictive value near 0.92. A notable finding was that among the lesions labeled "uncertain" by sonographers, the model would have correctly classified 83% of sarcomas and 78% of myomas. This shows that AI may add value specifically in cases where radiologists are unsure. However, the study was limited by conduction at a single institution with 20 sarcoma cases, utilization of synthetic oversampling to compensate for class imbalance (which may inflate performance estimates), and the requirement of manual segmentation, which further constrains the model's generalizability.[17]

In the MRI-based study by Roller and colleagues, 108 women with histologically confirmed uterine masses (69 fibroids and 39 sarcomas) were evaluated using 14 predefined MRI features scored independently by 2 abdominal radiologists, with age and tumor size as clinical variables. Radiomics features were then extracted from 3D segmentations of T2-weighted images using TexRAD software. Twelve of the 14 conventional MRI features and both clinical variables differed significantly between groups, and inter-reader agreement for most features was moderate to excellent, indicating that conventional MRI already carries strong discriminative information. The data set was split into a training cohort (n = 86) and a held-out test cohort (n = 22). Models were built using different combinations of clinical, conventional MRI, and radiomics variables. In the test cohort, the model using only conventional MRI features plus clinical features achieved an AUC of 0.956, the radiomics-only model achieved an AUC of 0.929, and the combined model integrating conventional MRI, clinical variables, and radiomics achieved an AUC of 0.989. The results reflect only a small incremental gain beyond MRI plus clinical features alone. However, the study was limited by its single-center retrospective design, a small test set, exclusion of diffusion sequences, reliance on manual segmentation, and lack of external validation.[19]

Clinical Application: Although conventional performs well, the use of AI on MRI would support radiologist decision making, which is particularly important given the rare incidence of sarcomas. On ultrasound, where indeterminate reads are more common, AI could be applied selectively to those uncertain cases to support whether additional imaging or specialist referral is needed. In this way, AI would function as a decision-support tool in selected scenarios rather than as a general triage method. For both modalities, AI could also be an important tool for efficiency given the frequent multiplicity of fibroids, particularly in follow-up of size for growth assessment.

Prediction of Treatment Success

MR-HIFU and AI Tools

MR-guided high intensity focused ultrasound or MR-HIFU is a uterine sparing treatment where noninvasive thermal ablation of tissue is achieved through the administration of high-intensity therapeutic ultrasound with MRI guidance.[18] This results in thermal necrosis of the targeted fibroid. Postprocedure symptoms of bleeding and bulk should be reduced.

Post-Treatment Response

Using imaging characteristics on MRI, treatment response can be evaluated using the ratio of nonperfused fibroid volume to the total volume (NPV/TPL), essentially, how much loss of function within the fibroid was obtained. The larger the nonperfused area of the fibroid post-treatment, the lower probability of postoperative recurrence and reintervention.[19] Calculation of this ratio has been a manual task for radiologists, and AI tools offer a workflow for automatic volume measurements and calculation of this ratio for assessment of procedure success.

Like other AI applications in imaging, the volumes of the uterus, fibroids, and nonperfused volume ratios can be measured through a deep learning-based segmentation pipeline with excellent to moderate reliability.[20] In a study by Slotman and colleagues, the mean difference between automatic and manual derived NPV/TFL ratio was 5%. There are limitations; however, segmentation of fibroids is harder than overall uterine volume measurements, and increased accuracy may rely on multiple MRI sequences (ie, T1 and T2) or the use of contrast.

Clinical application: Extended validation studies are needed in other data sets, but this type of automatic calculation offers promise to identify patients where re-treatment with HIFU could be considered, or preprocedurally to calculate fibroid burden and fibroid volume.

Copyright © 2026 Wolters Kluwer Health, Inc. All rights reserved.

Predicting Treatment Outcomes/Preoperative Evaluation

The efficacy of HIFU treatment can be predicted based on the histologic characteristics of uterine fibroids before treatment, which assists in patient counseling and selection for the procedure.[21] Specifically, some characteristics that limit HIFU treatment response are fibroids rich in blood supply or with rich cellular content. Those histologic characteristics can currently be estimated using preoperative MRI sequences.

In a study by Wei et al,[21] pretreatment MRI data was run through radiomics feature extraction and correlated with post-treatment NPV/TPL or NPV ratio. The performance of this model was compared with conventional image characteristics, such as signal intensity and enhancement and added value in predicting the post-treatment NPVR. Radiomics contributed 2 additional fibroid characteristics to improve predictive accuracy. Image nonuniformity and skewness of the T2WI signal intensity both negatively correlated with treatment response. Nonuniformity reflects heterogeneity of the tissue, and skewness reflects signal intensity with brighter pixels (higher signal) or positively skewed, reflecting an increase in water content and blood perfusion.

Clinical application: Data is preliminary and requires confirmation in larger and more heterogeneous data sets but does offer promise for preoperative patient selection and counseling on treatment outcomes.

ENDOMETRIOSIS AND ADENOMYOSIS

Diagnosis

Diagnosis of endometriosis and adenomyosis is often delayed due to varying clinical presentations, historical reliance on surgical pathology for definitive diagnosis, and diverse imaging characteristics. In the past 5 years, there has been additional work in optimizing imaging-based diagnosis of endometriosis and adenomyosis.[22]

Diagnosing endometriosis through imaging remains limited by the difficulty in visualizing the pathologic changes of stage I and II endometriosis through ultrasound and MRI and varying expertise and experience of sonologists and radiologists.[23–26] Both MRI and US perform better for the identification of endometriomas and deep infiltrative endometriosis.[23–25] One of the most recent systematic reviews analyzed both US and MRI with respect to their anatomic locations of disease, physician specialty, and geographic location. Interestingly, both modalities demonstrated superior performance with dedicated endometriosis protocols and when interpreted by experienced readers in practices dedicated to endometriosis.[27]

There has been progress in imaging-based diagnosis of adenomyosis with the development of the Morphologic Uterus Sonographic Assessment (MUSA) criteria for adenomyosis.[28,29] In a systematic review and meta-analysis of imaging modalities used to diagnosis adenomyosis, MRI and TVUS had sensitivities of 0.78 and 0.74, specificities of 0.88 and 0.76, and pooled area under the operator curve values of 0.77 and 0.7, respectively.[30] However, like endometriosis, ultrasound-based diagnosis of adenomyosis depends on operator experience. In addition, there is intraobserver and interobserver variability in ultrasound interpretation.[31] Unfortunately, MRI does not have as many features to assess compared with ultrasound and the reproducibility of criteria in research studies has been lacking. In addition, the cysts, which are a direct feature of adenomyosis, may not be apparent in the setting of hormonal treatment (tablets or intrauterine device).

Ultimately, imaging-based diagnosis of endometriosis and adenomyosis can be improved. Although the data is limited and the studies lack generalizability, current literature demonstrates the applicability of AI assisting with the diagnosis of endometriosis and adenomyosis by improving image quality and decreasing the dependence on the experience and expertise of the operator.

Endometriosis Diagnosis With AI

AI can be utilized to improve MRI imaging quality and, as a result, improve MRI-based diagnosis of endometriosis. Jiang and colleagues used a fuzzy c-means (FCM) clustering algorithm to process MRI images of ovarian endometriosis.[32,33] This clustering algorithm is a type of unsupervised machine learning that uses similarities between data points to cluster them together. Compared with the unprocessed MRI images, the images processed using the FCM algorithm showed clearer structures of the uterus, adjacent organs, and pelvic wall and clearer relationships between the endometriosis lesions and the vagina, bladder, and rectum.[32] So, the diagnostic accuracy of the MRI images processed with the FCM algorithm was higher (0.94) than for images processed with conventional MRI (0.61).[32]

Recently, a novel endometriosis MRI AI algorithm was developed on a large cohort of patients with surgically proven disease. The algorithm provides a general assessment of the presence or absence of disease, including all endometriosis phenotypes (superficial, endometriomas, and deep endometriosis). Seven hundred fifty-one patients were included in the cases and controls. The final 3D-DenseNet-121 classifier model demonstrated robust performance and findings indicated the most accurate predictions were obtained using T2W, T1W FS precontrast, and postcontrast images. Using an ensemble technique on the test set resulted in an F1 Score of 0.881, AUROCC of 0.911, sensitivity of 0.976, and specificity of 0.720. Seven radiology readers achieved 84.48% and 87.93% sensitivity without and with AI assistance in detecting endometriosis, respectively. This study introduced the first DL model to use multi-sequence MRI on a large cohort, showing results equivalent to human detection by trained readers in identifying endometriosis.[34]

Strengths of this study are its large sample size of 395 surgically confirmed cases of endometriosis with surgery within 3 months of MRI. Also, the expert level of subspecialty surgery, radiology and pathology provides robust ground truth data. Although a limitation of the study is that it is a single-center study, it has data derived from several MRI vendors, limiting scanner or sequence bias.[34]

AI can also be utilized to aid endometriosis diagnosis through a predictive model using ultrasound images and clinical symptoms. Nouri and colleagues created multiple binary machine learning models to classify cases as having endometriosis, or not, based on age, body mass index (BMI), history of infertility, state of sexual activity, and ultrasound findings consistent with endometriosis (kissing ovaries, sliding sign, endometrioma, and adhesions between the ovary and uterus). The models were validated with sensitivities ranging from 0.59 to 0.75, specificities ranging

from 0.71 to 0.83, and an area under the curve (AUC) ranging from 0.71 to 0.76.[35] The sensitivities are similar, and the specificities are lower than the pooled values from a 2016 Cochrane review.[23] This model was based on patients with Stage 3 or 4 endometriosis and therefore is not applicable to Stage 1 or 2 endometriosis. However, AI models such as this may provide an option for imaging-based endometriosis diagnosis that is less operator-dependent.

Clinical application: Existing research has focused more on the utilization of AI for the extent of endometriosis, rather than a general or categorical endometriosis diagnosis. Although studies are limited by a single-institution and retrospective design, they provide insight into the capabilities of AI in improving image quality and endometriosis diagnosis.

Adenomyosis Diagnosis With AI

Compared with endometriosis, more research has focused on utilizing AI to increase the ease and efficiency of adenomyosis diagnosis through MRI and ultrasound-based AI models.

Specifically, Burla and colleagues segmented and extracted radiomic features from MRIs of a small sample of patients with adenomyosis and patients without adenomyosis.[31] Radiomic features were tested for their diagnostic performance in differentiating adenomyosis from controls, and 11 features had AUC values ranging from 0.78 to 0.98. Two models were created using radiomics features. A model using all 11 features and a model using 5 features both had an AUC-ROC of 1.[36] The radiomics-based AI models had similar AUC values as ultrasound-based adenomyosis diagnosis without AI. However, the model was not validated with a test-set and therefore, the results are based only on the model training.

Furthermore, Zhao and colleagues created an end-to-end unified network framework (Adenomyosis Auto Diagnosis Network, A^2DNet), a type of machine learning that trains a model to link raw data (in this case, images) to an output (in this case, diagnosis of adenomyosis). Ultrasound examinations were performed on patients with adenomyosis and controls. The myometrium was evaluated for the presence of direct and indirect ultrasonographic signs of adenomyosis based on the MUSA criteria. A diagnosis of adenomyosis was made by the presence of one direct sign or 2 indirect signs of adenomyosis. The model was tested and diagnosed adenomyosis and normal controls with 0.92 accuracy.[37] The study illustrates the usefulness of machine learning in ultrasound-based diagnosis of adenomyosis using the MUSA criteria.

Similarly, to compare a deep learning model to trainees with intermediate ultrasound skills, Raimondo and colleagues created an end-to-end deep learning model using manually segmented ultrasound images that highlighted features of adenomyosis or fibroids. Diagnosis of adenomyosis was based on the presence of 2 or more sonographic findings of adenomyosis (indirect or direct) based on MUSA criteria. The adenomyosis diagnostic accuracy of the deep learning model was compared with the diagnostic accuracy of trainees with intermediate ultrasound skills, fourth-year Obstetrics and Gynecology residents, who analyzed stored ultrasound video clips of the uterus for each patient and reported a diagnosis. The deep learning machine had lower sensitivity (0.43 vs. 0.72), higher specificity (0.82 vs. 0.69), and lower accuracy (0.51 vs. 0.70) compared with intermediate-skilled trainees for adenomyosis diagnosis.[38] Although the deep learning model is promising, additional advancements are needed for the model to function at the level of expert radiologists.

Clinical application: These examples illustrate the capability of machine learning in assisting with the diagnosis of adenomyosis based on ultrasound. Additional studies and innovation are needed as the models may not be generalizable to other institutions or patient populations, as the models were based on data from single centers/institutions and a single ultrasound machine type.

Extent of Disease

Currently, there is limited research on the application of AI for imaging-based evaluation of the extent of adenomyosis. Most of the literature on the use of AI in imaging for endometriosis focuses on determining the extent of endometriosis disease, including the presence of deep endometriosis, posterior cul-de-sac obliteration, and bowel endometriosis.

There have been significant advancements in imaging-based evaluation of endometriosis with the creation and utilization of the IDEA (International Deep Endometriosis Analysis) protocol for ultrasound.[39] The IDEA protocol provides terms and measurements to standardize the sonographic evaluation and reporting of endometriosis. In a study of individuals, aged 18 to 50 years, with a history of chronic pelvic pain or endometriosis, transvaginal ultrasound scans were performed and reported by a single, experienced surgeon-sonologist, in accordance with the IDEA protocol. Imaging findings were compared with intraoperative findings and surgical pathology. The diagnostic performance of endometriomas had a sensitivity of 0.94 to 0.95 and a specificity of 1.0. For deep endometriosis of all sites, the ultrasound diagnostic performance sensitivity ranged from 0.84 to 1.0, and the specificity ranged from 0.97 to 1.0. For the pouch of Douglas obliteration, the sensitivity was 0.97 and the specificity was 0.97. For superficial endometriosis, the sensitivity was 0.04 to 0.43 and the specificity was 0.99 to 1.0.[40]

In comparison, in a systematic review and meta-analysis of different imaging modalities for diagnosing deep infiltrative endometriosis, among 21 studies, transvaginal ultrasound had a pooled sensitivity of 0.76, a pooled specificity of 0.95, and an area under the curve of 0.92. The meta-analysis was affected by significant heterogeneity (I^2 99%).[25] This difference in diagnostic performance between Mick and colleagues and Zhang and colleagues demonstrates how the evaluation of deep endometriosis by ultrasound is limited by the availability of sonographers and radiologists who are trained to perform the ultrasound evaluation based on the IDEA protocol and the time required to complete the more detailed evaluation.[22]

Like ultrasound, optimal endometriosis evaluation through MRI requires skilled interpretation and dedicated imaging protocols.[22] Zhang et al[25] systematic review and meta-analysis found MRI diagnostic accuracy among 13 studies with high heterogeneity (I^2 98%) to have a sensitivity of 0.82, specificity of 0.87, and area under the curve of 0.91 for deep infiltrative endometriosis. To improve standardization of MRI-based evaluation of endometriosis, an MRI-based protocol for evaluation of deep endometriosis using specific imaging sequences in addition to vaginal contrast gel and antiperistaltic agent (Glucagon) has been published by the Society of

Copyright © 2026 Wolters Kluwer Health, Inc. All rights reserved.

Abdominal Radiology Endometriosis Disease-Focused Panel.[41] Literature demonstrates that AI-based models may help circumvent the variability in diagnostic accuracy of imaging for endometriosis due to the variability in expertise and experience of sonologists and radiologists.

Classification and Extent of Deep Endometriosis With AI

Recent literature demonstrates the utility of AI in aiding the diagnosis of endometriosis-associated posterior cul-de-sac obliteration through MRI. Utilizing transvaginal ultrasound images of women of all ages with any indication for a pelvic ultrasound, a machine learning model was created to identify the presence or absence of a sliding sign. Sliding sign is a transvaginal ultrasound technique that utilizes transvaginal pressure to evaluate if the anterior rectum slides freely along the posterior vaginal, cervix, and uterus and is therefore used to evaluate for posterior cul-de-sac obliteration.[39] Ultrasounds were completed and interpreted by sonologists considered experts in sliding sign interpretation. Using the sonologists classification as a reference, during testing, the deep learning model's classification of posterior cul-de-sac obliteration had an AUC of 0.96, sensitivity of 0.88, and a specificity of 0.90.[42]

Similarly, machine learning models have been created to help diagnosis rectosigmoid deep endometriosis on ultrasound. Guerriero and colleagues created 7 models based on age, presence of ultrasound signs of uterine adenomyosis, presence of an endometrioma, adhesions of the ovary to the uterus, presence of "kissing ovaries" (aka when ovaries are pulled together by adhesions related to endometriosis), and absence of sliding sign among a data set of patients with clinical suspicion of deep endometriosis aged 16 to 60 years. The models had accuracies ranging from 0.69 to 0.75, sensitivities ranging from 0.66 to 0.84, specificity ranging from 0.71 to 0.77, and AUC ranging from 0.75 to 0.82.[43] Compared with the pooled diagnostic performance values in Zhang and colleagues systematic review and meta-analysis, the diagnostic performance of Maicas et al[42] model was similar and the performance of Guerriero et al[43] models was slightly lower.

In addition to assisting with ultrasound-based diagnosis of endometriosis, AI models can be used to assist junior sonologists during training. In a study of pelvic ultrasounds performed utilizing the IDEA protocol, an ultrasound-based AI model using YOLO (You Only Look Once), a deep learning system that can identify and locate objects in an image, was created to help detect deep pelvic endometriosis nodules. The diagnostic performance of junior sonologists improved with the assistance of the AI-based model for predicting deep endometriosis.[44]

Further innovation of imaging-based AI models for evaluating the extent of endometriosis has produced models combining different imaging modalities and models combining imaging and clinical characteristics. For example, a machine learning model trained using unpaired data from MRI and TVUS was created by matching MRI volumes with positive or negative posterior cul-de-sac labels from TVUS videos with positive or negative sliding sign labels. The model was tested with both MRI and TVUS to classify posterior cul-de-sac obliteration and the AUC was 0.80 in MRI testing and 0.89 in TVUS testing.[45] The ability to utilize the AI model with both MRI and TVUS images increases the generalizability of the model.

Lastly, a human-AI collaborative model was created by Wang and colleagues. This collaborative model utilizes the strengths of clinicians, MR imaging, and AI to aid the diagnosis of posterior cul-de-sac obliteration.[46] MRI images of the female pelvis were obtained from various centers with different MRI machines and different resolutions for pretraining. For training, the scans were from multiple clinical sites, each annotated by 3 clinicians experienced in imaging-based diagnosis of endometriosis. Scans used an endometriosis protocol and show a specific region around the uterus where signs of posterior cul-de-sac obliteration are more visible. In the testing phase, the surgical results served as "ground truth." The combination of an MRI-based AI model and multiple human labels had an accuracy of 0.8 and AUC of 0.89 which were higher than the clinician predictions and AI-only models.[46]

Clinical application: There has been significant innovation using AI-models to assist in the imaging-based evaluation of deep endometriosis. Additional studies are needed as the generalizability of most of the models are limited by small sample sizes, data from single institutions, reliance on images acquired by expert sonologists, and dependence on significant institutional resources.

Fertility and Endometriosis

There has been additional work in the field of infertility associated with endometriosis. Among 496 patients with endometriosis who underwent their first laparoscopy for endometriosis, He and colleagues adapted the Endometriosis Fertility Index (EFI) to predict natural pregnancy. The model used a combination of ultrasound radiomics and urinary proteomics to enhance the EFI.[47] The ultrasound radiomics aspect can automatically extract data, making the imaging evaluation more objective and consistent. The adapted model had an AUC of 0.909.[47]

Clinical Application: Although the model is limited by reliance on high-quality ultrasound images and sonographers, which limits its generalizability to institutions with less resources, the study illustrates the use of AI to enhance established predictive models focusing on endometriosis's impact on fertility.

CONCLUSION

Utilizing machine learning and radiomics, AI models based on imaging features alone and those combining imaging and clinical variables have been created to assist with the (1) recognition, segmentation, and localization of uterine fibroids, (2) differentiation of benign fibroids and uterine sarcomas, (3) prediction of HIFU success in the treatment of uterine fibroids, (4) diagnosis of adenomyosis and endometriosis, (5) identification of posterior cul-de-sac obliteration and rectosigmoid endometriosis, and (6) prediction of the impact of endometriosis on fertility. Studies evaluating these models are limited by single-institution designs, continued reliance on expert sonologists and radiologists, and dependence on organizational resources. Emerging trends to overcome these limitations include the development of large, multi-institutional imaging databases to support AI model training and validation.

REFERENCES

1. Pipes GM, Hung AJ, Kim KH. Artificial intelligence and gynecologic surgery. *Obstet Gynecol*. 2025;146:672–678.

2. Giuliani E, As-Sanie S, Marsh EE. Epidemiology and management of uterine fibroids. *Int J Gynaecol Obstet*. 2020; 149:3–9.
3. Baird DD, Dunson DB, Hill MC, et al. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. *Am J Obstet Gynecol*. 2003;188: 100–107.
4. Dueholm M, Lundorf E, Hansen ES, et al. Accuracy of magnetic resonance imaging and transvaginal ultrasonography in the diagnosis, mapping, and measurement of uterine myomas. *Am J Obstet Gynecol*. 2002;186:409–415.
5. Hindman N, Kang S, Fournier L, et al. MRI evaluation of uterine masses for risk of leiomyosarcoma: a consensus statement. *Radiology*. 2023;306:e211658.
6. Palheta MS, Medeiros FDC, Severiano ARG. Reporting of uterine fibroids on ultrasound examinations: an illustrated report template focused on surgical planning. *Radiol Bras*. 2023;56:86–94.
7. de Lima GCS, Torres US, Bueno LF, et al. Reproducibility of MRI features of uterine leiomyomas: a study on interobserver agreement and inter-method agreement with surgery. *Can Assoc Radiol J*. 2022;73:337–345.
8. Yang T, Yuan L, Li P, et al. Real-time automatic assisted detection of uterine fibroid in ultrasound images using a deep learning detector. *Ultrasound Med Biol*. 2023;49: 1616–1626.
9. Xi H, Wang W. Deep learning based uterine fibroid detection in ultrasound images. *BMC Med Imaging*. 2024;24:218.
10. Cai P, Yang T, Xie Q, et al. A lightweight hybrid model for the automatic recognition of uterine fibroid ultrasound images based on deep learning. *J Clin Ultrasound*. 2024;52:753–762.
11. Isensee F, Petersen J, Klein A, et al. nnU-net: self-adapting framework for U-Net-based medical image segmentation. *arXiv:180910486*. 2018:abs/1809.10486
12. Liu XY, Yuan ZL, Cong FZ, et al. Deep learning assisted detection and segmentation of uterine fibroids using multi-orientation magnetic resonance imaging. *Abdom Radiol (NY)*. 2025;50:4927–4938.
13. Liu Z, Sun C, Li C, et al. 3D segmentation of uterine fibroids based on deep supervision and an attention gate. *Front Oncol*. 2025;15:1522399.
14. Zhang J, Liu Y, Chen L, et al. DARU-Net: a dual attention residual U-Net for uterine fibroids segmentation on MRI. *J Appl Clin Med Phys*. 2023;24:e13937.
15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. Published May 18, 2015. Accessed December 2, 2025. https://arxiv.org/abs/1505.04597
16. Raffone A, Raimondo D, Neola D, et al. Diagnostic accuracy of ultrasound in the diagnosis of uterine leiomyomas and sarcomas. *J Minim Invasive Gynecol*. 2024;31:28–36 e1.
17. Chiappa V, Interlenghi M, Salvatore C, et al. Using rADio-MIcs and machine learning with ultrasonography for the differential diagnosis of myometRiAL tumors (the ADMIRAL pilot study). Radiomics and differential diagnosis of myometrial tumors. *Gynecol Oncol*. 2021;161:838–844.
18. Raffone A, Raimondo D, Neola D, et al. Diagnostic accuracy of MRI in the differential diagnosis between uterine leiomyomas and sarcomas: a systematic review and meta-analysis. *Int J Gynaecol Obstet*. 2024;165:22–33.
19. Roller LA, Wan Q, Liu X, et al. MRI, clinical, and radiomic models for differentiation of uterine leiomyosarcoma and leiomyoma. *Abdom Radiol (NY)*. 2024;49:1522–1533.
20. Slotman DJ, Bartels LW, Nijholt IM, et al. Development and validation of a deep learning-based method for automatic measurement of uterus, fibroid, and ablated volume in MRI after MR-HIFU treatment of uterine fibroids. *Eur J Radiol*. 2024;178:111602.
21. Wei C, Li N, Shi B, et al. The predictive value of conventional MRI combined with radiomics in the immediate ablation rate of HIFU treatment for uterine fibroids. *Int J Hyperthermia*. 2022;39:475–484.
22. VanBuren W, Feldman M, Shenoy-Bhangle AS, et al. Radiology state-of-the-art review: endometriosis imaging interpretation and reporting. *Radiology*. 2024;312:e233482.
23. Nisenblat V, Bossuyt PM, Farquhar C, et al. Imaging modalities for the non-invasive diagnosis of endometriosis. *Cochrane Database Syst Rev*. 2016;2:CD009591.
24. Chen-Dixon K, Uzuner C, Mak J, et al. Effectiveness of ultrasound for endometriosis diagnosis. *Curr Opin Obstet Gynecol*. 2022;34:324–331.
25. Zhang X, He T, Shen W. Comparison of physical examination, ultrasound techniques and magnetic resonance imaging for the diagnosis of deep infiltrating endometriosis: a systematic review and meta-analysis of diagnostic accuracy studies. *Exp Ther Med*. 2020;20:3208–3220.
26. Condous G, Gerges B, Thomassin-Naggara I, et al. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis and endometriosis classification systems: an International Consensus Statement. *Ultrasound Obstet Gynecol*. 2024;64:129–144.
27. Tong A, Cope AG, Waters TL, et al. Best practices: ultrasound versus MRI in the assessment of pelvic endometriosis. *AJR Am J Roentgenol*. 2024;223:e2431085.
28. Dosunmu SD, Sarno A, Lee E, et al. Validation of the revised MUSA criteria for sonographic detection of adenomyosis. *Abdom Radiol (NY)*. 2025. doi:10.1007/s00261-025-05039-y
29. Van den Bosch T, de Bruijn AM, de Leeuw RA, et al. Sonographic classification and reporting system for diagnosing adenomyosis. *Ultrasound Obstet Gynecol*. 2019;53:576–582.
30. Tellum T, Nygaard S, Lieng M. Noninvasive diagnosis of adenomyosis: a structured review and meta-analysis of diagnostic accuracy in imaging. *J Minim Invasive Gynecol*. 2020;27:408–418 e3.
31. Rasmussen CK, Van den Bosch T, Exacoustos C, et al. Intra- and inter-rater agreement describing myometrial lesions using morphologic uterus sonographic assessment: a pilot study. *J Ultrasound Med*. 2019;38:2673–2683.
32. Jiang N, Xie H, Lin J, et al. Diagnosis and nursing intervention of gynecological ovarian endometriosis with magnetic resonance imaging under artificial intelligence algorithm. *Comput Intell Neurosci*. 2022;2022:3123310.
33. Krasnov D, Davis D, Malott K, et al. Fuzzy C-means clustering: a review of applications in breast cancer detection. *Entropy (Basel)*. 2023;25:1021.
34. Moassefi M, Faghani S, Colak C, et al. Advancing endometriosis detection in daily practice: a deep learning-enhanced multi-sequence MRI analytical model. *Abdom Radiol (NY)*. 2025. doi:10.1007/s00261-025-04942-8
35. Nouri B, Hashemi SH, D JG, et al. Machine learning-based detection of endometriosis: a retrospective study in a population of iranian female patients. *Int J Fertil Steril*. 2024;18:362–366.
36. Burla L, Sartoretti E, Mannil M, et al. MRI-based radiomics as a promising noninvasive diagnostic technique for adenomyosis. *J Clin Med*. 2024;13:2344.
37. Zhao Q, Yang T, Xu C, et al. Automatic diagnosis for adenomyosis in ultrasound images by deep neural networks. *Eur J Obstet Gynecol Reprod Biol*. 2024;301:128–134.
38. 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;20:1724.
39. Guerriero S, Condous G, van den Bosch T, et al. Systematic approach to sonographic evaluation of the pelvis in women with suspected endometriosis, including terms, definitions and measurements: a consensus opinion from the International Deep Endometriosis Analysis (IDEA) group. *Ultrasound Obstet Gynecol*. 2016;48:318–332.
40. Mick I, Freger SM, Marien M, et al. Diagnostic accuracy of transvaginal ultrasonography for endometriosis according to the international deep endometriosis analysis consensus. *O G Open*. 2025;2:e061.
41. Tong A, VanBuren WM, Chamie L, et al. Recommendations for MRI technique in the evaluation of pelvic endometriosis: consensus

statement from the Society of Abdominal Radiology endometriosis disease-focused panel. *Abdom Radiol (NY)*. 2020;45:1569–1586.
42. Maicas G, Leonardi M, Avery J, et al. Deep learning to diagnose pouch of Douglas obliteration with ultrasound sliding sign. *Reprod Fertil*. 2021;2:236–243.
43. 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.
44. Xu J, Zhang A, Zheng Z, et al. Development and Validation an AI model to improve the diagnosis of deep infiltrating endometriosis for junior sonologists. *Ultrasound Med Biol*. 2025;51:1143–1147.
45. Zhang Y, Wang H, Butler D, et al. Unpaired multi-modal training and single-modal testing for detecting signs of endometriosis. *Comput Med Imaging Graph*. 2025;124:102575.
46. Wang H, Butler D, Zhang Y, et al. Human-AI collaborative multi-modal multi-rater learning for endometriosis diagnosis. *Phys Med Biol*. 2024;70:015008.
47. He Q, Zhang C, Hu Y, et al. The improved-EFI score: a multi-omics-based novel efficacy predictive tool for predicting the natural fertility of endometriosis patients. *Int J Gen Med*. 2025;18:881–895.