ISSN: 2835-7914

Research Article

Biomedical Science and Clinical Research

Artificial Intelligence in the Comparative Interpretation of Ultrasound and MRI DICOM Data for Endometriosis: A Commentary on Current Evidence, Interoperability Barriers and Research Priorities

Raouf Roshdy* and **REHAB ELSAID NOUR ELDIN YOUSSEF**

[1]*Egypt Health Foundation, Egypt*

[2]*Alexandria University, Egypt*

***Corresponding Author**
Raouf Roshdy, Egypt Health Foundation, Egypt.

Submitted: 2026, Aug 24; **Accepted:** 2026, Sep 22; **Published:** 2026, Sep 30

Citation: Roshdy, R., Youssef, R. E. N. E. (2026). Artificial Intelligence in the Comparative Interpretation of Ultrasound and MRI DICOM Data for Endometriosis: A Commentary on Current Evidence, Interoperability Barriers and Research Priorities. *Biomed Sci Clin Res, 5*(3), 01-04.

Abstract

Transvaginal ultrasound (TVUS) and pelvic magnetic resonance imaging (MRI) are both used, often in the same patient, to evaluate suspected endometriosis, and comparative diagnostic-accuracy studies indicate that the two modalities capture partly complementary information rather than one simply outperforming the other across all disease sites. Artificial intelligence (AI) has been applied with increasing frequency to endometriosis imaging, but a recent scoping review found that the evidence base remains concentrated in single-modality, single-centre, retrospective studies with internal validation only, and that models jointly using ultrasound and MRI DICOM data from the same patient are essentially absent from the published literature. This commentary summarizes what is empirically established about ultrasound-MRI comparison in endometriosis, what AI has and has not yet demonstrated, and the DICOM-level interoperability barriers that stand ahead of any same-patient multimodal AI system. Statements that extend beyond what cited sources report are explicitly flagged.

1. Introduction

Endometriosis is a chronic gynecological condition whose diagnosis has historically depended on laparoscopic visualization with histopathological confirmation, a pathway that by definition excludes women who are never referred for surgery. Repeated documentation of multi-year diagnostic delay from symptom onset to confirmed diagnosis has been summarized in the recent scoping-review literature; this commentary reports that summary rather than the underlying primary studies, which were not independently retrieved for this document. Imaging has become the principal non-surgical route to diagnosis, with transvaginal ultrasound recommended as the first-line modality and pelvic MRI used as a complementary second-line study, particularly for anterior-compartment and multi-compartmental disease and for preoperative planning [1].

Because both modalities are frequently performed in the same patient during a single diagnostic work-up, the two DICOM datasets they generate are, in principle, a natural substrate for a combined or comparative AI analysis. Whether that combination has actually been built, validated, and shown to add value beyond either modality alone is a separate, empirical question, and it is the question this commentary addresses.

2. What is Established about Ultrasound–MRI Comparison in the Same Patient

Several studies have evaluated TVUS and MRI in the same cohort against a surgical or histopathological reference standard. In a single-center series of 317 women, MRI showed the highest overall accuracy (85.4%) compared with transvaginal sonography (75.7%) and transrectal sonography (67.8%), although sensitivity for uterosacral ligament disease was, counter-intuitively, higher for transrectal and transvaginal sonography (82.8% and 70.9%) than for MRI (63.6%), while specificity favored MRI [2]. A 2025 systematic review pooling nine studies that compared MRI and TVUS in the same patients reported rectosigmoid sensitivities of 79–94% for TVUS and 86–94% for MRI, high specificity for both (84–95%), lower TVUS sensitivity for uterosacral-ligament and bladder lesions (25–83%), and higher MRI sensitivity for anterior and multi-compartmental disease (75–94%); the authors concluded that the two modalities are complementary rather than interchangeable [3]. A separate meta-analysis of ten studies on rectosigmoid deep infiltrating endometriosis (1,604 patients) found no statistically significant difference in pooled sensitivity (TVUS 0.85 vs MRI 0.83) or specificity (0.92 vs 0.95) between the modalities [4].

Taken together, this body of evidence supports a specific and fairly narrow claim: TVUS and MRI have broadly similar aggregate accuracy for deep infiltrating endometriosis, with site-specific strengths that favor ultrasound in the posterior compartment and MRI in anterior and multi-compartmental disease. It does not by itself establish anything about what an AI system trained on the two modalities together would add; that is a distinct and largely untested question, addressed in Section 3.

3. What AI has Demonstrated in Endometriosis Imaging and What it has Not

The most directly relevant evidence comes from a 2026 PRISMA-ScR scoping review of 32 primary AI studies in endometriosis imaging published between 2018 and 2025. Its findings are worth reporting in some detail because they define the actual, current state of the field rather than its aspirational one. Ultrasound was the dominant imaging modality (16/32 studies, 50%), followed by laparoscopic video (8/32, 25%) and MRI (7/32, 22%); classification of disease presence versus absence was the dominant AI task (25/32, 78%), typically using convolutional neural networks, most often ResNet variants (14/32, 44%). Reported diagnostic performance varied widely by task and reference standard, with area-under-curve values for classification tasks clustering around 0.80–0.99 and segmentation Dice scores ranging from roughly 0.29 to 0.96 depending on the anatomical target and architecture.

The methodological picture is less encouraging than the headline accuracy figures suggest. Internal validation only was used in 31 of 32 studies (97%); external validation, on a genuinely separate institution or population, was reported in a single study. Explicit confirmation that all images from a given patient were kept within a single training/test split a basic safeguard against data leakage was present in only 12 of 32 studies (38%), and was unclear in over half. Calibration reporting appeared in 2 of 32 studies (6%), and code availability in 4 of 32 (12%). The review's own conclusion, not a paraphrase of it, is that AI in endometriosis imaging is "rapidly evolving yet remains at an early stage of development," limited chiefly by retrospective single-center design and internal-validation-only evaluation.

On the specific question of models that use ultrasound and MRI from the same patient jointly, the evidence is thin. One conference study explored self-supervised pre-training for "multi-modal endometriosis classification"; [5] the accessible source material for that study does not specify, to the level of detail available for this commentary, whether the modalities combined were ultrasound and MRI DICOM series from the same patients, or a different combination of inputs. That detail is therefore as reported here, and the reader should not infer a confirmed ultrasound-plus-MRI same-patient fusion model from this citation alone. A separate 2025 resource describes a public multi-modal pelvic MRI dataset for deep-learning segmentation in endometriosis, but "multi-modal" there refers to multiple MRI sequences within one modality (e.g., T2-weighted, diffusion-weighted) rather than cross-modality fusion with ultrasound [6].

To my knowledge, and based on the searches conducted for this commentary, a validated AI system that ingests same-patient ultrasound and MRI DICOM data specifically for endometriosis and produces a joint or comparative diagnostic output has not been reported in the accessible peer-reviewed literature. This is a direct statement of an evidence gap, not a speculative claim, and it should be read as such.

4. The DICOM-Level Interoperability Barrier

Before a same-patient, cross-modality AI comparison can be built at any scale beyond a single research cohort, a more basic technical problem has to be addressed: ultrasound and MRI DICOM data are structurally heterogeneous, and MRI data are themselves heterogeneous across scanners and protocols. Work on radiomics reproducibility in a multicenter rectal MRI dataset found that 64.3% of the variation in a commonly used diffusion metric (apparent diffusion coefficient) was explained by differences in hardware and acquisition protocol across nine centers, versus 0.4% explained by patient-intrinsic factors such as age or sex, and that feature reproducibility between different segmentation approaches was only moderate (intraclass correlation 0.40–0.58 for single-slice versus whole-volume segmentation) [7].

This finding is from oncological rectal MRI, not endometriosis; it is cited here because the underlying technical problem that scanner and protocol variation can dominate over biological signal in extracted image features is a property of MRI radiomics pipelines generally, and applying it to endometriosis-specific same-patient models is A separate narrative review of the Image Biomarker Standardization Initiative (IBSI) describes it as complementary to DICOM and to the Quantitative Imaging Biomarkers Alliance (QIBA) acquisition standards, functioning as a shared framework intended to reduce inter-software variability in radiomics feature extraction and to support multicenter workflow comparability [8].

A broader review of barriers to clinical translation of AI-based radiomics identifies non-standardized data formats, incomplete metadata, and a lack of centralized, well-annotated repositories as recurring obstacles across cancer-imaging applications, alongside the more familiar problems of small training sets and overfitting [9]. A worked example from a cancer-imaging archive illustrates the practical scale of this problem: standard preprocessing for a radiomics or deep-learning pipeline required DICOM-to-NIfTI conversion, orientation harmonization, voxel-spacing normalization, and intensity normalization before the data were usable, specifically because DICOM stores each slice as a separate file with distributed metadata while most computational imaging libraries expect a single, geometrically consistent volume [10].

None of these sources describe endometriosis specifically, and none describe a same-patient ultrasound-MRI linkage problem, since ultrasound is not typically DICOM-volumetric in the way cross-sectional MRI is and carries its own separate set of format and metadata conventions. The general point that transfers with reasonable confidence is that harmonizing two structurally different DICOM sources from the same patient one cross-sectional and volumetric, one operator-dependent and largely 2D or Doppler-based adds a layer of preprocessing and standardization work beyond what any single-modality endometriosis AI study reviewed above has had to solve, since none of them combined the two modalities as model input.

5. A Research Agenda for Same-Patient Multimodal AI in Endometriosis

The following priorities follow from the gaps documented in Sections 3 and 4. They are presented as a proposed agenda, not as a set of established findings, and are marked [Speculation] as a block rather than sentence by sentence, since none of them has, to my knowledge, been implemented and reported for endometriosis specifically:

- Consistent linkage, at the metadata level, of same-patient ultrasound and MRI studies performed within a single diagnostic episode, so that paired data can be identified without manual chart review.
- Application of IBSI-aligned feature-extraction standards to any MRI radiomics component, alongside explicit reporting of scanner make, field strength, and sequence parameters, given the magnitude of hardware-driven variation documented in other radiomics domains.
- Patient-level (not image-level) partitioning of training, validation, and test sets, with explicit leakage-prevention reporting a safeguard that the 2026 scoping review found to be confirmed in well under half of existing endometriosis-imaging AI studies.
- External, multicenter validation as a prerequisite for any claim of generalizable accuracy, since only one of the 32 studies reviewed in the current evidence base performed this step.
- Prospective comparison of three arms unaided radiologist/sonographer interpretation, single-modality AI, and joint ultrasound-plus-MRI AI interpretation against a surgical or histopathological reference standard, to establish whether combining modalities in a model changes diagnostic yield relative to combining them only in the clinician's head, as currently occurs in routine multimodal work-up.

Whether such a system would improve on current practice is not something existing evidence can answer, and I am not able to state that it would; that outcome remains [Speculation] until a study of this design has been conducted and reported.

6. Limitations and Epistemic Status of This Commentary

This is a narrative commentary, not a systematic review; the literature search underlying it was targeted and non-exhaustive, and relevant studies may have been missed, particularly non-English-language work and unindexed grey literature. Several sources cited in Sections 3 and 4 concern imaging AI in oncology or endometrial cancer rather than endometriosis; where used, this is stated explicitly and the inference is flagged rather than presented as endometriosis-specific evidence. For two cited items (references 4 and 5), the full author list could not be confirmed from the sources accessible during preparation of this document, and they are referenced by title and source location rather than by author name, consistent with a preference for not asserting unconfirmed attribution. No claim in this commentary should be read as a guarantee of any AI system's future clinical performance; where AI-behavior claims are made, they are qualified as inference or flagged as unverified in the surrounding text, per the labelling convention stated in the abstract.

7. Conclusion

Ultrasound and MRI, performed in the same patient, provide diagnostically complementary not interchangeable information for endometriosis, and this is reasonably well established by direct comparative studies. AI applied to endometriosis imaging is an active but methodologically early field, concentrated in single-modality, single-center, internally validated models. A validated AI system that jointly interprets same-patient ultrasound and MRI DICOM data for endometriosis does not appear, from the evidence located for this commentary, to exist yet in the peer-reviewed literature, and the DICOM-level and protocol-level harmonization work that would need to precede such a system is itself only partly solved in adjacent imaging domains. The research agenda in Section 5 follows from these gaps but is not itself evidence that the agenda will succeed.

References

1. AlSaad, R., Farrell, T., Elhenidy, A., Albasha, S., & Thomas, R. (2026). Artificial Intelligence in Endometriosis Imaging: A Scoping Review. *AI, 7*(2), 43.
2. Alborzi, S., Rasekhi, A., Shomali, Z., Madadi, G., Alborzi, M., et al. (2018). Diagnostic Accuracy of Magnetic Resonance Imaging, Transvaginal, and Transrectal Ultrasonography in Deep Infiltrating Endometriosis. *Medicine, 97*(8), e9536.
3. O'Leary, M., Neary, C., & Lawrence, E. (2025). The Diagnostic Accuracy of Magnetic Resonance Imaging Versus Transvaginal Ultrasound in Deep Infiltrating Endometriosis and Their Impact on Surgical Decision-Making: A Systematic Review. *Diagnostics, 15*(22), 2856.
4. Xu, Z., Li, Y., Wang, Y., Wan, Y., & Chen, J. (2025). Transvaginal Ultrasound and Magnetic Resonance Imaging in Detecting Rectosigmoid Deep Infiltrating Endometriosis: A Comparative Meta-Analysis. *Frontiers in Medicine, 12*, 1552185.
5. Butler, D., Wang, H., Zhang, Y., To, M.-S., Condous, G., et al. (2023). The Effectiveness of Self-Supervised Pre-Training for Multi-Modal Endometriosis Classification. In *Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (*EMBC*). IEEE.
6. Liang, X., Alpuing Radilla, L. A., Khalaj, K., Dawoodally,

H., Mokashi, C., et al. (2025). A Multi-Modal Pelvic MRI Dataset for Deep Learning-Based Pelvic Organ Segmentation in Endometriosis. *Scientific Data, 12*, 1292.
7. Schurink, N. W., Van Kranen, S. R., Roberti, S., Van Griethuysen, J. J., Bogveradze, N., et al. (2022). Sources of Variation in Multicenter Rectal MRI Data and Their Effect on Radiomics Feature Reproducibility. *European radiology, 32*(3), 1506-1516.
8. Aguirre-Meneses, H., Stoehr-Muñoz, P., Molina-Gonzalez, M., Nuñez-Gaona, M. A., Roldan-Valadez, E., et al. (2025). Radiomics and the Image Biomarker Standardisation Initiative (IBSI): A Narrative Review Using A Six-Question Map and Implementation Framework for Reproducible Imaging Biomarkers. *Cureus, 17*(10).
9. Xu, Y., Li, Y., Wang, F., Zhang, Y., & Huang, D. (2025). Addressing the Current Challenges in the Clinical Application of AI-based Radiomics for Cancer Imaging. *Frontiers in Medicine, 12*, 1674397.
10. Rich, J., Kang, R., Jin, D., Subramanian, S., Duddalwar, V., et al. (2026). TCIA Radiology Image Processing for AI and Radiomics. *medRxiv*, 2026-06.