

Interobserver reliability and diagnostic accuracy of transvaginal ultrasound for adenomyosis: a multi-observer study

Raheem M^{1,2}, McClenahan P^{1,2}, Al Otaibi M^{1,2}, Espada M^{1,2}, Gill A^{1,2}, Eathorn A¹, Reid S¹, Condous G^{1,2}.
¹ Endometriosis and Advanced Endosurgery Unit
² Nepean hospital

Introduction

Adenomyosis is characterised by the presence of endometrial tissue within the myometrium and is associated with heavy menstrual bleeding, pelvic pain, infertility, and reduced quality of life. Transvaginal ultrasound (TVUS) is the primary imaging modality for diagnosis; however, interpretation of sonographic features may vary between observers. The Morphological Uterus Sonographic Assessment (MUSA) group established standardised ultrasound criteria for adenomyosis, including features such as myometrial cysts, hyperechoic islands, fan-shaped shadowing, asymmetrical myometrial thickening, and an irregular junctional zone. This study evaluates the interobserver reliability of identifying adenomyosis and its sonographic features using the MUSA criteria.

Aim

To evaluate the diagnostic accuracy and interobserver agreement of transvaginal ultrasound (TVUS) to predict adenomyosis.

Methods

Patients with complete preoperative transvaginal ultrasound and subsequent hysterectomy histopathology were included. Five independent observers with varying levels of expertise (senior, intermediate, and junior) reviewed stored ultrasound examinations consisting of still images and video clips. Observers were blinded to histopathological outcomes and assessed the presence or absence of adenomyosis according to the revised Morphological Uterus Sonographic Assessment (MUSA) criteria. Individual MUSA features assessed included myometrial cysts, hyperechogenic islands, echogenic subendometrial lines and buds, asymmetrical myometrial thickening, globular uterus, translesional vascularity, irregular junctional zone, interrupted junctional zone, and fan-shaped shadowing.

Diagnostic performance (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) was calculated using histopathology as the reference standard. Interobserver agreement was assessed using Fleiss' kappa and pairwise Cohen's kappa coefficients. Subgroup analyses compared diagnostic performance between still images and video clips.

Results and Discussion

- Moderate diagnostic accuracy was observed across observers (61–72%; mean 66.5%).
- Still images outperformed video clips overall (70.1% vs 62.9% accuracy), suggesting that selection of representative images may be more important than dynamic assessment.
- Sensitivity exceeded specificity for most observers, indicating a tendency to over-diagnose adenomyosis and generate false positives.
- Inter-observer agreement was poor (Fleiss' $\kappa = 0.18$), highlighting significant variability in interpretation of MUSA features and the need for further standardisation and training.

Conclusion

Ultrasound demonstrated moderate diagnostic accuracy for adenomyosis when compared with histopathology, but inter-observer agreement was low. Still images showed higher diagnostic accuracy than video assessment. These findings highlight the subjective nature of ultrasound diagnosis of adenomyosis and support the need for improved standardisation of image interpretation and observer training.

References

1. features of adenomyosis: results of a modified Delphi procedure. *Ultrasound Obstet Gynecol.* 2022;60(1):118–131. doi:10.1002/uog.24786.
2. Juan Luis Alcázar, Vara J, Usandizaga C, Ajossa S, Pascual MÁ, Guerriero S. *Transvaginal ultrasound versus magnetic resonance imaging for diagnosing adenomyosis: A systematic review and head-to-head meta-analysis.* *Int J Gynecol Obstet.* 2023;161(2):397–405.
3. Michele Tellum, Rasmussen CK, Petersson U, Mettler L. *Transvaginal Ultrasound for the Diagnosis of Adenomyosis: Systematic Review and Meta-Analysis.* *J Minim Invasive Gynecol.* 2018;25(2):257–264
4. Katherine Dartmouth. *A systematic review with meta-analysis: the common sonographic characteristics of adenomyosis.* *Ultrasound.* 2014;22(3):148–157.

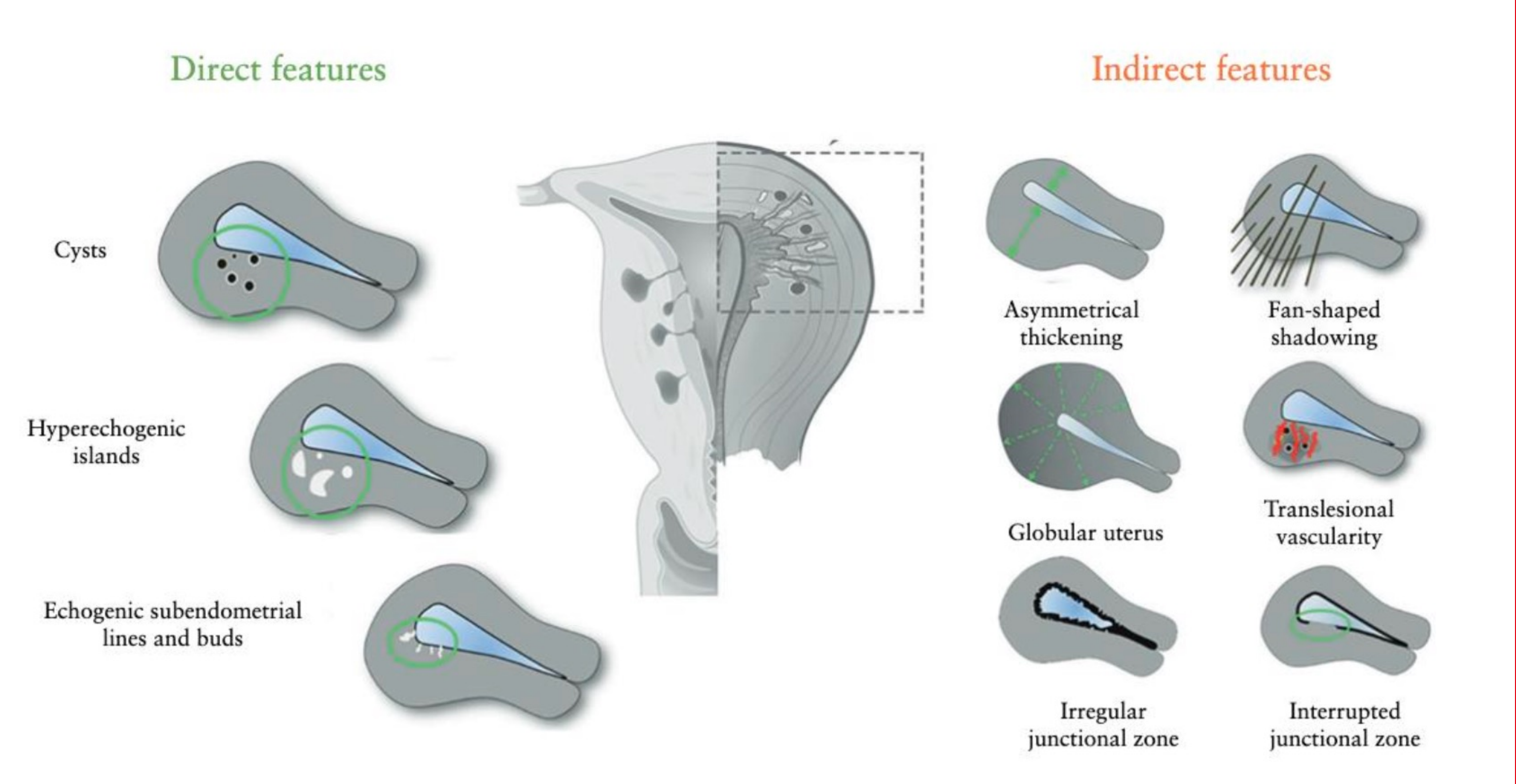


Figure 1. The Revised MUSA criteria

Outcome	Result
Overall diagnostic accuracy	61.3-72.1% (mean 66.5%)
Sensitivity	68.2-90.9%
Specificity	17.6-62.5%
Mean accuracy – still images	70.1%
Mean accuracy – video clips	62.9%
Interobserver agreement	0.18 (95% CI 0.16-0.19)
Pairwise agreement	0.10-0.34

Figure 2. Table showing diagnostic accuracy and interobserver reliability of ultrasound diagnosis of adenomyosis

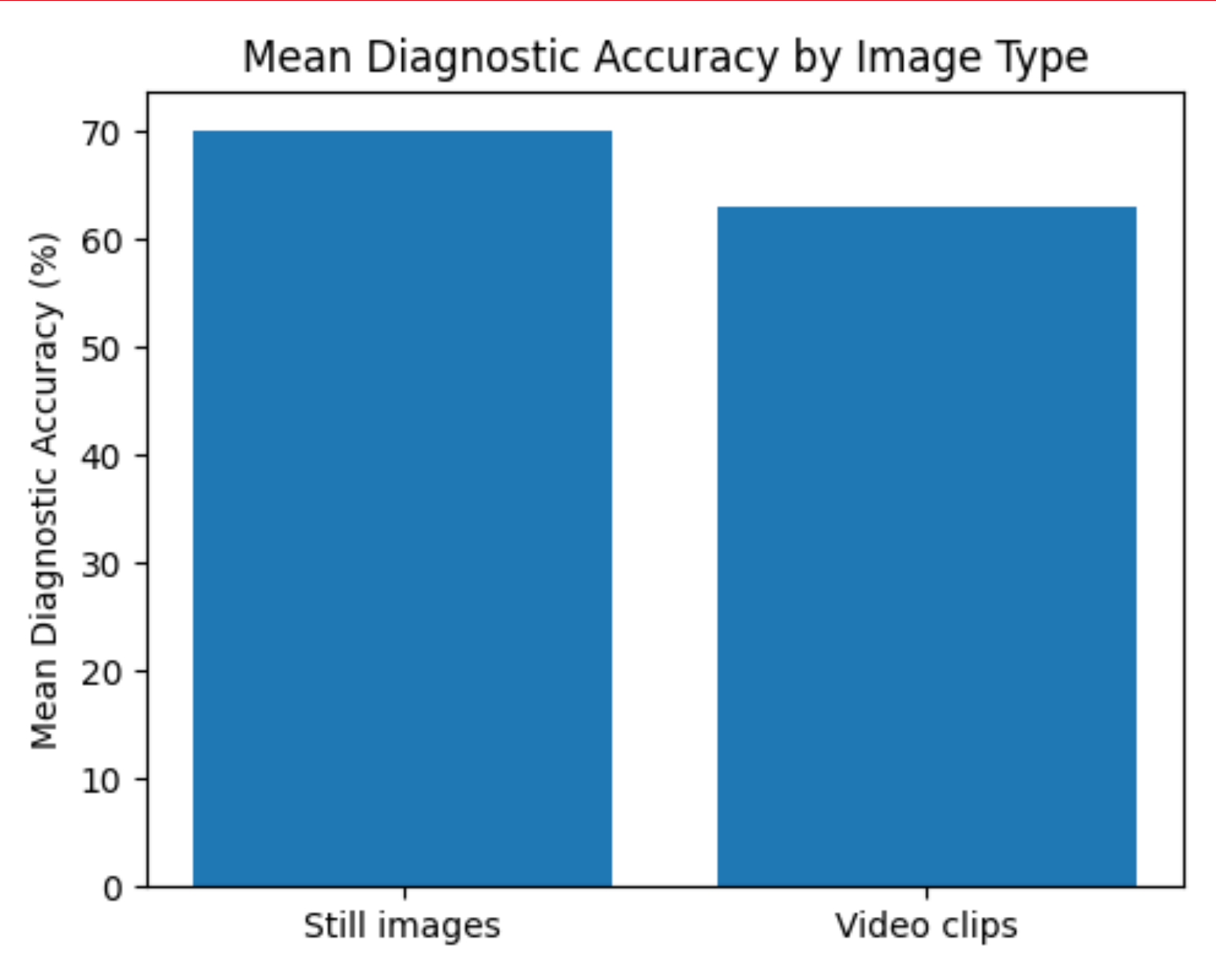


Figure 3. Figure showing diagnostic accuracy by image type