EISSN 1305-3612

Diagnostic and Interventional Radiology

TSR 1924
TURKISH SOCIETY OF RADIOLOGY

dirjournal.org

VOLUME 32
ISSUE 3
May 2026

Diagnostic and Interventional Radiology

Official Journal of the Turkish Society of Radiology

E-ISSN: 1305-3612
www.dirjournal.org

Diagnostic and Interventional Radiology

Official Journal of the Turkish Society of Radiology

E-ISSN: 1305-3612
www.dirjournal.org

Turkish Society of Radiology is one of the foremost medical specialty organizations in Türkiye. It was formed by the merger of the two main radiology societies of Türkiye, one of which was founded in 1924. The Society is based in Ankara, Türkiye.

TSR 1924
TURKISH SOCIETY OF RADIOLOGY

Diagnostic and Interventional Radiology

Official Journal of the Turkish Society of Radiology

E-ISSN: 1305-3612

www.dirjournal.org

Owner: Turkish Society of Radiology

Responsible Manager: Şükrü Mehmet Ertürk

Contact Information

Diagnostic and Interventional Radiology Turkish Society of Radiology

Hoşdere Cad., Güzelkent Sok., Çankaya Evleri, F/2, 06540
Ankara, Türkiye

E-mail: info@dirjournal.org

Phone: +90 (312) 442 36 53 **Fax:** +90 (312) 442 36 54

Publisher Contact

Address: Molla Gürani Mah. Kaçamak Sk.
No: 21/1 34093 İstanbul, Türkiye

Phone: +90 (530) 177 30 97

E-mail: info@galenos.com.tr/yayin@galenos.com.tr

Web: www.galenos.com.tr **Publisher Certificate Number:** 14521

Online Publication Date: May 2026

EISSN 1305-3612

International scientific journal published bimonthly.

TURKISH SOCIETY OF RADIOLOGY

Turkish Society of Radiology is one of the foremost medical specialty organizations in Türkiye. It was formed by the merger of the two main radiology societies of Türkiye, one of which was founded in 1924. The Society is based in Ankara, Türkiye.

Diagnostic and Interventional Radiology

Volume 32 • Issue 3

May 2026

www.dirjournal.org

Contents

Full text of these articles can be accessed online at www.dirjournal.org or through PubMed (https://www.ncbi.nlm.nih.gov/pmc/journals/2754/).

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253306

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Comparison of three-tesla magnetic resonance imaging with pathology in detecting deep myometrial invasion in endometrial cancer and revealing causes of discrepancy

Murat Ağırlar[1]
İlkay Çamlıdağ[2]
Murat Danacı[2]

[1]Çarşamba State Hospital, Clinic of Radiology, Samsun, Türkiye

[2]Ondokuz Mayıs University Faculty of Medicine, Department of Radiology, Samsun, Türkiye

PURPOSE

To evaluate the accuracy of three-tesla (3T) magnetic resonance imaging (MRI) in determining deep myometrial invasion (DMI) in endometrial cancer (EC) cases and to reveal possible causes of discrepancy.

METHODS

Patients with EC who underwent preoperative pelvic MRI examination at a tertiary hospital were independently and blindly reviewed for DMI by two radiologists with differing levels of expertise. On MRI, the invasion of the endometrial mass into the myometrium was defined as superficial or deep (<50% or ≥50%) and was compared with the gold standard pathology reports, evaluated using the same criteria. Cases with discrepancies between MRI and pathology were re-evaluated, and the causes of the discrepancies were identified.

RESULTS

A total of 226 patients were included. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of MRI in detecting DMI were calculated as 77%, 93%, 84%, 88%, and 87%, respectively. In cases where radiology and pathology results were incompatible, expansile masses were found in 38%, myomas located near the lesion in 27%, masses located in the uterine horn in 21%, and in 14%, a specific cause explaining the error could not be identified.

CONCLUSION

MRI is the definitive diagnostic method for determining DMI. Awareness of these causes of discrepancy in MRI reporting can increase the value of the examination and contribute to patient management.

CLINICAL SIGNIFICANCE

This study highlights that 3T MRI has a high accuracy (87%) in detecting DMI in EC and demonstrates that analyzing misinterpretations can enhance reliability in patient management.

KEYWORDS

Causes of discrepancy, deep myometrial invasion, diagnostic accuracy, endometrial carcinoma, magnetic resonance imaging

Corresponding author: Murat Ağırlar

E-mail: dr.muratagirlar@gmail.com

Received 28 February 2025; revision requested 23 March 2025; last revision received 11 June 2025; accepted 23 June 2025.

Epub: 21.07.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253306

Endometrial cancer (EC) is the most common gynecological cancer in developed countries. The majority of cases occur in individuals aged between 60 and 70 years.[1-3] EC is categorized into two histopathological subtypes: type 1 (80%–85%) and type 2 (10%–15%). Type 1 is estrogen-dependent and primarily affects younger women in the premenopausal or perimenopausal period. It consists of grade 1–2 endometrioid adenocarcinomas. In contrast, type 2 EC affects older women in the postmenopausal period. Histologically, it includes grade 3

You may cite this article as: Ağırlar M, Çamlıdağ İ, Danacı M. Comparison of three-tesla magnetic resonance imaging with pathology in detecting deep myometrial invasion in endometrial cancer and revealing causes of discrepancy. *Diagn Interv Radiol*. 2026;32(3):242-248.

endometrioid adenocarcinomas and other rare subtypes.[4-7] Approximately 67% of patients present with early-stage disease, and at this stage, the 5-year survival rate is 81%.[8] The prognosis depends on several factors, including histological type and grade, deep myometrial invasion (DMI), cervical invasion, and lymphovascular invasion.

The International Federation of Gynecology and Obstetrics (FIGO) classification system used for staging EC is a surgical staging procedure involving total hysterectomy with bilateral salpingo-oophorectomy, along with a thorough examination of the omentum, peritoneum, and intra-abdominal contents. Magnetic resonance imaging (MRI) is not included in the formal staging system for EC; however, it plays a critical role in treatment planning and determining the most appropriate initial therapy.[3] According to the classification updated by FIGO in 2023, DMI–defined as infiltration of 50% or more of the myometrium–is one of the criteria for stage IB EC. Additionally, a new stage IC category was introduced for aggressive histological subtypes, provided the tumor is confined to a polyp or the endometrium. Accordingly, the prognostic significance of DMI, a key factor in EC treatment and outcomes, has increased.[9-12] DMI is a predictor of extrauterine disease and lymph node metastasis. Lymph node metastasis occurs in 30% of DMI cases, compared with only 5% in cases of superficial myometrial invasion (SMI).[13,14]

In this context, lymph node metastasis–the strongest predictor of recurrence–is associated with DMI. Lymphadenectomy may be considered for grade 3 and other high-risk EC cases. Additionally, in patients with grade 1–2 tumors, lymph node dissection may be considered if DMI is present.[15] However, it has been shown that systematic lymphadenectomy does not improve overall survival in early-stage disease. Moreover, lymphadenectomy may lead to complications such as lymphedema and lymphocysts; therefore, it should be avoided in low-risk patients.[16,17] Identification of DMI through preoperative MRI assessment is thus clinically substantial for surgical planning.

In the literature, studies comparing MRI with intraoperative frozen section evaluation and postoperative pathological evaluation for determining DMI have reported high sensitivity, specificity, and accuracy for MRI, supporting its use as a superior diagnostic tool. MRI is the recommended imaging modality for preoperative radiological evaluation in EC.[9,16,18-20] However, to our knowledge, no studies have specifically focused on the underlying causes of diagnostic errors in MRI.

One objective of this study is to evaluate the sensitivity, specificity, and accuracy of three-tesla (3T) MRI in detecting DMI by comparing its findings with postoperative pathological evaluations in cases of EC. Another goal is to identify the causes of discordant results by re-evaluating cases with incorrect radiological assessments through a secondary review.

Main points

- Three-tesla magnetic resonance imaging (MRI) has a high accuracy (87%) in detecting deep myometrial invasion in endometrial cancer (EC).
- Identifying misinterpretations in imaging can improve diagnostic reliability and enhance patient management.
- MRI-based assessment of myometrial invasion can aid in preoperative planning, potentially reducing unnecessary extensive surgeries.
- Understanding the limitations and pitfalls of MRI interpretation is crucial for radiologists and clinicians in optimizing patient outcomes.
- A systematic approach to image analysis and error reduction can lead to better decision-making in EC management.

Methods

Patient selection

This retrospective study was approved by the Ondokuz Mayıs University Clinical Research Ethics Committee (protocol number: 2022/450, date: 13.10.2022). The requirement for informed consent was waived due to the retrospective nature of the study. Patients with a pathological diagnosis of EC who underwent pelvic staging MRI between January 2016 and June 2022, and who subsequently underwent hysterectomy as part of surgery following the MRI, were eligible for inclusion (n = 240).

Cases with tumors extending beyond the uterus and serosa (n = 8) and those with artifacts hindering proper evaluation due to intestinal peristalsis or hip prosthesis (n = 5) were excluded. Additionally, cases in which the interval between MRI and surgery exceeded 1 month (n = 2) were also excluded. The remaining patients underwent surgery within 1 month after MRI. The study workflow is summarized in the flowchart (Figure 1).

Multiparametric magnetic resonance imaging protocol

To reduce artifacts caused by peristalsis, the patients were instructed to fast for 6

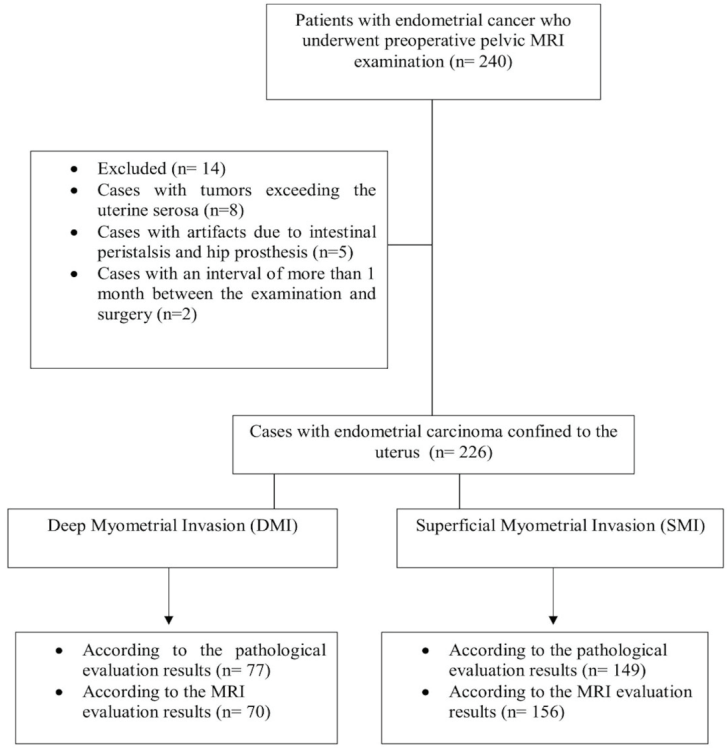


Figure 1. Flowchart of the study. MRI, magnetic resonance imaging.

hours and to empty their bladders approximately half an hour before the examination. All examinations were performed using a 3T MRI scanner (Philips Ingenia, Netherlands) with a 16-channel torso coil. The patients were scanned in the supine position. High-resolution oblique axial and coronal images, sagittal T2-weighted images, axial T1-weighted images, axial diffusion-weighted images, and dynamic contrast-enhanced 3D T1-weighted images were obtained. For contrast-enhanced imaging, 0.1 mmol/kg of gadoteric acid was intravenously injected at a rate of 2 mL/s using a power injector, and images were acquired at 30, 60, 90, and 120 seconds. Imaging parameters are summarized in Table 1.

Image analysis

All images were independently analyzed by an abdominal radiologist with 20 years of experience and a general radiologist with 5 years of experience in abdominal/pelvic MRI interpretation. Both were blinded to the histopathological results. Preoperative MRI staging was performed according to the revised 2023 FIGO criteria. The degree of myometrial invasion was primarily assessed on high-resolution T2-weighted images obtained at a specific angle. Correlation was then made with non-angled diffusion-weighted imaging and contrast-enhanced series. The component of the endometrial lesion invading the myometrium was measured in millimeters and normalized to the myometrial thickness (Figure 2). It was documented whether myometrial invasion was <50% or ≥50%. During comparison of MRI results with histopathological findings, cases with discordant assessments were re-evaluated jointly by the two radiologists, and reasons for discrepancies were identified.

Histopathological examination

All pathological examinations were performed by an expert gynecological pathologist. Diagnoses were determined according to the 2014 World Health Organization classification of tumors of the uterine corpus. In the hysterectomy specimens, the depth of tumor invasion into the myometrium and the corresponding myometrial thickness were measured microscopically in a manner consistent with MRI assessment. DMI was considered present when the depth of invasion was ≥50%.

Statistical analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences version 22.0 (IBM, New York, USA). Descriptive statistics were expressed as numbers and percentages for categorical variables and as mean, standard deviation, minimum, and maximum for continuous variables.

The chi-square test was used to assess differences between categorical variables. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of MRI in detecting the depth of myometrial invasion were calculated. A *P* value of <0.05 was considered statistically significant.

Results

A total of 226 patients with EC were eligible for the study. The mean age of the patients was 60 ± 10 years (range: 36–94 years). The average size of mass lesions was 27 ± 14 mm (range: 10–92 mm). Following histopathological evaluation, 25% of patients had endometrial adenocarcinoma, endometrioid type, grade 1 (n = 58); 42% had grade 2 (n = 94); 11% had grade 3 (n = 24); and 22% had other less commonly observed histopathological diagnoses (n = 50).

According to EC types, the mean age of 152 patients (67%) with type 1 was 58 ± 9.5 years, and the mean age of 74 patients (33%) with type 2 was 64 ± 9.7 years.

In type 1 EC, 24.4% (n = 38) of patients had DMI and 75.6% (n = 114) had SMI. In type 2 EC, 56.6% (n = 43) had DMI and 43.4% (n = 33) had SMI. The difference between type 1 and type 2 with regard to DMI was statistically significant ($P < 0.001$). There was no statistically significant difference between type 1 and type 2 clinical classifications regarding discordant results between MRI and histopathology in detecting DMI ($P = 0.307$).

According to histopathological examination, DMI was detected in 77 patients (34%), whereas 149 patients (66%) had SMI.

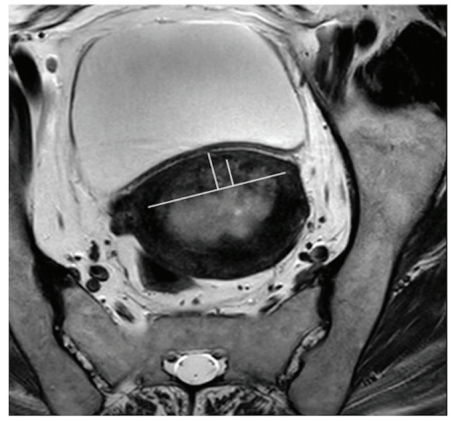


Figure 2. The long vertical line perpendicular to the plane and parallel to the inner border of the myometrium illustrates the thickness of the myometrium, whereas the short line represents the depth of extension of the mass into the myometrium in millimeters. A ratio of 0.5 or greater between the depth of extension and the thickness of the myometrium indicates deep myometrial invasion.

Table 1. Magnetic resonance imaging protocol for imaging of endometrial cancer

	TR (ms)	TE (ms)	ST (mm)	SS (mm)	FOV (mm)	FA (degrees)	Matrix	*b* value (s/mm²)
T2 TSE sagittal	3000–5000	90	5	1	300	90	300 × 264	
T2 TSE oblique coronal	3000–5000	90	3	0.3	200	90	308 × 252	
T2 TSE oblique axial	3000– 5000	90	3	0.3	200	90	308 × 253	
mDIXON–IP/OP axial	3.2	TE1: 1.12 TE2: 2	3	-1.5	350	none	196 × 144	
DWI axial	5704	77	5	0	375	none	124 × 106	0, 200, 800
mDIXON axial (contrast-enhanced, fat-suppressed)*	3.2	TE1: 1.13 TE2: 2	3	-1.5	350	none	200 × 165	

*0.1 mmol/kg gadoteric acid was administered intravenously with a power injector at a rate of 2 mL/sec.
TR, recovery time; TE, echo time; ST, slice thickness; SS, slice spacing; FOV, field of view; FA, flip angle; TSE, turbo spin echo; IP, in-phase; OP, out-of-phase; DWI, diffusion-weighted imaging.

On MRI, 70 patients (31%) were found to have DMI, whereas 156 patients (69%) had SMI. In 11 patients with SMI on histopathological evaluation, MRI was interpreted as DMI. In 18 patients with DMI identified on histopathological evaluation, MRI was incorrectly interpreted as not having DMI, leading to understaging (Table 2). MRI had a sensitivity of 77%, specificity of 93%, PPV of 84%, NPV of 88%, and an accuracy of 87% in detecting DMI.

In 197 of 226 patients evaluated with MRI, the results were compatible with pathology in terms of DMI. In 29 patients with discrepant results, we investigated the causes of discrepancy through secondary retrospective evaluation (Table 3).

In 11 patients with incorrect staging on MRI, mass lesions were observed, causing thinning of the myometrium due to expansion. As a result, six of these patients were understaged and five were overstaged (Figure 3).

Eight patients had fibroids located in contact with or within 1 cm of the endometrial mass. Uterine fibroids disrupted the homogeneous structure of the myometrium, making it difficult to determine the extent of the lesion's invasion. As a result, six patients were understaged and two were overstaged (Figure 4).

In 6 patients, the mass lesions were located in the uterine horn. The thinning of the myometrium at the level of the uterine horn made it difficult to determine the depth of myometrial invasion. As a result, three patients were understaged and three were overstaged (Figure 5).

In MRI evaluations of 4 patients where misclassification occurred, we could not identify any substantial factor that might have caused the discrepancy. We classified the cause in this group as "personal factors" (Figure 6).

We analyzed the frequency of previously identified factors leading to erroneous evaluations in the 197 cases where radiology and pathology results were consistent. Expansive masses were observed in 19 cases, masses near the uterus in 8 cases, and masses in the uterine horn in 5 cases. The frequencies of these findings were statistically significant in the group with erroneous evaluations ($P < 0.001$). The interobserver agreement in MRI evaluation was assessed using Cohen's kappa coefficient, which was calculated as 0.831, indicating excellent agreement.

Table 2. Comparison of magnetic resonance imaging with pathology in the detection of deep myometrial invasion

		Pathology		
		SMI	DMI	Total
MRI	**SMI**	138	18	156
	DMI	11	59	70
	Total	149	77	226

MRI, magnetic resonance imaging; SMI, superficial myometrial invasion; DMI, deep myometrial invasion.

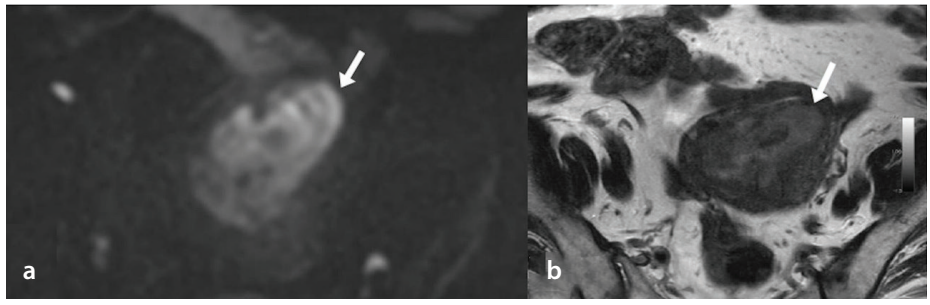


Figure 3. In a 69-year-old patient with a mixed tumor consisting of grade 1 endometrial adenocarcinoma of the endometrioid type and undifferentiated carcinoma, a mass lesion causing expansion in the uterine cavity is observed on MRI (white arrows). On DWI, the mass shows high signal intensity **(a)**, whereas in T2-weighted oblique axial sections **(b)**, it shows slightly higher signal intensity than the muscle. The lesion was concluded to have SMI, but pathological evaluation revealed DMI. MRI, magnetic resonance imaging; DWI, diffusion-weighted imaging; SMI, superficial myometrial invasion; DMI, deep myometrial invasion.

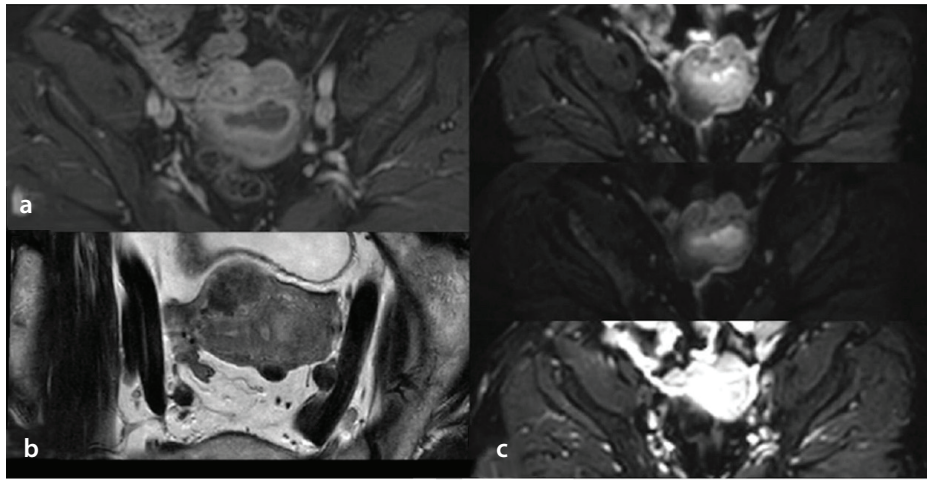


Figure 4. In a 63-year-old patient with a pathological diagnosis of grade 1 endometrial adenocarcinoma of the endometrioid type, the contrast-enhanced series **(a)** shows a weakly enhancing endometrial mass relative to the myometrium in the arterial phase and heterogeneous subserosal myomas at the uterine fundus. On T2-weighted oblique axial series **(b)**, a mildly hyperintense mass lesion is observed; on DWI **(c)**, a high signal intensity mass is seen. Based on MRI evaluation, the lesion was concluded to have DMI; however, the pathological evaluation showed SMI. DMI, deep myometrial invasion; DWI, diffusion-weighted imaging; MRI, magnetic resonance imaging; SMI, superficial myometrial invasion.

Table 3. Causes of discrepancy between magnetic resonance imaging and pathology with regard to deep myometrial invasion in patients with endometrial cancer

Causes	Understaged	Overstaged	Total
Expansile lesions	6	5	11
Myoma associated with mass	6	2	8
Mass located in the uterine horn	3	3	6
Personal factors	3	1	4

Discussion

In this study, which was conducted with EC cases, we compared MRI and histopathology results for determining DMI and successfully identified DMI with MRI in most patients, consistent with other studies in the literature.

In most cases where we misidentified myometrial invasion on MRI, there were mass lesions causing expansion. A mass lesion expanding within the uterine cavity results in thinning of the myometrium. In thin myometrium, it is expected that there will be a higher likelihood of measurement errors than in the normally thickened myometrium. In other studies, a frequently mentioned factor contributing to DMI evaluation errors is the presence of an expansile mass. Our review of the literature also identified postmenopausal atrophic myometrium as a potential error factor in some studies. In our cases, however, the error factors did not include isolated atrophic myometrium.

Lesions located in the uterine horn have also been identified as error factors in other studies. Myometrial thickness is low at the horn level, and its flexibility is limited. Therefore, even small masses can cause expansion and may hinder accurate evaluation. Subserosal or intramural myomas can disrupt the homogeneous structure and zonal anatomy of the myometrium, leading to possible mismeasurement of myometrial thickness and incorrect determination of DMI. Although some studies have cited the presence of adenomyosis as an error factor, it did not appear in the cases where we made incorrect staging.[4,6,21-25]

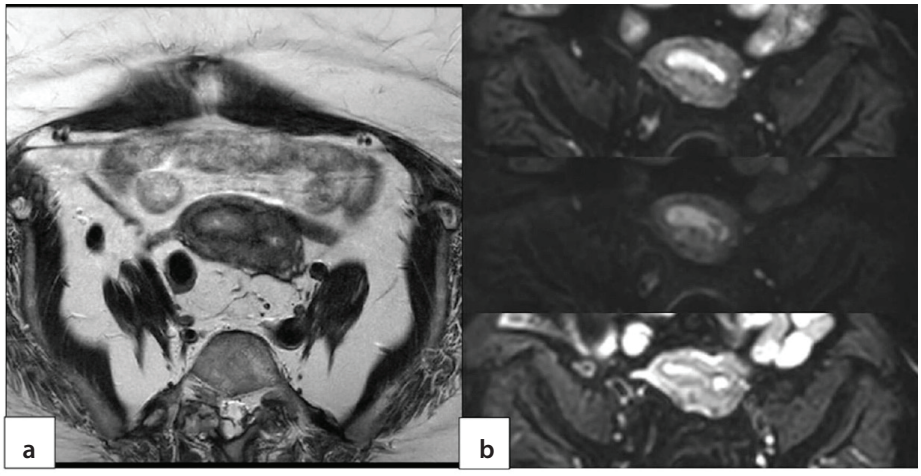


Figure 5. In a 54-year-old patient with a pathological diagnosis of grade 2 endometrial adenocarcinoma of the endometrioid type, a mildly hyperintense mass lesion is observed on T2-weighted oblique axial series (**a**), and a high signal intensity mass is seen on diffusion-weighted imaging (**b**), located in the right uterine horn and causing expansion at this level. Based on the MRI examination, the lesion was interpreted as not having DMI. However, the pathological evaluation revealed the presence of DMI. MRI, magnetic resonance imaging; DMI, deep myometrial invasion.

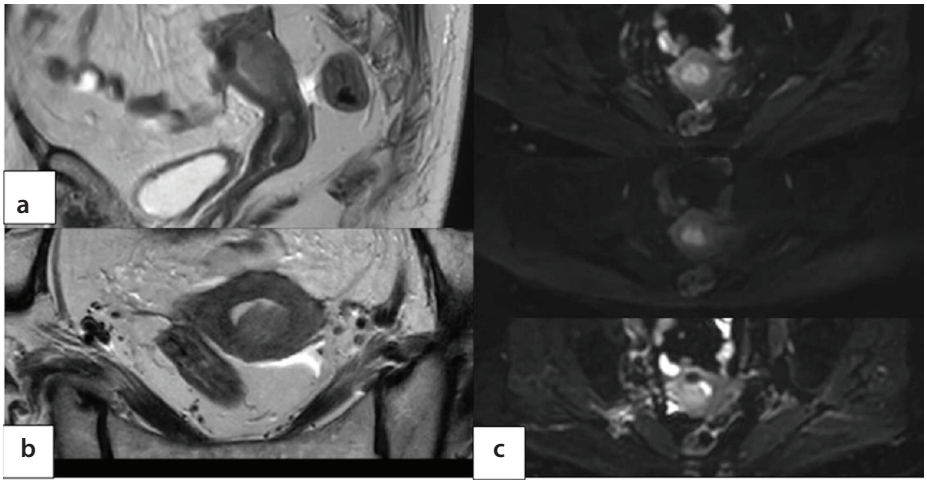


Figure 6. In a 64-year-old patient with a pathological diagnosis of high-grade mixed carcinoma, a mildly hyperintense endometrial mass lesion is observed on T2-weighted sagittal (**a**) and oblique axial (**b**) series, and it appears as a high signal intensity lesion on diffusion-weighted imaging (**c**). Based on the MRI examination, the lesion was concluded not to exhibit DMI. However, pathological evaluation confirmed DMI. No specific factors were identified on MRI that could explain the incorrect assessment. MRI, magnetic resonance imaging; DMI, deep myometrial invasion.

In a recent study, Liyanage et al.[26] analyzed 235 women with EC and reported that MRI detected DMI with an overall accuracy of 86%, sensitivity of 72%, and specificity of 91%. They also showed that the presence of subserosal or intramural leiomyomas or adenomyosis reduced sensitivity while leaving specificity essentially unchanged.

In the literature, apart from the present study, there are numerous studies comparing preoperative MRI with postoperative histopathological assessment in cases of EC in terms of DMI. In these studies, the sensitivity of MRI in detecting DMI ranged from 50% to 91%, specificity from 72% to 94%, accuracy from 77% to 93%, PPV from 63% to 92%, and NPV from 81% to 92%.[9,16,19,27-31] In our study, although the sensitivity rate is lower than in some studies, we observed high specificity and accuracy rates.

In the 2023 revision of the FIGO staging system, stage 1 was further detailed, and additional subcategories were introduced. In this classification, both the presence of DMI and the aggressiveness of the histological subtype are taken into consideration.[12] In our study, we found that aggressive histological subtypes did not negatively affect the accuracy of MRI assessments.

We demonstrated the high accuracy of MRI in determining myometrial invasion in EC cases, consistent with previous studies. We believe that considering the error factors identified in this study during MRI evaluation can enhance the reliability of the examination. Accordingly, recognizing MRI–pathology mismatches is essential: overstaging can expose patients to unnecessary lymphadenectomy, whereas understaging may delay adjuvant therapy, potentially affecting survival. Since DMI is an important marker for determining the risk of lymph node metastasis, it can assist in deciding whether lymphadenectomy should be performed prior to surgery. This approach may help prevent unnecessary lymphadenectomy and its associated complications.

In the literature, several studies have investigated the success of MRI in determining DMI, and as we mentioned in our article, various factors contributing to errors have been discussed. However, we did not find any other studies that specifically analyzed the causes of MRI errors and provided statistical data on these factors. Based on our literature review, our study includes one of the largest patient groups, which we consider a strength.

Our study has some limitations. First, it is a retrospective study. Second, in our MRI protocol, T2-weighted series were acquired in oblique axial and coronal views, but diffusion-weighted and dynamic contrast-enhanced series were not obtained in oblique planes. This may be considered a limitation, as it reduces the contribution of contrast-enhanced and diffusion-weighted sequences in determining DMI.

In conclusion, MRI has high accuracy in determining DMI in EC cases. However, erroneous interpretation of myometrial invasion may occur due to large expansile lesions, accompanying large myomas, and lesions located in the horn. We think that these factors should be mentioned in the radiology report when present, as they may act as misleading factors in evaluating DMI. This would enable more reliable treatment planning for the patient based on MRI findings.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Faria SC, Devine CE, Rao B, Sagebiel T, Bhosale P. Imaging and staging of endometrial cancer. *Semin Ultrasound CT MR*. 2019;40(4):287-294. [Crossref]
2. Cubo-Abert M, Díaz-Feijoo B, Bradbury M, et al. Diagnostic performance of transvaginal ultrasound and magnetic resonance imaging for preoperative evaluation of low-grade endometrioid endometrial carcinoma: prospective comparative study. *Ultrasound Obstet Gynecol*. 2021;58(3):469-475. [Crossref]
3. Maheshwari E, Nougaret S, Stein EB, et al. Update on MRI in evaluation and treatment of endometrial cancer. *Radiographics*. 2022;42(7):2112-2130. [Crossref]
4. Dokter E, Anderson L, Cho SM, et al. Radiology-pathology correlation of endometrial carcinoma assessment on magnetic resonance imaging. *Insights Imaging*. 2022;13(1):80. [Crossref]
5. Meissnitzer M, Forstner R. MRI of endometrium cancer - how we do it. *Cancer Imaging*. 2016;16:11. [Crossref]
6. Otero-García MM, Mesa-Álvarez A, Nikolic O, et al. Role of MRI in staging and follow-up of endometrial and cervical cancer: pitfalls and mimickers. *Insights Imaging*. 2019;10(1):19. [Crossref]
7. Amant F, Mirza MR, Koskas M, Creutzberg CL. Cancer of the corpus uteri. *Int J Gynaecol Obstet*. 2018;143(Suppl 2):37-50. [Crossref]
8. Makker V, MacKay H, Ray-Coquard I, et al. Endometrial cancer. *Nat Rev Dis Primers*. 2021;7(1):88. [Crossref]
9. Yin XH, Jia HY, Shi M, Wu H, Li YM. Magnetic resonance imaging for detection of depth of myometrial invasion and cervical invasion in patients with endometrial carcinoma. *Int J Clin Exp Med*. 2015;8(10):19501-19505. [Crossref]
10. Aracki-Trenkic A, Stojanov D, Petric A, Benedeto-Stojanov D, Trenkic M, Ignjatovic J. The role of magnetic resonance imaging in the evaluation of endometrial carcinoma. *J BUON*. 2016;21(3):542-548. [Crossref]
11. Berek JS, Matias-Guiu X, Creutzberg C, et al. FIGO staging of endometrial cancer: 2023. *Int J Gynaecol Obstet*. 2023;162(2):383-394. [Crossref]
12. Avesani G, Bonatti M, Venkatesan AM, Nougaret S, Sala E. RadioGraphics update: 2023 FIGO staging system for endometrial cancer. *Radiographics*. 2024;44(7):e240084. [Crossref]
13. Larson DM, Connor GP, Broste SK, Krawisz BR, Johnson KK. Prognostic significance of gross myometrial invasion with endometrial cancer. *Obstet Gynecol*. 1996;88(3):394-398. [Crossref]
14. Iitsuka C, Asami Y, Hirose Y, et al. Preoperative magnetic resonance ımaging versus ıntraoperative frozen section diagnosis for predicting the deep myometrial ınvasion in endometrial cancer: our experience and literature review. *J Obstet Gynaecol Res*. 2021;47(9):3331-3338. [Crossref]
15. Ota T, Tsuboyama T, Onishi H, et al. Diagnostic accuracy of MRI for evaluating myometrial invasion in endometrial cancer: a comparison of MUSE-DWI, rFOV-DWI, and DCE-MRI. *Radiol Med*. 2023;128(6):629-643. [Crossref]
16. Karataşlı V, Çakır İ, Şahin H, Ayaz D, Sancı M. Can preoperative magnetic resonance imaging replace intraoperative frozen sectioning in the evaluation of myometrial invasion for early-stage endometrial carcinoma? *Ginekol Pol*. 2019;90(3):128-133. [Crossref]
17. Wright JD, Huang Y, Burke WM, et al. Influence of lymphadenectomy on survival for early-stage endometrial cancer. *Obstet Gynecol*. 2016;127(1):109-118. [Crossref]
18. Zamani F, Goodarzi S, Hallaji F, et al. Diagnostic value of pelvic MRI for assessment of the depth of myometrial invasion and cervical involvement in endometrial cancer: comparison of new versus old FIGO staging. *Iran J Radiol*. 2012;9(4):202-208. [Crossref]
19. Cade TJ, Quinn MA, McNally OM, Neesham D, Pyman J, Dobrotwir A. Predictive value of magnetic resonance imaging in assessing myometrial invasion in endometrial cancer: is radiological staging sufficient for planning conservative treatment? *Int J Gynecol Cancer*. 2010;20(7):1166-1169. [Crossref]
20. Rei M, Rodrigues I, Condeço P, Igreja F, Veríssimo C, Mendinhos G. Endometrial cancer: preoperative versus intraoperative staging. *J Gynecol Obstet Hum Reprod*. 2019:101647. [Crossref]
21. Hricak H, Stern JL, Fisher MR, Shapeero LG, Winkler ML, Lacey CG. Endometrial carcinoma staging by MR imaging. *Radiology*. 1987;162(2):297-305. [Crossref]
22. Andreano A, Rechichi G, Rebora P, Sironi S, Valsecchi MG, Galimberti S. MR diffusion imaging for preoperative staging of myometrial invasion in patients with endometrial cancer: a systematic review and meta-analysis. *Eur Radiol*. 2014;24(6):1327-1338. [Crossref]
23. Kinkel K, Forstner R, Danza FM, et al. Staging of endometrial cancer with MRI: guidelines of the European Society of Urogenital Imaging. *Eur Radiol*. 2009;19(7):1565-1574. [Crossref]
24. Lee JH, Dubinsky TJ, Andreotti RF, et al. ACR appropriateness Criteria(R) pretreatment evaluation and follow-up of endometrial cancer of the uterus. *Ultrasound Q*. 2011;27(2):139-145. [Crossref]
25. Nougaret S, Lakhman Y, Vargas HA, et al. From staging to prognostication: achievements and challenges of MR imaging in the assessment of endometrial cancer. *Magn Reson Imaging Clin N Am*. 2017;25(3):611-633. [Crossref]
26. Liyanage A, Cardoza S, Kasabia D, Moore H. Accuracy of MRI in predicting deep myometrial invasion in endometrial cancer and the influence of leiomyoma, adenomyosis and the microcystic elongated and fragmented tumour pattern. *J Med Imaging Radiat Oncol*. 2024;68(3):235-242. [Crossref]
27. Chung HH, Kang SB, Cho JY, et al. Accuracy of MR imaging for the prediction of myometrial invasion of endometrial carcinoma. *Gynecol Oncol*. 2007;104(3):654-659. [Crossref]

28. Cabrita S, Rodrigues H, Abreu R, et al. Magnetic resonance imaging in the preoperative staging of endometrial carcinoma. *Eur J Gynaecol Oncol*. 2008;29(2):135-137. [Crossref]

29. Sato S, Itamochi H, Shimada M, et al. Preoperative and intraoperative assessments of depth of myometrial invasion in endometrial cancer. *Int J Gynecol Cancer*. 2009;19(5):884-887. [Crossref]

30. McComiskey MH, McCluggage WG, Grey A, Harley I, Dobbs S, Nagar HA. Diagnostic accuracy of magnetic resonance imaging in endometrial cancer. *Int J Gynecol Cancer*. 2012;22(6):1020-1025. [Crossref]

31. Wu CY, Tai YJ, Shih IL, et al. Preoperative magnetic resonance imaging predicts clinicopathological parameters and stages of endometrial carcinomas. *Cancer Med*. 2022;11(4):993-1004. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253612

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Evaluation of the congenital absence of the vas deferens with magnetic resonance imaging: preliminary findings

Hakan İmamoğlu[1]
Erdem Sarı[1]
Özgür Karabıyık[1]
Emre Can Akınsal[2]
Gökmen Zararsız[3,4]
Oğuz Ekmekçioğlu[2]

[1]Erciyes University Faculty of Medicine, Department of Radiology, Kayseri, Türkiye

[2]Erciyes University Faculty of Medicine, Department of Urology, Kayseri, Türkiye

[3]Erciyes University Faculty of Medicine, Department of Biostatistics, Kayseri, Türkiye

[4]Erciyes University, Drug Application and Research Center (ERFARMA), Kayseri, Türkiye

PURPOSE

To date, no study provides definitive evidence for the pathogenesis of congenital absence of the vas deferens (CAVD). This study aims to evaluate the vas deferens (VD), particularly the intra-abdominal part and accompanying seminal vesicle (SV) pathologies, in search of an explanation for the pathogenesis of the disease using magnetic resonance imaging (MRI) in patients clinically diagnosed with CAVD.

METHODS

MRI scans of patients admitted to our center with clinically diagnosed unilateral CAVD (CUAVD) or bilateral CAVD (CBAVD) in the infertility clinic were retrospectively evaluated. SV hypoplasia, SV agenesis, the distal part of the VD close to the ampulla, and the intra-abdominal part of the VD were investigated. Additionally, the association of CAVD and SV pathologies was assessed.

RESULTS

Clinically and confirmed with scrotal sonography by evaluating the proximal part of the VD, 32 patients (62.7%) had CBAVD, and 19 patients (37.3) had CUAVD. In MRI, the intra-abdominal part of the VD was visible in 52.9% of all patients. The association between the intra-abdominal part of the VD and CAVD was statistically significant in the CBAVD patient group compared with the CUAVD group (Bonferroni-adjusted *P* value = 0.006). The intra-abdominal part of the VD dilatation is a new finding in CAVD and was not found in patients with CUAVD. Only 2 out of 51 patients (3.9%) had a standard SV.

CONCLUSION

In the assessment of CAVD and accompanying SV pathologies, detailed findings are obtained by MRI even in the evaluation of the intra-abdominal part of the VD. Preliminary findings in this study are consistent with the theory of acquired vasal agenesis in CBAVD.

CLINICAL SIGNIFICANCE

The detailed findings of an MRI may contribute to a better understanding of the disease.

KEYWORDS

Congenital absence of vas deferens, infertility, magnetic resonance imaging, seminal vesicle, vas deferens

Corresponding author: Emre Can Akınsal

E-mail: emreakinsal@hotmail.com

Received 19 August 2025; revision requested 15 September 2025; accepted 05 November 2025.

Epub: 19.11.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253612

Congenital absence of the vas deferens (CAVD) is one of the critical etiological causes of male infertility.[1] It is frequently observed in young and middle-aged men. Bilateral CAVD (CBAVD) has been identified in approximately 1.3% of infertile men, and unilateral CAVD (CUAVD) has been detected in 1% of cases.[2] Specifically, in cases of CBAVD, a mutation in the *cystic fibrosis transmembrane conductance regulator (CFTR)* gene associated with cystic fibrosis is frequently detected, and the majority of these cases exhibit CAVD.[3] To date, no study provides definitive evidence or a clear explanation for the mechanism of CAVD. In recent years, studies have focused on the genetic etiology of CAVD. More than 2.000 CFTR

You may cite this article as: İmamoğlu H, Sarı E, Karabıyık Ö, Akınsal EC, Zararsız G, Ekmekçioğlu O. Evaluation of the congenital absence of the vas deferens with magnetic resonance imaging: preliminary findings. *Diagn Interv Radiol*. 2026;32(3):249-254.

mutations have been identified as closely associated with CBAVD.[4] Although it has been suggested that CFTR and ADGRG2 mutations are the genetic cause of the majority of CAVD cases, a precise genetic diagnosis cannot be established in 10%–20% of patients with CBAVD and 60%–70% of patients with CUAVD.[4,5] Embryologically, the seminal vesicle (SV) originates from the ampulla of the vas deferens (VD) as a diverticular structure. Due to the interrelated developmental mechanisms of the VD and SV during embryological life, SV pathologies are also commonly observed in cases where CAVD is detected.[6] The most common SV pathologies are agenesis and developmental disorders. The most frequently observed clinical finding in these patients is azoospermia; however, in cases with unilateral vasal agenesis, oligospermia, normospermia, and spontaneous fertilization are possible.[7]

Physical examination and scrotal ultrasonography can evaluate the proximal part of the VD. In cases where the etiology of infertility is being investigated in men, various radiological imaging methods are used to evaluate the intra-abdominal and distal part of the VD and SV structures, especially in cases where post-testicular pathologies leading to obstructive azoospermia are considered as a preliminary diagnosis.[8,9] Transrectal ultrasound examination is often the initial approach.[9] However, transrectal ultrasound cannot evaluate the intra-abdominal part of the VD. Moreover, the examination may cause discomfort for the patient; when patients are informed about how the transrectal ultrasound procedure will be performed, they may not accept it. In recent years, studies have been published emphasizing the importance of magnetic resonance imaging (MRI) and computed tomography (CT) in SV and VD imaging.[10,11] However, SV and VD structures in men investigated for the etiology of infertility have rarely been evaluated using CT.[6] Patients investigated for infertility are mostly young men, and the use of CT in these patients may not be appropriate due to radiation exposure. MRI does not involve exposure to ionizing radiation and has superior soft tissue resolution compared with CT.[8] It has several essential features, including its exceptional ability to detect the intra-abdominal part of the VD, higher resolution, and superior soft tissue contrast, which allow for a more detailed evaluation of SV morphology.[8,12] Despite the significant advantages of MRI in CAVD, there is a limited number of studies in the literature on the use of MRI in demonstrating CAVD and possibly associated SV pathology.[8,13,14]

This study aims to evaluate the intra-abdominal part of the VD and accompanying SV pathologies in search of an explanation for the pathogenesis of the disease using MRI in patients clinically diagnosed with CAVD.

Main points

- Magnetic resonance imaging (MRI) provides detailed findings for the assessment of congenital absence of the vas deferens (CAVD) and accompanying seminal vesicle pathologies, including the evaluation of the intra-abdominal part of the vas deferens (VD). This study identified the intra-abdominal part of the VD dilatation as a new finding.
- In the MRI of the patients with CAVD, the intra-abdominal part of the VD was observed in more than half of all patients.
- The association between the intra-abdominal part of the VD and CAVD was statistically significant in the bilateral CAVD (CBAVD) patient group compared with the unilateral CAVD (CUAVD) group. The intra-abdominal part of the VD dilatation was not found in patients with CUAVD. These preliminary findings may shed light on the pathogenesis of the CBAVD.

Methods

The Institutional Review Board approval for this single-center retrospective study was obtained from the Erciyes University Ethics Committee (decision number: 2023/171, date: March 8, 2023). Informed consent was waived for retrospective analysis.

Study population

All participants were infertile patients who applied to the infertility clinic. The study exclusion criteria were as follows: (1) prior pelvic surgery, (2) prior scrotal infection, (3) patients with uncertain diagnosis, or (4) poor image quality in the MRI. The MRI images were of patients who refused transrectal ultrasound examination to investigate the etiology of obstructive azoospermia and oligospermia. The diagnosis of CAVD was made clinically and by using scrotal ultrasonography. The MRIs of patients with clinically diagnosed CUAVD or CBAVD between January 2016 and January 2023 were retrospectively evaluated.

Magnetic resonance imaging examinations

The MRI was performed using a 1.5-T system (Signa, GE Medical Systems; Milwaukee, USA) with a 16-channel body coil, without the use of an endorectal coil and contrast material. The images were evaluated on T2-weighted sequences in axial, coronal, and sagittal planes (Table 1). CUAVD, CBAVD, SV hypoplasia, SV agenesis, the distal part of the VD close to the ampulla, and the intra-abdominal part of the VD were investigated. Normal range of the SV length and diameter was defined as 22–38 mm and 11–19 mm, respectively.[15] Normal diameter of VD in the ampulla was defined as 3–5 mm.[15] SV hypoplasia was defined as a maximum diameter of the SV being < 50% of normal or < 5 mm.[16] VD dilatation was described as a maximum diameter of the VD being > 5 mm.[15] The prevalence of these clinical conditions and the associations of CAVD and SV pathologies were examined. Imaging analysis was performed by two radiologists (H. İ., Ö. K.) using the center's picture archiving and communication system (Sectra Workstation IDS7, Teknikringen, Sweden) in consensus. Each radiologist was blinded to the clinical diagnosis. Measurements were made in axial, coronal, and sagittal planes, and the maximum diameter and length of the SV and the maximum diameter of the VD were evaluated for the diagnosis (Figure 1).

Power analysis

A post-hoc power analysis based on the Pearson chi-squared test used to evaluate the primary hypothesis indicated a statistical

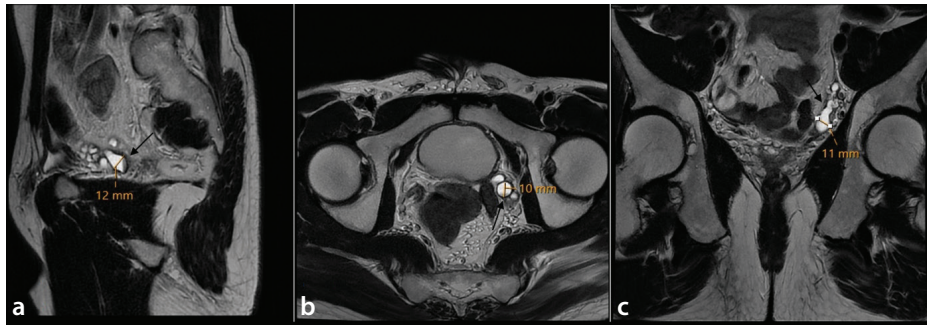


Figure 1. A 25-year-old patient with bilateral vasal agenesis and azoospermia. The T2-weighted sagittal **(a)**, axial **(b)**, and coronal **(c)** images show the intra-abdominal part of the left vas deferens dilatation (black arrows). Diagnosis was made by measuring the maximum diameter of the vas deferens.

power of 0.835 at an alpha level of 0.05 and an effect size of 0.411. Power analyses were conducted using PASS 11.0 (NCSS Inc., USA).

Statistical analysis

Statistical analyses were performed using TURCOSA (Turcosa Analytics Ltd., www.turcosa.com.tr) statistical software. Differences in MRI findings between patients with CUAVD and CBAVD and the relationship between SV pathologies and CAVD were assessed using Pearson's chi-squared test, Fisher's exact test, and age-adjusted logistic regression analysis. Data are summarized as frequencies, percentages, and corresponding confidence intervals. Confidence intervals of the estimated proportions were calculated using the Wald or Fisher approach based on the small-sample assumption (npq ≥ 5). Odds ratios were calculated using 95% confidence intervals; *P* values were adjusted using the Bonferroni approach to control for multiple testing. Adjusted *P* values of < 5% are considered statistically significant.

Results

Fifty-one patients were included in the study. The mean age of the patients was 29.53 ± 5.36 years. Clinically and confirmed with scrotal sonography by evaluating the proximal part of the VD, 32 patients (62.7%) had CBAVD, and 19 patients (37.3%) had CUAVD. The clinical and radiologic characteristics of the patients are shown in Table 2. In one case, the VD was palpable at the scrotal level as a blind-ending tubular structure, whereas

Table 1. Technical data of the imaging protocol used on the 1.5-T MRI system

	T2-weighted axial imaging	T2-weighted coronal imaging	T2-weighted sagittal imaging
FOV	240	240	240
Matrix	352 × 352	352 × 352	288 × 288
TE (ms)	141	168	147
TR (ms)	3.470	6.862	3.175
Section thickness (mm)	3	3	3.5
Number of sections	23	20	20
NEX	4	4	1.5
Bandwidth (kHz)	62.5	62.5	50
Slice gap (mm)	0.5	0.5	1
Acquisition time (min)	5.28	10.06	1.37

MRI, magnetic resonance imaging; FOV: field of view, TE: echo time; TR, repetition time; NEX, number of excitations.

Table 2. Clinical and radiologic characteristics of the patients (n = 51)

Parameter	Findings
Agenesis status	
• Unilateral, n (%)	19 (37.2)—Right: 8 (42.1%), Left: 11 (57.9%)
• Bilateral, n (%)	32 (62.8)
Right testicular volume, median (min–max), mL*	18 (2–28)
Left testicular volume, median (min–max), mL*	18 (2–24)
Right epididymal morphology	
• Normal, n (%)	38 (74.6)
• Absence of the epididymal body and tail, n (%)	13 (25.4)
Left epididymal morphology	
• Normal, n (%)	36 (70.6)
• Absence of the epididymal body and tail, n (%)	15 (29.4)
Renal agenesis	
• Present, n (%)	10 (19.6)—Right: 5 (50), Left: 5 (50)
• Absent, n (%)	41 (80.3)
Semen analysis	
• Oligospermia, n (%)	4 (7.8)
• Azoospermia, n (%)	47 (92.2)
Genetic findings	
• Normal karyotype, n (%)	28 (54.9)
• Abnormal karyotype, n (%)	12 (23.5)
• Not available, n (%)	11 (21.6)

*One patient had both congenital vasal agenesis and Klinefelter syndrome, and the testicular volumes in this case were 2 mL on each side.
Epididymal and vasal structures were classified based on imaging and physical examination findings. Renal agenesis status was confirmed radiologically. Testicular volumes are presented as median and range. min, minimum; max, maximum.

in another case, it was palpable as a fibrotic band. In all other patients, the VD could not be identified at the scrotal level. All patients with oligospermia had CUAVD. Renal agenesis was detected in 5 cases in the CUAVD patient group (26.3%) and 5 cases in the CBAVD group (15.6%). In MRI, the intra-abdominal part of the VD was seen in 52.9% of all patients. The intra-abdominal part of the VD was found in 22 (68.7%) of the patients with CBAVD (Figure 2). In patients with CUAVD, the intra-abdominal part of the VD was detected in 5 patients (26.3%) (Figure 3). The association between the intra-abdominal part of the VD and CAVD was statistically significant in the CBAVD patient group compared with the CUAVD group (Bonferroni-adjusted *P* value = 0.0066) (Table 3). The CBAVD patient group had a 6.168 (1.737–21.905)-fold increased risk of an intra-abdominal part of the VD as compared with the CUAVD patient group. The intra-abdominal part of the VD dilatation was found in 5 patients (% 15.6%) with CBAVD (Figure 4), whereas no intra-abdominal part of the VD dilatation was found in patients with CUAVD (Bonferroni-adjusted *P* value = 0.2870) (Table 3).

In 49 patients with CAVD (96.1%), accompanying SV hypoplasia and/or agenesis was detected. Among the 32 patients with CBAVD, bilateral SV agenesis was found in 15 patients (46.8%). Ipsilateral SV agenesis was found in 17 of 19 patients (89.4%) with CUAVD. Only 2 patients (3.9%) had a normal SV; 1 patient had CBAVD, and the other had CUAVD.

Discussion

In the MRI of the patients with CAVD, the intra-abdominal part of the VD was observed in more than half of all patients. Unlike the few similar studies on this subject in the literature,[13] our current study showed a higher detection rate of the intra-abdominal part of the VD in patients with CBAVD than in those with CUAVD. Furthermore, this study identified the intra-abdominal part of the VD dilatation as a new finding.

Two separate theories have been proposed for the pathogenesis of CAVD. The first theory suggests that CAVD occurs as a result of an organogenesis disorder. This theory is supported by the high prevalence of renal agenesis observed simultaneously in CUAVD.[4] In this study, the incidence of renal agenesis was proportionally higher in the CUAVD patient group. The second theory, more closely associated with CFTR variants, proposes that acquired vasal agenesis develops over time, with an increase in fluid viscosity and development of obstruction and degeneration in the lumen of the VD.[17] The association with cystic fibrosis in CBAVD cases is well known.[18] Conversely, literature data indicate that normal VD is detected in cases with CFTR gene expression in fetal life.[19] Additionally, experimental studies on animals without cystic fibrosis gene mutation have shown that defects occurred in the SV and/or VD secondary to obstruction and inflammation.[18-20] In the present study, the CBAVD patient group had a 6.168 (1.737–21.905)-fold increased risk of intra-abdominal part of the VD as compared with the CUAVD patient group, and although not statistically significant, a high rate of intra-abdominal part of the VD dilatation was found in CBAVD. The intra-abdominal part of the VD dilatation was not found in patients with CUAVD. We think these preliminary findings, along with some of the literature data above, may be related to the theory of acquired vasal agenesis in CBAVD.

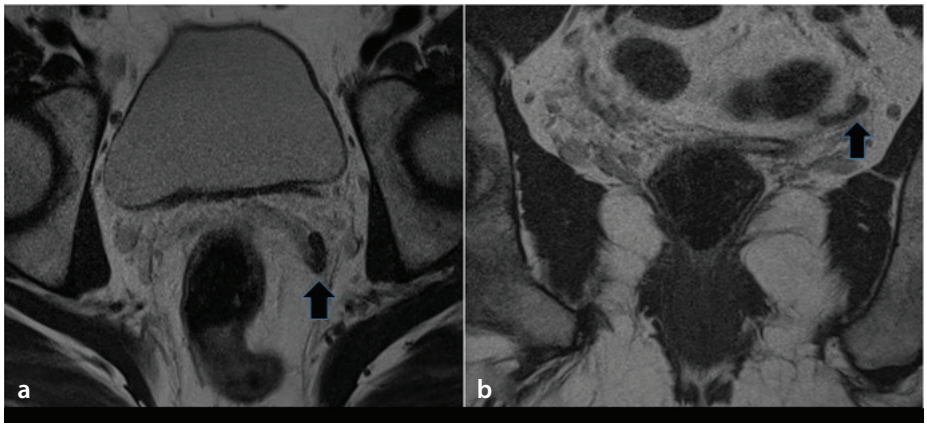


Figure 2. A 24-year-old patient with bilateral vasal agenesis and azoospermia. The T2-weighted axial **(a)** and coronal **(b)** images show the intra-abdominal part of the left vas deferens (black arrows).

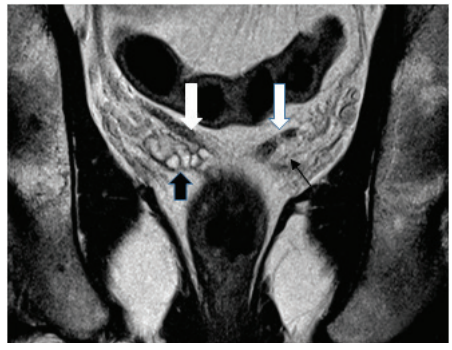


Figure 3. A 35-year-old patient with left vas deferens agenesis and azoospermia. The T2-weighted coronal image shows the bilateral distal part of the vas deferens close to the ampulla (white arrows). Although the right seminal vesicle (SV) of the same patient appears normal (thick black arrow), agenesis is present in the left SV (thin black arrow).

Table 3. The association of the intra-abdominal part of the VD and CAVD

Variables	CAVD		Bonferroni-adjusted *P* value	Age-adjusted OR (95% CI)
	CBAVD (n = 32)	CUAVD (n = 19)		
Intra-abdominal VD				
Present	22 [68.8% (95% CI: 51.3%–82.1%)]	5 [26.3% (95% CI: 11.5%–49.1%)]	0.0066†	6.168 (1.737–21.905)
Absent	10 [31.2% (95% CI: 17.8%–48.7)]	14 [73.7% (95% CI: 50.9%–88.6%)]		
Intra-abdominal VD dilatation				
Present	5 [15.6% (95% CI: 6.4%–32.2%)]	0 [0.0% (95% CI: 0.0%–17.7%)]	0.2870‡	NC
Absent	27 [84.4% (95% CI: 67.8%–93.6%)]	19 [100.0% (95% CI: 82.4%–100.0%)]		

Data values are presented as n (%). † *P* value is calculated using the Pearson chi-squared test. ‡ *P* value is calculated using Fisher's exact test. Percentages are presented with their 95% confidence intervals. Significant *P* values are shown in bold. VD, vas deferens; CAVD, congenital absence of the vas deferens; CUAVD, congenital unilateral absence of the vas deferens; CBAVD, congenital bilateral absence of the vas deferens; OR, odds ratio; CI, confidence interval; NC, not computed due to zero counts.

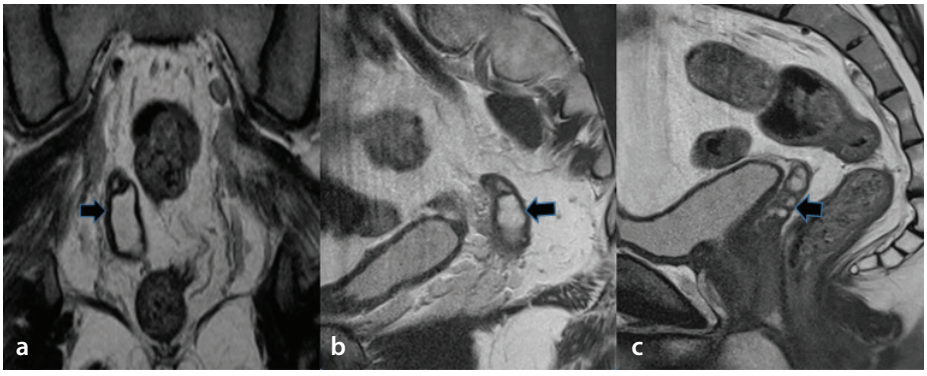


Figure 4. A 27-year-old patient with bilateral vasal agenesis and azoospermia. The T2-weighted coronal **(a)** and sagittal **(b)** images show the intra-abdominal part of the right vas deferens dilatation (black arrows). The T2-weighted sagittal image **(c)** shows a normal ipsilateral seminal vesicle (black arrow).

SV pathologies (agenesis, hypoplasia) were found in 90%–100% of CAVD cases in the literature.[6,9,13,21] In the current study, this rate was 96.1%, which is consistent with the literature data. Among the 32 patients with CBAVD in this study, bilateral SV agenesis was found in 15 patients (46.8%). In the literature, the detection rate of bilateral SV agenesis in patients with CBAVD has been reported to range widely, from 9%–83%.[13,21-24] This wide range may be attributed to variations in the number of patients examined, diagnostic methods, and diagnostic criteria for SV pathologies in different studies. In the CUAVD group, SV agenesis on the same side was found in 17 out of 19 patients (89.4%). A similar rate was found in a study conducted by AbdElnaser et al.[22] (70.45% for right CUAVD–SV agenesis association and 80% for left). There is a close embryological relationship between VD and SV.[6,9,25] This process may play a significant role in the development of SV pathologies.

In this current study, only 2 patients (3.9%) had a standard (both morphology and size) SV;1 patient had CBAVD, and the other had CUAVD. Previous studies have reported a normal SV rate of 6.4%–55% in CAVD.[18,21,26,27] MRI was not used in these studies to detect SV pathologies. This may be the reason for the difference between the current study and these studies. With the increased use of MRI in patients with CAVD and accompanying SV pathologies, these rates can be determined more accurately.

The current study has some limitations, the most impactful of which is the limited number of patients. Nevertheless, the findings of this study can be confirmed and generalized through multicenter prospective studies with a large number of patients. The second limitation is selection bias from including only patients who refused transrectal ultrasound. Studies examining all patients may yield more accurate and different results. The third limitation is the retrospective design of this study, which results in a lack of genetic evaluation. Although CFTR mutation analysis was performed in a subset of our cases, the testing methods have evolved, and in some instances, patients who were initially reported as mutation-negative were later found to harbor CFTR mutations upon re-evaluation with updated techniques. Because of this inconsistency, we refrained from including these heterogeneous genetic data in the final analysis; however, we shared the genetic data in a subset of patients in Table 2. This limitation, along with the retrospective design, prevents us from definitively distinguishing congenital agenesis from acquired obstruction. Nevertheless, we included available clinical data, such as renal anomalies and semen analysis, to provide additional context. The fourth limitation is the use of 1.5-T MRI without an endorectal coil and in the absence of a standardized measurement protocol in this subject. Since there was no 3-T MRI in our hospital when patient imaging started, a 1.5-T MRI system was used. An endorectal coil was not used because patients in this study did not want uncomfortable procedures, such as a transrectal probe or an endorectal coil. A higher field strength, such as 3-T with an endorectal coil, may improve spatial resolution and provide better examination for SV and VD.

In conclusion, MRI provides valuable information on CAVD and associated SV pathologies. Our findings are consistent with—but do not prove—the theory of acquired vasal agenesis in CBAVD. Given the absence of uniform genetic testing and the evolving nature of CFTR mutation analyses, these results should be interpreted with caution. Further prospective, multicenter studies with a larger cohort, standardized genetic testing, and comprehensive clinical correlation are required to confirm these preliminary findings.

Acknowledgment

We would like to thank the Proofreading and Editing Office of the Dean for Research at Erciyes University for copyediting and proofreading service for this manuscript.

Footnotes

Conflict of interest disclosure

The authors declared that they have no conflict of interest.

References

1. Jequier AM, Ansell ID, Bullimore NJ. Congenital absence of the vasa deferentia presenting with infertility. *J Androl*. 1985;6(1):15-19. [Crossref]
2. Anguiano A, Oates RD, Amos JA, et al. Congenital bilateral absence of the vas deferens. A primarily genital form of cystic fibrosis. *JAMA*. 1992;267(13):1794-1797. [Crossref]
3. Chiang HS, Lu JF, Liu CH, Wu YN, Wu CC. CFTR (TG)m(T)n polymorphism in patients with CBAVD in a population expressing low incidence of cystic fibrosis. *Clin Genet*. 2009;76(3):282-286. [Crossref]
4. Bieth E, Hamdi SM, Mieusset R. Genetics of the congenital absence of the vas deferens. *Hum Genet*. 2021;140(1):59-76. [Crossref]
5. Fang J, Wang X, Sun X, Cui Y, Diao F, Yang X. Congenital absence of the vas deferens with hypospadias or without hypospadias: phenotypic findings and genetic considerations. *Front Genet*. 2022;13:1035468. [Crossref]
6. Goldstein M, Schlossberg S. Men with congenital absence of the vas deferens often have seminal vesicles. *J Urol*. 1988;140(1):85-86. [Crossref]
7. Mieusset R, Bieth E, Daudin M, et al. Male partners of infertile couples with congenital unilateral absence of the vas deferens are mainly non-azoospermic. *Andrology*. 2020;8(3):645-653. [Crossref]
8. Kim B, Kawashima A, Ryu JA, Takahashi N, Hartman RP, King BF Jr. Imaging of the seminal vesicle and vas deferens. *RadioGraphics*. 2009;29(4):1105-1121. [Crossref]
9. Kuligowska E, Baker CE, Oates RD. Male infertility: role of transrectal US in diagnosis and management. *Radiology*. 1992;185(2):353-360. [Crossref]
10. Ocal O, Karaosmanoglu AD, Karcaaltıncaba M, Akata D, Ozmen M. Imaging findings of congenital anomalies of seminal vesicles. *Pol J Radiol*. 2019;84:e25-e31. [Crossref]

11. Catania R, Dasyam N, Furlan A, Borhani AA. Cross sectional imaging of seminal vesicles and vasa deferentia. *Abdom Radiol (NY)*. 2020;45(7):2049-2062. [Crossref]
12. Chen HW, Huang SC, Li YW, Chen SJ, Sheih CP. Magnetic resonance imaging of seminal vesicle cyst associated with ipsilateral urinary anomalies. *J Formos Med Assoc*. 2006;105(2):125-131. [Crossref]
13. Chiang HS, Lin YH, Wu YN, Wu CC, Liu MC, Lin CM. Advantages of magnetic resonance imaging (MRI) of the seminal vesicles and intra-abdominal vas deferens in patients with congenital absence of the vas deferens. *Urology*. 2013;82(2):345-351. [Crossref]
14. Wang H, Peng Y, Fu W, Hu X, Li C, Guan J. MRI findings of obstructive azoospermia: lesions in and out of pelvic cavity. *Abdom Radiol (NY)*. 2020;45(3):851-864. [Crossref]
15. Kim ED, Lipshultz LI, Howards SS. Male infertility. In: Gillenwater JY, Grayhack JT, Howards SS, Mitchell ME, editors. Adult and pediatric urology. 4th ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2002. p. 1658–1683.
16. Raviv G, Mor Y, Levron J, et al. Role of transrectal ultrasonography in the evaluation of azoospermic men with low-volume ejaculate. *J Ultrasound Med*. 2006;25(7):825-829. [Crossref]
17. Cuppens H, Cassiman JJ. CFTR mutations and polymorphisms in male infertility. *Int J Androl*. 2004;27(5):251-256. [Crossref]
18. Mickle J, Milunsky A, Amos JA, Oates RD. Congenital unilateral absence of the vas deferens: a heterogeneous disorder with two distinct subpopulations based upon aetiology and mutational status of the cystic fibrosis gene. *Human Reprod*. 1995;10(7):1728-1735. [Crossref]
19. Gaillard DA, Carre-Pigeon F, Lallemand A. Normal vas deferens in fetuses with cystic fibrosis. *J Urol*. 1997;158(4):1549-1552. [Crossref]
20. Reynaert I, Van Der Schueren B, Degeest G, Manin M, Cuppens H, Scholte B, Cassiman JJ. Morphological changes in the vas deferens and expression of the cystic fibrosis transmembrane conductance regulator (CFTR) in control, deltaF508 and knock-out CFTR mice during postnatal life. *Mol Reprod Dev*. 2000;55(2):125-135. [Crossref]
21. Daudin M, Bieth E, Bujan L, Massat G, Pontonnier F, Mieusset R. Congenital bilateral absence of the vas deferens: clinical characteristics, biological parameters, cystic fibrosis transmembrane conductance regulator gene mutations, and implications for genetic counseling. *Fertil Steril*. 2000;74(6):1164-1174. [Crossref]
22. AbdElnaser T, Elkhiat YI, El-Azizi HM, Fatah E, Abd M, Elshibany AM, GamalEl Din SF. A cross-sectional study of the genital duct and renal anomalies in Egyptian cases of congenital absence of the vas deferens. *Hum Fertil (Camp)*. 2022;25(4):738-744. [Crossref]
23. Holsclaw DS, Perlmutter AD, Jockin H, Shwachman H. Genital abnormalities in male patients with cystic fibrosis. *J Urol*. 1971;106(4):568-574. [Crossref]
24. Valman HB, France NE. The vas deferens in cystic fibrosis. *Lancet*. 1969;13(2):566-567. [Crossref]
25. Kolettis PN, Sandlow JI. Clinical and genetic features of patients with congenital unilateral absence of the vas deferens. *Urology*. 2002;60(6):1073-1076. [Crossref]
26. de la Taille A, Rigot JM, et al. Correlation between genito-urinary anomalies, semen analysis and CFTR genotype in patients with congenital bilateral absence of the vas deferens. *Br J Urol*. 1998;81(4):614-619. [Crossref] Erratum in: *Br J Urol*. 1998;82(5):777.
27. Schlegel PN, Shin D, Goldstein M. Urogenital anomalies in men with congenital absence of the vas deferens. *J Urol*. 1996;155(5):1644-1648. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253396

ABDOMINAL IMAGING

LETTER TO THE EDITOR

Solitary plasmacytoma: a rare and unusual tumor of the liver

Ayşe Erden[1]
Koray Ceyhan[2]

[1]Ankara University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

[2]Ankara University Faculty of Medicine, Department of Pathology, Ankara, Türkiye

Dear Editor,

We read with interest the review article by Stanietzky et al.[1] entitled "Unusual liver tumors: spectrum of imaging findings with pathologic correlation" in the recent issue of Diagnostic and Interventional Radiology. In this article, the authors provide both textual and visual portrayals of hepatic neoplasms that are both unusual and rare. The authors also discussed the subject of "solitary plasmacytoma" under the subtitle of "Multiple myeloma and solitary plasmacytoma" and stated that this rare lesion has variable imaging findings. In this context, we would like to share the liver imaging findings of our case that was histopathologically diagnosed as "hepatic solitary plasmacytoma."

A 50-year-old woman presented with diffuse abdominal pain that had intensified at the right upper quadrant for around 3 months. Aside from laparoscopic cholecystectomy, her medical history was unremarkable, and her physical examination was normal. Hemogram, liver function, and renal function test results were in normal range. Erythrocyte sedimentation rate was 33 mm/h (normal: 0–25 mm/h). Viral hepatitis markers and tumor markers (alpha fetoprotein, CEA, CA 15-3, CA19-9, CA125) were negative.

The patient subsequently underwent contrast-enhanced dynamic computed tomography (CT) of the liver. A hypervascular hepatic mass at segment 2 that showed progressive washout was detected on CT (Figure 1). For further characterization of the hepatic lesion, magnetic resonance imaging of the upper abdomen was performed using a 3.0-T system, which showed a solid hepatic mass with thin septations and irregular lobulated borders, measuring 4.5 × 3.5 cm at segment 2. The lesion was hypointense on T1-weighted images and hyperintense relative to the liver on T2-weighted images (Figure 2a). The lesion remained hyperintense on the diffusion weighted images obtained, with a *b*-value of 1,000 s/mm^2. It was slightly hyperintense on the apparent diffusion coefficient map, except for a small hypointense component located at its lower part (Figure 2b, c, d). On dynamic imaging performed following intravenous injection of a hepatospecific contrast agent, the mass showed early enhancement in the arterial phase and progressive wash-out in the venous phases (Figure 2e, f). The lesion was hypointense relative to the liver on hepatobiliary phase (Figure 2g).

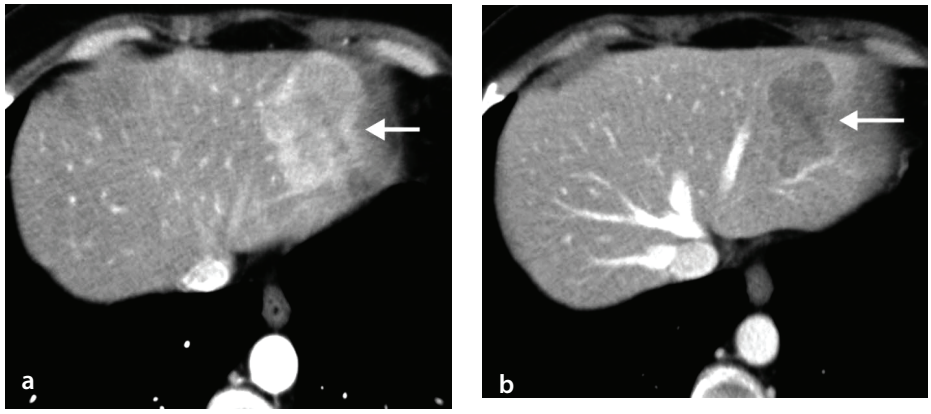


Figure 1. (a) Arterial phase axial computed tomography image demonstrates a hypervascular hepatic mass (arrow) at segment 2. **(b)** The lesion shows washout (arrow) at portal venous phase.

Corresponding author: Ayşe Erden

E-mail: ayse.erden@medicine.ankara.edu.tr

Received 09 April 2025; revision requested 12 May 2025; accepted 13 June 2025.

Epub: 18.08.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253396

You may cite this article as: Erden A, Ceyhan K. Solitary plasmacytoma: a rare and unusual tumor of the liver. *Diagn Interv Radiol*. 2026;32(3):255-257.

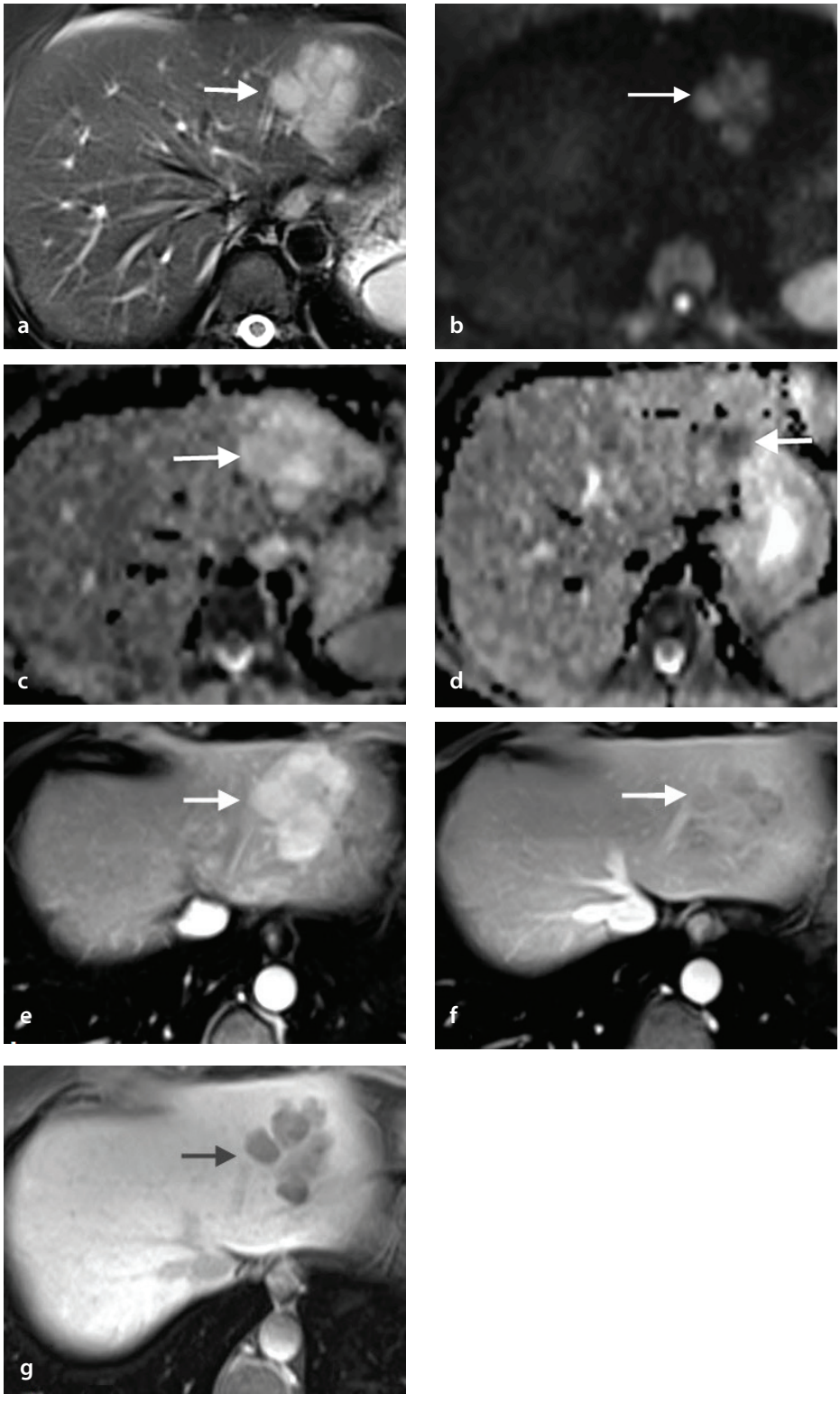


Figure 2. (a) Axial T2-weighted, fat-suppressed (SPAIR) turbo spin echo magnetic resonance image shows a hyperintense hepatic mass (arrow) at segment 2. **(b)** Axial diffusion weighted image at b = 1,000 s/mm^2 and **(c)** apparent diffusion coefficient (ADC) map display slightly hyperintense mass (arrow). **(d)** On the ADC map immediately caudate to this slice, a small hypointense component was present at the lower part of the lesion (arrow). Axial T1-weighted fat-suppressed gradient-recalled echo VIBE images during **(e)** arterial and **(f)** portal venous phases show a hepatic lesion that demonstrates arterial enhancement with washout in the venous phase (arrows). **(g)** The lesion appears to be hypointense (arrow) on hepatobiliary phase. SPAIR, spectral attenuated inversion recovery.

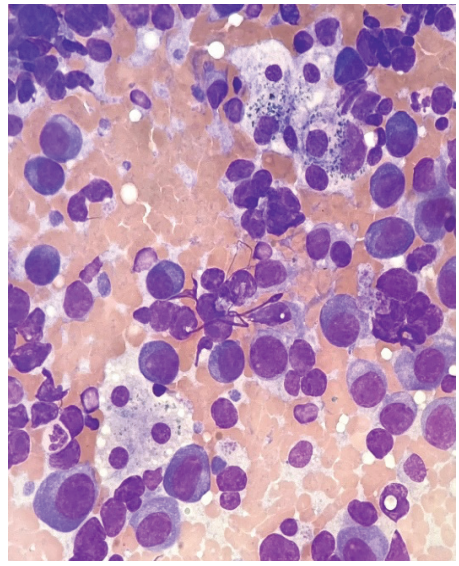


Figure 3. Fine-needle aspiration cytology of hepatic mass. Isolated atypical plasma cells and two non-neoplastic hepatocyte clusters with cytoplasmic bile pigments can be seen in the background (May–Grünwald–Giemsa stain).

Based on the imaging findings, the differential diagnosis included hypervascular liver lesions. Hepatocellular carcinoma was not considered the primary diagnosis due to the absence of a history of chronic liver disease. The lesion's morphology and hepatobiliary phase images were not typical for focal nodular hyperplasia. The lack of intralesional fat made the diagnosis of hepatic adenoma less likely. As the imaging findings were non-specific and inconclusive, and a possible hypervascular metastasis could not be ruled out, a liver biopsy was deemed appropriate.

Ultrasound-guided fine needle aspiration biopsy result was compatible with plasmacytoma (Figure 3). The definitive diagnosis was made via immunocytochemical analysis of cell block material. Immunocytochemical study showed that tumor cells were strongly positive for CD38 and CD138 antibodies and displayed lambda monoclonality.

Serum electrophoresis showed immunoglobulin G (IgG) monoclonal gammopathy. The free light chain assay ratio was abnormal (free kappa/free lambda ratio: 0.25; normal, 0.26–1.65). The immunofixation of the serum showed IgG lambda monoclonal gammopathy. Concentrated urine immunofixation (×50) revealed the presence of lambda light chain.

The imaging findings of the entity known as "solitary plasmacytoma" are nonspecific.[2-4] Based on our observations, it can be stated

that, in the relevant clinical setting, it would be appropriate to keep the possibility of plasmacytoma in mind in the differential diagnosis of an arterially enhancing liver mass.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Stanietzky N, Salem AE, Elsayes KM, et al. Unusual liver tumors: spectrum of imaging findings with pathologic correlation. *Diagn Interv Radiol*. 2025;31(2):58-67. [Crossref]
2. Ooi GC, Chim JC, Au WY, Khong PL. Radiologic manifestations of primary solitary extramedullary and multiple solitary plasmacytomas. *AJR Am J Roentgenol*. 2006;186:821-827. [Crossref]
3. Ng P, Slater S, Radvan G, Price A. Hepatic plasmacytomas: case report and review of imaging features. *Australas Radiol*. 1999;43(1):98-101. [Crossref]
4. Ueda K, Matsui H, Watanabe T, et al. Spontaneous rupture of liver plasmacytoma mimicking hepatocellular carcinoma. *Intern Med*. 2010;49(7):653-657. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253566

ABDOMINAL IMAGING

LETTER TO THE EDITOR

Reply: Solitary plasmacytoma: a rare and unusual tumor of the liver

Nir Stanietzky
Khaled M. Elsayes

The University of Texas, MD Anderson Cancer Center, Department of Abdominal Imaging, Texas, USA

Dear Editor,

Thank you for the opportunity to review this fascinating case. Solitary hepatic plasmacytomas are exceptionally rare tumors, with only a small number described in the literature.[1] This case is an outstanding example demonstrating this rare entity on multiphasic computed tomography and magnetic resonance imaging. The imaging features are variable, and the tumor often resembles a metastasis or a primary hepatic neoplasm.[2,3] In this case, the mass demonstrates hypervascularity with progressive washout and an area of restricted diffusion that could easily be mistaken for other, more common neoplasms. Pathologic confirmation is required in almost every case. However, as the authors state, the diagnosis should be considered as a remote possibility in liver masses with atypical presentations.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Chalopin T, Barillot I, Biny JP, et al. Primary solitary plasmacytoma of the liver - successful treatment with fractionated stereotactic radiotherapy (Cyberknife®): a case report. *J Med Case Rep*. 2017;11(1):193. [Crossref]
2. Ooi GC, Chim JC, Au WY, Khong PL. Radiologic manifestations of primary solitary extramedullary and multiple solitary plasmacytomas. *AJR Am J Roentgenol*. 2006;186(3):821-827. [Crossref]
3. Stanietzky N, Salem AE, Elsayes KM, et al. Unusual liver tumors: spectrum of imaging findings with pathologic correlation. *Diagn Interv Radiol*. 2025;31(2):58-67. [Crossref]

Corresponding author: Nir Stanietzky

E-mail: nstanietzky@mdanderson.org

Received 17 July 2025; accepted 23 July 2025.

Epub: 19.08.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253566

You may cite this article as: Stanietzky N, Elsayes KM. Reply: solitary plasmacytoma: a rare and unusual tumor of the liver. *Diagn Interv Radiol*. 2026;32(3):258.

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253445

ARTIFICIAL INTELLIGENCE AND INFORMATICS

REVIEW

Foundation models for radiology: fundamentals, applications, opportunities, challenges, risks, and prospects

Tugba Akinci D'Antonoli[1,2]
Christian Bluethgen[3]
Renato Cuocolo[4]
Michail E. Klontzas[5-7]
Andrea Ponsiglione[8]
Burak Kocak[9]

[1]University Hospital Basel, Department of Diagnostic and Interventional Neuroradiology, Basel, Switzerland

[2]University Children's Hospital Basel, Department of Pediatric Radiology, Basel, Switzerland

[3]University Hospital Zurich, Institute for Diagnostic and Interventional Radiology, Zurich, Switzerland

[4]Department of Medicine, Surgery and Dentistry, University of Salerno, Baronissi, Italy

[5]University of Crete School of Medicine, Artificial Intelligence and Translational Imaging (ATI) Lab, Department of Radiology, Crete, Greece

6Foundation for Research and Technology (ICS-FORTH), Institute of Computer Science, Computational Biomedicine Lab, Crete, Greece

7Karolinska Institute, Department of Clinical Science, Intervention and Technology (CLINTEC), Division of Radiology, Stockholm, Sweden

8University of Naples Federico II, Department of Advanced Biomedical Sciences, Naples, Italy

9Başakşehir Çam and Sakura City Hospital, Department of Radiology, İstanbul, Türkiye

Corresponding author: Tugba Akinci D'Antonoli

E-mail: tugba.akincidantonoli@unibas.ch

Received 06 May 2025; revision requested 10 June 2025; accepted 22 June 2025.

Epub: 08.07.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253445

ABSTRACT

Foundation models (FMs) represent a significant evolution in artificial intelligence (AI), impacting diverse fields. Within radiology, this evolution offers greater adaptability, multimodal integration, and improved generalizability compared with traditional narrow AI. Utilizing large-scale pre-training and efficient fine-tuning, FMs can support diverse applications, including image interpretation, report generation, integrative diagnostics combining imaging with clinical/laboratory data, and synthetic data creation, holding significant promise for advancements in precision medicine. However, clinical translation of FMs faces several substantial challenges. Key concerns include the inherent opacity of model decision-making processes, environmental and social sustainability issues, risks to data privacy, complex ethical considerations, such as bias and fairness, and navigating the uncertainty of regulatory frameworks. Moreover, rigorous validation is essential to address inherent stochasticity and the risk of hallucination. This international collaborative effort provides a comprehensive overview of the fundamentals, applications, opportunities, challenges, and prospects of FMs, aiming to guide their responsible and effective adoption in radiology and healthcare.

KEYWORDS

Artificial intelligence, deep learning, foundation models, ChatGPT, large language models, medical imaging, radiology

Artificial intelligence (AI), particularly deep learning (DL), has demonstrated considerable efficacy in medical image analysis across various imaging modalities.[1,2] Traditionally, however, AI models in healthcare have been mostly developed for narrow tasks that are highly specific and limited. The recent emergence of foundation models (FMs) represents a significant paradigm shift.[3,4] These large DL models exhibit broad adaptability to a wide range of downstream tasks with minimal task-specific modification.[5,6]

A notable example of FMs is represented by large language models (LLMs), optimized for language-centric tasks, such as summarization, translation, and answering questions.[7] Although LLMs primarily process text, the broader category of FMs can encompass multiple modalities, including text, images, audio, and a diverse spectrum of unstructured data.[8,9] This inherent multimodality aligns well with the diverse data types encountered in modern medicine, such as imaging, clinical narratives, laboratory results, and genomic information.[10-12]

Although current radiology workflows predominantly utilize task-specific models, the multimodal capabilities of FMs make them particularly promising for this field, offering potential support across various interpretative and non-interpretative scenarios (Figure 1).[13,14] The capabilities of LLMs have already been explored for several radiology-related tasks, including report generation,[15] multilingual report translation,[16] information extraction from free-text reports,[17] and the assessment of domain-specific radiological knowledge.[18] Despite growing interest, the use of FMs in radiology is still in the early stages, with ongoing active research and development.[13,14,19-21]

You may cite this article as: Akinci D'Antonoli T, Bluethgen C, Cuocolo R, Klontzas ME, Ponsiglione A, Kocak B. Foundation models for radiology: fundamentals, applications, opportunities, challenges, risks, and prospects. *Diagn Interv Radiol*. 2026;32(3):259-272.

To facilitate the development and potential adoption of FMs, this narrative review synthesizes current knowledge about FMs and aims to provide a comprehensive overview of FMs in the context of radiology. It introduces the fundamental concepts behind FMs, examines their potential applications in radiology, highlights emerging opportunities, outlines key challenges, and suggests future directions in both research and practice.

Fundamental concepts of foundation models

FMs mark a fundamental shift within the conceptual hierarchy of AI (Figure 2), moving beyond conventional, narrowly focused AI systems. They are a class of large-scale AI models developed through training on vast and diverse datasets.[3] A defining feature is their pre-trained nature; unlike conventional models engineered for a single, narrow task (e.g., solely lung nodule detection or lung segmentation), FMs serve as versatile base models (created through pre-training), adaptable to numerous downstream applications (through fine-tuning, continued training on smaller, task-specific datasets).

This inherent adaptability results from several key characteristics (Figure 3). First, the pre-training stage of FMs usually leverages self-supervised learning, allowing the model to learn rich data representations from the data itself (e.g., by solving pretext tasks, such as predicting masked portions of an image or a text) using unstructured, unlabeled, or weakly labeled data (Figure 4).[22,23] This contrasts sharply with conventional methods, which typically require significant amounts of high-quality (manually) labeled data for each distinct task, a major bottleneck due to cost and expert time. Although FMs still require labeled data for fine-tuning, the reliance on specific data for each application can be substantially reduced compared with conventional AI methods (Figure 5).

Main points

- Foundation models (FMs) are versatile artificial intelligence (AI) systems pre-trained on large, diverse datasets, enabling them to adapt to many tasks with minimal fine-tuning.
- FMs with multimodal capacities offer powerful tools for complex radiological applications, such as report generation and diagnostic decision-making.
- FMs have the potential to democratize AI in healthcare by requiring less local data for fine-tuning, helping under-resourced centers.
- Major challenges to FM use in imaging include stochasticity, hallucinated outputs, transparency, bias, sustainability, and regulations.

Second, the reduced need for labeled data allows FM development on a large scale, referring to model size, computational resources, and dataset size.[24] This scale enables the models to learn more generalizable and robust representations that, in turn, support the scalability of the FMs themselves (larger models become feasible with more training data, and more generalizable representations apply to more possible downstream applications). This differentiates FMs from conventional models, which often exhibit limited generalizability beyond the precise conditions (e.g., patient populations or tasks) for which they were trained. Additionally, scaling models has led to the emergence of functionalities beyond the explicit training objectives,[3] such as instruction following, capabilities unprecedented (or at least hardly detectable) in smaller-scale models.[25-27]

Finally, self-supervised learning and the scale of FMs equip them with strong transfer learning capabilities.[28-31] The general knowledge acquired during the resource-intensive pre-training phase can be effectively utilized for new, specific tasks through minimal fine-tuning. This facilitates few-shot learning (where only a small number of task-specific examples are provided) and zero-shot learning (using no examples),[32-34] where models adapt with substantially less specific data than conventional approaches demand. For instance, an FM pre-trained via self-supervised learning on large chest X-ray datasets may be fine-tuned for rib fracture detection using only dozens of cases, whereas a conventional model may require thousands to reach comparable performance.

Developing multimodal foundation models

FMs first took shape in natural language processing (NLP) in the form of LLMs, such as Generative Pre-trained Transformer (GPT)-4 (OpenAI) and Claude (Anthropic). Although FMs can be unimodal, focusing exclusively on one data type, such as text (in the case of LLMs) or images,[35,36] a development of particular importance for radiology is their potential to be multimodal by being able to process and integrate diverse data types, including images [e.g., X-rays,[37-40] computed tomography (CT),[41,42] and magnetic resonance imaging (MRI)[43]], text (e.g., reports and other electronic health record documents), and potentially many more (Figure 1).

Key concepts and modules of FMs concerning radiological applications are presented in Figure 6. Although architectures vary, the transformer design is a frequently used backbone.[44] Its central feature, the attention mechanism, allows it to focus on specific elements of the input sequence. This enables the model to capture long-range dependencies and contextual relationships within data effectively, which gave rise to its initial success in NLP and subsequent adaptation for vision and multimodal scenarios.[45,46] A key concept in handling diverse inputs is the use of modality-specific encoders.[47] These components compress high-dimensional inputs (such as CT scans or text reports) into lower-dimensional embeddings (i.e., numerical vector representations), capturing essential features (e.g., tissue density, anatomical structures, radiological terms). Common encoder architectures include vision transformers and convolutional neural networks for images, and transformers for processing text data. To enable the model to understand relationships across different data types, techniques such as contrastive learning are often employed during pre-training (Figure 7). For instance, the model learns that a specific chest X-ray and its corresponding report describe the same case. Model weights are adjusted so that the embeddings for a matching image–report pair are pulled closer together in a conceptual "shared space," whereas embeddings for unrelated pairs (e.g., the same chest X-ray paired with a report from a different patient) are pushed further apart to learn meaningful cross-modal associations. The Contrastive Language–Image Pre-Training (CLIP) model is a dual neural network trained on a variety of image and text data pairs and is an early example of FMs created this way.[45]

After the individual encoders have processed their respective inputs, fusion modules are used to combine this information, which can happen in several ways (Figure 8). Mechanisms such as cross-attention are particularly powerful here, allowing the model to weigh dynamically the relevance of different parts of one modality based on the content of another–for example, attending to specific words in a report when analyzing a corresponding slice in a CT scan.

Finally, decoders transform these fused representations into desired outputs, which could range from generating text (e.g., report summaries) and predicting classes or outcomes to segmenting relevant image

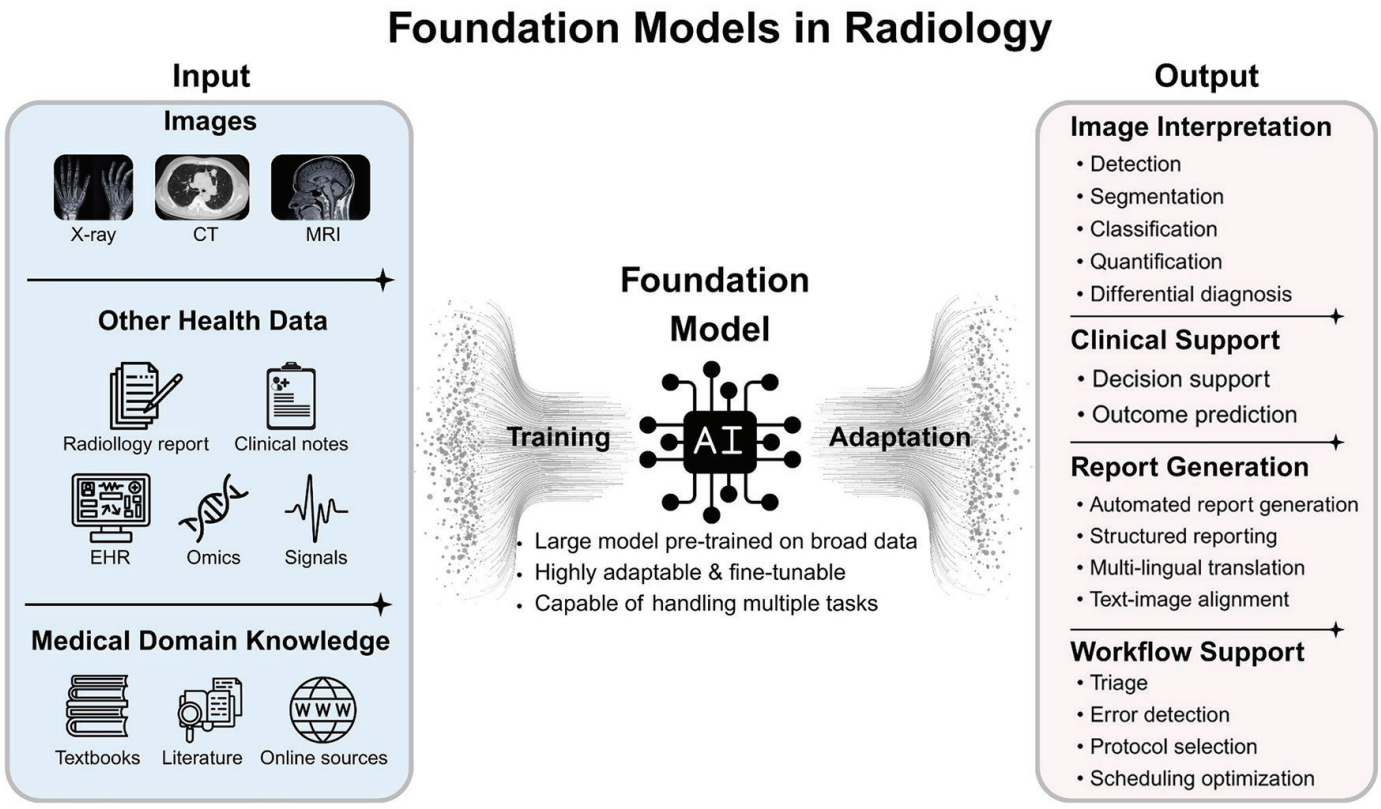


Figure 1. Foundation models in radiology with diverse inputs to perform various interpretative and non-interpretative tasks through training and adaptation. CT, computed tomography; MRI, magnetic resonance imaging; AI, artificial intelligence; EHR, electronic health record.

regions, depending on the specific application. Adapting existing FMs to a specific task can be achieved through full fine-tuning (updating all model parameters) or less computationally expensive parameter-efficient fine-tuning techniques that update only a small number of parameters (Figure 9).[48]

These characteristics make FMs versatile, adaptable, and data-efficient AI models that can integrate multimodal data and capture long-range dependencies within high-dimensional data and across different modalities that may elude narrower models. This uniquely positions FMs to tackle complex problems from the field of radiology by providing potentially richer, more contextualized insights that reflect clinical reality better than conventional AI models.[47]

Applications in radiology

Creating a radiology-specific FM from scratch could be highly cost-intensive, as radiology consists of a wide range of imaging modalities, including X-rays, ultrasound, nuclear imaging, and MRI, that have significant variations in their underlying technologies and data characteristics.[49] Nevertheless, recent advances have shown promising pathways to adapt or fine-tune general-purpose models for domain-specific tasks and modalities, leading to a growing number of successful applications in radiology.

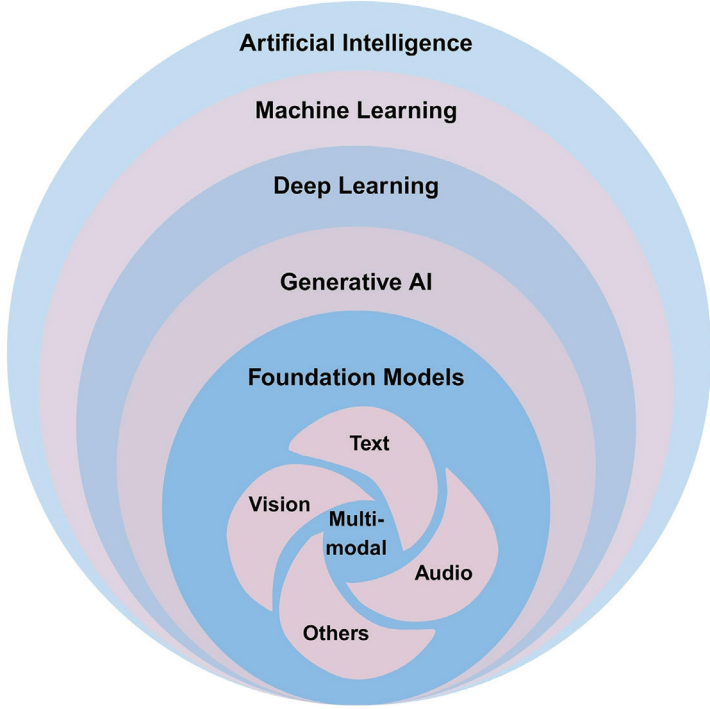


Figure 2. Oversimplified conceptual hierarchy, illustrating the relationship and progression of key domains in artificial intelligence towards foundation models with multimodal capabilities. AI, artificial intelligence.

One core application is medical image segmentation, which aims to delineate regions of interest, such as lesions or organs, automatically. DL models, especially those using the nnU-Net architecture, have shown high accuracy in normal anatomical seg-

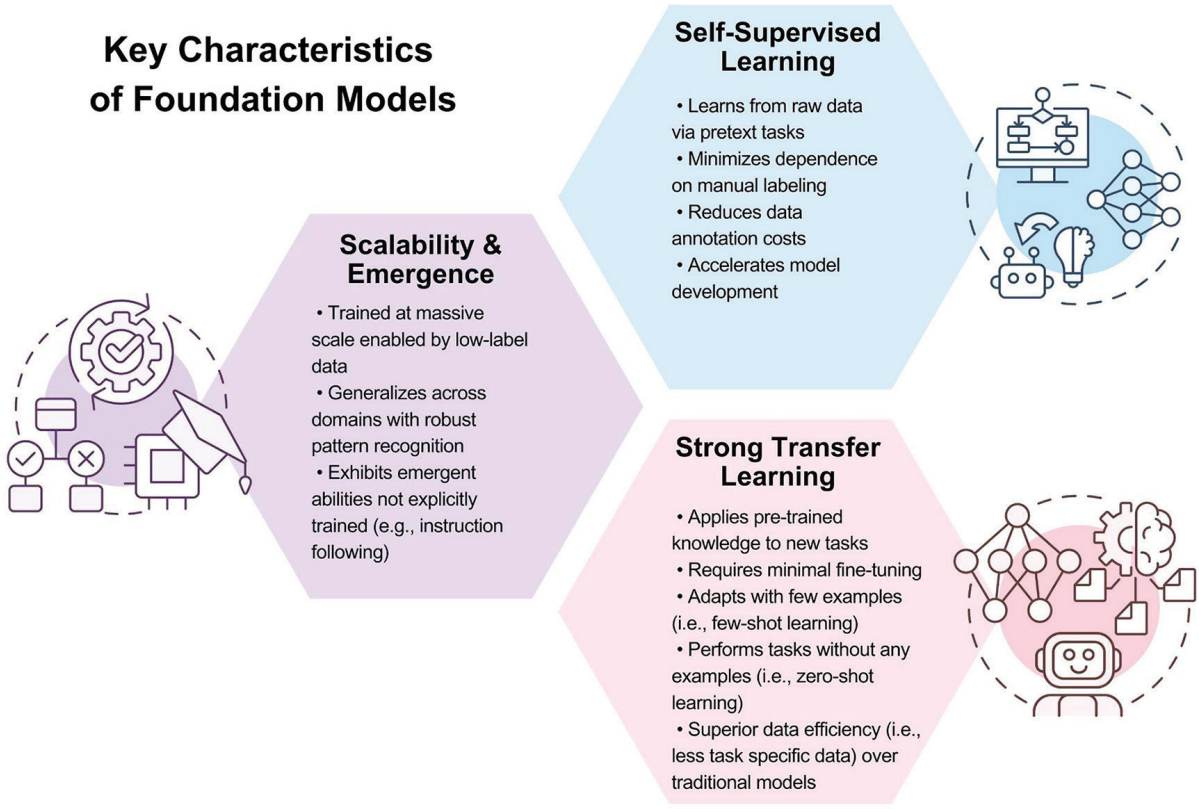


Figure 3. Summary of three key characteristics of foundation models.

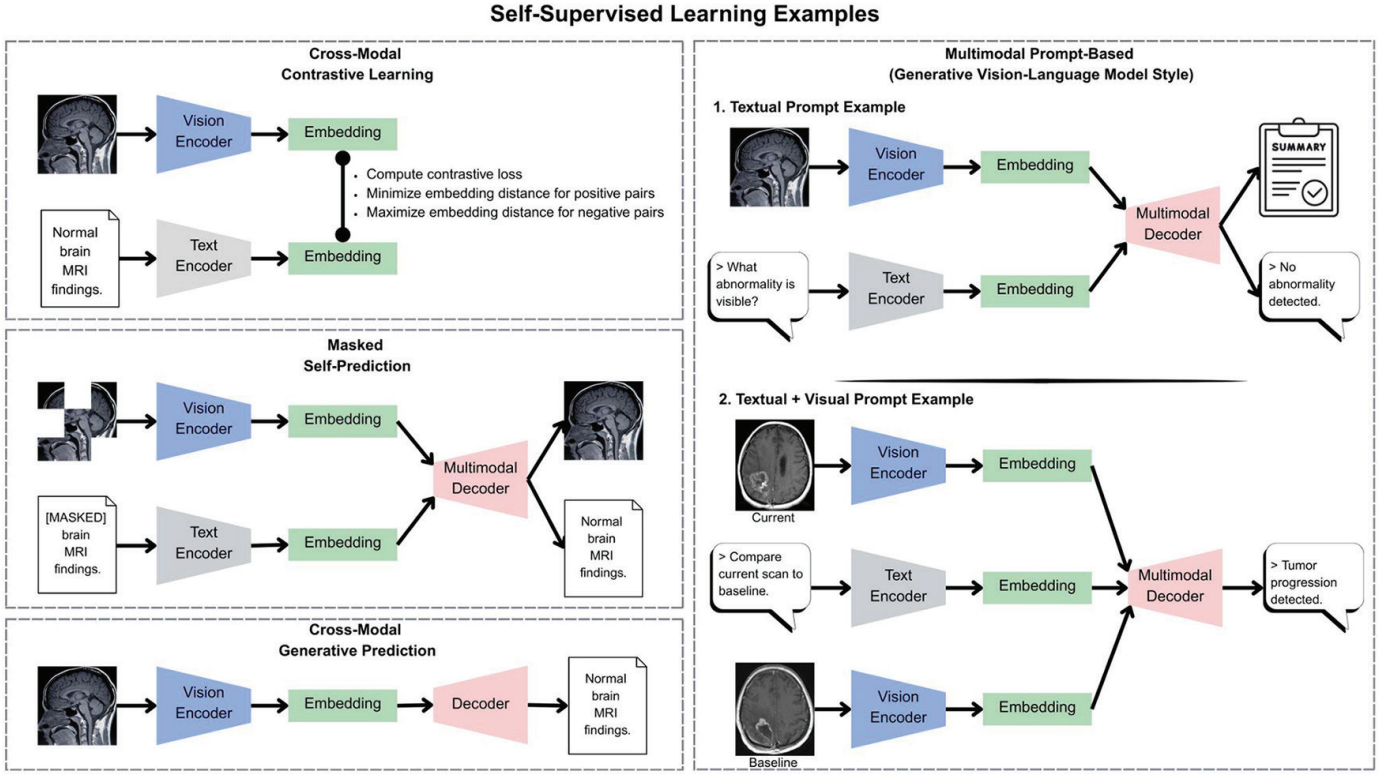


Figure 4. Simplified multimodal self-supervised learning examples. (Top left) Contrastive learning aligns paired image-text embeddings [e.g., brain magnetic resonance imaging (MRI) and report], usually by minimizing the distance between positive pairs and maximizing the distance between negative pairs. (Middle left) Masked self-prediction reconstructs masked regions of an image or text using cross-modal context. (Bottom left) Cross-modal generation predicts one modality from another (e.g., full text reports from image embeddings). (Top right) Text-only prompt-based learning uses a natural language query (e.g., "What abnormality is visible?") to guide generative outputs based on an MRI image input. (Bottom right) Multimodal prompt-based learning uses both current and baseline imaging along with text prompts (e.g., "Compare current scan to baseline") to support complex clinical tasks, such as disease progression assessment or change detection. The baseline image serves as a visual prompt.

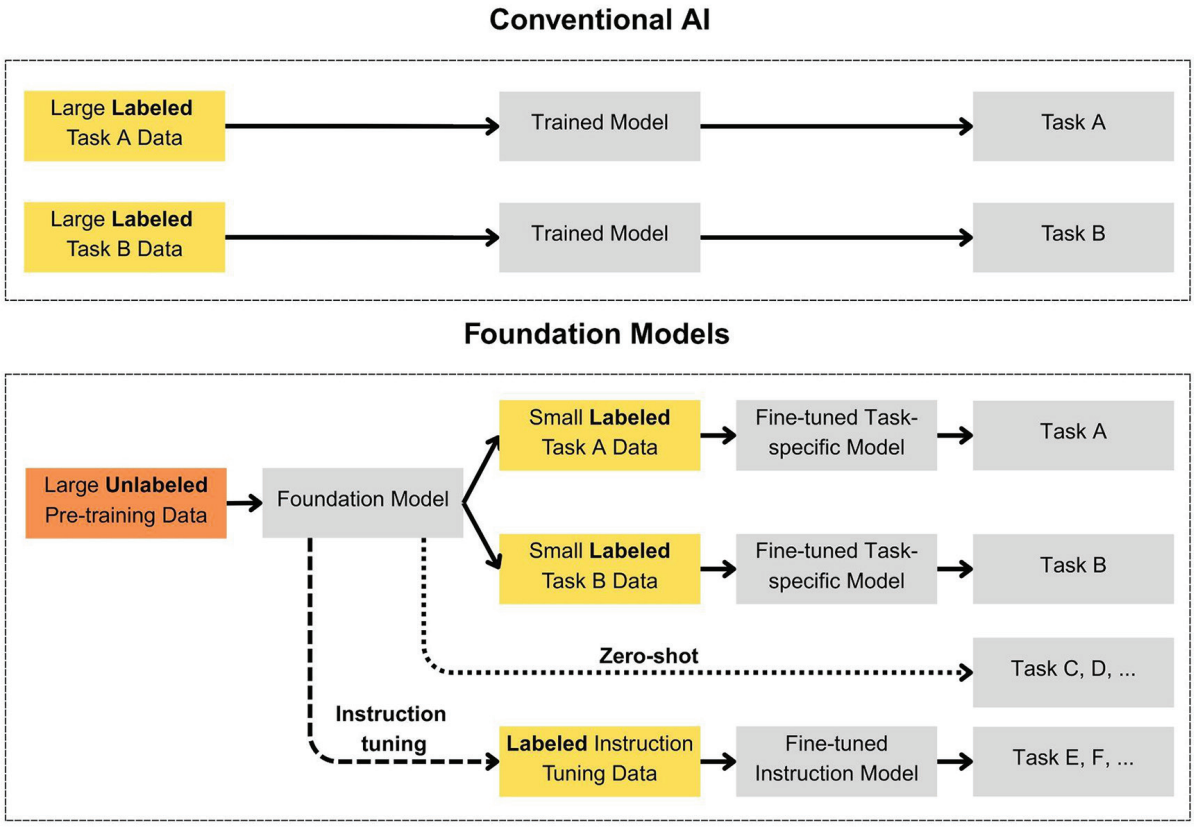


Figure 5. Comparison of conventional artificial intelligence (AI) and foundation models (FMs). Conventional AI requires large, labeled datasets tailored to each specific task and produces models limited to that task. For example, Task A (e.g., lung nodule detection on chest X-ray) and Task B (e.g., lung opacity classification) each require separate models trained on task-specific labeled data. In contrast, FMs are pre-trained on massive unlabeled data and can be adapted to various downstream tasks using smaller labeled datasets. FMs also support zero-shot learning, where the model performs new tasks, such as Task C (e.g., pneumonia detection) and Task D (e.g., pleural effusion identification) without fine-tuning or labeled data. Instruction tuning allows the model to learn from instruction–output pairs (i.e., prompt > output or prompt + image > output) and match expected outputs by adjusting itself when errors occur. This enables it to perform complex tasks, such as Task E (e.g., automated lung report generation) and Task F (e.g., triage of abnormal chest findings) without needing labeled examples for each.

mentation tasks.[50-52] However, the challenge arises when it comes to pathologies, where separate models need to be trained for each one of them; for example, a model trained for liver tumor segmentation cannot be directly applied to the lung tumor or prostate cancer segmentation tasks.

In April 2023, the Segment Anything Model (SAM) was introduced, demonstrating the potential for a single model to handle various segmentation tasks across different domains without needing retraining or fine-tuning.[53] Despite this, SAM's performance on complex medical segmentation tasks, such as those involving the pancreas and spine, has not been satisfactory.[51,54,55] Vision FMs, such as SAM, can serve as starting points, which are then adapted into modality/task-specific FMs (e.g., Medical SAM 2) optimized by leveraging the unique characteristics of medical imaging modalities.[37,56,57]

Another notable development is UniverSeg, a single task-agnostic model trained using a large and diverse set of open-access medical datasets.[58] This model can generalize to the anatomies and segmentation tasks that were not in the training set or were never seen by the model previously. Notably, UniverSeg significantly outperformed existing few-shot methods across all held-out datasets. However, it is important to note that the model was only applied to two-dimensional data and single-label segmentation, and its performance for three-dimensional medical image data remains unclear.

FMs have also shown promise for lesion identification and characterization in different clinical scenarios.[19,59] For example, CXR-Base was developed using a large collection of unlabeled chest X-ray images through self-supervised learning.[59] This approach was sequentially applied to both natural images from ImageNet-1k and chest X-ray images from various public datasets, encompassing a total of 1.1 million chest X-ray images. The model demonstrated good performance across multiple datasets from different centers for diagnosing diseases such as coronavirus disease 2019, pneumonia, and tuberculosis.

Radiology report generation and comprehension represent further promising areas for multimodal FMs. In the task of generating radiology reports, these models can identify abnormalities within images from various modalities while incorporating the patient's medical history and clinical examination findings.[60] By integrating both text and images, these models can generate precise radiology reports, help standardize report quality by detecting inconsistencies or omissions, and subsequently reduce the workload for radiologists.[61] Additionally, these models can generate reports in multiple languages and adjust the complexity of the language to suit the target audience, providing detailed content for specialists and simplified versions for general practitioners.[16,62]

For the comprehension task, physicians can also use multimodal models to enhance case comprehension by engaging in text-based dialogues that focus on specific image sections, allowing for detailed descriptions of those areas.[63] Furthermore, the reports generated by FMs can offer preliminary diag-

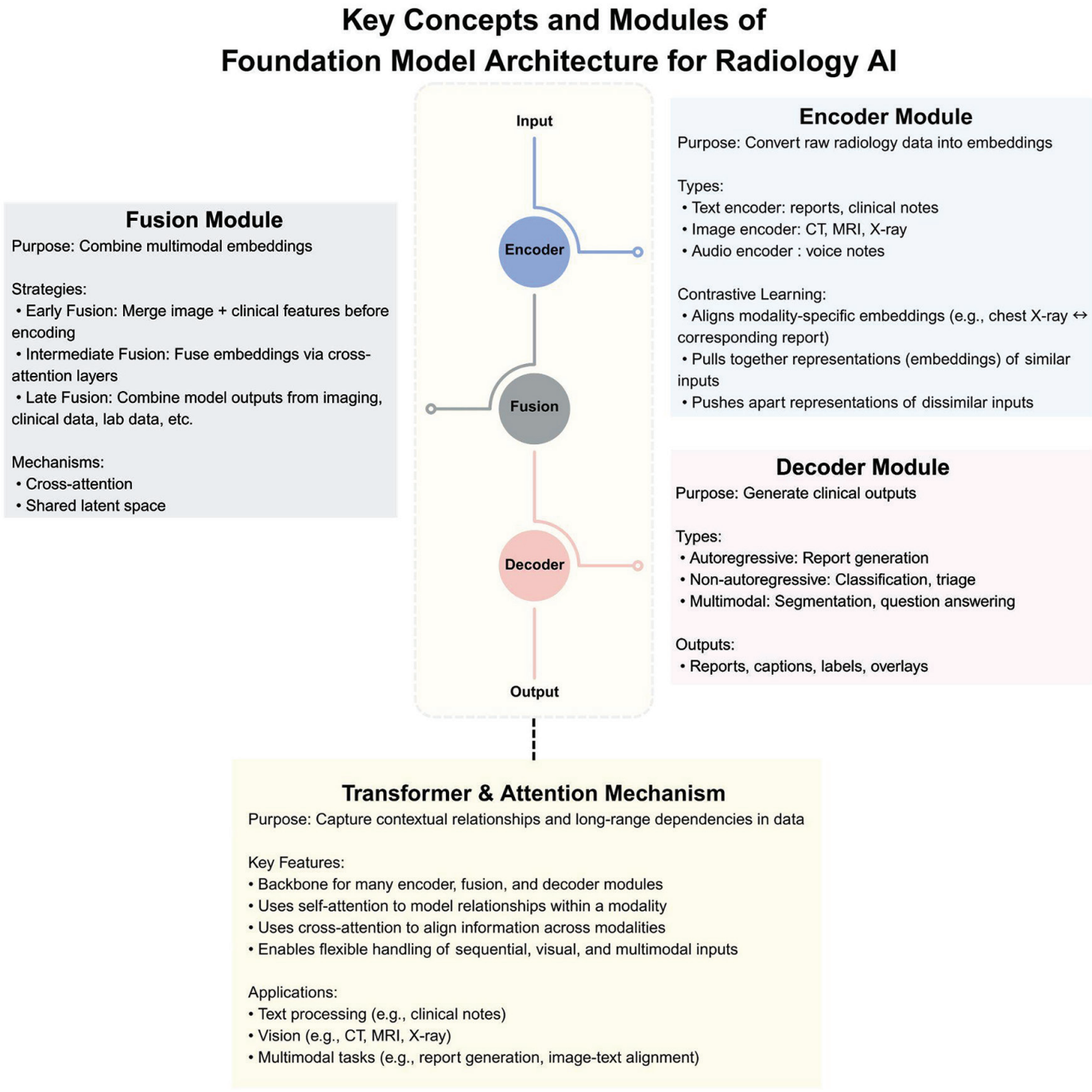


Figure 6. Key concepts and modules of foundation models, with respect to radiological applications. CT, computed tomography; MRI, magnetic resonance imaging; AI, artificial intelligence.

noses, supporting radiologists and clinicians in their decision-making process.[64] These models can potentially propose treatment options or recommend additional diagnostic tests, enhancing the overall clinical workflow for personalized medicine.[65]

Opportunities for radiology

Building on the capabilities outlined above, the adoption of FMs in radiology presents several strategic opportunities (Figure 10).

A key opportunity lies in the ability to fine-tune pre-trained FMs with smaller, local datasets.[66] This can reduce the inequalities related to the availability of data,[67] democratizing access to these applications for healthcare systems with limited data or access to infrastructure. The lower reliance on large datasets can allow their use in centers with limited funding or limited population coverage, which prevents the collection of a high number of cases.

Fine-tuning the models with local datasets can mitigate biases related to underrepresented population characteristics or local peculiarities related to equipment or radiological protocols. At the same time, leveraging their pre-training on large datasets, FMs trained on diverse populations can provide more equitable care recommendations, reducing diagnostic errors in underrepresented groups such as children, ethnic minorities, or patients with rare conditions.[68]

To further address data imbalance, techniques such as synthetic data generation can be used. FMs have the potential to create synthetic medical images, such as CT scans,

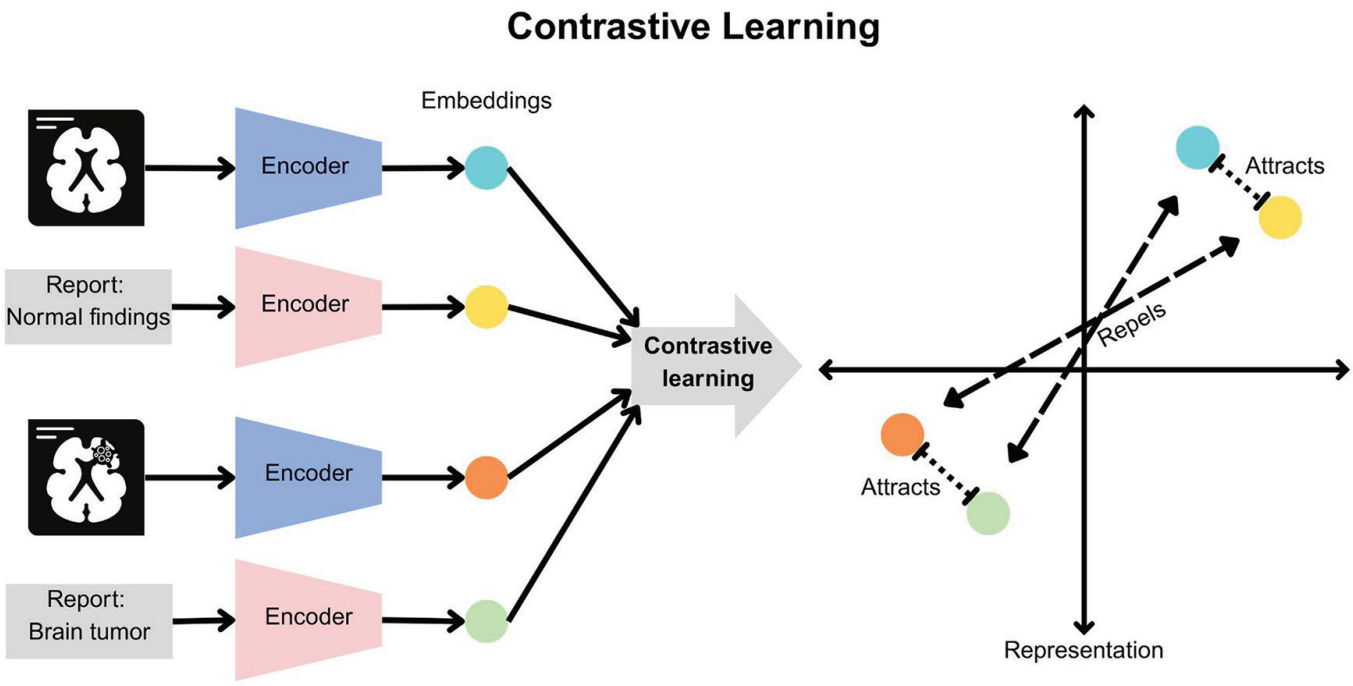


Figure 7. Simplified illustration of contrastive learning. Paired medical images and reports are encoded into embeddings (i.e., numerical vector representations), which are trained to attract matching pairs and repel mismatched ones. This process aligns similar representations in the shared embedding space to improve cross-modal understanding.

MRIs, and X-rays, that resemble real-world data.[69,70] These artificially generated datasets can serve as valuable supplements to existing image collections, particularly when access to patient data is restricted due to privacy issues or limited availability. By generating variations of medical images, these models can help address imbalances in datasets, effectively representing a broader spectrum of pathologies.

The ability to build upon pre-trained backbones has the potential to shorten the innovation-to-implementation cycle significantly. Researchers can build on top of models trained to capture broad medical imaging features and clinical context, rather than creating new models from scratch.[71] Reducing the duration of the innovation-to-implementation cycle can accelerate the development of novel applications, simplify cross-institutional collaborations, and allow innovations developed in academic settings to be rapidly tested and adapted in hospitals, startups, or public health agencies.

FMs also offer unique educational benefits. Automatically annotating synthetic or real images with detailed descriptions, such as the identification of lesions, tumors, or anatomical landmarks, can help radiology residents and healthcare professionals quickly understand complex images. Furthermore, FMs can enhance educational content by not only annotating images but also offering

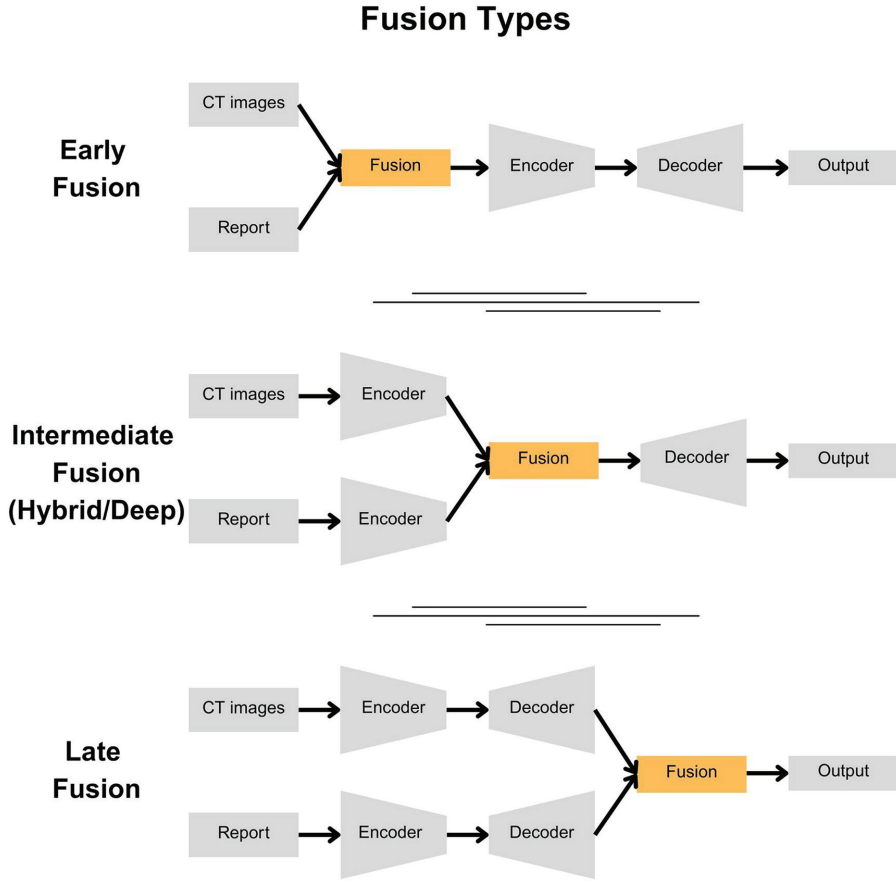


Figure 8. Simplified illustration of multimodal fusion strategies. Illustration of early, intermediate, and late fusion methods for combining medical images and their respective reports. Intermediate fusion (i.e., hybrid or deep fusion) is more common in foundation models due to its flexibility and effectiveness in capturing cross-modal interactions. CT, computed tomography.

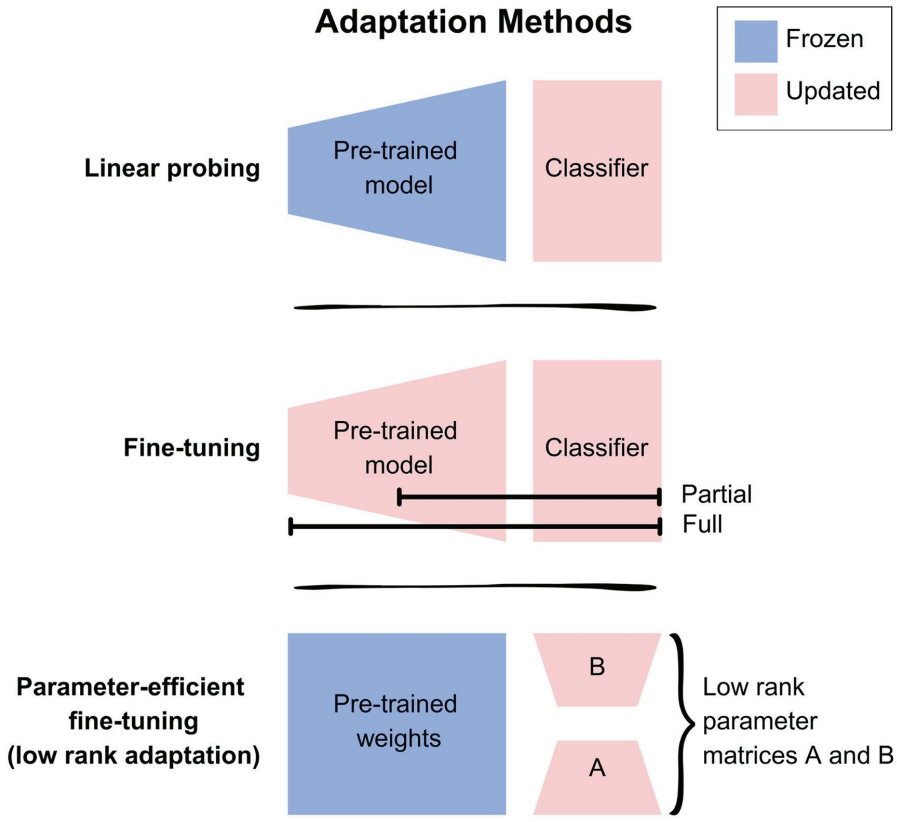


Figure 9. Simplified overview of adaptation methods for pre-trained models. Linear probing updates only the classifier, fine-tuning updates the classifier and optionally parts or all of the pre-trained model, whereas parameter-efficient fine-tuning (e.g., LoRA) inserts low-rank matrices (A and B) without updating original weights.

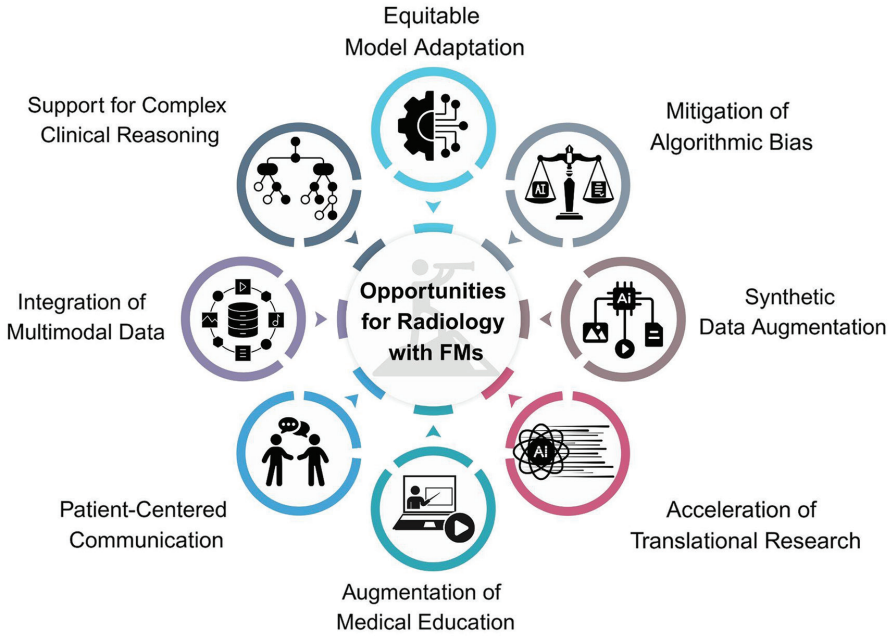


Figure 10. Potential opportunities of foundation models in radiology. FM, foundation models.

detailed explanations of pathologies, their clinical importance, and treatment options, creating an engaging and interactive learning experience.[21]

Moreover, patients will be empowered as their health data, such as their medical history and imaging results, could be translated into personalized, simplified explanations of their condition, treatment options, and possible outcomes with FMs, thus helping patients gain a clearer understanding of their specific health situation.[72,73]

FMs introduce a new paradigm in precision medicine by enabling the integration of diverse data types, such as the combination of diagnostic images, omics, clinical, and laboratory data.[10] Radiologists interpret images in light of clinical information and questions and combine or harmonize these different types of information without much effort in their day-to-day jobs. However, existing AI applications have been inherently less accurate than radiologists, wherever such data harmonization was required. With the advance of FMs, there has been a significant leap towards the combination of multimodal data, which enables more accurate prognostication, risk stratification, and treatment planning.[74]

Beyond data integration, FMs also have the capacity to support complex diagnostic reasoning in uncertain or ambiguous clinical situations. In real clinical practice, radiologists often deal with cases that are not clear-cut, where the diagnosis is not obvious, and decisions have to be made with incomplete information, taking into account follow-up data. FMs can execute different tasks that can help in these situations by offering insights that take the full clinical context into account, rather than just giving a simple yes/no answer.[47] Such models can be trained on a variety of tasks representing real diagnostic scenarios, where the radiologist is presented with imaging examinations that depict a series of pathological conditions and require complex reasoning.

To provide a perspective on emerging opportunities, a comparative overview of conventional AI (single or multimodal), multimodal FMs, radiologist interpretation, and a combined radiologist-FM approach is presented in Table 1.

Challenges and risks for radiology

As discussed so far, although FMs hold promise for transforming radiology, they also introduce multifaceted challenges. Radiologists must remain aware of these issues and proactively address them to ensure the safe, ethical, and effective implementation of FMs in radiology (Figure 11).

One of the main concerns is the stochastic nature of these models, where their outputs may vary every time they are executed, and another is that they can generate plausi-

Table 1. Comparative overview of hypothetical glioma diagnostic pathways using conventional artificial intelligence, multimodal foundation models, radiologist interpretation, and a combined radiologist–FM approach, with an optimistic perspective

Aspect	Conventional AI (single or multimodal)	Multimodal FM	Radiologist (human)	Radiologist + multimodal FM
Inputs	Brain magnetic resonance imaging only + structured clinical data (e.g., age, simple symptoms)	MRI + clinical notes (e.g., seizures, cognitive decline) + structured labs (e.g., IDH status if known) + prior imaging	MRI + full clinical history + prior scans + surgical reports, pathology if available	Same as FM (MRI + notes + labs + prior imaging)
Processing	Task-specific detection/ classification (e.g., identification of mass, segmentation of lesion, basic grading)	Integration of MRI patterns + textual symptoms + past treatments to predict diagnosis, with multimodal reasoning	Correlation of imaging patterns with clinical context and history (e.g., tumor vs. abscess vs. gliosis)	FM preliminary synthesis + human critical review, correction, and clinical judgment
Clinical understanding	Limited contextual understanding detects and segments tumor but lacks the ability to interpret complex clinical nuances not trained for	Broader contextual reasoning by combining multiple modalities, such as text and image; may suggest differential diagnoses, but reliability varies, particularly in rare or atypical cases	Deep contextual understanding built from experience and training; able to integrate subtle imaging findings with complex clinical backgrounds	Combines FM rapid synthesis with human clinical judgment; radiologist can verify, contextualize, and override FM suggestions when needed
Decision support	Reports lesion size, volume, location, edema; sometimes estimates basic tumor type (low-grade vs. high-grade)	Provides full contextual report (e.g., type prediction, treatment suggestions such as re-operation, re-irradiation)	Provides full diagnostic impression (e.g., tumor type suggestion, progression vs. pseudoprogression analysis, biopsy guidance)	FM provides draft; radiologist validates, corrects, adds nuanced interpretation, finally providing an optimized report
Adaptability	Low to moderate: Cannot dynamically adapt to unexpected cases (e.g., rare glioma subtypes, atypical presentations)	High: Generalizes from large data, but can still hallucinate or miss rare conditions	Very high: Experienced ones adapt to rare, complex, or conflicting clinical/imaging situations	Potentially highest: However, real world performance depends on the quality of human–AI interaction (e.g., hesitance to override AI outputs)
Interaction	No interaction: Static outputs	Emerging: Can answer prompts (e.g., "What is differential diagnosis?")	High: interacts with clinicians, neurosurgeons, oncologists, and pathologists; discusses complex cases	Highest: Radiologist queries FM, adjusts outputs, discusses with clinicians dynamically
Limitations	Misses context (e.g., prior radiation therapy effects); cannot explain atypical findings	Hallucination risk; lack of explainability in complex or rare cases	Human variability, fatigue, long interpretation times	Human-in-the-loop improves safety, but challenges remain (e.g., bias, liability concerns, need for clear definitions of radiologists' role in AI-assisted decision support)

AI, artificial intelligence; FM, foundation models; MRI, magnetic resonance imaging; IDH, isocitrate dehydrogenase.

ble-sounding yet incorrect or entirely fabricated information (a phenomenon known as "hallucination").[7] Beyond these inherent issues, FMs present several broader challenges, especially in radiology and healthcare in general. These include challenges related to sustainability, transparency, ethics, cybersecurity, privacy, standardization, and validation.

Another concern is sustainability and environmental impact, as the development and deployment of FMs are highly resource-intensive. They demand vast computational power, energy, and even water. For example, generating a single image with a generative AI model can consume the equivalent of half a smartphone's battery charge,[75] while producing 10–50 medium-length chatbot responses may require up to half a liter of fresh water (Figure 12).[76] Since radiology already depends on energy-intensive imaging equipment, implementing these models in radiology could further exacerbate environmental burdens.[77,78] Addressing these issues and promoting sustainable practices are essential to reducing the environmental impact of FMs in radiology.

FMs often operate as "black boxes," producing outputs without providing clear explanations of their reasoning processes.[79] In radiology, transparency and explainability are critical. Diagnostic decisions made by radiologists must be evidence-based to guide treatment plans and ensure patient safety. If a radiologist cannot justify a diagnosis aided by these tools, their trust in this decision may be undermined. Implementing models with reasoning capabilities, such as OpenAI's GPT o1 series, or adopting frameworks that are designed to facilitate reasoning may help improve trust in model outputs.[80]

Moreover, FMs are prone to perpetuating or even amplifying biases present in the training data, which can contribute to healthcare disparities.[67] In addition, unequal access to such technologies may further disadvantage under-resourced institutions.[67] The legal framework surrounding AI in radiology is also still evolving, and questions about liability, especially in cases where diagnostic errors result from following or ignoring AI-aided recommendations, remain unresolved.[81] Clear ethical guidelines and legal standards are needed to navigate these challenges responsibly.

Training FMs requires large datasets that may contain sensitive patient information. This raises substantial privacy concerns; therefore, ensuring rigorous data anonymization practices during the model training, as well as not using patient data directly as input during model implementation, is essential.[79] FMs may also pose cybersecurity threats, and these tools could be exploited by malicious actors to extract sensitive patient data through techniques such as

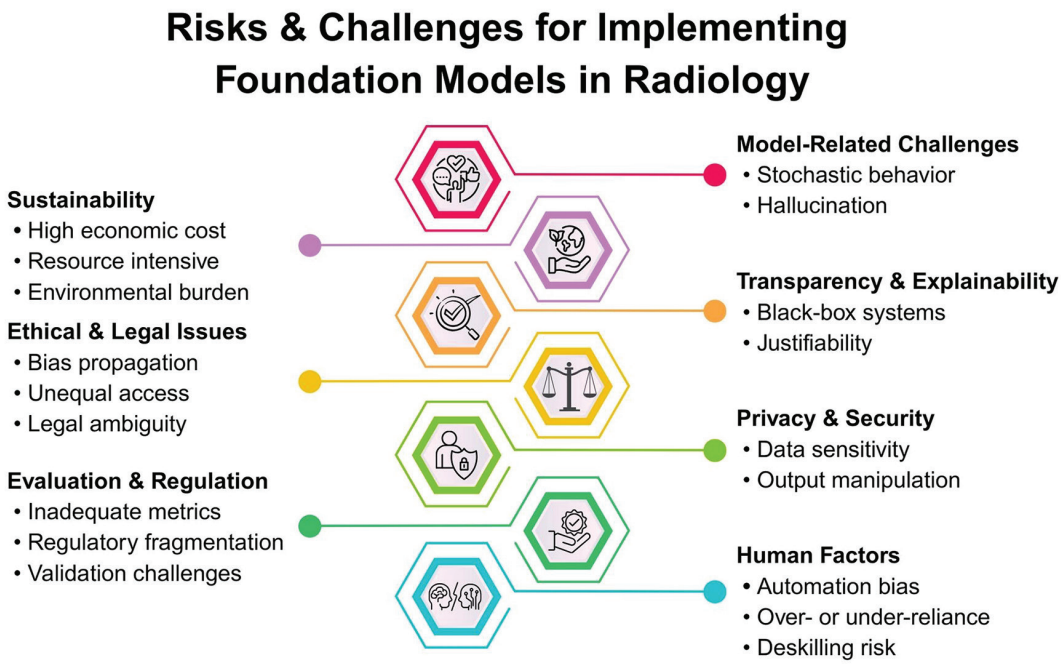


Figure 11. Potential risks and challenges ahead for implementing foundation models in radiology.

jailbreaking or to manipulate model output through techniques such as backdoor attacks.[82,83] Ensuring robust security protocols and continuous monitoring is essential to safeguard patient data and maintain trust when implementing these tools in radiology.

Evaluating the performance of FMs presents another set of challenges. Traditional metrics, such as accuracy or F1 score, may be inadequate for assessing generative outputs or for evaluating the model's generation quality due to the lack of a reference standard.[47] Moreover, regulatory guidelines for ensuring the clinical safety and efficacy of these models are still in their infancy, and each country or region creating its own set of frameworks makes it harder to disseminate these tools (e.g., the AI Act across Europe and the Food and Drug Administration medical device law across the USA).[84] Rigorous validation processes and international regulatory alignment are necessary to overcome this hurdle. Besides the aforementioned challenges, foundational models introduce additional risks. For example, over-reliance on AI tools may lead to the deskilling of radiologists, weakening their ability to assess critically the AI-aided recommendations.[85] Ongoing education and training for radiologists are essential to mitigate deskilling and ensure appropriate use of these technologies.

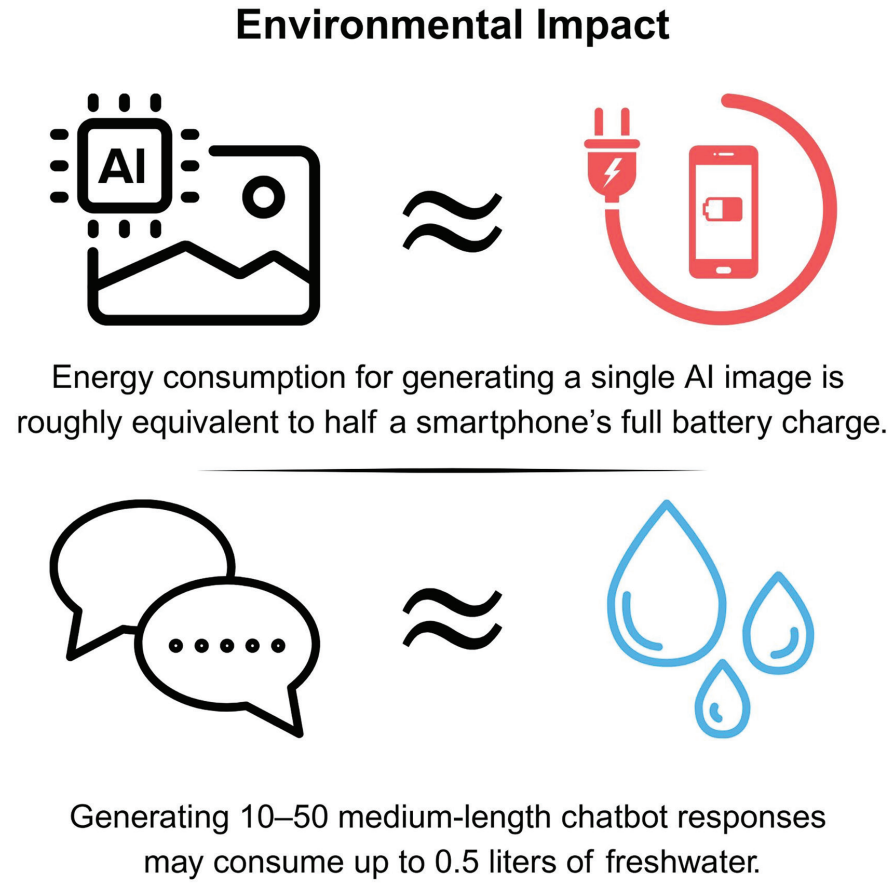


Figure 12. Environmental impact of foundation models. AI, artificial intelligence.

Prospects

As previously discussed, attention-based FMs have represented a technological leap in AI capabilities, with a wide range of potential applications in medical imaging. However, it should be noted that AI research is in continuous development, and even as LLMs and FMs are just starting to be employed in the radiology domain, novel technologies are already aiming to complement or substitute current architectures and improve upon

their performance, alignment, and other limitations.

Among DL developments, state space models and recurrent neural networks currently represent promising architectures in the context of FMs, although the latter is a relatively mature technology.[86-88] Both methods, with different specific implementations, including hybrid approaches,[89] incorporate recursive computation and representation of longer data sequences compared with "pure" transformer-based models. Furthermore, as is often the case in AI, these represent, at least in part, implementations of concepts initially theorized decades ago,[90] which have found new applications due to an increase in computational power and data availability.

Independent of the chosen neural network architecture, in the future, radiologists should expect (and increase their demand for) greater use of open-source software in the setting of generative AI and FMs. At the moment, this domain is largely dominated by proprietary (i.e., closed) technologies, which obfuscate the data used to train these tools, the network's architecture, and the specific weights stored within the trained model. As previously mentioned, this lack of transparency represents a limitation to the implementation of FMs in healthcare, as well as running contrary to the principles contained within the European Union's AI regulatory framework.[91] Nevertheless, high-performance and large-scale FMs, which are also open-source, are already available, with Meta's LLaMA being the most well-known. On the other hand, the increase in transparency afforded by open-source software comes with different considerations and potential tradeoffs. Although open-sourcing the model itself does not inherently compromise the privacy of the original training data (which are usually kept separate) or the input data, ensuring model security, preventing model misuse, and mitigating potential risks such as cyberattacks requires careful governance; this may represent a significant issue in sensitive contexts, such as medical imaging. Furthermore, although open source does not represent an outright impediment to patenting, it does present a greater degree of challenge in protecting the technology behind a medical device and allowing a company to extract the economic value necessary to justify the large-scale investments required to develop such devices and the models running in the backend. This tension between private companies and public interest is not new to healthcare and has been the object of long debates in, for example, the setting of pharmaceutics.[92-94]

AI and FMs can certainly look to these lessons to further establish the appropriate ethical and regulatory framework as these technologies increase their footprint in medical imaging, rather than attempting to reinvent the wheel. A clear sign of the relevance of these considerations is represented by the EU Commission's recently announced intention to withdraw the proposed AI Liability Directive in its 2025 work program, demonstrating the regulator's difficulties in balancing patient protection and incentivizing innovation.[95,96]

AI and FMs will almost certainly impact healthcare, especially medical imaging,[97] in the future. In this setting, radiologists will need to be ready to increase their involvement in multidisciplinary teams. Deployment (and development) of FMs will require the expansion of the expertise requirements in imaging departments and closer collaboration with information technology, data science, and machine learning operations professionals. It could also be argued that the current vision in this profession regarding the implementation of this type of AI is still limited and mostly based on "adding on" FM to the current clinical workflow.[98] However, it is also possible that this may not be the best strategy to implement this technology and may lead to unmet expectations and low impact on patient outcomes.[99,100] Rather, the time may soon come to face the reality that FMs will require a radical rethinking of parts of the medical imaging practice: for example, regarding the scale of service delivery and role of the radiologist.[101]

Final thoughts

FMs represent a potential paradigm shift in AI, offering broad adaptability, multimodal integration, and improved generalizability across a wide range of tasks. In radiology, FMs have an immense potential to enable applications spanning image analysis, report generation, and integrative diagnostics across heterogeneous data sources. However, realizing this potential requires addressing key challenges, including issues of transparency, sustainability, data privacy, regulatory complexity, and ethical implementation. The inherent stochasticity and risk of bias in these models necessitate rigorous validation and continuous monitoring. Successful integration will require not only technical advancement but also adaptive clinical workflows and absolute transparency, potentially facilitated through open-source frameworks. Radiologists (along with other stakeholders) must play a central role in guiding the responsible development and deployment of FMs to ensure they augment, rather than undermine, the quality, safety, and equity of patient care. To this end, the authors of this international collaborative effort provide the radiology community with a set of practical recommendations based on the content extensively discussed in this work, to facilitate the better integration of FMs into clinical practice (Table 2). The authors hope that both the review and the accompanying recommendations will serve as a solid foundation for radiologists in adapting to rapidly evolving AI technologies, specifically FMs.

Table 2. Practical recommendations for radiologists and other healthcare stakeholders regarding the integration of artificial intelligence foundation models into clinical practice

Category	Recommendation
Safety and localization	• Rigorously test foundation models on your data/protocols before any clinical use • Ensure strict patient privacy, anonymization, and regulatory compliance are provided by the vendor
Clinical use and trust	• Prioritize transparent models (e.g., with explainable AI methods) to understand AI reasoning • Integrate AI thoughtfully to genuinely improve, not just add to, processes • Always retain final clinical judgment; AI assists, radiologists decide • Evaluate AI outputs in real-world settings (e.g., different scanners, different patient demographics) to assess robustness • Develop clear communication strategies for disclosing AI-assisted interpretations to referring physicians and patients
Responsible AI	• Be watchful for potential biases and regularly assess following updates for fairness • Demand vendor transparency • Push for third-party validation of AI tools (not just vendor-reported metrics) • Participate in and support open science initiatives for shared model evaluation • Be mindful of AI's environmental/computational, economic, and social footprint
Future-proofing	• Continuously learn AI principles, data science, and ethics to build AI literacy to help make informed decisions • Follow evolving AI regulations (e.g., AI Act) and best practices • Build internal AI governance structures or committees to oversee and audit ongoing usage and updates

AI, artificial intelligence.

Acknowledgements

ChatGPT (4o) and Gemini 2.5 Pro were partially used under strict author supervision for language refinement, summarization, and brainstorming to enhance the content originally written by the authors. CB received research support, independent of this work, from Promedica Foundation, Chur, CH.

Footnotes

Conflict of interest disclosure

Tuğba Akıncı D'Antonoli is a Section Editor of Diagnostic and Interventional Radiology. She had no involvement in the peer-review or decision-making process of this article and had no access to information regarding its peer review.

Burak Koçak, MD, served as a Section Editor of Diagnostic and Interventional Radiology (Section: Artificial Intelligence and Informatics; 2023–2025) during the submission and evaluation of this article. He had no involvement in the peer-review or decision-making process of this article and had no access to information regarding its peer review.

The other authors declare no conflicts of interest.

References

1. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. [Crossref]
2. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digital Med*. 2021;4(1):65. [Crossref]
3. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. *arXiv*. 2021. [Crossref]
4. Azad B, Azad R, Eskandari S, et al. Foundational models in medical imaging: a comprehensive survey and future vision. *arXiv*. 2023. [Crossref]
5. Zhou C, Li Q, Li C, et al. A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT. *arXiv,* v2. 1 May 2023. [Crossref]
6. Myers D, Mohawesh R, Chellaboina VI, et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput*. 2024;27(1):1-26. [Crossref]
7. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol*. 2024;30(2):80-90. [Crossref]
8. Li C, Gan Z, Yang Z, et al. Multimodal foundation models: from specialists to general-purpose assistants. *arXiv*. 2023. [Crossref]
9. Fei N, Lu Z, Gao Y, et al. Towards artificial general intelligence via a multimodal foundation model. *Nat Commun*. 2022;13(1):3094. [Crossref]
10. Simon BD, Ozyoruk KB, Gelikman DG, Harmon SA, Türkbey B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. *Diagn Interv Radiol*. October 1, 2024. [Crossref]
11. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773-1784. [Crossref]
12. Tu T, Azizi S, Driess D, et al. Towards generalist biomedical AI. *NEJM AI*. 2024;1(3). [Crossref]
13. Zhang S, Xu Y, Usuyama N, et al. A multimodal biomedical foundation model trained from fifteen million image-text pairs. *NEJM AI*. 2025;2(1). [Crossref]
14. Strotzer QD, Nieberle F, Kupke LS, et al. Toward foundation models in radiology? Quantitative assessment of GPT-4V's multimodal and multianatomic region capabilities. *Radiology*. 2024;313(2):240955. [Crossref]
15. Busch F, Hoffmann L, Dos Santos DP, et al. Large language models for structured reporting in radiology: past, present, and future. *Eur Radiol*. 2025;35(5):2589-2602. [Crossref]
16. Meddeb A, Lüken S, Busch F, et al. Large language model ability to translate CT and MRI free-text radiology reports into multiple languages. *Radiology*. 2024;313(3):241736. [Crossref]
17. Fink MA, Bischoff A, Fink CA, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. *Radiology*. 2023;308(3):231362. [Crossref]
18. Patil NS, Huang RS, van der Pol CB, Larocque N. Comparative performance of ChatGPT and bard in a text-based radiology knowledge assessment. *Can Assoc Radiol J*. 2024;75(2):344-350. [Crossref]
19. Chen M, Zhang M, Yin L, et al. Medical image foundation models in assisting diagnosis of brain tumors: a pilot study. *Eur Radiol*. 2024;34(10):6667-6679. [Crossref]
20. Ali L, Alnajjar F, Swavaf M, Elharrouss O, Abd-Alrazaq A, Damseh R. Evaluating segment anything model (SAM) on MRI scans of brain tumors. *Sci Rep*. 2024;14(1):21659. [Crossref]
21. Wu C, Zhang X, Zhang Y, Wang Y, Xie W. Towards generalist foundation model for radiology. *arXiv*. 2023. [Crossref]
22. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. *Nat Biomed Eng*. 2022;6(12):1346-1352. [Crossref]
23. Huang SC, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification:

a systematic review and implementation guidelines. *npj Digital Med*. 2023;6(1):74. [Crossref]

24. Hoffmann J, Borgeaud S, Mensch A, et al. Training compute-optimal large language models. *arXiv*. 2022. [Crossref]
25. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. *arXiv*. 2022. [Crossref]
26. Liu H, Li C, Wu Q, Lee YJ. Visual Instruction Tuning. *Advances in Neural Information Processing Systems*. December 15, 2023. [Crossref]
27. Schaeffer R, Miranda B, Koyejo S. Are emergent abilities of large language models a mirage? *arXiv*. 2023. [Crossref]
28. Thrun S, Pratt L. Learning to learn: introduction and overview. In: Thrun S, Pratt L, eds. Learning to Learn. Boston, MA: Springer;1998. [Crossref]
29. Caruana R. Multitask Learning. *Springer Science and Business Media LLC*. 1997;28:41-75. [Crossref]
30. Raina R, Ng AY, Koller D. Constructing informative priors using transfer learning. In: *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*. ACM Press; 2006:713-720. [Crossref]
31. Wu P, Dietterich TG. Improving SVM accuracy by training on auxiliary data sources. In: *Twenty-First International Conference on Machine Learning - ICML '04*. ACM Press; 2004:110. [Crossref]
32. Palatucci M, Pomerleau D, Hinton GE, Mitchell TM. Zero-shot learning with semantic output codes. *Adv Neural Inf Process Syst*. 2009. [Crossref]
33. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. *arXiv*. 2022. [Crossref]
34. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv*. 2020. [Crossref]
35. Oquab M, Darcet T, Moutakanni T, et al. DINOv2: learning robust visual features without supervision. *arXiv*. 2023. [Crossref]
36. Pai S, Hadzic I, Bontempi D, et al. Vision foundation models for computed tomography. *arXiv*. 2025. [Crossref]
37. Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat Biomed Eng*. 2022;6(12):1399-1406. [Crossref]
38. Chen Z, Varma M, Delbrouck JB, et al. CheXagent: Towards a foundation model for chest X-ray interpretation. *arXiv*. 2024. [Crossref]
39. Deperrois N, Matsuo H, Ruipérez-Campillo S, et al. RadVLM: a multitask conversational vision-language model for radiology. *arXiv*. 2025. [Crossref]
40. Bluethgen C, Chambon P, Delbrouck JB, et al. A vision-language foundation model for the generation of realistic chest X-ray images. *Nat Biomed Eng*. 2025;9(4):494-506. [Crossref]
41. Blankemeier L, Cohen JP, Kumar A, et al. Merlin: a vision language foundation model for 3D computed tomography. *arXiv*. 2024. [Crossref]
42. Hamamci IE, Er S, Wang C, et al. Developing generalist foundation models from a multimodal dataset for 3D computed tomography. *arXiv*. 2024. [Crossref]
43. Sun Y, Wang L, Li G, Lin W, Wang L. A foundation model for enhancing magnetic resonance images and downstream segmentation, registration and diagnostic tasks. *Nat Biomed Eng*. 2025;9(4):521-538. [Crossref]
44. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv*. 2017. [Crossref]
45. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. *arXiv*. 2021. [Crossref]
46. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv*. 2020. [Crossref]
47. Paschali M, Chen Z, Blankemeier L, et al. Foundation models in radiology: what, how, why, and why not. *Radiology*. 2025;314(2):e240597. [Crossref]
48. Han Z, Gao C, Liu J, Zhang J, Zhang SQ. Parameter-efficient fine-tuning for large models: a comprehensive survey. *arXiv*. 2024. [Crossref]
49. Zhang S, Metaxas D. On the challenges and perspectives of foundation models for medical image analysis. *Med Image Anal*. 2024;91:102996. [Crossref]
50. Rayed E, Islam SMS, Niha SI, Jim JR, Kabir M, Mridha MF. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Inform Med Unlocked*. 2024;47:101504. [Crossref]
51. Akinci D'Antonoli T, Berger LK, Indrakanti AK, et al. TotalSegmentator MRI: Robust Sequence-independent Segmentation of Multiple Anatomic Structures in MRI. *Radiology*. 2025;314(2):e241613. [Crossref]
52. Wasserthal J, Breit HC, Meyer MT, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell*. 2023;5(5):e230024. [Crossref]
53. Kirillov A, Mintun E, Ravi N, et al. Segment anything. *arXiv*. 2023. [Crossref]
54. He S, Bao R, Li J, et al. Computer-Vision Benchmark Segment-Anything Model (SAM) in Medical Images: Accuracy in 12 Datasets. *arXiv*. 2023. [Crossref]
55. Shi P, Qiu J, Abaxi SMD, Wei H, Lo FP-W, Yuan W. Generalist vision foundation models for medical imaging: a case study of Segment Anything Model on zero-shot medical segmentation. *Diagnostics (Basel)*. 2023;13(11). [Crossref]
56. Chen S, Ma K, Zheng Y. Med3D: Transfer learning for 3D medical image analysis. *arXiv*. 2019. [Crossref]
57. Zhu J, Hamdi A, Qi Y, Jin Y, Wu J. Medical SAM 2: segment medical images as video via Segment Anything Model 2. *arXiv*. 2024. [Crossref]
58. Butoi VI, Ortiz JJG, Ma T, Sabuncu MR, Guttag J, Dalca AV. UniverSeg: Universal medical image segmentation. *arXiv*. 2023. [Crossref]
59. Xu L, Ni Z, Sun H, Li H, Zhang S. A foundation model for generalizable disease diagnosis in chest X-ray images. *arXiv*. 2024. [Crossref]
60. Yan A, McAuley J, Lu X, et al. RadBERT: adapting transformer-based language models to radiology. *Radiol Artif Intell*. 2022;4(4):e210258. [Crossref]
61. Chen Z, Song Y, Chang TH, Wan X. Generating radiology reports via memory-driven transformer. *arXiv*. 2020. [Crossref]
62. Luo L, Vairavamurthy J, Zhang X, et al. ReXplain: translating radiology into patient-friendly video reports. *arXiv*. 2024. [Crossref]
63. Chen Q, Hu X, Wang Z, Hong Y. MedBLIP: Bootstrapping language-image pre-training from 3D medical images and texts. *arXiv*. 2023. [Crossref]
64. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-265. [Crossref]
65. He Y, Huang F, Jiang X, et al. Foundation model for advancing healthcare: challenges, opportunities, and future directions. *arXiv*. 2024. [Crossref]
66. Wolf D, Payer T, Lisson CS, et al. Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging. *Sci Rep*. 2023;13(1):20260. [Crossref]
67. Koçak B, Ponsiglione A, Stanzione A, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn Interv Radiol*. 2025;31(2):75-88. [Crossref]
68. Tripathi S, Gabriel K, Dheer S, et al. Understanding biases and disparities in radiology AI datasets: a review. *J Am Coll Radiol*. 2023;20(9):836-841. [Crossref]
69. Sizikova E, Badal A, Delfino JG, et al. Synthetic data in radiological imaging: current state and future outlook. *BJR Artificial Intelligence*. 2024;1(1). [Crossref]
70. Chambon P, Bluethgen C, Delbrouck JB, et al. RoentGen: Vision-language foundation model for chest X-ray generation. *arXiv*. 2022. [Crossref]
71. He Y, Huang F, Jiang X, et al. Foundation model for advancing healthcare: challenges,

opportunities and future directions. *IEEE Rev Biomed Eng*. 2025;18:172-191. [Crossref]

72. Cai P, Yao Z, Liu F, et al. PaniniQA: enhancing patient education through interactive question answering. *arXiv*. 2023. [Crossref]
73. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. [Crossref]
74. Sun K, Xue S, Sun F, et al. Medical multimodal foundation models in clinical diagnosis and treatment: applications, challenges, and future directions. *arXiv*. 2024. [Crossref]
75. Luccioni S, Jernite Y, Strubell E. Power hungry processing: watts driving the cost of AI deployment? In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM; 2024:85-99. [Crossref]
76. Li P, Yang J, Islam MA, Ren S. Making AI Less "Thirsty": uncovering and addressing the secret water footprint of AI models. *arXiv*. 2023. [Crossref]
77. Doo FX, Vosshenrich J, Cook TS, et al. Environmental sustainability and ai in radiology: a double-edged sword. *Radiology*. 2024;310(2):e232030. [Crossref]
78. Kocak B, Ponsiglione A, Romeo V, Ugga L, Huisman M, Cuocolo R. Radiology AI and sustainability paradox: environmental, economic, and social dimensions. *Insights Imaging*. 2025;16(1):88. [Crossref]
79. Akinci D'Antonoli T. Ethical considerations for artificial intelligence: an overview of the current radiology landscape. *Diagn Interv Radiol*. 2020;26(5):504-511. [Crossref]
80. Bhattacharjya D, Lee J, Agravante DJ, Ganesan B, Marinescu R. Foundation model sherpas: guiding foundation models through knowledge and reasoning. *arXiv*. 2024. [Crossref]
81. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med*. 2021;4(1):31. [Crossref]
82. Yao Y, Duan J, Xu K, Cai Y, Sun Z, Zhang Y. A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly. *High-Confidence Computing*. 2024:100211. [Crossref]
83. Ponsiglione A, Cuocolo R. Radiology under siege? Adversarial attacks against deep learning algorithms. *Eur J Radiol*. 2023;169:111156. [Crossref]
84. Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. *NPJ Digital Med*. 2024;7(1):237. [Crossref]
85. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: the impact of artificial intelligence BIRADS suggestions on reader performance. *Radiology*. 2023;307(4):e222176. [Crossref]
86. Sieber J, Alonso CA, Didier A, Zeilinger MN, Orvieto A. Understanding the differences in foundation models: attention, state space models, and recurrent neural networks. *arXiv*. 2024. [Crossref]
87. Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. *arXiv*. 2021. [Crossref]
88. Stanić A, Ashley D, Serikov O, et al. The languini kitchen: enabling language modelling research at different scales of compute. *arXiv*. 2023. [Crossref]
89. De S, Smith SL, Fernando A, et al. Griffin: mixing gated linear recurrences with local attention for efficient language models. *arXiv*. 2024. [Crossref]
90. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780. [Crossref]
91. Kotter E, D'Antonoli TA, Cuocolo R, et al. Guiding AI in radiology: ESR's recommendations for effective implementation of the European AI Act. *Insights Imaging*. 2025;16(1):33. [Crossref]
92. Statman M, Tyebjee TT. Trademarks, patents, and innovation in the ethical drug industry. *J Mark*. 1981;45(3):71-81. [Crossref]
93. De George RT. Intellectual property and pharmaceutical drugs: an ethical analysis. *Bus Ethics Q*. 2005;15(4):549-575. [Crossref]
94. Millum J. Are pharmaceutical patents protected by human rights? *J Med Ethics*. 2008;34(11):e25. [Crossref]
95. Ruschemeier H. The de-regulatory turn of the EU Commission. *Verfassungsblog*. 2025. [Crossref]
96. VOSS A. Carriages preview | Legislative Train Schedule. Accessed April 26, 2025. [Crossref]
97. Kocak B, Baessler B, Cuocolo R, Mercaldo N, Pinto Dos Santos D. Trends and statistics of artificial intelligence and radiomics research in radiology, nuclear medicine, and medical imaging: bibliometric analysis. *Eur Radiol*. 2023;33(11):7542-7555. [Crossref]
98. Linguraru MG, Bakas S, Aboian M, et al. Clinical, cultural, computational, and regulatory considerations to deploy AI in radiology: perspectives of RSNA and MICCAI experts. *Radiol Artif Intell*. 2024;6(4):e240225. [Crossref]
99. Del Gaizo AJ, Osborne TF, Shahoumian T, Sherrier R. Deep learning to detect intracranial hemorrhage in a national teleradiology program and the impact on interpretation time. *Radiol Artif Intell*. 2024;6(5):e240067. [Crossref]
100. Savage CH, Tanwar M, Elkassem AA, et al. Prospective evaluation of artificial intelligence triage of intracranial hemorrhage on noncontrast head CT examinations. *AJR Am J Roentgenol*. 2024;223(5):e2431639. [Crossref]
101. Pierre K, Haneberg AG, Kwak S, et al. Applications of artificial intelligence in the radiology roundtrip: process streamlining, workflow optimization, and beyond. *Semin Roentgenol*. 2023;58(2):158-169. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.252851

ARTIFICIAL INTELLIGENCE AND INFORMATICS

COMMENTARY

The AI Act: responsibilities and obligations for healthcare professionals and organizations

Kicky Gerhilde van Leeuwen[1,2]
Leon Doorn[3,4]
Erik Gelderblom[5]

[1]Romion Health, Utrecht, The Netherlands
[2]Health AI Register, Utrecht, The Netherlands
[3]MedQAIR, Amsterdam, The Netherlands
[4]QAIR, Amsterdam, The Netherlands
[5]Radboud University Medical Center, Department of Medical Technology and Clinical Physics, Nijmegen, The Netherlands

Corresponding author: Kicky van Leeuwen

E-mail: Kicky.vanleeuwen@romionhealth.com

Received 23 March 2025; revision requested 08 April 2025; accepted 12 May 2025.

Epub: 29.05.2025

Publication date: 04.05.2026

DOI:10.4274/dir.2025.252851

On August 1, 2024, the artificial intelligence (AI) Act 2024/1689 officially came into force within the European Union (EU). Since the United States Executive Order 14110 on AI from 2023 was recently revoked, it sets the global standard as a regulatory framework to govern AI systems.[1] The Act applies across all sectors and, as such, also introduces requirements and controls for the use of AI in healthcare. Although medical devices (MDs) (with and without AI) have long been subject to the rules and requirements of the MD Regulation (MDR) (preceded by the MD directive) and the *in vitro* diagnostics regulation (IVDR) (preceded by the *in vitro* diagnostic MDs directive), these requirements primarily focus on the manufacturers.[2,3] The AI Act extends this dynamic by introducing AI-specific requirements for manufacturers (providers), as well as additional responsibilities for the users (deployers) of AI-enabled MDs.

Central to the AI Act is the classification of AI systems based on their level of risk: prohibited, high-risk, limited-risk, minimal-risk, and general-purpose AI models (with and without systemic risk) or systems. MDs incorporating AI are generally classified as "high-risk" because AI often serves as a key functionality or safety component, and most software-based MDs require a conformity assessment, per their assigned risk classification, by a notified body under the MDR or IVDR before they can be placed on the EU market. High-risk AI systems must meet stringent requirements for design, risk management, performance, transparency, human oversight, logging, and monitoring under the AI Act to ensure their safe and effective use.[4]

The additional requirements for the providers do not exempt healthcare organizations and individual users, designated deployers, from keeping pace with the new regulations.[5] Some requirements are already covered by the MDR and IVDR, such as ensuring the MD is used according to its intended purpose and reporting incidents. Other regulatory frameworks, such as the General Data Protection Regulation 2016/679, may require healthcare organizations to conduct data protection impact assessments to ensure privacy is adequately protected.[6]

This commentary highlights the most important additional requirements for deployers of high-risk AI solutions in healthcare, as summarized in Figure 1 and Table 1. It explores the boundaries of responsibility between the MD industry, healthcare organizations, and individual users. We reflect on how the AI Act reshapes accountability and places new demands on healthcare professionals as users of AI systems.

Obligations for healthcare organizations and users (deployers)

Ensuring artificial intelligence literacy among healthcare staff

Healthcare organizations are expected to ensure the AI literacy of their staff to support the safe and responsible use of AI systems (AI Act, article 4). The level of AI literacy required depends on context and role. For clinical users, this may involve general AI knowledge (understanding the capabilities and risks of AI) and system-specific knowledge (understanding how to interpret the AI system's output and detect malfunctioning). This responsibility applies not only at the time of deployment but also over the entire product lifecycle, as updates with new functionalities may occur over time. Other responsibilities include input data control,

You may cite this article as: van Leeuwen KG, Doorn L, Gelderblom E. The AI Act: responsibilities and obligations for healthcare professionals and organizations. *Diagn Interv Radiol*. 2026;32(3):273-275.

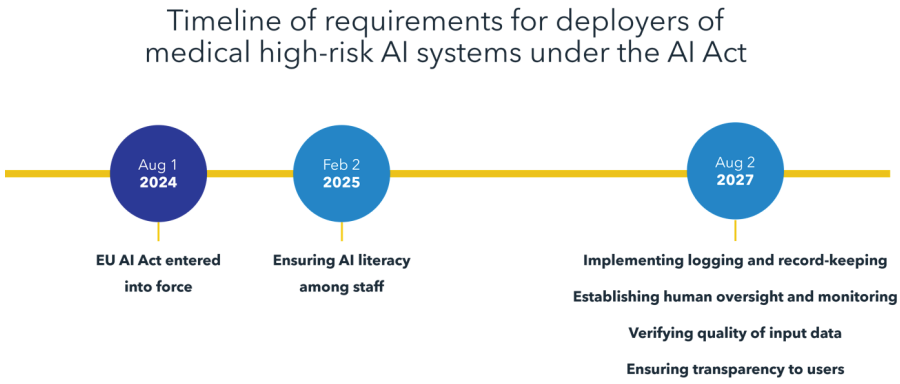


Figure 1. Timeline of requirements for deployers of medical high-risk AI systems under the AI Act. AI, artificial intelligence.

Table 1. Responsibilities for deployers (healthcare organizations) of AI systems under the AI Act

Responsibilities for deployers of AI systems under the AI Act	Related articles
Ensuring AI literacy among staff	Article 4
Implementing logging and record-keeping	Article 26(6)
Establishing human oversight and monitoring	Articles 26(2), 26(5)
Verifying the quality of input data	Article 26(4)
Ensuring transparency to users	Article 50

AI, artificial intelligence.

record keeping, and monitoring for automation bias, which will require IT personnel and medical physicists to become AI literate. Management and leadership roles may also need to enhance their AI literacy, as they are often involved in implementation decisions and governance (AI Act, Preamble 20 and 91).

The topic of AI is still often lacking in formal educational programs.[7] It is therefore up to healthcare organizations to judge what level of AI literacy is sufficient and for whom. Suitable materials and supplementary training may be obtained through professional societies, conferences, external parties, or AI system providers. It is important to note that this requirement applies as of February 2, 2025, as shown in Figure 1.

The AI Act also promotes the AI literacy of "affected persons," who, in the healthcare context, are likely to be patients or their relatives. It states that the European AI Board should support the commission in promoting AI literacy tools and public awareness (AI Act, Preamble 20). Some hospitals contribute by informing and educating patients on how AI is applied, for example, through posters or information on their websites.

Providers' role: Providers also have a responsibility to ensure sufficient AI literacy. They must ensure their own staff are AI literate and facilitate adequate knowledge among their users (AI Act, article 4). Appropriate instructions for use are often obligatory under the MDR and IVDR and are always obligatory under the AI Act (AI Act, article 13). Providers also frequently organize user training to support the responsible use of their systems.

Implementing logging and record-keeping

Logging and record-keeping responsibilities under the AI Act are shared between the AI system provider and the deployer.

Deployers are responsible for managing and storing logs once the system is in use within their organization. They must ensure that logs are retained for a period appropriate to the AI system's intended purpose, with a minimum duration of 6 months, unless determined otherwise by another EU or national law. These logs should be accessible for internal reviews, audits, and the reporting of incidents to relevant authorities when necessary (AI Act, article 26(6)).

Providers' role: Providers are responsible for embedding technical capabilities within the AI system to allow for automatic event recording over its entire lifecycle. These logs must capture key events related to system functionality, including identifying potential risks and data necessary to facilitate human oversight and monitoring. Some AI vendors and platforms already provide dashboards with longitudinal insights on system functioning to facilitate monitoring and human oversight (AI Act, article 12).

Establishing human oversight and monitoring

Human oversight is required to minimize risks to health, safety, and fundamental rights (AI Act, Article 26(2), 26(5)). This involves defining clear workflows to ensure that anomalies or unexpected performance are detected. For example, healthcare professionals must have the authority to override AI predictions when clinical judgment contradicts the AI output. Human oversight must also safeguard against automation bias by ensuring that healthcare professionals critically assess the recommendations provided by the AI system. If deployers suspect that using a high-risk AI system according to its instructions could pose a risk, they must immediately suspend its use and notify the provider and relevant authorities without delay. In the case of a serious incident, they must first inform the provider, followed by the importer or distributor and the market surveillance authority. Depending on the nature of the incident, the surveillance authority may be either the traditional MD competent authority for safety incidents or the appointed national surveillance authority under the AI Act for infringements of fundamental rights (AI Act, article 3[49(c)]).

Human oversight may range from reviewing individual results generated by an AI system to more holistic oversight, in which trends are monitored over time to allow early detection of performance drift, bias, or malfunctioning of the AI system. Logging and record-keeping capabilities play a key role in supporting such oversight measures. Although the AI Act does not exclude the possibility of autonomous AI use, it requires that appropriate oversight measures be in place to ensure safe deployment. As the concept of autonomy remains open to interpretation, future guidelines will be essential to clarify what level of human oversight is suitable for different degrees of autonomy.

Providers' role: Providers of high-risk AI systems must ensure that their systems are designed to enable effective human oversight through appropriate human–machine interface tooling. These measures should be

built into the system or be possible for the deployer to implement. Oversight measures must be proportionate to the system's risks, autonomy, and context (AI Act, article 14).

Verifying the quality of input data

Deployers are responsible for verifying that the data input into AI systems complies with the requirements specified by the provider (AI Act, article 26(4)). Poor quality or incomplete input data could lead to erroneous AI predictions, posing risks to patient safety. Where AI systems may continue to learn from data in clinical practice, proper quality control by the healthcare provider becomes even more important, as it may affect the overall accuracy of the AI system. Healthcare organizations may consider implementing automated procedures to pre-screen data for quality assurance and to ensure it adheres to the requirements outlined by the AI system's provider. For example, in radiological AI, data orchestration is often used to ensure that images meet specified criteria, such as resolution and metadata, to allow the system to process them correctly.

Providers' role: In the instructions to users, providers must clearly specify the input data requirements (AI Act, article 13(3)).

Ensuring transparency to users

The responsibility for providing transparency on the AI system predominantly lies with the provider; however, it is up to deployers to ensure that this information reaches the users. Users are most often healthcare providers, for example, when an AI system supports a physician in the diagnostic process. However, users can also be patients, such as when they use an AI system for (chronic) disease management. Healthcare organizations must ensure that users of AI systems are adequately informed that they are using an AI-based product and are aware of its capabilities, limitations, and potential risks to health, safety, and fundamental rights (AI Act, Preamble 27, 72; article 50).

Providers' role: Providers are responsible for supplying information about the AI system through an instruction for use (AI Act, article 13), which is already mandatory under the MDR for most MDs. Additionally, the AI Act explicitly states that users must be informed when they are interacting with an AI system (AI Act, Preamble 72; article 50(1)).

Impact on in-house developed AI

In-house developed MDs, used exclusively for a healthcare organization's own patients and not placed on the market, may be exempt from third-party (notified body) conformity assessments under the MDR. Without this obligation, such devices are not classified as high-risk under the AI Act, Article 6 (b). However, article 43(3) suggests that certain AI systems, contrary to Article 6(b), may still qualify as high-risk even though they are exempt from third-party conformity assessments under Union Harmonisation Legislation listed in Annex I. To facilitate uniform implementation of the AI Act for in-house developed AI-enabled MDs, further clarification or guidance from the European Commission is desired.

In the meantime, healthcare organizations could apply the MDR concept for in-house developed products. This means they should aim to ensure safety, security, and the protection of fundamental rights. This can be achieved by following the requirements for high-risk AI systems, including risk management, quality management system requirements, and post-market monitoring, potentially through the use of harmonized standards.

General-purpose models and administrative AI tools

Software solutions using general-purpose AI models, such as large language models, are gaining popularity. These systems can support administrative work, automate note-taking, summarization, or report generation. The intended purpose of the AI system utilizing a general-purpose AI model determines its risk classification under the AI Act. When there is no medical intended purpose (and therefore no qualification as a MD under the MDR), these systems are generally considered minimal-risk under the AI Act. There are no specific obligations for deployers of AI systems classified as minimal-risk. However, providers of general-purpose AI systems face additional requirements, mostly related to effectiveness, interoperability, robustness, reliability, transparency, and model evaluation (AI Act, article 50).

In conclusion, the AI Act represents a substantial shift in regulating AI systems used in healthcare, extending responsibilities to healthcare organizations as deployers. By emphasizing AI literacy, data quality, human oversight, transparency, and monitoring, the Act promotes the safe and effective use of AI in clinical practice. Healthcare organizations must rise to the challenge of implementing these systems responsibly, balancing innovation with patient safety, even as many standards and guidance documents are still under development.[8] Ultimately, the success of AI in healthcare depends on collaboration between providers and deployers, along with a shared commitment to compliance, education, and ethical use.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. European Artificial Intelligence Act, 2024/1689 (2024). Accessed: December 30, 2024. [Crossref]
2. European Medical Device Regulation, 2017/745 (2017). European Commission. Accessed: December 30, 2024. [Crossref]
3. European *in-vitro* diagnostics regulation, 2017/746 (2017). European Commission. Accessed December 30, 2024. [Crossref]
4. Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. *NPJ Digit Med.* 2024;7(1):237. [Crossref]
5. Kotter E, D'Antonoli TA, Cuocolo R, et al. Guiding AI in radiology: ESR's recommendations for effective implementation of the European AI Act. *Insights Imaging*. 2025;16(1):33. [Crossref]
6. European General Data Protection Regulation, 2016/679 (2016). Accessed December 30, 2024. [Crossref]
7. Gilbert A, Pizzolla E, Palmieri S, Briganti G. Artificial intelligence in healthcare and regulation challenges: a mini guide for (mental) health professionals. *Psychiatr Danub*. 2024;36(Suppl 2):348-353. [Crossref]
8. van Kolfschooten H, van Oirschot J. The EU artificial intelligence Act (2024): implications for healthcare. *Health Policy*. 2024;149:105152. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253449

ARTIFICIAL INTELLIGENCE AND INFORMATICS

COMMENTARY

When AI reviews your work: author-centered reflections on LLMs in peer review

Burak Koçak[1]
Mehmet Ruhi Onur[2]

[1]Başakşehir Çam and Sakura City Hospital, Department of Radiology, İstanbul, Türkiye

[2]Hacettepe University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

The integration of large language models (LLMs), such as Chat Generative Pre-trained Transformer (ChatGPT) and Gemini, into peer review has recently emerged as a critical and rapidly evolving issue, raising serious concerns.[1,2] LLMs can be used in various ways during the review process, including language refinement, drafting initial feedback, and even generating full review reports from scratch, yet the extent of their involvement remains unclear.[2,3] Although journals, editors, and reviewers have been the focus of most previous discussions about the use of LLMs in peer review, this commentary shifts attention to authors–the individuals whose unpublished work is being evaluated. Although the core concerns may be shared, authors might experience them from a distinct perspective, shaped by their limited control over the review process and their reliance on it for a fair, expert, and confidential evaluation of their work (Figure 1 and Table 1).

Importantly, and reflecting these very concerns, major academic publishers and journals generally prohibit the use of LLMs in the peer review process, particularly the uploading of manuscripts into such tools.[4,5] However, because these tools are easily accessible, there is a risk that reviewers might use them without disclosure, which would breach editorial policies and bypass oversight. From an author's viewpoint, this potential for unacknowledged LLM use adds another layer of uncertainty to an already non-transparent peer review system.

For authors, one of the most important concerns is the potential breach of confidentiality surrounding their unpublished work.[1,6] The peer review process is conventionally built on a foundation of trust and strict confidentiality, intended to safeguard novel data and ideas from premature or unauthorized disclosure. However, the use of LLMs–particularly general-purpose, widely available models that may store or externally process input–poses a serious risk. If a reviewer inputs all or part of a confidential manuscript into such a model, sensitive content could inadvertently become part of future training data. For authors who have invested substantial time, intellectual effort, and resources into their research, the idea that their findings could be exposed or repurposed before publication is deeply concerning. Even though some LLMs or chat modes claim to offer secure data handling through temporary chat sessions or offline use, authors have no assurance that reviewers will choose or correctly implement these options.

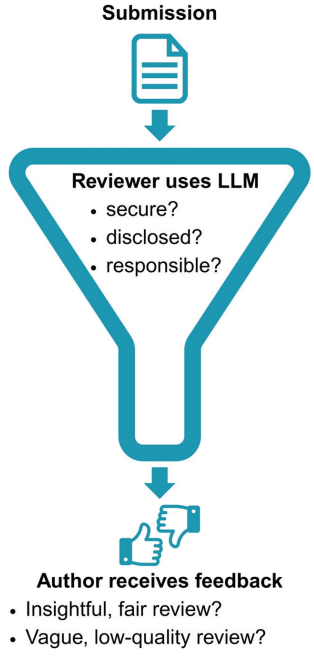


Figure 1. In the context of LLM-involved peer review, the outcome for the author depends on the often unclear and undisclosed extent of LLM use by reviewers. LLM, large language model.

Corresponding author: Burak Koçak

E-mail: drburakkocak@gmail.com

Received 07 May 2025; revision requested 20 May 2025; accepted 25 May 2025.

Epub: 02.06.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253449

You may cite this article as: Koçak B, Onur MR. When AI reviews your work: author-centered reflections on LLMs in peer review. *Diagn Interv Radiol*. 2026;32(3):276-278.

Table 1. Key author-centered concerns regarding LLM-involved peer review and their implications

Concern	Description	Implications for authors
Confidentiality	Reviewer may input the manuscript content into general-purpose LLMs	Potential breach of confidentiality; unauthorized reuse of unpublished ideas or data
Feedback quality	Feedback may be overly generic, superficial, or context-insensitive	Limited value in improving the manuscript; lack of expert-level critique; the possibility of misguidance
Hallucination risk	LLMs may introduce fictitious references or identify non-existent flaws	Authors may waste effort addressing invalid or fabricated concerns
Inconsistency	Responses may be internally inconsistent or conflict with other reviewers' comments	Challenges in interpreting and responding to contradictory or incoherent feedback
Bias and manipulation	LLMs may favor prestigious authors or verbose texts; vulnerable to prompt manipulation	Risk of unfair assessments and unintentional reinforcement of systemic biases
Lack of transparency	Reviewers may not disclose their use of LLMs	Authors may be unaware of AI-generated content and unprepared to interpret LLM-specific issues. Familiarity with AI-generated generic, irrelevant, or fabricated language may erode trust in reviewers and the integrity of journal reviews

LLM, large language model; AI, artificial intelligence.

Beyond confidentiality, the quality and reliability of the feedback generated by LLMs pose a major challenge for authors.[7,8] Authors submit their manuscripts expecting insightful, expert critique that helps refine their arguments, methodology, and findings. However, LLMs often lack the nuanced, critical insight of human reviewers.[2] They may generate generic praise or criticism and struggle to evaluate complex or niche academic topics effectively, failing to produce a properly balanced review.[9] Authors receiving such superficial or generic reviews may feel their work has not been truly assessed by an expert, hindering their ability to revise the manuscript effectively. This use of LLMs may result in a notable shift in the author–reviewer dynamic, where authors may develop serious criticisms of the reviewers' reports. Furthermore, authors who focus solely on publishing their work by adhering to reviewer feedback–without questioning its validity–may inadvertently weaken their submission by incorporating misguided or irrelevant revisions, potentially leading to a decline in the quality of the first manuscript draft rather than improvement.

In addition, the potential for inconsistencies, contradictions, and bias in LLM-shaped reviews can create confusion and frustration for authors.[10,11] LLMs are highly sensitive to prompt variations, meaning that even slight changes in phrasing can produce markedly different responses. This variability may lead to internally inconsistent reviews or comments that contradict feedback from other reviewers. LLMs can also exhibit sycophancy, aligning with a reviewer's biased phrasing rather than the manuscript's objective content. From an author's perspective, receiving contradictory or unclear feedback makes it difficult to identify valid points for revision. Compounding this, LLMs may demonstrate bias–potentially favoring papers from well-known authors or prestigious institutions if the review is not blinded.[11] This raises concerns about fairness and equity in the evaluation process, particularly for authors from less prominent backgrounds.

Another notable concern is the tendency of LLMs to generate irrelevant or fabricated content, including fictitious references.[9] Authors may receive comments based on non-existent issues or be asked to address points supported by fabricated citations. Identifying these "hallucinations" requires authors–or editors–to critically scrutinize every detail of the review, adding another layer of burden to the already demanding process of manuscript revision.

Perhaps the most fundamental problem from the author's viewpoint is the lack of transparency regarding LLM use in review.[12] Reviewers may not disclose their use of AI tools, and the inherent opacity of LLMs–combined with tools designed to make AI-generated text appear human-like–makes detection challenging for editorial teams. This means authors may receive a review shaped or even generated by an LLM without knowing it. Without this knowledge, authors are ill-equipped to interpret the feedback appropriately or to advocate for their work in response to potential LLM idiosyncrasies such as hallucinations or contradictions.

Recognizing these challenges, one key recommendation is to notify authors if the peer review process involves LLM assistance.[2] This disclosure is crucial, as it allows authors to understand the potential influence of the tool and to respond accordingly to feedback that may reflect LLM limitations. It empowers authors to critically evaluate the review and address possible flaws attributable to AI rather than blindly accepting potentially inaccurate or irrelevant comments.

In conclusion, while often unspoken, LLM involvement in peer review may be more common than acknowledged and is likely to increase with the widespread availability of these tools. For authors, the integrity of peer review depends on receiving expert, objective, reliable, and confidential evaluations. Rather than pursuing an unrealistic ban, the focus should shift toward managing LLM use responsibly–ensuring strong human oversight and critical judgment so that LLMs support, rather than undermine, the peer review process.[13] Safeguarding the integrity of peer review requires clear journal policies, targeted training for editors and reviewers, and transparency with authors to enable informed responses. Authors, both as contributors and community members, play a critical role in upholding peer review standards amid increasing LLM involvement. They, in turn, should remain vigilant and adopt best practices to protect the integrity of their work (Figure 2).

Figure 2. Author responsibilities and recommended actions in the context of LLM-involved peer review. Although individual measures are generally feasible, community-level recommendations require collective action, such as by editorial boards or professional societies. LLM, large language model.

Footnotes

Acknowledgments

Language of this manuscript was checked and improved by ChatGPT (4o). The authors conducted strict supervision when using this tool.

Conflict of interest disclosure

Burak Koçak, MD, served as a Section Editor of Diagnostic and Interventional Radiology (Section: Artificial Intelligence and Informatics; 2023–2025) during the submission and evaluation of this article. He had no involvement in the peer-review or decision-making process of this article and had no access to information regarding its peer review.

Mehmet Ruhi Onur, MD, served as Editor-in-Chief of Diagnostic and Interventional Radiology from 2023 to 2025 and during the submission and evaluation of this article. He had no involvement in the peer-review or decision-making process of this article and had no access to information regarding its peer review. The manuscript was handled by an independent editor.

References

1. Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res Integr Peer Rev*. 2023;8(1):4. [Crossref]
2. Kocak B, Onur MR, Park SH, Baltzer P, Dietzel M. Ensuring peer review integrity in the era of large language models: a critical stocktaking of challenges, red flags, and recommendations. *European Journal of Radiology Artificial Intelligence*. 2025;2:100018. [Crossref]
3. Zhou L, Zhang R, Dai X, Hershcovich D, Li H. Large language models penetration in scholarly writing and peer review. Published online February 16, 2025. [Crossref]
4. Hamm B, Marti-Bonmati L, Sardanelli F. ESR Journals editors' joint statement on Guidelines for the use of large language models by authors, reviewers, and editors. *Insights Imaging*. 2024;15(1):18. [Crossref]
5. Moy L. Guidelines for use of large language models by authors, reviewers, and editors: considerations for imaging journals. *Radiology*. 2023;309(1):e239024. [Crossref]
6. Zhuang Z, Chen J, Xu H, Jiang Y, Lin J. Large language models for automated scholarly paper review: a survey. Published online January 17, 2025. [Crossref]
7. Liang W, Zhang Y, Cao H, et al. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*. 2024;1(8):AIoa2400196. [Crossref]
8. Ou J, Walden WG, Sanders K, et al. CLAIMCHECK: how grounded are LLM critiques of scientific papers? Published online March 27, 2025. [Crossref]
9. Donker T. The dangers of using large language models for peer review. *Lancet Infect Dis*. 2023;23(7):781. [Crossref]
10. Borji A. A Categorical archive of ChatGPT failures. Published online April 3, 2023. [Crossref]
11. von Wedel D, Schmitt RA, Thiele M, Leuner R, Shay D, Redaelli S, Schaefer MS. Affiliation bias in peer review of abstracts by a large language model. *JAMA*. 2024;331(3):252-253. [Crossref]
12. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K. Guidance for authors, peer reviewers, and editors on use of ai, language models, and chatbots. *JAMA*. 2023;330(8):702-703. [Crossref]
13. Ebadi S, Nejadghanbar H, Salman AR, Khosravi H. Exploring the impact of generative AI on peer review: insights from journal reviewers. *J Acad Ethics*. Published online February 11, 2025. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253391

BREAST IMAGING

ORIGINAL ARTICLE

Comparison of changes in dynamic contrast-enhanced magnetic resonance imaging and flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography parameters from baseline to post-neoadjuvant therapy in predicting pathological response in breast cancer

Hüseyin Akkaya[1]
Aygül Polat Kelle[2]
Tuba Dalgalar Akkaya[3]
Selim Özdemir[4]
Kübra Karaaslan Erişen[5]
Bozkurt Gülek[5]

[1]Ondokuz Mayıs University Faculty of Medicine, Department of Radiology, Samsun, Türkiye

[2]University of Health Sciences Türkiye, Adana City Training and Research Hospital, Clinic of Nuclear Medicine, Adana, Türkiye

[3]Samsun University Faculty of Medicine, Department of Radiology, Samsun, Türkiye

[4]Düziçi State Hospital, Clinic of Radiology, Osmaniye, Türkiye

[5]University of Health Sciences Türkiye, Adana City Training and Research Hospital, Clinic of Radiology, Adana, Türkiye

Corresponding author: Hüseyin Akkaya

E-mail: dr.hsynakkaya@gmail.com

Received 14 April 2025; revision requested 10 June 2025; last revision received 22 June 2025; accepted 21 July 2025.

Epub: 18.08.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253391

PURPOSE

This study aimed to compare the value of differences (Δ) in parameters obtained via both dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) and flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) between baseline and post-neoadjuvant therapy in predicting the pathological response to neoadjuvant therapy in breast cancer.

METHODS

A total of 109 patients who underwent both baseline and post-neoadjuvant therapy DCE-MRI and ^{18}F-FDG PET/CT examinations were retrospectively analyzed. The DCE-MRI parameters and ^{18}F-FDG PET/CT parameters [metabolic tumor volume (MTV), standardized uptake value $(SUV)_{max}$, SUV_{mean}, and total lesion glycolysis] were recorded at both time points. Additionally, the Δs between these parameters were calculated. Postsurgical pathology reports were documented, and the patients were subsequently categorized into two groups: those exhibiting pathologic complete response (pCR) and those exhibiting partial response. Parameters from DCE-MRI and ^{18}F-FDG PET/CT were compared to determine which predicted pathological response to neoadjuvant therapy more effectively.

RESULTS

Patients with partial response demonstrated a higher rate of histologic grade 3 than those with pCR ($P = 0.030$). The only DCE-MRI parameter to indicate a significant difference between the two groups ($P = 0.024$) was the Δ(%)wash-out rate. Among the baseline parameters, only MTV successfully predicted pathological response ($P = 0.033$). The only post-neoadjuvant therapy parameter to be predictive of pathological response ($P = 0.003$) was SUV_{mean}. In receiver operating characteristic analysis, ΔSUV_{mean} emerged as the most significant parameter for predicting pathological response, followed by post-neoadjuvant SUV_{mean} [area under the curve: 0.724 (95% confidence interval: 0.630–0.805) and 0.673 (0.577–0.760), respectively].

CONCLUSION

The Δ^{18}F-FDG PET/CT parameters are better than ΔDCE-MRI in predicting pathologic response to neoadjuvant therapy. Among these parameters, ΔSUV_{mean} is the most successful.

CLINICAL SIGNIFICANCE

Neoadjuvant chemotherapy (NAC) response is one of the most important criteria in breast cancer prognosis. The two most important imaging modalities in breast cancer diagnosis and follow-up protocols are MRI and ^{18}F-FDG PET/CT. However, it is not clear which of these two modalities is more successful in predicting the difference in treatment response between baseline and post-NAC.

KEYWORDS

Breast cancer, metabolic tumor volume, total lesion glycolysis, dynamic magnetic resonance imaging, positron emission tomography

You may cite this article as: Akkaya H, Polat Kelle A, Dalgalar Akkaya T, Özdemir S, Karaaslan Erişen K, Gülek B. Comparison of changes in dynamic contrast-enhanced magnetic resonance imaging and flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography parameters from baseline to post-neoadjuvant therapy in predicting pathological response in breast cancer. *Diagn Interv Radiol*. 2026;32(3):279-286.

Breast cancer is a significant public health concern, and women in particular face high breast cancer incidence and mortality rates.[1] Prognostic factors are critical for estimating recurrence risk and providing increasingly personalized treatment for patients.[2] Current therapeutic approaches in breast cancer focus on neoadjuvant chemotherapy (NAC)–also referred to as primary or induction chemotherapy–which is administered prior to surgical intervention.[3] Histological response following NAC remains the most important factor associated with recurrence risk.[2-4] In addition, early prediction of NAC response can reduce patient costs and protect patients from unnecessary side effects and time loss, leading to the use of alternative methods. Therefore, additional and more precise methods of predicting NAC response and estimating prognosis are necessary.

Although novel imaging parameters have been developed to characterize the biological features of breast tumors, factors affecting the quality of the therapeutic response remain uncertain.[5,6] Dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) shows tissue vascularity and perfusion due to blood flow by analyzing signal intensity.[7] Semi-quantitative parameters derived from DCE-MRI can be calculated based on the signal intensity curve.[8] Although studies investigating the relationship between these semi-quantitative parameters and therapeutic responses in various tumor tissues are not uncommon in the literature,[7-9] research examining the relationship of the differences (Δ) in these parameters between diagnosis and post-neoadjuvant therapy with NAC response remains limited.

Flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) is a valuable tool in the staging of breast cancer at diagnosis and in evaluating therapeutic response and suspicion of recurrence using standardized uptake value (SUV) and its associations with various predictive and prognostic factors.[10] The SUV_{max} parameter corresponds to the point with the highest SUV value measured within a region of interest (ROI); in other words, it defines the most active part of the tumor and does not provide insight into the overall tumor status. Metabolic tumor volume (MTV) and total lesion glycolysis (TLG) have been introduced as semi-quantitative measures of ^{18}F-FDG uptake within a heterogeneous tumor mass. The MTV represents the volume of tumoral tissue demonstrating a specific ^{18}F-FDG uptake activity beyond the intensity of FDG uptake by tumor tissues.[11] Several studies have also compared the efficacy of TLG with that of MTV in predicting therapeutic responses.[2,10,11]

This study aims to evaluate whether semi-quantitative ΔDCE-MRI parameters and Δ^{18}F-FDG PET/CT indices (ΔMTV, ΔSUV_{max}, ΔSUV_{mean}, and ΔTLG) could predict pathological responses to NAC in patients with breast cancer during initial staging. These parameters are also compared with other predictive and prognostic factors [age, cellular proliferation index, histologic grade, hormone receptor expression, and human epidermal growth factor receptor 2 (HER2) status] in terms of the pathological response after NAC.

Main points

- Of the dynamic contrast-enhanced magnetic resonance imaging parameters, only Δ(%) wash-out values differed between the two groups with pathologic complete response (pCR) and partial response, with higher Δ(%) wash-out in the pCR group ($P = 0.024$).
- Delta standardized uptake value (ΔSUV_{mean}) demonstrated the best diagnostic test performance at predicting treatment response among these parameters [area under the curve: 0.724 (95% confidence interval: 0.630–0.805)].
- Metabolic tumor volume (MTV), SUV_{mean}, ΔMTV, ΔSUV_{mean}, delta total lesion glycolysis (ΔTLG), and Δwash-out values after neoadjuvant treatment and the distribution between groups were similar.
- In patients with pCR, baseline MTV, Δ(%) MTV, Δ(%) SUV_{mean}, and Δ(%) TLG values were significantly higher, whereas post-neoadjuvant chemotherapy SUV_{mean} values were significantly lower than those with a partial response.

Methods

Patient selection and study design

The present study was conducted in full accordance with the guidelines of the Declaration of Helsinki, revised in 2000 in Edinburgh. Approvals for the study were obtained from the Ethics Committee of University of Health Sciences Türkiye, Adana City Training and Research Hospital and the Turkish Ministry of Health (28.03.2024/3232). The requirement for informed consent from the patients was waived due to the retrospective nature of the study. Of the 262 cases diagnosed during this period, 109 female patients who underwent both preoperative DCE breast MRI and ^{18}F-FDG PET/CT at both diagnosis and after the completion of NAC were included in the study, following the application of the exclusion criteria (Figure 1). All patients underwent clinical tumor staging via baseline clinical examination, mammography, and ultrasonography along with a tru-cut biopsy. Both ^{18}F-FDG PET/CT and cranial MRI examinations were performed for tumor–node–metastasis staging. To avoid the partial volume effect during measurement of ^{18}F-FDG PET/CT uptake, all lesions in the study had a long axis >15 mm.

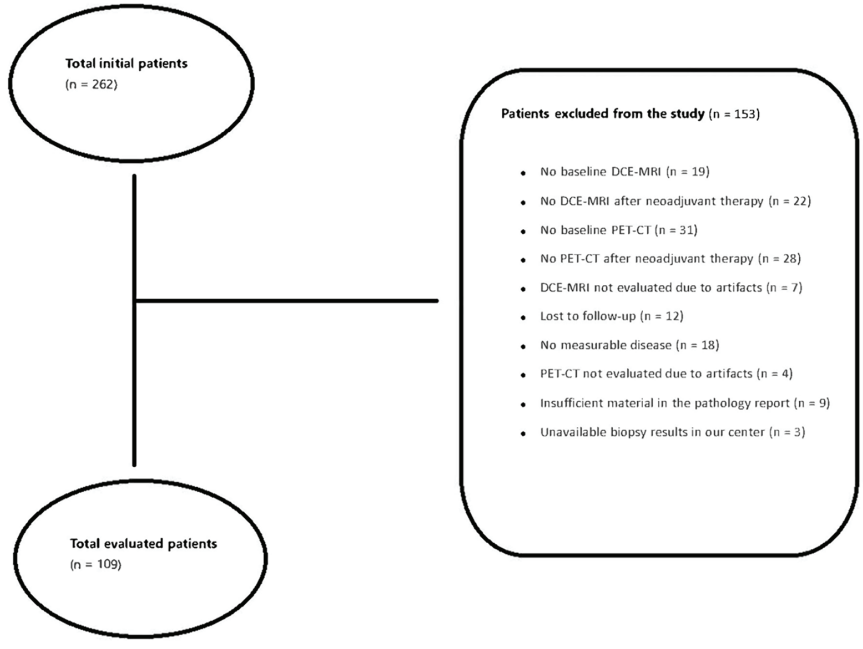


Figure 1. Flowchart of the study. DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging; PET-CT, positron emission tomography-computed tomography.

Semi-quantitative parameters from DCE-MRI at the time of diagnosis were recorded. Subsequently, semi-quantitative DCE-MRI parameters obtained after NAC were documented, including (1) maximum enhancement, (2) maximum relative enhancement, (3) T0 (s), (4) time to peak (s), (5) wash-in rate (s^{-1}), and (6) wash-out rate (s^{-1}). Differences between the measurements obtained at diagnosis and those recorded after NAC were calculated. Additionally, the SUV_{max}, SUV_{mean}, TLG, and MTV values obtained from ^{18}F-FDG PET/CT at diagnosis and after NAC were recorded, and the Δs between these parameters were calculated. The percentage changes (Δ%) for all ^{18}F-FDG PET/CT and MRI parameters were calculated using the following formula:

Percentage change (Δ%) = (delayed parameter − baseline parameter) / baseline parameter × 100.

Magnetic resonance imaging acquisition

The MRI protocols for this study adhered to the standards established by the breast MRI accreditation program of the American College of Radiology.[12] A three-dimensional T1-weighted fast gradient echo-based DCE series with one pre-contrast and three sequential post-gadolinium contrast-enhanced sequences, each with a scan duration of approximately 3 minutes, was imaged. The examinations were conducted using a 3-Tesla scanner (Ingenuity; Philips Healthcare, the Netherlands) with a 16-channel breast coil. The DCE-MRI was performed using a fat-saturated three-dimensional fast gradient echo sequence with the following parameters: repetition time/echo time, 5.8/3 ms; flip angle, 15°; spatial resolution, 0.5 × 0.6 × 1.4 mm; voxel size, 0.8 × 0.8 × 2; temporal resolution, 60 s; number of sections, 200; and field of view, 360 × 360 mm. Fat suppression was achieved via the short-scar periareolar inferior pedicle reduction technique. All images were acquired in axial orientation. The contrast agent used was gadobutrol (Gadovist, Bayer Schering Pharmaceuticals, Mississauga, Canada) administered at a dose of 0.1 mmol/kg body weight and a rate of 2 cc/s, followed by a 15 cc saline flush.

Magnetic resonance imaging evaluation

All obtained DCE-MRI were retrospectively reviewed based on consensus by two experienced radiologists with 9 and 7 years of experience in breast MRI. The radiologists were blinded to clinicopathological findings and clinical follow-up data except for data on the presence of breast cancer. All images obtained using the software were analyzed. Regions of interest were defined as areas demonstrating abnormal signals on the MRI and were manually delineated using the oval-shaped function (Figure 2). Time–intensity curves and time-to-peak values were automatically generated based on these ROIs.

In accordance with previously published methodologies, multiple circular ROIs measuring 25 mm^2 (approximately 5.6 mm in diameter) were carefully placed within the solid portion of the tumor, as indicated by DCE-MRI and T2-weighted imaging.[2,7,9] The ROIs were meticulously positioned within the tumor to avoid cystic regions or visual artifacts, and average values within each ROI were recorded. These analyses were performed using a dedicated workstation (IntelliSpace, Philips, Netherlands).

Acquisition of flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography

Patients were placed in the supine position during imaging, and CT transmission scanning was performed without intravenous contrast enhancement. The scan utilized a low tube current (135 kVp, 46–79 mAs), a slice thickness of 4.0 mm, a gantry rotation of 0.53 s, and a collimator width of 5 × 3 mm. Following a minimum fasting period of 6 hours, an intravenous injection of 370–555 MBq ^{18}F-FDG was administered. Whole-body images were acquired using a PET/CT scanner (Ingenuity; Philips Healthcare, the Netherlands) 40–60 minutes after injection.

Visual analysis employed a four-point certainty scoring system (definitely negative, equivocal: probably negative, equivocal: probably positive, and definitely positive) and included assessments of anatomical site and lesion size. The emission data were acquired for 2.5 min per bed (six to seven beds), which were later attenuation-corrected with the digital CT data. Image reconstruction used the ordered subsets expectation maximization algorithm of two iterations and eight subsets.

Flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography evaluation

All ^{18}F-FDG PET/CT images were retrospectively evaluated by a nuclear medicine specialist with 17 years of experience in PET-CT. This specialist was blinded to clinicopathological and clinical follow-up data that did not cover the presence of breast cancer. The foci of FDG uptake were identified if FDG accumulation exceeded that of comparable normal contralateral or surrounding tissues. The workstation automatically generated SUV_{max}, SUV_{mean}, and MTV (cm^3) from the volume of interest (VOI) (Figure 3). The target lesion margins inside the VOI were also automatically defined, and voxels exceeding 41% of the SUV_{max} within the VOI were included in the MTV and SUV_{mean} calculations. Among the relative thresholds of 40% SUV_{max} for tumor delineation and contouring on PET/CT in patients with breast cancer, it is the most appropriate and popular cut-off value and prognostic marker.[1-4]

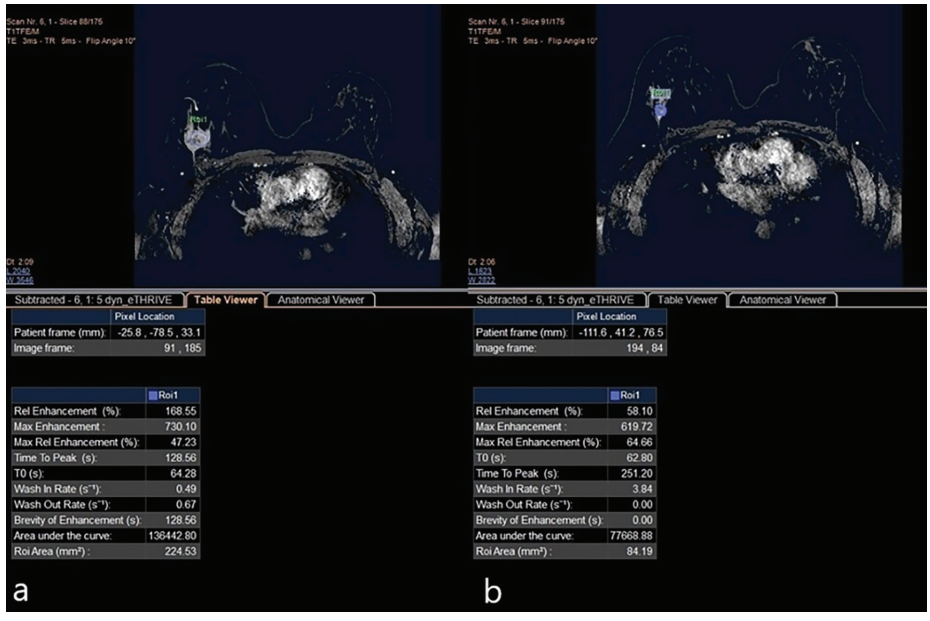


Figure 2. Changes in dynamic contrast-enhanced magnetic resonance imaging baseline **(a)** and post-neoadjuvant therapy **(b)** parameters of the same patient. Although there was a significant change in all dynamic parameters in this patient, the pathology result was a partial response.

The TLG was calculated by multiplying MTV by SUV_{mean}. Due to the study's retrospective nature, semi-quantitative measurements were retrospectively reproduced by the nuclear medicine specialist taking part in the study, who was blinded to all other data.

Chemotherapy regimen and histological evaluation

The NAC regimens included combinations such as doxorubicin and cyclophosphamide, doxorubicin and cyclophosphamide followed by taxane, doxorubicin and docetaxel, and taxane combined with anti-HER2 agents.

The pathological characteristics of the tumors examined in this work, including their histological type, histological grade, Ki67 proliferation index, hormone receptor expression, and HER2 status, were analyzed based on biopsies obtained at diagnosis prior to NAC. Immunohistochemical (IHC) analysis was used to determine whether the tumor tissue was positive or negative for HER2 and hormone receptors [estrogen receptor (ER) and progesterone receptor (PR)]. For hormone receptor expression positivity, ≥10% was accepted. The HER2 expression results were grouped as negative, 1+, 2+, or 3+. Then, the fluorescence *in situ* hybridization (FISH) method was used for gene amplification in cases with HER2 IHC staining score 2+. Cases with an IHC staining score of 3+ were defined as HER2 positive, or in the case of an IHC staging score of 2+, FISH positive.The Elston and Ellis[13] criteria (grades 1, 2, and 3) were used to determine histological grades. The IHC study of the tumor tissue obtained positive or negative results for ER and PR and stratified these results according to the modified Allred scale.[14] Histological responses to antibodies in the breast were assessed using Miller–Payne grading.[15] Pathologic complete response (pCR) is defined as the complete disappearance of cancer cells on histopathologic examination of mastectomy material after neoadjuvant treatment. Since the presence of residual ductal carcinoma *in situ* (DCIS) has prognostic significance and affects the comparability of the results with other studies, in this study, the pCR group was considered to be the absence of any tumor cells, including DCIS. The patients were categorized into five subgroups based on the recommendations of the 12th International Breast Conference: (1) luminal A; (2) luminal B-HER2(–); (3) luminal B-HER2(+); (4) HER2(+); (5) triple-negative.[16] The threshold for Ki67 positivity was 14%.

Statistical analysis

Statistical analysis was conducted using SPSS version 23.0 statistical software (SPSS, Chicago, IL, USA). Categorical variables were summarized as counts and percentages, and continuous variables were presented as means and standard deviations or medians (Q1–Q3) where applicable. Comparisons of categorical variables were conducted using the chi-squared test. The Kolmogorov–Smirnov test was applied to assess the normality of parameter distributions. For non-normally distributed parameters, the Mann–Whitney U test was employed. The sensitivity and specificity of the relevant parameters in predicting patient response to treatment were calculated. Moreover, the area under the receiver operating characteristic curve (AUC) was analyzed to determine the cut-off values. A P value of <0.05 was considered statistically significant in all tests.

Results

The rate of patients with histological grade 3 was found to be higher in those exhibiting a partial response than those achieving pCR ($P = 0.030$). No significant differences were observed between the treatment groups in relation to the remaining parameters ($P > 0.05$) (Table 1).

In patients with pCR, baseline MTV, Δ(%)MTV, Δ(%)SUV_{mean}, and Δ(%)TLG values were significantly higher ($P = 0.033$; $P = 0.014$; $P < 0.001$; and $P = 0.014$, respectively), whereas post-NAC SUV_{mean} values were significantly lower than those with a partial response ($P = 0.003$) (Table 2).

Tumor size and/or changes in tumor size were not associated with response to NAC. However, patients who achieved pCR exhibited significantly higher Δ(%)wash-out values ($P = 0.024$). No significant differences were observed between the two groups in terms of the remaining DCE-MRI parameters (Table 3).

Baseline MTV (AUC: 62.5%), post-NAC SUV_{mean} (AUC: 67.3%), ΔMTV (AUC: 64.5%), ΔSUV_{mean} (AUC: 72.4%), ΔTLG (AUC: 64.4%), and the DCE-MRI parameter Δwash-out (AUC: 63.5%) were found to be statistically significant predictors of treatment response ($P < 0.05$). The ΔSUV_{mean} parameter demonstrat-

Table 1. Comparison of demographic and pathological findings between the two groups

	Partial response (n = 72) n (%)	pCR (n = 37) n (%)	*P* value
Age (years), median (range)	47 (25–70)	53 (34–72)	0.345
Lymphovascular invasion	28 (38.9)	11 (29.7)	0.345
Histological grade			
1	34 (47.2)	12 (32.4)	**0.030***
2	22 (30.6)	22 (62.2)	
3	16 (22.2)	3 (5.4)	
CERB2			
0	12 (16.7)	9 (24.3)	0.355
1	12 (16.7)	6 (16.2)	
2	37 (51.4)	13 (35.1)	
3	11 (15.3)	9 (24.3)	
High Ki67 expression	45 (62.5)	20 (54.1)	0.395
ER	54 (75)	30 (81.1)	0.475
PR	44 (61.1)	23 (62.2)	0.915
HER2	25 (34.7)	13 (35.1)	0.966
Luminal A	15 (20.8)	10 (27.0)	0.801
Luminal B HER2-negative	26 (36.1)	12 (32.4)	0.677
Luminal B HER2-positive	18 (25.0)	11 (29.7)	0.597
HER2-positive	7 (9.7)	2 (5.4)	0.438
Triple-negative	6 (8.3)	2 (5.4)	0.579

*$P < 0.05$, chi-squared test. pCR, pathological complete response; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

ed the best diagnostic test performance at predicting treatment response among these parameters, with a cut-off value of >30.5 and an AUC of 72.4 (Table 4).

The MTV, SUV_{mean}, ΔMTV, ΔSUV_{mean}, ΔSUV_{mean}, ΔTLG, and Δwash-out values after neoadjuvant treatment and the distribution between biological subtypes were similar (P = 0.837; P = 0.325; P = 0.897; P = 0.664; P = 0.782; P = 0.425, respectively) (Figure 4).

Discussion

The response to NAC is the most significant parameter for predicting breast cancer prognosis.[17,18] Therefore, the ability to predict the response to NAC can be considered equivalent to anticipating the course and prognosis of the disease.[19] DCE breast MRI and ^{18}F-FDG PET/CT are the two most utilized imaging modalities for patients with breast cancer worldwide. Therefore, the current study aimed to identify the most effective parameter among these two imaging techniques for predicting the response to NAC.

Although SUV is a widely accepted parameter in clinical practice, SUV_{max} remains controversial.[20] This is because SUV_{max} reflects the single-pixel value of the maximum intensity of ^{18}F-FDG activity rather than the entire tumor mass. Furthermore, SUV_{max} is affected by body weight.[21] In the current study, no SUV_{max} value (baseline, post-NAC, or Δ) successfully predicted the response to NAC. On the other hand, SUV_{mean}, as an average measurement, avoids these limitations. Accordingly, ΔSUV_{mean} and post-NAC SUV_{mean} were determined to be the most valuable ^{18}F-FDG PET/CT parameters for predicting histopathological response in this study.

By adjusting for lean body mass, MTV reflects metabolic activity across the entire tumor mass.[22,23] In addition, MTV appears to be correlated with total tumor burden and tumor aggressiveness in various solid tumors, thus providing useful prognostic information.[24] Both TLG and MTV are volumetric parameters used to estimate total radioactivity within a tumor above a set threshold.[25] They have both proven to be useful in assessing the therapeutic effects and prognosis of chemotherapy.[26] The size of the tumor is also included in the calculation of TLG and MTV. It is not surprising that TLG and MTV increase with increasing tumor size, as found in some recent studies. In addition, some studies have shown that poorly differentiated tumors have significantly higher TLG and larger MTV than well-differentiated tumors.[27,28] In this study, both ΔMTV and baseline MTV predicted response to NAC, whereas MTV after

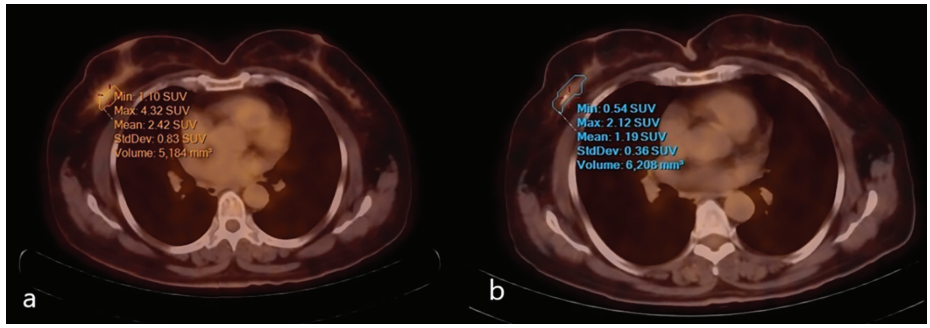


Figure 3. Changes in flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography parameters at **(a)** baseline and **(b)** after neoadjuvant treatment of the same patient. This patient, whose ΔSUV_{max}, ΔSUV_{mean} changes were evident, had a pathologic complete response according to pathology. SUV, standardized uptake value.

Table 2. Comparison of flourine-18 fluorodeoxyglucose positron emission tomography/computed tomography parameters and Δ(%) values between the two groups

	Partial response (n = 72)	pCR (n = 37)	*P* value
	Median (Q1–Q3)	Median (Q1–Q3)	
Baseline MTV (cc)	155 (84–254)	19.2 (11.9–44.8)	**0.033***
Post-NAC MTV (cc)	387 (305–532)	4.25 (2.68–6.4)	0.659
ΔMTV (%)	683 (541–835)	82.3 (67–90.8)	**0.014***
Baseline SUV_{max}	742 (5.21–122)	8.19 (5.42–22.69)	0.206
Post-NAC SUV_{max}	328 (2.98–452)	3.69 (2.78–4.98)	0.554
ΔSUV_{max} (%)	517 (295–663)	59.8 (37.7–77.8)	0.192
Baseline SUV_{mean}	3.89 (3.1–7.27)	4.15 (3.1–9.19)	0.481
Post-NAC SUV_{mean}	3.05 (2.45–3.65)	2.17 (1.73–3.18)	**0.003****
ΔSUV_{mean} (%)	26.4 (19.6–43.2)	48.9 (32.6–64.9)	**<0.001****
Baseline TLG	61.8 (29.4–154.3)	117.9 (41.2–293.4)	0.051
Post-NAC TLG	11.5 (7.42–18.73)	11.9 (6.36–29.7)	0.687
ΔTLG (%)	79.1 (65.4–90.3)	88.2 (75.2–94.1)	**0.014***

*P < 0.05, **P < 0.01, Mann–Whitney U test. MTV, metabolic tumor volume; SUV, standardized uptake value; TLG, total lesion glycolysis; pCR, pathologic complete response; NAC, neoadjuvant chemotherapy.

Table 3. Comparison of dynamic contrast-enhanced magnetic resonance imaging parameters and Δ(%) values between the two groups

	Partial response (n = 72)	pCR (n = 37)	*P* value
	Median (Q1–Q3)	Med (Q1–Q3)	
Baseline tumor size on MRI (mm)	36 (29–45)	36 (22–52)	0.773
Post-NAC tumor size on MRI (mm)	19 (14–26)	20 (14–27)	0.875
ΔTumor size on MRI (mm)	45.1 (27.8–60)	47.6 (31–62.4)	0.569
Baseline maximum enhancement	963 (786–1080)	952 (711–1023)	0.113
Post-NAC maximum enhancement	442 (214–768)	475 (254–672)	0.766
ΔMaximum enhancement (%)	55.4 (18.6–77.9)	41.8 (14.6–71.4)	0.768
Baseline maximum reel enhancement	51 (44–58)	52 (41–59)	0.820
Post-NAC maximum reel enhancement	32 (12–50)	29 (12–46)	0.883
ΔMaximum reel enhancement	37.2 (2.5–74.5)	25.4 (11.5–76.9)	0.825
Baseline TTP	139 (124–163)	154 (124–188)	0.355
Post-NAC TTP	168 (140–187)	174 (144–192)	0.238
ΔTTP (%)	-13.9 (-40.3–11)	-3.03 (-46.9–13.5)	0.908
Baseline wash-in	5.86 (4.05–7.88)	5.56 (3.75–7.44)	0.957
Post-NAC wash-in	3.42 (1.69–6.2)	3.71 (2.3–4.67)	0.683
ΔWash-in (%)	23.4 (-5.6–68.3)	38.7 (17.8–60.9)	0.754
Baseline wash-out	0.55 (0.23–1.58)	0.67 (0.36–1.84)	0.129
Post-NAC wash-out	0.3 (0.18–1.1)	0.25 (0.14–0.54)	0.118
ΔWash-out (%)	35.19 (-91.5–66.4)	41.67 (31.8–75.0)	**0.024***

*P < 0.05, ** P < 0.01, Mann-Whitney U test. pCR, pathologic complete response; TTP, time to peak; NAC, neoadjuvant chemotherapy; MRI, magnetic resonance imaging.

Table 4. Receiver operating characteristic analysis of parameters predictive of pathological response to neoadjuvant therapy

	AUC (95% CI)	Cut-off	Sensitivity	Specificity	LR+	LR–	PPV	NPV
Baseline MTV	0.625 (0.527–0.717)	>9.92	86.49	35.21	0.75	2.61	41	83.3
Post-NAC SUV_{mean}	0.673 (0.577–0.760)	≤2.32	56.76	84.72	3.71	0.51	65.6	79.2
ΔMTV	0.645 (0.547–0.734)	>70.2	72.97	54.93	1.62	0.49	45.8	79.6
ΔSUV_{mean}	0.724 (0.630–0.805)	>30.5	81.08	56.94	1.88	0.33	49.2	85.4
ΔTLG	0.644 (0.546–0.734)	>74.6	81.08	45.07	1.48	0.42	43.5	82.1
ΔWash-out	0.635 (0.536–0.726)	>-9.52	100	38.89	1.64	0	44.3	100

< positive diagnostic test for values under the optimal cut-off > positive diagnostic test for values above the optimal cut-off, *P < 0.05. AUC, area under the curve; CI, confidence interval; LR+, positive likelihood ratio; LR–, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value; MTV, metabolic tumor volume; TLG, total lesion glycolysis; NAC, neoadjuvant chemotherapy; SUV, standardized uptake value.

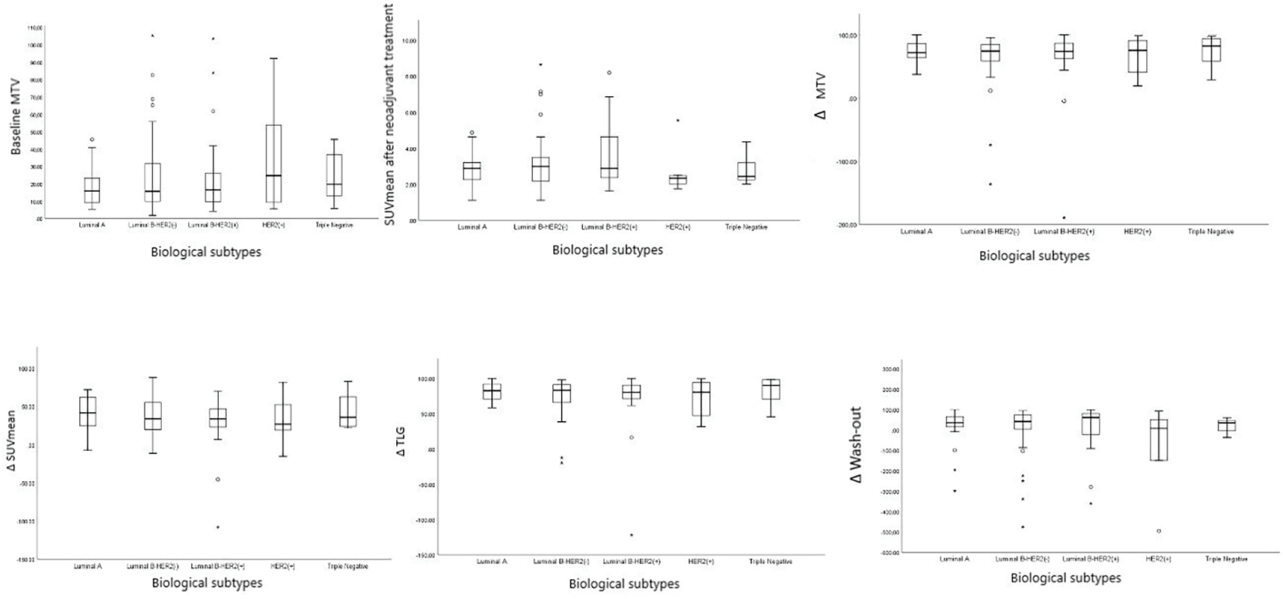


Figure 4. Baseline metabolic tumor volume, SUV_{mean} after neoadjuvant treatment, ΔMTV, ΔSUV_{mean}, ΔTLG, Δwash-out in relation to currently used biological subtypes. SUV, standardized uptake value.

NAC was not significantly associated with treatment response. In this study, the superior value of SUV_{mean} compared with MTV and TLG likely stems from its independence from tumor size, as tumor size was found to be unrelated to NAC response. The study's independent evaluation of ^{18}F-FDG PET/CT parameters and overall analysis determined ΔSUV_{mean} to be the most effective parameter for predicting NAC response.

The finding that ΔSUV_{mean} is the most significant parameter, both within the ^{18}F-FDG PET/CT parameters and overall, demonstrates that changes in tumor biology (i.e., Δ values) are the most effective predictors of response to NAC. These variations are better represented by the differences between baseline assessments and post-NAC evaluations, particularly in ΔSUV_{mean} and ΔMTV, which can predict pCR and, consequently, influence surgical procedures.

Invasive lobular carcinoma (ILC) also differs from invasive ductal carcinoma in terms of receptor expression.[29] In particular, ILC is almost always (95%) ER-positive, raising the possibility of increased utility of ER-targeting PET tracers for patients with ILC.[29] Sixteen alpha ^{18}F-fluoroestradiol is an ER-targeting PET tracer with high sensitivity and specificity for the detection of ER-positive tumors. It can be used as a predictive biomarker to demonstrate ER heterogeneity, evaluate the pharmacokinetics of ER-targeted agents, quantify residual ER during endocrine therapy, and determine the biological optimal dose of novel ER-targeted drugs.

One factor that may influence the response to systemic chemotherapy is tumor perfusion. Tumor tissue with relatively poor perfusion is believed to be less exposed to chemotherapeutic agents and therefore may not respond adequately to NAC.[29,30] This deficiency could be a factor contributing to poor responses to intravenous chemotherapy. Prior studies on large breast tumors using various techniques, including histologic analysis of vessel density, PET with ^{15}O-water, DCE-MRI, single-photon imaging with 99mTc-sestamibi, and Doppler sonography, all suggest that blood flow in LABC is highly variable.[31-33] Quantitative DCE-MRI, which assumes two chambers representing the extravascular extracellular space and blood plasma, provides absolute and thus objective measures of perfusion.[3,20,21] However, among the DCE-MRI parameters evaluated both at baseline and post-NAC in this study, only Δwash-out was able to predict the treatment response. Nevertheless, the specificity of Δwash-out was found to be significantly lower than the parameters of ^{18}F-FDG PET/CT.

Many factors may influence ^{18}F-FDG uptake into tumor cells, including the up-regulation of glucose transporters and hexokinase enzymes and neo-angiogenesis, which in turn are related to the number of viable tumor cells, their aggressiveness, and proliferative activity.[1,7,15] Angiogenesis is one of the most important of these, and one of the biggest effects of NAC drugs is that they af-

fect tumor perfusion; that is, angiogenesis. However, as shown in this study, the disruption of angiogenesis is only one of the factors affecting the tumor's pathological response. For this reason, DCE MRI parameters, whose main mechanism of action is to show tumor perfusion, show considerable variability due to the disruption of angiogenesis after neoadjuvant treatment, but do not show this pathological response. However, the main mechanism of action of PET-CT parameters is metabolic activity. Metabolic activity is higher in tumors with high perfusion/angiogenesis; however, its effect on metabolic activity, which is the main indicator of PET-CT parameters, is limited. Therefore, the rate of being affected by perfusion disorders caused by NAC drugs is limited.

Previous studies have suggested that advanced T and N stages and ER-negative status are associated with high SUV_{max} values in patients with breast cancer. Although lymph node response was not examined in this study, these parameters, which are effective in predicting the response of mass to neoadjuvant treatment, were found to be unaffected by biological subtypes.

This study had several limitations. First, the current literature describes DCE-MRI parameters using k_{trans}, v_e, and k_{ep} calculations. The authors were unable to use these parameters, since no application capable of calculating these values is available in the hospital; this can be regarded as a limitation concerning the integration of the study with existing literature. Second, FDG PET/CT was performed in the supine position, and no additional imaging was performed specifically for the position of the breasts. This may have been a misleading factor, especially in parameters such as MTV and TLG, where tumor volumes are also calculated. Some recent studies have examined the effect of breast position on PET-CT results.[34] Furthermore, the small sample size due to stringent exclusion criteria and the retrospective nature of the study were additional limitations.

In conclusion, in patients with LABC, ^{18}F-FDG PET/CT performed both at baseline and post-NAC demonstrated the metabolic changes induced by treatment. The ΔSUV_{mean} parameter is the most successful predictor of pathological response to NAC among all PET/CT parameters.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Hyun SH, Ahn HK, Park YH, et al. Volume-based metabolic tumor response to neoadjuvant chemotherapy is associated with an increased risk of recurrence in breast cancer. *Radiology*. 2015;275(1):235-244. [Crossref]
2. Diwanji D, Onishi N, Hathi DK, et al. ^{18}F-FDG dedicated breast PET complementary to breast MRI for evaluating early response to neoadjuvant chemotherapy. *Radiol Imaging Cancer*. 2024;6(2):230082. [Crossref]
3. Akin M, Orguc S, Aras F, Kandiloglu AR. Molecular subtypes of invasive breast cancer: correlation between PET/computed tomography and MRI findings. *Nucl Med Commun*. 2020;41(8):810-816. [Crossref]
4. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA*. 2019;321(3):288-300. [Crossref]
5. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752. [Crossref]
6. Korde LA, Somerfield MR, Carey LA, et al. Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO guideline. *J Clin Oncol*. 2021;39(13):1485-1505. [Crossref]
7. Reig B, Lewin AA, Du L, et al. Breast MRI for evaluation of response to neoadjuvant therapy. *Radiographics*. 2021;41(3):665-679. [Crossref]
8. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164-72. Erratum in: *Lancet*. 2019;393(10175):986. [Crossref]
9. Pace L, Nicolai E, Luongo A, et al. Comparison of whole-body PET/CT and PET/MRI in breast cancer patients: lesion detection and quantitation of ^{18}F-deoxyglucose uptake in lesions and in normal organ tissues. *Eur J Radiol*. 2014;83(2):289-296. [Crossref]
10. Pak K, Seok JW, Kim HY, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis in breast cancer: a meta-analysis. *Nucl Med Commun*. 2020;41(8):824-829. [Crossref]
11. Yousefirizi F, Klyuzhin IS, O JH, et al. TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images - a multi-center generalizability analysis. *Eur J Nucl Med Mol Imaging*. 2024;51(7):1937-1954. [Crossref]
12. American College of Radiology Breast Accreditation Program: Modalities. [Crossref]
13. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19:403-410. [Crossref]
14. Allred DC, Harvey JM, Berardo M, Clark GM. Prognosis and predictive factors in breast cancer by inmunohistochemical analysis. *Mod Pathol*. 1998;11:155-168. [Crossref]
15. Wang L, Luo R, Lu Q, et al. Miller-Payne grading and 70-gene signature are associated with prognosis of hormone receptor-positive, human epidermal growth factor receptor 2-negative early-stage breast cancer after neoadjuvant chemotherapy. *Front Oncol*. 2021;11:735670. [Crossref]
16. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ; Panel members. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011;22(8):1736-1747. [Crossref]
17. Mghanga FP, Lan X, Bakari KH, Li C, Zhang Y. Fluorine-18 fluorodeoxyglucose positron emission tomography-computed tomography in monitoring the response of breast cancer to neoadjuvant chemotherapy: a meta-analysis. *Clin Breast Cancer*. 2013;13(4):271-279. [Crossref]
18. Jiménez-Ballvé A, García García-Esquinas M, Salsidua-Arroyo O, et al. Prognostic value of metabolic tumour volume and total lesion glycolysis in ^{18}F-FDG PET/CT scans in locally advanced breast cancer staging. *Rev Esp Med Nucl Imagen Mol*. 2016;35(6):365-372. [Crossref]
19. Arici S, Karyagar SS, Karyagar S, et al. The predictive role of metabolic tumor volume on no response to neoadjuvant chemotherapy in patients with breast cancer. *J Oncol Pharm Pract*. 2020;26(6):1415-1420. [Crossref]
20. Choi EK, Yoo IR, Kim SH, Park SY, O JH, Kang BJ. The value of pre- and post-neoadjuvant chemotherapy F-18 FDG PET/CT scans in breast cancer: comparison with MRI. *Acta Radiol*. 2018;59(1):41-49. [Crossref]
21. Nguyen-Thu H, Hanaoka H, Nakajima T, et al. Early prediction of triple negative breast cancer response to cisplatin treatment using diffusion-weighted MRI and ^{18}F-FDG-PET. *Breast Cancer*. 2018;25(3):334-342. [Crossref]
22. Gerke O, Naghavi-Behzad M, Nygaard ST, et al. Diagnosing bone metastases in breast cancer: a systematic review and network meta-analysis on diagnostic test accuracy studies of 2-[^{18}F]FDG-PET/CT, ^{18}F-NaF-PET/CT, MRI, contrast-enhanced CT, and bone scintigraphy. *Semin Nucl Med*. 2025;55(1):137-151. [Crossref]
23. Kitajima K, Miyoshi Y, Yamano T, Odawara S, Higuchi T, Yamakado K. Prognostic value of FDG-PET and DWI in breast cancer. *Ann Nucl Med*. 2018;32(1):44-53. [Crossref]
24. Lee SM, Bae SK, Kim TH, et al. Value of 18F-FDG PET/CT for early prediction of pathologic response (by residual cancer burden criteria) of locally advanced breast cancer to neoadjuvant chemotherapy. *Clin Nucl Med*. 2014;39(10):882-886. [Crossref]

25. An YS, Kang DK, Jung Y, Kim TH. Volume-based metabolic parameter of breast cancer on preoperative 18F-FDG PET/CT could predict axillary lymph node metastasis. *Medicine (Baltimore)*. 2017;96(45):8557. [Crossref]

26. Capobianco N, Meignan M, Cottereau AS, et al. Deep-learning ^{18}F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-Cell lymphoma. *J Nucl Med*. 2021;62(1):30-36. [Crossref]

27. Wen W, Xuan D, Hu Y, Li X, Liu L, Xu D. Prognostic value of maximum standard uptake value, metabolic tumor volume, and total lesion glycolysis of positron emission tomography/computed tomography in patients with breast cancer: a systematic review and meta-analysis. *PLoS One*. 2019;14(12):0225959. [Crossref]

28. El-Galaly TC, Villa D, Cheah CY, Gormsen LC. Pre-treatment total metabolic tumour volumes in lymphoma: does quantity matter? *Br J Haematol*. 2022;197(2):139-155. [Crossref]

29. Ulaner GA, Jhaveri K, Chandarlapaty S, et al. Head-to-head evaluation of 18F-FES and 18F-FDG PET/CT in metastatic invasive lobular breast cancer. *J Nucl Med*. 2021;62(3):326-331. [Crossref]

30. Vercellino L, Cottereau AS, Casasnovas O, et al. High total metabolic tumor volume at baseline predicts survival independent of response to therapy. *Blood*. 2020;135(16):1396-1405. [Crossref]

31. Chen W, Zhu L, Yu X, Fu Q, Xu W, Wang P. Quantitative assessment of metabolic tumor burden in molecular subtypes of primary breast cancer with FDG PET/CT. *Diagn Interv Radiol*. 2018 Nov;24(6):336-341. [Crossref]

32. Im HJ, Bradshaw T, Solaiyappan M, Cho SY. Current methods to define metabolic tumor volume in positron emission tomography: which one is better? *Nucl Med Mol Imaging*. 2018;52(1):5-15. [Crossref]

33. Higuchi T, Fujimoto Y, Ozawa H, et al. Significance of metabolic tumor volume at baseline and reduction of mean standardized uptake value in ^{18}F-FDG-PET/CT imaging for predicting pathological complete response in breast cancers treated with preoperative chemotherapy. *Ann Surg Oncol*. 2019;26(7):2175-2183. [Crossref]

34. Kupik O, Tuncel M, Özgen Kıratlı P, et al. Value of dynamic 18F-FDG PET/CT in predicting the success of neoadjuvant chemotherapy in patients with locally advanced breast cancer: a prospective study. *Mol Imaging Radionucl Ther*. 2023;32(2):94-102. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253578

BREAST IMAGING

LETTER TO THE EDITOR

Letter to the Editor: Perinodal signal in breast magnetic resonance imaging: flare sign and extracapsular spread

Deniz Esin Tekcan Şanlı[1]
Ahmet Necati Şanlı[2]

[1]Gaziantep University Faculty of Medicine, Department of Radiology, Gaziantep, Türkiye

[2]Abdulkadir Yüksel State Hospital, Clinic of General Surgery, Gaziantep, Türkiye

Dear Editor,

We read with great interest the study by Özgür et al.[1] entitled "Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging," which proposes the "flare sign" as a promising imaging marker for extracapsular extension (ECE) on breast magnetic resonance imaging (MRI). The authors should be acknowledged for systematically characterizing this novel radiological feature and evaluating its predictive accuracy in a pathologically confirmed cohort of axillary metastases.

The concept of T2 hyperintense perinodal signal has been sporadically discussed in previous studies under various terminologies, such as "perinodal edema" or "perifocal signal abnormality." For example, Baltzer et al.[2] described perifocal edema on MRI as a specific but low-sensitivity indicator of nodal malignancy, while Kim et al.[3] reported strand-like or circumferential T2 signal increases around axillary nodes but without histopathological ECE correlation. Moreover, in head and neck cancer imaging, Hiyama et al.[4] employed the term "flare sign" to describe perinodal spread, suggesting its potential utility in evaluating extranodal invasion. However, to the best of our knowledge, the study by Özgür et al.[1] represents the first application of the "flare sign" as a specifically defined, histopathologically validated diagnostic feature of ECE in metastatic axillary lymph nodes on breast MRI.

Despite this strength, we believe several methodological aspects warrant further clarification. First, the average interval of 48 days (range: 4–92) between MRI and surgery raises concerns regarding the temporal stability of imaging findings, especially for dynamic features, such as perinodal edema. Importantly, perinodal and peritumoral edema are more commonly observed in biologically aggressive tumors, such as triple-negative and human epidermal growth factor receptor 2 (HER2)-positive subtypes, or in high-grade lesions.[5] Therefore, this extended preoperative interval may have introduced bias due to the rapid progression of nodal infiltration or signal changes, particularly in these aggressive phenotypes. Furthermore, while the authors presented molecular subtype distributions (e.g., HER2-positive, triple-negative), they did not investigate the association between the flare sign and tumor biology. Histopathologic or molecular subtype stratification (e.g., ductal vs. lobular carcinoma) may have elucidated differential patterns of perinodal signal, especially considering that invasive lobular carcinomas are known to spread differently and more diffusely.[6]

Second, the study lacks interobserver agreement metrics (e.g., kappa statistics), which are essential to establish the reproducibility of flare sign interpretation.

Third, although the authors excluded patients who received neoadjuvant therapy, they did not report whether other perinodal pathologies (e.g., extranodal tumor deposits, extranodal vascular emboli) were assessed histologically, factors that may contribute to T2 signal elevation and potentially confound the flare sign.[7,8]

Additionally, the omission of ECE extent (e.g., ≤2 vs. >2 mm), which has shown prognostic and therapeutic relevance,[9] limits further stratification of risk. Finally, although the authors state that 35 patients underwent sentinel lymph node biopsy only, they do not elaborate on

Corresponding author: Deniz Esin Tekcan Şanlı

E-mail: tekcandenizesin@gmail.com

Received 23 July 2025; accepted 28 July 2025.

Epub: 29.09.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253578

You may cite this article as: Tekcan Şanlı DE, Şanlı AN. Letter to the editor: perinodal signal in breast magnetic resonance imaging: flare sign and extracapsular spread. *Diagn Interv Radiol*. 2026;32(3):287-288.

how the known false-negative rate of sentinel lymph node biopsy (approximately 8.3%) may have impacted the detection of ECE in this subset.

Furthermore, the generalizability of the findings is limited, as the study was conducted at a single institution using only a 1.5T MRI system. External validation across multiple centers and imaging platforms is essential to confirm the reproducibility of the flare sign as a reliable biomarker.

Finally, the imaging analysis was restricted to T2-weighted sequences; dynamic contrast-enhanced MRI parameters, such as peak enhancement, washout kinetics, or time–intensity curve types, were not included. A multiparametric MRI approach could provide a more comprehensive assessment and improve diagnostic performance by integrating functional and morphological information.

In conclusion, Özgür et al.[1] provide compelling preliminary evidence that the flare sign could serve as a high-specificity biomarker for ECE, with potential implications for surgical planning and axillary staging. Future prospective multicenter studies incorporating volumetric ECE assessment, multiparametric MRI features, and interobserver reliability analyses would be valuable to confirm and extend these findings.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Özgür C, Sunal BS, Hereklioğlu S, Öznur M, Özkan Gürdal S. Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging. *Diagn Interv Radiol*. 2025;31(4):333-339. [Crossref]
2. Baltzer PA, Dietzel M, Burmeister HP, et al. Application of MR mammography beyond local staging: is there a potential to accurately assess axillary lymph nodes? evaluation of an extended protocol in an initial prospective study. *AJR Am J Roentgenol*. 2011;196(5):W641-647. [Crossref]
3. Kim WH, Kim HJ, Park CS, et al. Axillary Nodal Burden Assessed with Pretreatment Breast MRI is associated with failed sentinel lymph node identification after neoadjuvant chemotherapy for breast cancer. *Radiology*. 2020;295(2):275-282. [Crossref]
4. Hiyama T, Kuno H, Nagaki T, et al. Extra-nodal extension in head and neck cancer: how radiologists can help staging and treatment planning. *Jpn J Radiol*. 2020;38(6):489-506. [Crossref]
5. Park NJ, Jeong JY, Park JY, et al. Peritumoral edema in breast cancer at preoperative MRI: an interpretative study with histopathological review toward understanding tumor microenvironment. *Sci Rep*. 2021;11(1):12992. [Crossref]
6. Limaiem F, Khan M, Lotfollahzadeh S. Lobular breast carcinoma. 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. [Crossref]
7. Yang YS, Feng F, Qiu YJ, Zheng GH, Ge YQ, Wang YT. High-resolution MRI-based radiomics analysis to predict lymph node metastasis and tumor deposits respectively in rectal cancer. *Abdom Radiol (NY)*. 2021;46(3):873-884. [Crossref]
8. Xu Q, Xu Y, Wang J, Sun H, Lin J, Xie S. Distinguishing mesorectal tumor deposits from metastatic lymph nodes by using diffusion-weighted and dynamic contrast-enhanced magnetic resonance imaging in rectal cancer. *Eur Radiol*. 2023;33(6):4127-4137. [Crossref]
9. Mamtani A, Barrio AV, Goldman DA, Wen HY, Vincent A, Morrow M. Extranodal tumor deposits in the axillary fat ındicate the need for axillary dissection among T1-T2cN0 patients with positive sentinel nodes. *Ann Surg Oncol*. 2020;27(10):3585-3592. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253598

BREAST IMAGING

LETTER TO THE EDITOR

Reply: Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging

Cihan Özgür[1]
Baran Serdar Sunal[1]
Savaş Hereklioğlu[1]
Meltem Öznur[2]

[1]Tekirdağ Namık Kemal University Faculty of Medicine, Department of Radiology, Tekirdağ, Türkiye

[2]Tekirdağ Namık Kemal University Faculty of Medicine, Department of Pathology, Tekirdağ, Türkiye

Dear Editor,

We would like to thank the authors for their interest in our study entitled "Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging".[1] We appreciate their thoughtful comments[2] and the opportunity to respond.

Our study was designed as a preliminary one focusing on the radiologic–pathologic correlation of extracapsular extension (ECE) using breast magnetic resonance imaging (MRI). In our study, we evaluated the flare sign as a practical imaging marker that could be interpreted easily in routine clinical settings, independent of advanced multiparametric analysis. Rather than analyzing extensive dynamic or functional parameters, our aim was to assess whether a single, morphologic feature—interpreted globally across the axilla—could predict ECE with acceptable diagnostic performance.

Regarding the patient population, it is important to clarify that the frequent use of neoadjuvant chemotherapy (NAC) in locally advanced, triple-negative and HER-2 positive breast cancers substantially influenced the composition of our cohort. In our institutional experience, patients referred for NAC exhibited pronounced flare signs and/or direct extracapsular spread on MRI. However, we made a deliberate methodological decision to exclude patients who received NAC, as NAC is known to significantly alter both the imaging and histopathologic appearance of nodal structures (Figure 1). Including such patients would have introduced bias and obscured the natural imaging–pathology relationship.

As shown in our study flow diagram, the majority of excluded cases had received NAC. Furthermore, the strong tendency to refer patients with triple-negative and HER-2 positive breast cancers for NAC resulted in an inhomogeneous distribution across molecular subtypes, making meaningful group comparisons unfeasible.

Although we did not assess inter-reader variability statistically, all imaging was interpreted in consensus by two experienced radiologists who were blinded to histopathologic results. This methodology is consistent with previous studies in the literature, such as those by Kimura et al.[3] and Bode et al.[4], both of which adopted consensus reading protocols in similar diagnostic settings.

We believe that these design choices were essential to preserve internal consistency and to enable a clear and reliable radiologic–pathologic correlation. We acknowledge that the variability in the MRI-to-surgery interval and the absence of ECE size categorization are among the main limitations of our study. Further prospective research involving multiparametric imaging and molecular subtype analysis will be essential to confirm and extend these early observations.

Corresponding author: Cihan Özgür

E-mail: cihanozguritf@gmail.com

Received 05 August 2025; accepted 12 August 2025.

Epub: 29.09.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253598

You may cite this article as: Özgür C, Sunal BS, Hereklioğlu S, Öznur M. Reply: diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging. *Diagn Interv Radiol*. 2026;32(3):289-290.

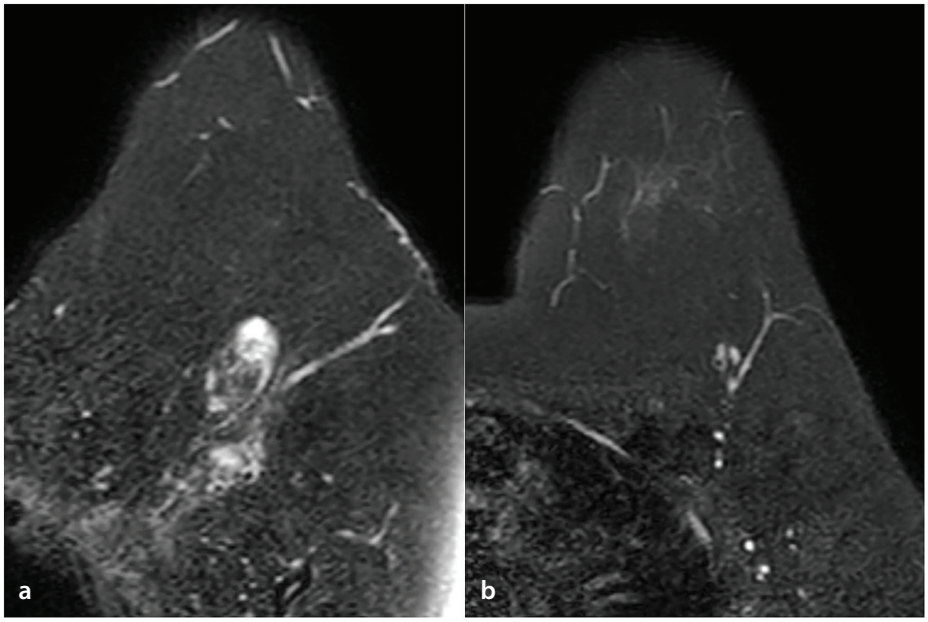


Figure 1. Axillary lymph node response to neoadjuvant chemotherapy in a patient with invasive ductal carcinoma. **(a)** Pre-treatment axial fat-suppressed T2-weighted breast MRI demonstrates multiple axillary lymphadenopathies with perinodal T2 hyperintensity (flare sign). **(b)** Post-treatment image shows complete nodal radiologic response, with resolution of lymphadenopathies and disappearance of the flare sign. Pathologic staging following mastectomy revealed ypT1a, ypN0 disease with residual cancer burden class II. Four axillary lymph nodes showed reactive hyperplasia without metastasis. Pathologic confirmation of possible extracapsular extension was not feasible due to complete nodal response. This case is presented for illustrative purposes and was not part of the analyzed study cohort. MRI, magnetic resonance imaging.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Özgür C, Sunal BS, Hereklioğlu S, Öznur M, Özkan Gürdal S. Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging. *Diagn Interv Radiol*. 2025;31(4):333-339. [Crossref]
2. Şanlı DET, Şanlı AN. Letter to the editor: perinodal signal in breast magnetic resonance imaging: flare sign and extracapsular spread. *Diagn Interv Radiol*. 2025. [Crossref]
3. Kimura Y, Sumi M, Sakihama N, Tanaka F, Takahashi H, Nakamura T. MR imaging criteria for the prediction of extranodal spread of metastatic cancer in the neck. *AJNR Am J Neuroradiol*. 2008;29(7): 1355-1359. [Crossref]
4. Bode M, Schrading S, Masoumi A, Morscheid S, Schacht S, Dirrichs T, Gaisa N, Stickeler E, Kuhl CK. Abbreviated MRI for comprehensive regional lymph node staging during pre-operative breast MRI. *Cancers (Basel)*. 2023;15(6):1859. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253249

CHEST IMAGING

INVITED REVIEW

Imaging in chronic thromboembolic pulmonary hypertension: review of the current literature

Koray Hekimoglu[1]
Deepa Gopalan[2]
Mehmet Ruhi Onur[3]
Gökhan Kahraman[3]
Tankut Akay[4]

[1]TOBB University of Economics and Technology, Department of Radiology, Ankara, Türkiye

[2]Imperial College Healthcare NHS Trust, Department of Radiology, United Kingdom, England

[3]Hacettepe University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

[4]Başkent University Faculty of Medicine, Department of Cardiovascular Surgery, Ankara, Türkiye

ABSTRACT

Chronic thromboembolic pulmonary hypertension (CTEPH) is a severe, life-threatening complication of pulmonary embolism with pulmonary hypertension (PH). The combination of insufficient resolution of thrombi following pulmonary emboli and accompanying microvascular disease results in PH. Advances in imaging can offer better insight into CTEPH diagnosis and management, but lack of disease awareness among radiologists has been shown to be a cause of CTEPH misdiagnosis or delayed diagnosis. This review highlights features pertinent to CTEPH diagnosis. The primary focus is on different modalities with their distinctive signs and newly developed technologies employing artificial intelligence systems.

KEYWORDS

Chronic thromboembolic pulmonary diseases, chronic thromboembolic pulmonary hypertension , pulmonary embolism, pulmonary hypertension, pulmonary angiography, computed tomography, dual-energy computed tomography, magnetic resonance imaging, four-dimensional magnetic resonance imaging, artificial intelligence

Chronic thromboembolic pulmonary hypertension (CTEPH) is characterized by incomplete resolution of thrombus in the pulmonary arteries. It originates from deep veins in the lower limbs and is associated with pulmonary hypertension (PH). The accumulation of residual chronic thrombus in the pulmonary arteries triggers fibrotic proliferation and secondary microvasculopathy, leading to an elevation in pulmonary arterial pressure and pulmonary vascular resistance (PVR).[1] The majority of thrombosis resolves following acute pulmonary thromboembolism (PE), but almost 5% of cases persist as chronic PE.[2] Known and unknown factors are responsible for incomplete clot resolution. Abnormal fibrinogen structure and function, inhibition of thrombus angiogenesis, and inflammatory thrombosis are the main factors triggering microvasculopathy.[3] There are also certain risk factors for CTEPH, including antiphospholipid syndrome and elevated factor VIII, splenectomy, chronic inflammatory processes, vascular shunts, infections, cancer, and non-O blood groups.[4]

As per the 2022 European Society of Cardiology/European Respiratory Society (ESC/ERS) PH guidelines, CTEPH, placed in Group 4 PH, is a form of precapillary PH that is hemodynamically characterized by mean pulmonary arterial pressure (mPAP) measured at rest >20 mmHg, PVR >2 Wood units, and pulmonary artery wedge pressure <15 mmHg.[3] In the presence of comorbidities, mPAP can exceed 15 mmHg. These parameters should be supported with imaging modalities to establish an accurate diagnostic approach. Right heart catheterization (RHC) is therefore an integral component in CTEPH evaluation (Graphic 1).

CTEPH is defined by the presence of perfusion defects at least 3 months after anticoagulation following acute PE. Abnormal vascular organization and disorganized vascular remodeling lead to characteristic abnormalities in pulmonary arteries, including eccentric clots, stenosis, webs, bands, pruning, and pouch defects. Increased arterial tortuosity is also a feature of CTEPH.[5] A cohort of patients with chronic PE without PH is in the "Chronic Thromboembolic Pulmonary Disease" category.[3,6]

Corresponding author: Koray Hekimoglu

E-mail: korayhekim@gmail.com

Received 20 May 2025; revision requested 11 June 2025; last revision received 11 August 2025; accepted 02 September 2025.

Epub: 23.09.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253249

You may cite this article as: Hekimoglu K, Gopalan D, Onur MR, Kahraman G, Akay T. Imaging in chronic thromboembolic pulmonary hypertension: review of the current literature. *Diagn Interv Radiol*. 2026;32(3):291-303.

The diagnosis of CTEPH can be challenging even with contemporary diagnostic modalities.

Although it may be a relatively uncommon disease, the diagnosis is often underestimated due to non-specific signs and symptoms as well as a lack of awareness among the radiological community.[7]

The initial period of CTEPH, commencing approximately 3 months post-acute PE and extending into the first year, is characterized as the "honeymoon period." Right ventricular hypertrophy and interventricular septal bowing may initially compensate for rising pulmonary pressures. During this phase, the pulmonary arteries may not attain critical values for mPAP and PVR. Prolonged elevation of mPAP measured at rest >30 mmHg leads to cor pulmonale and right heart failure, with low survival.[8]

A distinctive feature of CTEPH that sets it apart from other PH groups is the potential for cure with surgical pulmonary endarterectomy (PEA). There is generally a dramatic response to PEA surgery, with a rapid and substantial decrease in mPAP and permanent improvement of pulmonary hemodynamics.

Methods

This review was conducted to synthesize and critically evaluate current imaging techniques and emerging technologies in the diagnosis and management of CTEPH. A systematic search was conducted across the PubMed, Scopus, and Google Scholar Databases, focusing on the literature published between January 2000 and April 2025. The search terms included "CTEPH," "dual-energy computed tomography," "photon-counting CT," "4D MRI,". "cardiac MRI in PH," "Artificial Intelligence in pulmonary hypertension," and "pulmonary endarterectomy." Studies were chosen based on their clinical significance, methodological quality, and relevance to current imaging practices. Inclusion criteria encompassed peer-reviewed English-language articles, clinical trials, reviews, and guideline documents. Exclusion criteria included case reports, non-peer-reviewed abstracts, and non-English texts. Major sections of the manuscript are structured thematically according to modality, with integration of the recent ESC/ERS guidelines (2022) and expert consensus recommendations. All authors independently performed the literature screening, and any possible disagreement was resolved by consensus.

Main points

- Chronic thromboembolic pulmonary hypertension (CTEPH) should be excluded in all patients presenting with pulmonary hypertension.
- CTEPH can be easily overlooked on imaging; hence, a high degree of suspicion is needed during imaging review.
- In challenging cases, novel diagnostic methodologies such as dual-energy computed tomography can be instrumental in achieving an accurate diagnosis.
- Consulting with CTEPH expert centers is highly beneficial for making informed decisions.

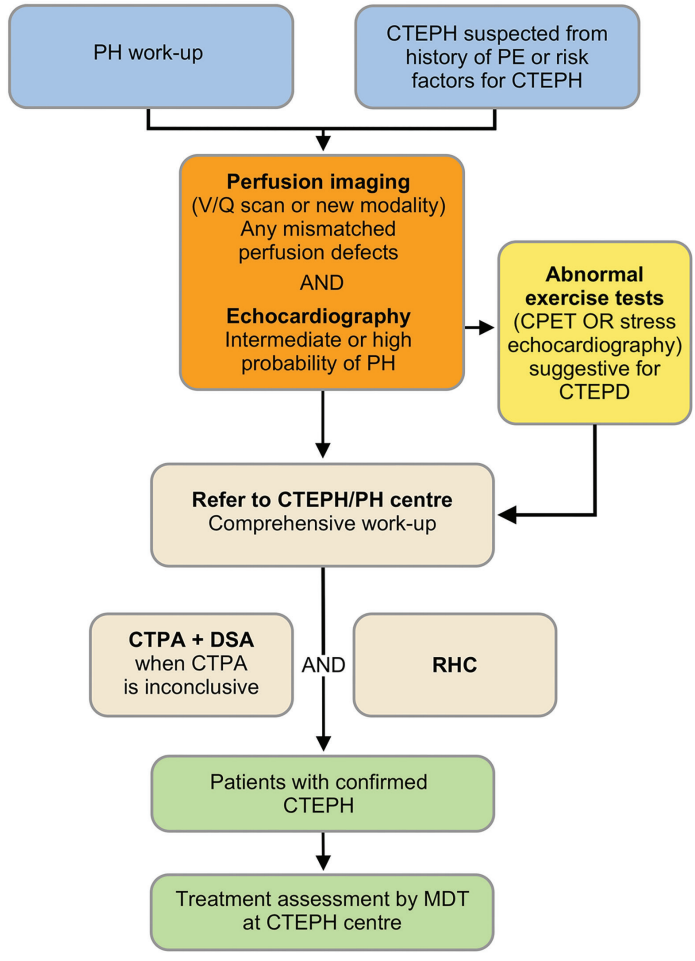
Diagnostic approach to chronic thromboembolic pulmonary hypertension

Ventilation/perfusion scan

Ventilation/perfusion (V/Q) scintigraphy is the widely accepted screening modality for excluding CTEPH. Guidelines still affirm a V/Q scan as a first step in the diagnostic process in CTEPH. Abnormal perfusion with preserved V/Q mismatch is suspicious for CTEPH, but as perfusion defects can also be seen in acute PE, large vessel vasculitis, pulmonary arterial malignancy, veno-occlusive disease, and vascular stenosis,[9] it is necessary to perform anatomical imaging to confirm CTEPH. Single-photon emission computed tomography (SPECT) imaging is preferred over planar scintigraphy due to its multiplanar imaging capability and better diagnostic capacity (Figures 1 and 2). Low radiation dose exposure is a major advantage of V/Q scintigraphy. It has been shown that approximately 27–136-times less radiation exposure is required than with other diagnostic modalities using X-ray sources.[10]

Computed tomography

Computed tomography (CT) pulmonary angiography (CTPA) plays a central role in



Graphic 1. Chronic thromboembolic pulmonary hypertension diagnosis algorithm adapted from the 2022 European Society of Cardiology/European Respiratory Society guidelines.[12] PH, pulmonary hypertension; CTEPH, chronic thromboembolic pulmonary hypertension; V/Q, ventilation/perfusion; CPET, chronic pulmonary hypertension; DSA, digital subtraction angiography; RHC, right heart catheterization; MDT, CTEPH Multidisciplinary Team.

the diagnosis of both acute thromboembolic pulmonary disease and CTEPH. In the setting of acute pulmonary embolism, CTPA is the first-line imaging modality recommended by the 2019 ESC/ERS guidelines, providing rapid and accurate identification of thrombi in the pulmonary arteries. In acute PE, CTPA not only allows evaluation of pulmonary arterial patency but also provides prognostic information through detection of right ventricular strain [e.g., right ventricle (RV)/left ventricle (LV) ratio, septal bowing] and extra-vascular findings. In the chronic setting, according to the 2022 ESC/ERS guidelines, echocardiography is the recommended first-line diagnostic modality (recommendation IB), whereas CTPA is included in the diagnostic algorithm for CTEPH (recommendation IC). The CTPA technique is a valuable anatomical tool in the assessment of CTEPH, particularly for visualizing eccentric wall-adherent thrombi, vessel stenosis, webs, and bands. Right ventricular remodeling is a key feature that develops as a consequence of long-standing pressure overload. Remodeling manifests as right ventricular hypertrophy due to increased afterload, followed by progressive dilatation in advanced stages, which ultimately leads to impaired contractility and right heart failure. The CTPA method can indirectly demonstrate these changes through findings such as increased RV free wall thickness, RV dilation, interventricular septal flattening or bowing toward the LV, and enlargement of the right atrium and inferior vena cava (IVC). These features, when combined with vascular abnormalities, strongly support the diagnosis of CTEPH and provide valuable prognostic information.[11,12]

To provide optimal evaluation of the pulmonary and cardiovascular system, CTPA should be performed in the pulmonary arterial phase. Structured evaluation of CT includes assessment of vasculature, cardiac chambers, and lung parenchyma. Although high-quality CTPA can effectively demonstrate proximal CTEPH, it is still possible to miss distal CTEPH with conventional CTPA. One of the major limitations of CTPA in the context of CTEPH is its reduced sensitivity in evaluating subsegmental pulmonary arteries, primarily due to vessel tapering and suboptimal opacification in the distal pulmonary vasculature. These technical limitations can lead to underestimation of the distal disease burden, particularly in chronic thromboembolic disease affecting small-caliber vessels.

In addition to assessing the extent of the clot burden, CTPA offers key advantages in evaluating associated underlying lung and/or cardiac diseases, as well as other potential causes of acute chest pain.[13]

Besides diagnosis, CTPA also has the potential to predict patient outcomes. Pulmonary arterial occlusion exceeding 30% considerably elevates vascular resistance, leading to PH and subsequent elevated right ventricular afterload as well as an increased RV/LV ratio.[14] It is crucial to recognize that a substantial clot burden in the pulmonary bed poses a severe risk for right ventricular and cardiac dysfunction. During every CTPA evaluation process, the pulmonary trunk and main branch diameters and the RV/LV diameter ratio should be measured to determine

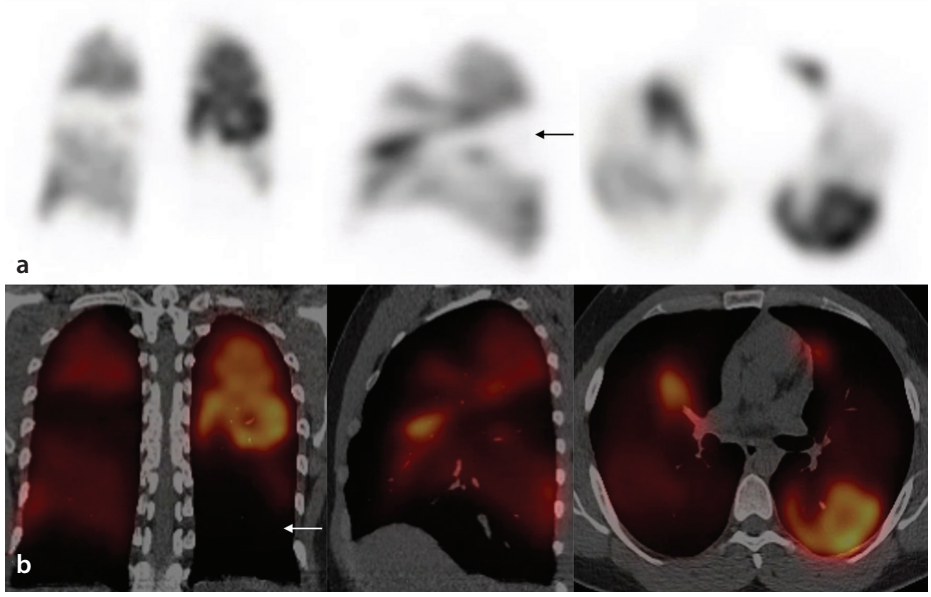


Figure 1. Selected images of coronal, sagittal, and axial single-photon emission computed tomography (SPECT) perfusion (**a**, top panel) and corresponding fusion images of SPECT perfusion and CT (**b**, bottom panel). Multiple bilateral lobar perfusion defects (white arrow) and segmental (black arrow) distribution in a 45-year-old man with chronic thromboembolic pulmonary hypertension.

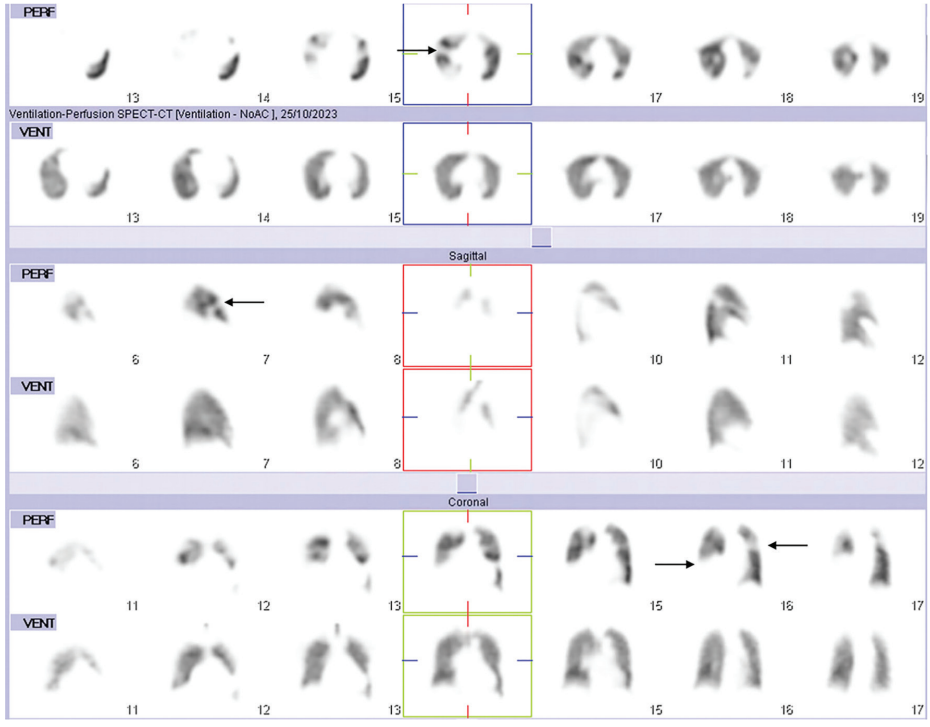


Figure 2. Selected images of coronal, sagittal, and axial perfusion (top) and ventilation (bottom) from a ventilation/perfusion single-photon emission computed tomography series in a 40-year-old woman with chronic thromboembolic pulmonary hypertension. Multiple bilateral mismatched perfusion defects (black arrows) are observed.

whether PH is present. A main pulmonary trunk diameter >29 mm, an RV/LV ratio >0.9 on a four-chamber view, and leftward septal bowing are findings that indicate a high sensitivity for predicting PH on CTPA. These measurements should be interpreted in the context of CTEPH (Figures 3 and 4).[2,6,15] In addition, CTEPH is often associated with secondary changes in right heart morphology, including right ventricular hypertrophy, right atrial enlargement, and IVC dilation features, that reflect chronic pressure overload.

Besides these PH findings, obtuse angle-shaped, eccentrically positioned thrombus; asymmetric pruning or cut-off pulmonary artery branches; irregular linear filling defects, webs, or bands; and compensatory bronchial artery hypertrophy are all signs of CTEPH (Table 1, Figure 5). Generally, bronchial arteries originate from the systemic arterial circulation, most commonly from the descending thoracic aorta. Hypertrophy of the bronchial arteries, which often arise from the aorta or systemic vessels, represents compensatory collateral circulation in CTEPH. These enlarged bronchial arteries may contribute to retrograde systemic-to-pulmonary flow. During PEA, this retrograde flow poses a substantial risk of intraoperative bleeding, making recognition of bronchial artery hypertrophy on imaging crucial for surgical planning. Pulmonary artery calcification, although relatively uncommon, may be observed in patients with chronic, long-standing CTEPH and is considered a marker of disease chronicity.[1]

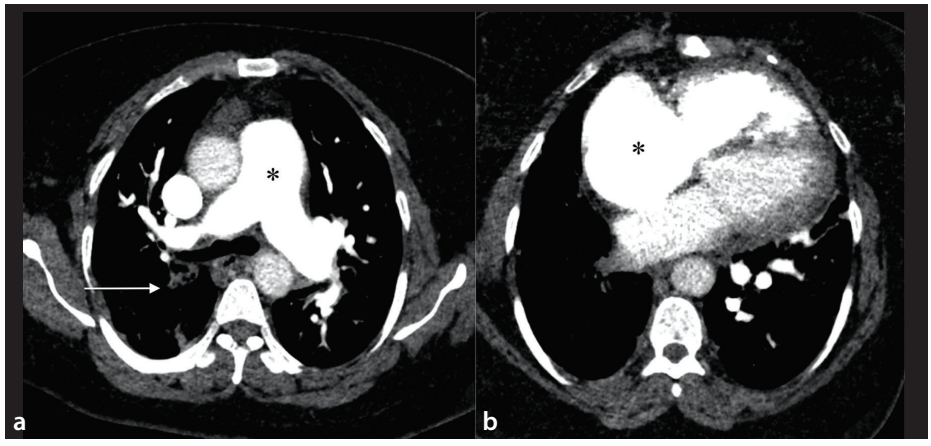


Figure 3. Axial computed tomography pulmonary angiography images. **(a)** Marked dilatation of the main pulmonary artery (star), exceeding the diameter of the adjacent ascending aorta due to pulmonary hypertension. Absence of visible vascular structures in the right lower lobe, suggestive of segmental occlusion due to chronic thromboembolic disease (white arrow). **(b)** Cardiomegaly with prominent right atrial and ventricular enlargement in pulmonary hypertension (star).

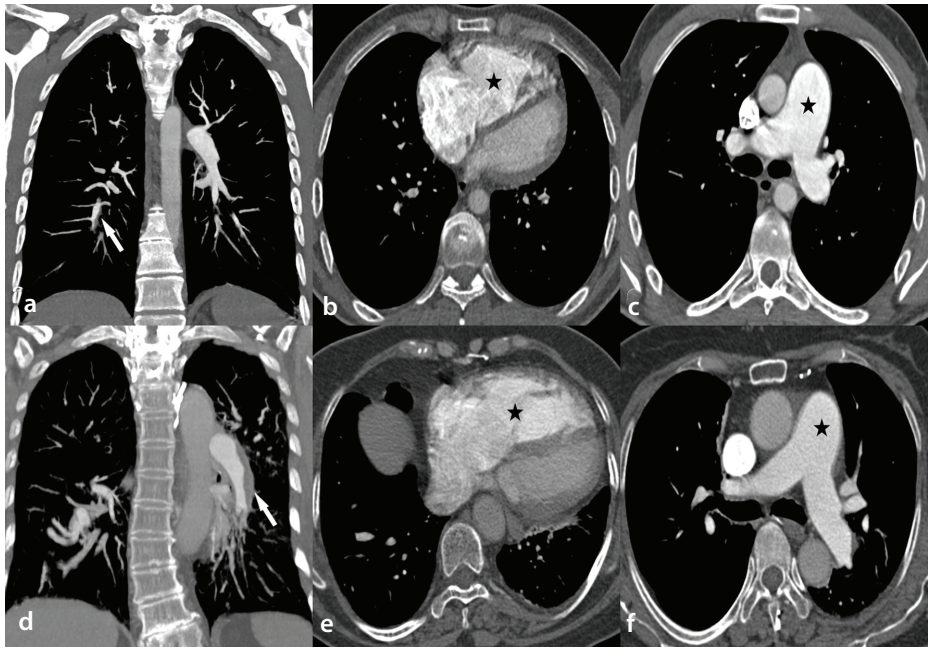


Figure 4. Computed tomography pulmonary angiography (CTPA) findings in patients with chronic thromboembolic pulmonary hypertension. Top row **(a-c)**: patient 1. **(a)** Coronal reformatted CTPA image demonstrating chronic intraluminal thrombi within a segmental pulmonary artery (arrow); **(b)** axial image showing right ventricular dilatation secondary to pressure overload (star); **(c)** axial image illustrating main pulmonary artery enlargement consistent with pulmonary hypertension (star). Bottom row **(d-f)**: patient 2. **(d)** Coronal reformatted CTPA image demonstrating chronic intraluminal thrombi (arrow); **(e)** axial image showing right ventricular dilatation (star); **(f)** axial image illustrating main pulmonary artery enlargement (star).

It is crucial to acknowledge that a portion of patients with CTEPH may also have concurrent acute PE. Consequently, CTPA can effectively reveal the presence of acute–subacute and chronic thrombi combinations and associated complications.[16]

Although CTPA is a foundational imaging modality in diagnosing acute pulmonary embolism and is also pivotal in evaluating CTEPH, its reliance on ionizing radiation necessitates strict adherence to the "as low as reasonably achievable" principle to balance diagnostic benefits with patient safety. Practical strategies for dose reduction include justification of each CT scan, automatic tube current modulation, lowering kilovolt (kV) peak, adjusting pitch, delaying scan initiation to limit unnecessary exposure, and optimizing scan range.[17] Additionally, the use of iodinated contrast media entails a risk of contrast-induced acute kidney injury, particularly in patients with pre-existing renal dysfunction. Pre-hydration protocols and contrast dose minimization are advised in at-risk groups.[18]

In the imaging evaluation of CTEPH, standardized CTPA protocols are crucial for ensuring optimal diagnostic accuracy and reproducibility. Acquisition should be conducted using low-kV settings (typically 80–140 kV), selected based on body habitus or through automatic tube voltage selection systems to improve vascular contrast and reduce radiation dose. The tube current should be automatically modulated to maintain consistent image quality across varying anatomical thicknesses. Short gantry rotation times are critical to reduce motion artifacts, particularly in patients who are unstable or dyspneic. The use of thin collimation and 1-mm slice thickness reconstructions is recommended to ensure high spatial resolution. Scanning may be performed in either craniocaudal or caudocranial directions; the latter is preferable in patients with respiratory compromise to reduce basal motion artifacts. Although deep inspiratory breath-hold remains the

Table 1. Imaging findings in acute versus chronic thromboembolic disease and imaging modalities

Imaging finding	Imaging modality
Acute	
Centrally located thrombus	CTPA, CPA
Acute angle with vessel wall	CTPA, CPA
Polo-mint sign, floating	CTPA, CPA
Convex margins	CTPA, CPA
Preserved or increased caliber	CTPA, CPA
Parenchymal infarcts, reverse halo sign	Chest CT
Chronic	
Total occlusion— "cut-off" sign	CTPA, CPA
Obtuse angle with vessel wall	CTPA, CPA
Linear filling defects, bands, webs	CTPA, CPA
Concave margins	CTPA, CPA
Eccentric thrombus abutting the vessel wall	CTPA, CPA
Asymmetric pruning/decreased vessel caliber	CTPA, CPA
Bronchial artery hypertrophy	CTPA, CPA
Mosaic attenuation pattern	Dual-energy CTPA, V/Q scan

Note: Both CTPA and Chest CT refer to standard protocols as well as their dual-energy CT implementations. CTPA, computed tomography pulmonary angiography; Chest CT, thorax computed tomography; CPA, conventional pulmonary angiography; V/Q, ventilation/perfusion.

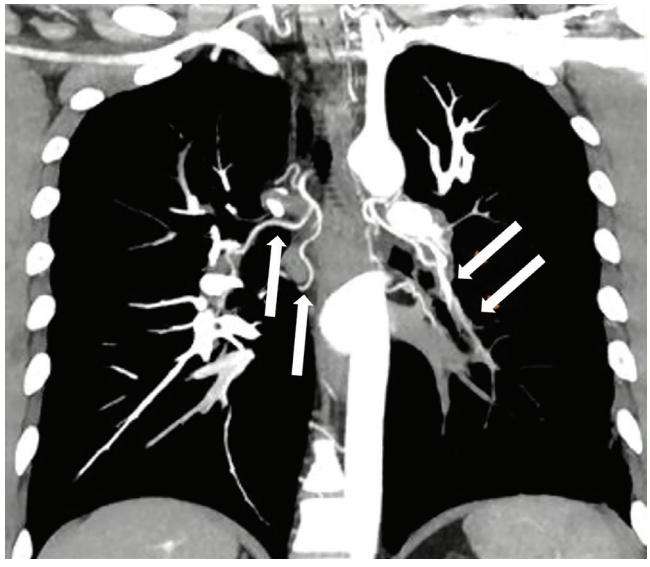


Figure 5. Coronal reformatted computed tomography pulmonary angiography image in chronic thromboembolic pulmonary hypertension. Segmental occlusion of the left lower lobe pulmonary artery is evident. Segmental occlusions are also present in the right lung. Marked bronchial artery hypertrophy (white arrows) is best observed on the coronal maximum intensity projection image.

preferred respiratory maneuver, shallow breathing or breath-holding at vital capacity may be necessary for symptomatic patients.

Intravenous contrast administration should be tailored to patient size, with a preferred flow rate of 3–5 mL/s and contrast volumes ranging from 80 to 100 mL, using agents with iodine concentrations of 300–370 mg/mL. Bolus triggering, ideally with region-of-interest placement in the pulmonary artery or ascending aorta, ensures optimal timing of image acquisition. For comprehensive assessment in suspected CTEPH, the use of a generalized chest CT angiography protocol that includes simultaneous opacification of both pulmonary and systemic arteries is recommended. This approach is not a conventional triple-rule-out protocol aimed at coronary artery evaluation; rather, the acquisition is optimized in patients with CTEPH to achieve simultaneous opacification of the pulmonary arteries and systemic arterial supply, allowing for evaluation of bronchial artery hypertrophy and systemic-to-pulmonary collaterals. Recognizing these systemic collaterals is crucial for both diagnosis and surgical planning in CTEPH. Advanced techniques such as dual-energy (DE) CT or photon-counting (PC) CT offer additional perfusion mapping capabilities, highlighting perfusion defects not easily seen on standard morphological imaging. Subtraction imaging may also be employed when pre-contrast data is available, particularly for assessing iodine distribution in the lung parenchyma. In special populations such as pregnant patients, modifications include the use of low tube voltage, caudocranial scanning, and enhanced contrast injection protocols to counteract increased blood volume and cardiac output while minimizing radiation exposure, particularly to the fetus. The application of model-based iterative or deep-learning reconstruction techniques is strongly encouraged to reduce image noise without increasing the dose.[7]

A potential pitfall in CTPA interpretation is transient interruption of contrast (TIC). This occurs when deep inspiration or abrupt intrathoracic pressure changes during injection lead to an admixture of unopacified IVC blood with contrast-opacified superior vena cava blood, which produces heterogeneous opacification of the pulmonary arteries and may mimic true filling defects. In addition to TIC, several other artifacts may complicate image interpretation. Respiratory motion can generate pseudo-filling defects, and beam hardening from dense contrast in the superior vena cava or subclavian veins may obscure adjacent pulmonary arteries. Inadequate contrast timing may result in insufficient pulmonary arterial enhancement, rendering the study non-diagnostic, and partial volume averaging—particularly adjacent to calcified vessel walls—may falsely appear as intraluminal thrombus. Furthermore, in patients with right heart dysfunction, delayed or heterogeneous contrast distribution can create flow-related artifacts that resemble vascular obstruction. Awareness of these pitfalls and the application of optimized scanning protocols—including proper bolus tracking, patient instruction, and thin-slice reconstructions—are crucial to avoid misinterpretation and ensure accurate diagnosis.[19]

Dual-energy computed tomography

Dual-energy CT (DECT) lung perfusion has gained widespread recognition as a valuable tool for identifying perfusion abnormalities.

Iodine maps can increase diagnostic confidence for CTEPH identification, and DECT perfusion imaging has been shown to reach close to 100% sensitivity and specificity for the diagnosis of CTEPH when combined with CTPA. The perfusion defects are typically wedge-shaped and located peripherally, distinguishing them from the more diffuse or central perfusion abnormalities often observed in conditions such as emphysema;[20] DECT perfusion defects provide pixel-wise quantification of the spatial area of hypoperfusion.[21] The presence of perfusion defects in cases where there is no morphological evidence of chronic thromboembolism should raise suspicion of distal CTEPH (Figures 6 and 7).[22]

Furthermore, DECT scanners possess an additional capability for contrast media reduction through virtual monoenergetic imaging (VMI). The reconstruction of images at low energies (40–50 keV) with VMI application enhances vascular opacity, enabling the utilization of low-contrast media on DECT. Mosaic perfusion of the lungs is an integral feature of CTEPH, characterized by geographical areas of hypoperfused and normal or hyperperfused areas (Figure 8). Mosaic attenuation is not specific to CTEPH and may also be seen in small airway diseases, such as chronic obstructive pulmonary disease (COPD), hypersensitivity pneumonitis, and constrictive bronchiolitis. In CTEPH, the vascular etiology is suggested by a paucity of vessels in the hypoperfused zones, whereas in airway diseases, the affected areas often demonstrate air trapping on expiratory CT, which may be subtle in COPD but is usually more conspicuous in hypersensitivity pneumonitis or constrictive bronchiolitis. Recognizing these differences is essential in the differential diagnosis of mosaic attenuation patterns. It should also be noted that similar imaging findings—though typically less pronounced—may be encountered in bronchial obstructive pathologies, particularly in advanced stages of emphysema. Other lung abnormalities in CTEPH include parenchymal infarcts and related sequelae.

Photon-counting computed tomography

Photon-counting CT (PCCT) with novel detector technology permits ultra-high resolution (UHR) imaging of the lungs and pulmonary vasculature. This enables the visualization of subtle abnormalities at the microcirculatory level, providing detailed morphological insights and differentiating chronic thromboembolic lesions from artifacts. The PCCT method offers key radiation dose reductions while maintaining or improving image quality.[23] In CTEPH, the subpleural region of the lungs is the most affected area, characterized by major perfusion defects due to the presence of small-vessel disease. Poor subpleural perfusion can be precisely assessed using the UHR capability of PCCT, whereas CTEPH microvasculopathy, characterized by thin arterioles, venules, and ill-defined micronodules in the subpleural region of the lungs, which is not detectable with other imaging modalities, could be identified using PCCT.[24]

Additionally, four-dimensional (4D)-CT perfusion imaging with PCCT allows dynamic visualization of blood flow, enabling quantification of time-dependent perfusion parameters. This method provides complementary

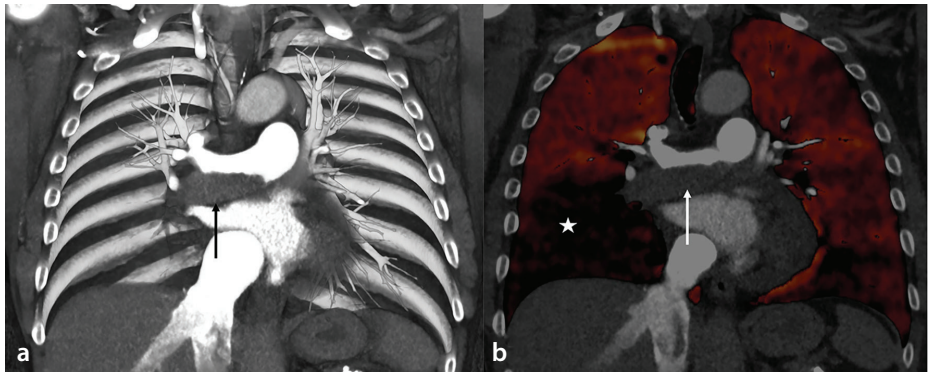


Figure 6. (a) Coronal computed tomography pulmonary angiography shows eccentric thrombus in the right main pulmonary artery (black arrow) and occlusion of the right lower lobe vessels. **(b)** Dual-energy iodine map demonstrates a corresponding lobar perfusion defect in the right lung (star). Perfusion is also abnormal in the left lung due to subsegmental thrombi.

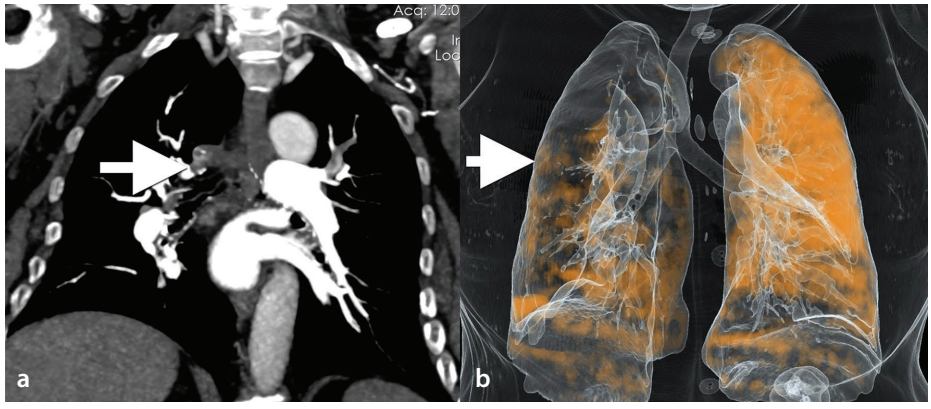


Figure 7. Coronal reformatted dual-energy computed tomography pulmonary angiography and volume-rendered perfusion images. **(a)** Lobar occlusion of the right upper lobe pulmonary artery due to chronic thromboembolic pulmonary hypertension (arrow); **(b)** corresponding three-dimensional volume-rendered perfusion image reveals marked perfusion defects in the right lung (arrow).

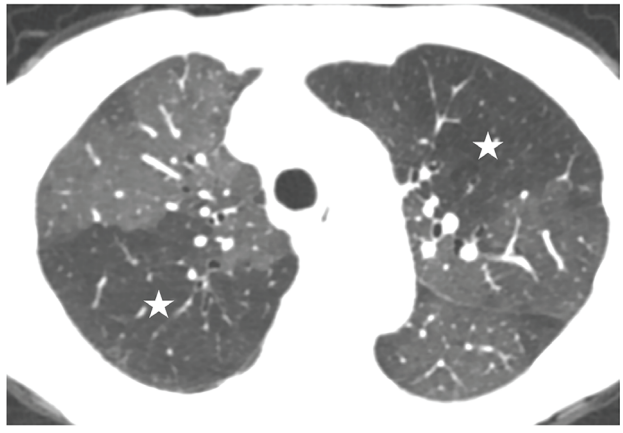


Figure 8. Geographical areas of normal (gray areas) and reduced (black areas, star) perfusion, known as mosaic attenuation in chronic thromboembolic pulmonary hypertension.

functional data alongside anatomical CTPA findings, particularly useful in patients with suspected distal or microvascular disease. Although 4D perfusion imaging with PCCT offers valuable functional data, it is not yet part of routine clinical practice. Its current use remains largely investigational, and further multicenter validation studies are required before clinical implementation.

Catheter pulmonary angiography

Catheter pulmonary angiography (CPA) remains a cornerstone in the diagnostic workup of CTEPH, particularly when planning interventions such as PEA or balloon pulmonary angioplasty (BPA) (Figure 9). The CPA technique enables dynamic, high-resolution visualization of pulmonary arterial anatomy and is especially valuable for detecting chronic changes such as pouch defects, webs, bands, and abrupt vascular cutoffs. These findings help differentiate CTEPH from acute thromboembolic disease. However, CPA is highly operator dependent and subject to projectional limitations, which can lead to underestimation of distal disease burden. Overlapping vascular structures or suboptimal contrast timing may also obscure subtle lesions. Thus, high-quality biplane imaging and expert interpretation are essential. Furthermore, CPA offers the advantage of being combined with RHC in a single session, allowing for both anatomical and hemodynamic assessment. However, it is an invasive modality associated with procedural risks, including contrast nephropathy, vascular injury, arrhythmias, and, rarely, perforation. Its use should be balanced against clinical necessity and patient comorbidities.

Despite advances in non-invasive imaging modalities, RHC is still mandatory for hemodynamic evaluation. It is possible to combine RHC with CPA for detailed assessment of pulmonary circulation.[11,25]

Optical coherence tomography and intravascular ultrasound

Optical coherence tomography (OCT) and intravascular ultrasound (IVUS) are real-time intravascular high-resolution imaging modalities for detecting or characterizing PE and simultaneously measuring the luminal diameter.[26] The resolution capacity of the OCT system is almost 10-fold higher than that of IVUS for the evaluation of subsegmental branches of pulmonary arteries. Although not used routinely, both systems have been shown to be useful for the identification of BPA targets;[27] however, caution must be exercised, as it is possible to induce pulmonary injury during OCT by forceful injection of a contrast jet. Forceful injection of contrast during OCT may result in pulmonary vascular injury, including barotrauma, vessel dissection, or even rupture, particularly in fragile or chronically diseased arterial segments.[28]

Magnetic resonance imaging

Cardiac magnetic resonance imaging (MRI) is a comprehensive modality that combines basic morphologic imaging with advanced functional and tissue characterization techniques. Its value is particularly evident in CTEPH, where chronic pressure overload leads to progressive right ventricular remodeling. Cine steady-state free precession sequences are the cornerstone of cardiac MRI and provide dynamic visualization of cardiac motion, allowing accurate measurement of right ventricular and LV volumes, ejection fraction, and wall thickness. In CTEPH, these sequences typically reveal right ventricular hypertrophy and dilatation, interventricular septal flattening or leftward bowing, and right atrial enlargement, which together reflect the chronic hemodynamic stress imposed by elevated pulmonary pressures.

Beyond morphology and function, contrast-enhanced sequences add critical tissue information. Late gadolinium enhancement is a robust marker of focal fibrosis, most commonly observed at the right ventricular insertion points in patients with long-standing PH, including CTEPH. This localized fibrosis is thought to result from repetitive mechanical stress at the interventricular junction and has been associated with adverse outcomes.

More advanced characterization is possible with mapping techniques. Specifically, T1 mapping and extracellular volume quantification provide a quantitative measure of diffuse interstitial expansion and fibrosis, changes that cannot be captured by late gadolinium enhancement alone. By contrast, T2 mapping is primarily sensitive to myocardial edema and can highlight reversible injury or acute strain superimposed on chronic remodeling. In the context of CTEPH, T1 mapping values and extracellular volume fractions are often elevated, reflecting diffuse myocardial changes secondary to pressure overload, whereas T2 mapping may help identify additional acute components of myocardial stress.

These MRI-derived parameters have particular importance in CTEPH because they correlate strongly with invasive hemodynamic measurements. Increased right ventricular volumes, reduced ejection fraction, abnormal strain, and elevated T1 or extracellular volume values are associated with relatively high PVR and mPAP.

Although mPAP itself is determined only by RHC, cardiac MRI provides reliable non-invasive surrogates that mirror hemodynamic burden and help in risk stratification. Moreover, cardiac MRI is valuable for follow-up after interventions such as PEA or BPA. Cine imaging can demonstrate reverse remodel-

Figure 9. Digital subtraction angiography imaging of the right lung in a patient with chronic thromboembolic pulmonary hypertension with segmental occlusions and abrupt truncations (arrows), stenosis in the right lower lobe artery (arrowhead), and reduced parenchymal perfusion (star) in the capillary phase.

ing, with a reduction in right ventricular size and improved function, whereas mapping techniques may show partial normalization of previously elevated T1 or extracellular volume values. These changes parallel clinical and hemodynamic improvement and underscore the role of MRI not only in diagnosis but also in monitoring therapeutic response in CTEPH.[29]

Recent studies have shown that newly improved MRI scanners with faster image acquisition and perfusion capabilities have a 97% sensitivity and 92% specificity for the diagnosis of CTEPH.[30] To evaluate slow-flowing blood in pulmonary arteries, it is useful to combine the spin-echo double inversion recovery (black-blood) sequence with the half-Fourier single-shot turbo spin-echo sequence.[31] Dynamic contrast-enhanced (DCE) MRI with three-dimensional (3D) T1-weighted gradient-echo sequences with more speedy acquisitions (Figure 10) is used for the delineation of vascular changes associated with CTEPH and can help discriminate CTEPH from other causes of PH.[32] Quantitative evaluation, such as blood flow, volume, and mean transit time values, are additional benefits of DCE-MRI for better evaluation of microcirculation in patients with CTEPH in terms of progress or response to treatment.[33]

Compared with pulmonary angiography and CTPA, DCE-MRI had a slightly lower sensitivity than CTPA between 83% and 86% for depicting total occlusion and a sensitivity of almost 70% for the diagnosis of chronic thrombi, webs, and linear bands.[34]

Ferumoxytol (FE) and oxygen-enhanced MRI (oeMRI) studies have recently been used as alternative methods for lung perfusion. FE exhibits exceptional superparamagnetic properties and is gaining popularity as a potential alternative to gadolinium-based contrast agents due to safety concerns. Consequently, pulmonary MRI with FE could serve as a viable diagnostic approach, particularly in pregnant women and individuals with chronic kidney disease. FE-enhanced MRI offers a wide range of vascular indications, including CTEPH, and can be further enhanced with 4D imaging techniques. The oeMRI method facilitates the visualization of primary lung function, gas exchange, and perfusion mechanisms. In comparison to other imaging techniques, oeMRI exhibits moderate diagnostic compatibility with V/Q scans. Notably, oeMRI demonstrates superior sensitivity to ventilation scintigraphy.[35,36] Although FE-enhanced MRI may serve as an alternative in patients with renal dysfunction, the American College of Radiology recommends avoiding its use during pregnancy unless absolutely necessary, following a thorough risk–benefit analysis.[37]

Four-dimensional magnetic resonance imaging

4D flow MRI can provide a comprehensive and real-time objective evaluation of different blood-flow parameters, such as wall shear stress, vorticity, kinetic energy, and pressure gradients.[38] This application holds strong potential for evaluation of the pulmonary vascular system (Figure 11). Early onset of retrograde flow in the main pulmonary artery is an important abnormality in PH. A linear correlation between mPAP and vortex value in PH has been shown; a relative vortex flow time over 14.3% of the cardiac cycle has 97% sensitivity and 96% specificity for PH.[39] Vorticity in the main pulmonary artery also correlates with PVR, and helicity reveals main pulmonary artery stiffness with right ventricular outflow function.[40,41] Despite these promising findings, multicenter validation is required to establish the utility of 4D flow in PH diagnosis.

Phase-resolved functional lung magnetic resonance imaging

Currently, there is a focus on lung perfusion imaging without using intravenous contrast media. The phase-resolved functional lung (PREFUL) technique is a novel, non-contrast MRI application that utilizes artificial intelligence (AI)-driven subtraction images of lungs. It possesses the potential to serve as a diagnostic aid in a spectrum of lung diseases, particularly in patients with PH and CTEPH. The primary objective of the PREFUL tech-

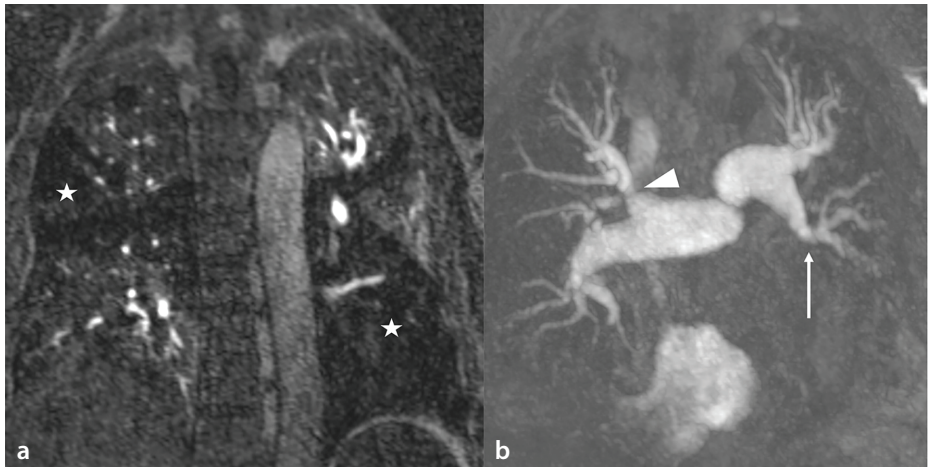


Figure 10. (a) Magnetic resonance perfusion and **(b)** angiography imaging in chronic thromboembolic pulmonary hypertension (CTEPH). Occlusion of the left lower lobe (arrow) and origin stenosis of the right upper lobe (arrowhead) are associated with corresponding perfusion defects on the left side in CTEPH (star).

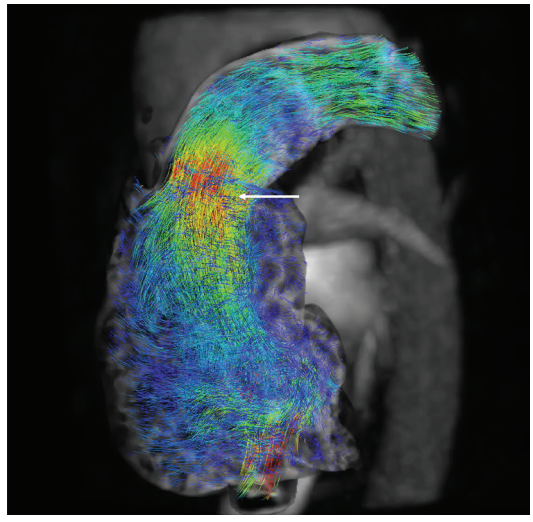


Figure 11. Magnetic resonance four-dimensional flow image of the right ventricle outflow tract and pulmonary artery, with abnormal flow vortex (arrow) in the main pulmonary artery due to pulmonary hypertension in chronic thromboembolic pulmonary hypertension.

nique is to discern variations in perfusion defects between inspiration and expiration image groups.[42] Prospective multicenter validation is necessary prior to the adoption of this technique in routine clinical practice.

Positron emission tomography

Fluorine-18 fluorodeoxyglucose (^{18}F-FDG) is a radiopharmaceutical analog that is predominantly utilized in positron emission tomography (PET)-CT to evaluate glucose metabolism. The PET-CT approach is a valuable imaging modality employed in tumor diagnostics, primarily based on the elevated glucose metabolism observed within tumor cells.

Uptake of ^{18}F-FDG in PH has been reported in the RV, central pulmonary artery, and lung parenchyma. Increased ^{18}F-FDG uptake in these areas may reflect underlying inflammation, vascular remodeling, or increased metabolic demand due to pressure overload. These findings may help differentiate CTEPH from other pulmonary vascular diseases and provide insight into disease activity or prognosis.[43] The most critical role of ^{18}F-FDG in Group 4 PH is in the differential diagnosis between CTEPH and pulmonary artery tumors; pulmonary artery malignancy has a poor prognosis and is occasionally misdiagnosed as PE, leading to incorrect treatment or surgery.

In certain cases, distinguishing CTEPH from tumors (particularly angiosarcomas) and vasculitis can be challenging due to the location and shape of the thrombus (Figures 12 and 13). Consequently, the high cellularity of tumors is generally associated with elevated uptake of ^{18}F-FDG, which serves as a valuable marker for correct diagnosis and may allow surgery or chemotherapy to be performed promptly.[44]

Another important role of PET-CT is its capability to diagnose large-vessel vasculitis, such as Behçet's, Takayasu, and giant cell arteritis. Early diagnosis of vasculitis plays a valuable role in reducing complications such as aneurysm, PE, and CTEPH.[45]

Evidence-based evaluation of imaging techniques for chronic thromboembolic pulmonary hypertension

The assessment of imaging modalities in pulmonary vascular disease, particularly in CTEPH, reveals important variability in both evidence strength and methodological quality.

A critical review of imaging methods used for CTEPH shows that the strength of evidence supporting each technique varies between studies and that many studies have methodological limitations. This emphasizes the need to clearly understand the diagnostic accuracy, potential sources of bias, and clinical applicability of each imaging modality. Although V/Q scintigraphy remains the gold-standard first-line test for CTEPH screening, with reported sensitivity of 90%–100% and specificity of 94%–100%,[46] a systematic review using the Quality Assessment of Diagnostic Accuracy Studies-2 tool revealed that most V/Q SPECT studies have significant methodological bias, with 11 of 13 studies rated high risk in ≥2 domains with important applicability concerns.[47] The diagnostic performance of CTPA in detecting CTEPH is limited, with reported sensitivity and specificity of 76% and 96%, respectively, at the patient level. However, its accuracy improves when high-resolution, multi-detector CT scanners are employed and images are interpreted by experienced radiologists. The CTPA approach may miss distal disease (potential risk of underdiagnosis), and evaluation of image quality and reader expertise remains critical. Although DECT also offers perfusion insights, formal evidence for diagnostic accuracy is still emerging.[12] In a study involving 74 patients with suspected CTEPH, DCE-MRI demonstrated a sensitivity of 100% for screening, compared with 97% for V/Q SPECT. The level of evidence supporting this finding was classified as level 3, Stage 3 technical efficacy. Given its radiation-free nature, DCE-MRI represents a promising imaging

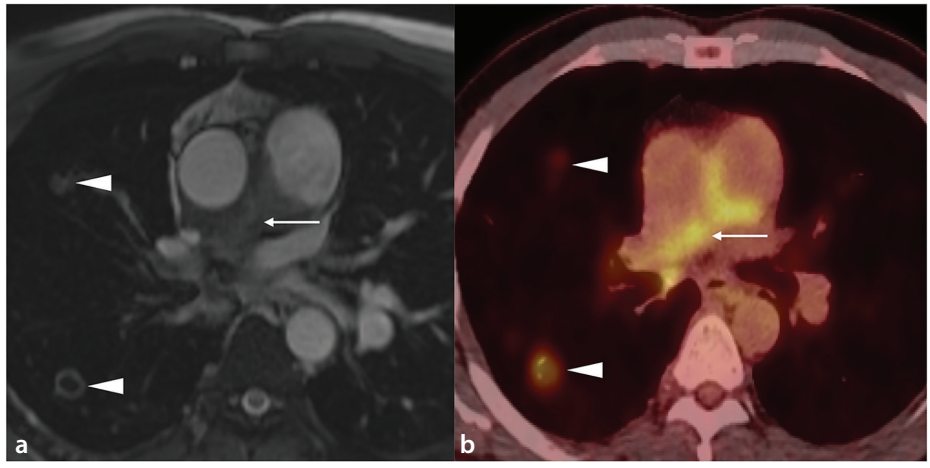


Figure 12. Large vessel vasculitis with pulmonary arterial involvement. **(a)** Axial magnetic resonance image with soft tissue thickening involving the right pulmonary artery (thin arrow) and small cavitating and non-cavitating lesions in the right lung (arrowheads); **(b)** corresponding axial positron emission tomography/computed tomography shows both pulmonary vascular and lung parenchymal lesions to be fluorodeoxyglucose avid.

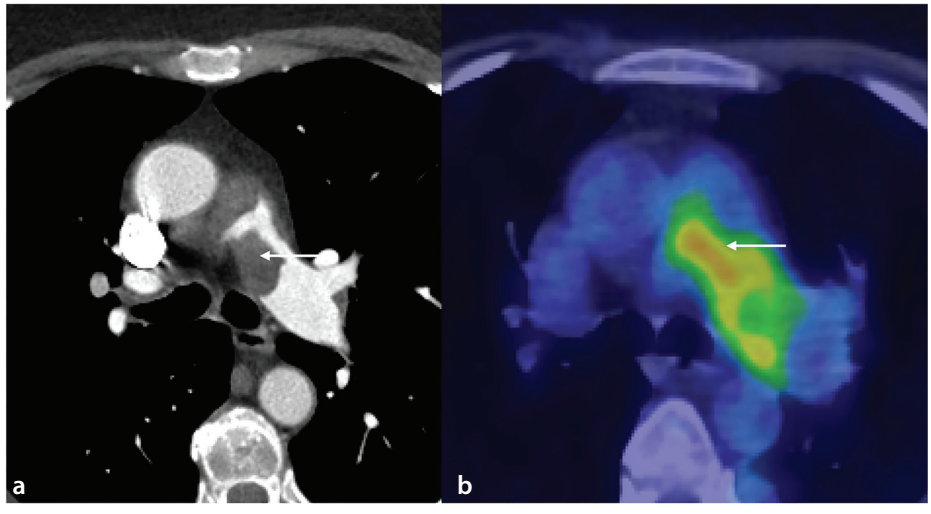


Figure 13. Pulmonary artery sarcoma. **(a)** Computed tomography pulmonary angiography image shows expansile soft tissue lesion in the main and left pulmonary arteries (thin arrow); **(b)** corresponding positron emission tomography/computed tomography imaging shows intense fluorodeoxyglucose avidity in the left pulmonary artery.

modality; however, its clinical utility requires confirmation through larger, prospective studies.[48] Digital subtraction angiography (DSA) is currently regarded as the reference standard when CTPA findings are inconclusive. However, no formal assessment of risk of bias or grading of the level of evidence is provided in the current guidelines. Its use remains standard but is mostly supported by expert consensus and procedural experience rather than comparative trials.

The 2022 ESC/ERS guidelines for PH, which also encompass CTEPH as a distinct subtype, provide structured diagnostic recommendations. In these guidelines, CTPA is explicitly recommended in the diagnostic work-up of patients with suspected PH to evaluate for CTEPH (class I, level C). This grading reflects a strong consensus regarding the central role of CTPA. Beyond CTPA, the guidelines emphasize a multimodality and multidisciplinary approach. Perfusion lung or V/Q scintigraphy remains a cornerstone for screening unexplained PH and is also listed as a class I, level C recommendation to exclude or confirm the probability of CTEPH. Additional imaging modalities, such as DSA and chest CT, are considered useful in selected contexts (class IIa), particularly for anatomical delineation and interventional planning. According to the 2022 ESC/ERS guidelines, the diagnostic approach to CTEPH in symptomatic patients should follow a structured and multidisciplinary pathway. The entire diagnostic process should be coordinated within experienced, high-volume CTEPH centers, where Multidisciplinary Team (MDT) comprising PH specialists, radiologists, surgeons, and interventionalists can guide appropriate management decisions.

Reporting recommendations

Reports should clearly state the extent of pulmonary artery analyzability; in cases where only central arteries are evaluable or image quality is compromised, the study should be labeled as indeterminate or non-diagnostic. Quantitative assessment of contrast enhancement is advisable, and attenuation values should be reported when suboptimal opacification is suspected. Simultaneous evaluation of mediastinal and lung window images is essential to differentiate true emboli from artifacts. The interpretation and communication of CTPA findings in suspected CTEPH should follow a standardized reporting structure that encompasses both diagnostic and prognostic information, facilitating clinical decision-making and longitudinal patient care. Central to this approach is the inclusion of a core set of imaging findings—derived through expert consensus—that have demonstrated prognostic significance. These include the RV/LV diameter ratio measured on axial images, central embolus location, the presence of an isolated subsegmental PE, septal bowing, right ventricular hypertrophy, bronchial artery dilatation, intravascular webs or bands, and pulmonary artery retraction. Supplementary findings such as main pulmonary artery diameter, complete arterial occlusion, and organized mural thrombi are considered clinically useful and should be reported. For perfusion studies (e.g., DECT or V/Q), quantification of the perfusion defect extent (e.g., as % of total lung volume) should be reported. Further evaluation with MRI for verification of CTEPH diagnosis should primarily focus on right ventricular functional and morphological parameters, including the ventricular mass index, RV ejection fraction, and RV longitudinal strain. Additionally, MRI can provide indirect estimations of hemodynamic parameters, such as mPAP, ventriculoarterial coupling (RV–PA coupling), end-systolic elastance, and arterial elastance; however, these values are not directly measured but rather estimated through derived indices and are ideally confirmed by invasive RHC.[7,49]

Multidisciplinary team and expert centers for chronic thromboembolic pulmonary hypertension

The idea of a "CTEPH MDT" was first postulated at the fifth World Symposium on PH in 2013 and has subsequently been permanently embedded into the CTEPH treatment algorithm. Recent studies increasingly support the concept that CTEPH-MDTs should be based in high-volume centers with established experience in pulmonary PEA and BPA. Concentrating expertise in such centers has been associated with improved diagnostic accuracy, optimized patient selection, and superior long-term outcomes. A "CTEPH expert center" should include an experienced CTEPH cardiovascular surgeon, PH pulmonologist, and cardiologist (preferentially a BPA specialist), as well as a cardiothoracic radiologist with specialist expertise in CTEPH.[50] Given the diagnostic complexities, expert radiological opinion has assumed a pivotal role in the implementation of a successful CTEPH program. The specialist MDT is essential to ensure accurate diagnosis, appropriate timing, and precise surgical and BPA planning. In addition, a considerable number of patients with CTEPH require long-term follow-up with CTEPH-experienced medical professionals to ensure optimal and prolonged survival.

In addition to surgical and interventional options, medical therapies such as Riociguat treatment, an approved soluble guanylate cyclase stimulator, play a crucial role in patients who are inoperable, have persistent PH following PEA, or are undergoing BPA.[51] A combined approach integrating pharmacologic treatment with procedural interventions is increasingly considered to improve hemodynamics and clinical status in selected patients.[12,50]

Pulmonary endarterectomy

PEA is one of the treatment options for CTEPH in the absence of life-threatening serious comorbidities. This open-heart surgery necessitates a true bilateral endarterectomy through the pulmonary arteries. The proximal extension of the cast defines the level of disease, and the classification is made according to the material removed surgically. Level I defines a disease in the right or left main pulmonary artery; level II is a disease in the lobar branches; level III is in the proximal segmental arteries; and level IV is in the distal segmental and subsegmental pulmonary branches (Figure 14).[52]

Residual thrombi, particularly in the distal branches of the pulmonary artery, should be reported following PEA surgery.

Cardiopulmonary bypass is followed by circulatory arrest in deep hypothermia, which is necessary to keep the surgical field free of blood via the bronchopulmonary collaterals and protect brain function. Identification of the correct plane of dissection in the segmental and sub-segmental branches and removing the material en bloc with its tail is pivotal.

PEA remains a challenging surgery and is determined by multiple factors, such as technical operability (e.g., accessibility of the obstructions and surgical expertise) and the risk–benefit ratio, as determined by preoperative hemodynamics and comorbidities.

Following PEA surgery, the literature reported a survival rate exceeding 90% within 3 years. However, if patients with CTEPH had not undergone surgery, the survival rate during this period was only 70%.[8,53]

Balloon pulmonary angioplasty

BPA has emerged as a viable alternative treatment option for patients with CTEPH who experience thrombi formation within the distal–proximal segmental locations of level

Figure 14. Pulmonary endarterectomy specimen of chronic thrombus demonstrates typical yellowish-white chronic thrombus with tails.

III–IV PAs. An alternative option for BPA is inoperable CTEPH or in patients with residual lesions following PEA.[12] The BPA method requires extensive training and has been shown to have a learning curve to reduce complications. Before and after the BPA procedure, pressure gradient analysis should be performed to evaluate the procedural success.[54] Vascular injury due to wire perforation is the most feared complication of BPA, with a rate of approximately 7.7%.[55] An evolving trend is the combined use of PEA and BPA as a "hybrid approach."

Artificial intelligence in chronic thromboembolic pulmonary hypertension imaging

Applying AI in diagnostic imaging involves utilizing machine learning (ML) algorithms and methodologies to analyze images, thereby assisting in the interpretation and optimal diagnosis. It is important to note that ML is a subset of AI, which broadly encompasses various computational techniques designed to simulate human decision-making. Given the rarity and diagnostic complexity of CTEPH, AI-driven tools—particularly ML-based algorithms—could provide vital support to radiologists by enhancing the detection of subtle vascular abnormalities and improving diagnostic consistency across centers. ML can assist in quantifying vascular morphology by automatically segmenting pulmonary vessels on CT imaging. Additionally, vascular density, volume fraction, volume ratio of the vessels, and pulmonary artery tortuosity can be measured in patients with CTEPH using ML.[56]

Promising diagnostic methods for AI-based CT techniques include CT morphometry (pulmonary vasculature analysis) and computational fluid dynamics (CFD).

AI-based CT morphometry is a specialized technique employed to detect vascular remodeling. It utilizes AI-driven vasculature analysis derived from CT scans to quantify the varying sizes of vessels. Vascular pruning in CTEPH has been associated with the loss of vascular volume and density.[57]

CFDs focuses on detecting the flow velocity and wall shear stress in pulmonary artery branches with the assistance of AI. There is a reduction in wall shear stress in patients with PH compared with healthy individuals. This technique is potentially beneficial in determining the therapeutic response to PEA and BPA.[58]

The application of ML in CTEPH research presents several issues, primarily due to the condition's rarity, resulting in limited patient cohorts. Additionally, assessing the substantial heterogeneity of vascular pathologies associated with CTEPH poses considerable difficulties. Leveraging the major advancements in ML technology, further research on CTEPH imaging utilizing AI has the potential to be of immense value in CTEPH diagnosis, risk stratification, and therapy evaluation but requires validation in large cohorts.

In conclusion, this review highlighted the known and newly developed imaging modalities, including AI-based techniques, for the diagnosis of CTEPH. Despite advances in diagnostic modalities over the past two decades, CTEPH remains challenging to diagnose and manage. Radiologists play a crucial role in improving the outcomes of patients with CTEPH. The decision-making process in CTEPH diagnosis and management requires an expert MDT. Although new modalities and techniques have many advantages, their utilization is limited in many centers. In the future, CTEPH expert centers should connect with each other to provide second opinions for complex cases as well as to expand radiologist training. These improvements could be beneficial for pooling all resources and gaining optimal expertise.

Future research on CTEPH imaging should focus on the integration of advanced modalities, such as PCCT, DECT, 4D-flow MRI, and AI-driven image analysis. Prospective multicenter trials are required to validate their diagnostic and prognostic utility in diverse patient populations. Standardized endpoints, including the 6-minute walk distance, mPAP, PVR, and overall survival, should be adopted.

In addition, AI-based techniques such as CT morphometry and CFD have shown potential in quantifying pulmonary vascular pruning, tortuosity, and altered hemodynamics. These technologies could support personalized therapy planning and early identification of disease progression.

International collaboration between expert centers is essential for establishing imaging-based biomarkers, creating open-access datasets, and developing consensus protocols to advance the field.

Footnotes

Conflict of interest disclosure

Mehmet Ruhi Onur, MD, served as Editor-in-Chief of Diagnostic and Interventional Radiology from 2023 to 2025 and during the submission and evaluation of this article. He had no involvement in the peer-review or decision-making process of this article and had no access to information regarding its peer review. The manuscript was handled by an independent editor.

The other authors declare no conflicts of interest.

References

1. Ghani H, Weir-McCall JR, Ruggiero A, Pepke-Zaba J. Imaging in chronic thromboembolic pulmonary disease: current practice and advances. *Int J Cardiol Congenit Heart Dis.* 2024;17:100536. [Crossref]
2. Hahn LD, Chung JH. Imaging of chronic thromboembolic pulmonary hypertension. *Radiol Clin North Am.* 2025;63(2):223-234. [Crossref]
3. Delcroix M, Torbicki A, Gopalan D, et al. ERS statement on chronic thromboembolic

pulmonary hypertension. *Eur Respir J.* 2021;57(6):2002828. [Crossref]

4. Delcroix M, Kerr K, Fedullo P. Chronic thromboembolic pulmonary hypertension. Epidemiology and risk factors. *Ann Am Thorac Soc.* 2016;13(Suppl 3):201-206. [Crossref]
5. Rahaghi FN, Ross JC, Agarwal M, et al. Pulmonary vascular morphology as an imaging biomarker in chronic thromboembolic pulmonary hypertension. *Pulm Circ.* 2016;6(1):70-81. [Crossref]
6. Cerrone E, Hameed AG, Kiely DG, et al. Do hemodynamic definitions of chronic thromboembolic pulmonary hypertension distinguish between distinct phenotypes of chronic thromboembolic pulmonary disease? *Ann Am Thorac Soc.* 2025;22(3):331-338. [Crossref]
7. Klok FA, Barco S, Bertoletti L, et al. Optimal approach to performing and reporting computed tomography angiography for suspected acute pulmonary embolism: a clinical consensus statement of the ESC working group on pulmonary circulation & right ventricular function, the Fleischner Society, the Association for Acute Cardiovascular Care (ACVC) and the European Association of Cardiovascular Imaging (EACVI) of the ESC, Endorsed by European Respiratory Society (ERS), Asian Society of Thoracic Radiology (ASTR), European Society of Thoracic Imaging (ESTI), and Society of Thoracic Radiology (STR). *Radiology.* 2025;315(3):e243833. [Crossref]
8. Kim NH, D'Armini AM, Delcroix M, et al. Chronic thromboembolic pulmonary disease. *Eur Respir J.* 2024;64(4):2401294. [Crossref]
9. Narechania S, Renapurkar R, Heresi GA. Mimickers of chronic thromboembolic pulmonary hypertension on imaging tests: a review. *Pulm Circ.* 2020;10(1):2045894019882620. [Crossref]
10. Isidoro J, Gil P, Costa G, Pedroso de Lima J, Alves C, Ferreira NC. Radiation dose comparison between V/P-SPECT and CT-angiography in the diagnosis of pulmonary embolism. *Phys Med.* 2017;41:93-96. [Crossref]
11. Konstantinides SV, Meyer G, Becattini C, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *Eur Heart J.* 2020;41(4):543-603. [Crossref]
12. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J.* 2022;43(38):3618-3731. [Crossref] Erratum in: *Eur Heart J.* 2023;44(15):1312. [Crossref]
13. Ghaye B, Ghuysen A, Bruyere PJ, D'Orio V, Dondelinger RF. Can CT pulmonary angiography allow assessment of severity and prognosis in patients presenting with pulmonary embolism? What the radiologist needs to know. *Radiographics.* 2006;26(1):39-40. [Crossref]
14. Wood KE. Major pulmonary embolism: review of a pathophysiologic approach to the golden hour of hemodynamically significant pulmonary embolism. *Chest.* 2002;121(3):877-905. [Crossref]
15. Grosse C, Grosse A. CT findings in diseases associated with pulmonary hypertension: a current review. *Radiographics.* 2010;30(7):1753-1777. [Crossref]
16. Guérin L, Couturaud F, Parent F, et al. Prevalence of chronic thromboembolic pulmonary hypertension after acute pulmonary embolism. Prevalence of CTEPH after pulmonary embolism. *Thromb Haemost.* 2014;112(3):598-605. [Crossref]
17. McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for reducing radiation dose in CT. *Radiol Clin North Am.* 2009;47(1):27-40. [Crossref]
18. Davenport MS, Perazella MA, Yee J, et al. Use of intravenous iodinated contrast media in patients with kidney disease: consensus statements from the American College of Radiology and the National Kidney Foundation. *Kidney Med.* 2020;2(1):85-93. [Crossref]
19. Wittram C, Yoo AJ. Transient interruption of contrast on CT pulmonary angiography: proof of mechanism. *J Thorac Imaging.* 2007;22(2):125-129. [Crossref]
20. Dournes G, Verdier D, Montaudon M, et al. Dual-energy CT perfusion and angiography in chronic thromboembolic pulmonary hypertension: diagnostic accuracy and concordance with radionuclide scintigraphy. *Eur Radiol.* 2014;24(1):42-51. [Crossref]
21. Bird E, Hasenstab K, Kim N, et al. Mapping the spatial extent of hypoperfusion in chronic thromboembolic pulmonary hypertension using multienergy CT. *Radiol Cardiothorac Imaging.* 2023;5(4):e220221. [Crossref]
22. Celtikci P, Hekimoglu K, Kahraman G, Bozbas S, Gultekin B, Akay HT. Diagnostic impact of quantitative dual-energy computed tomography perfusion imaging for the assessment of subsegmental pulmonary embolism. *J Comput Assist Tomogr.* 2021;45(1):151-156. [Crossref]
23. Kerber B, Flohr T, Ulrich S, Lichtblau M, Frauenfelder T, Franckenberg S. Photon-counting CT iodine maps for diagnosing chronic pulmonary thromboembolism: a pilot study. *Invest Radiol.* 2025;60(5):328-333. [Crossref]
24. Remy-Jardin M, Duhamel A, Delobelle M, Bervar JF, Flohr T, Remy J. Lung microvasculopathy in chronic thromboembolic pulmonary hypertension: high-resolution findings with photon-counting detector CT in 29 patients. *Eur Radiol.* 2025;35(10):6369-6381. [Crossref]
25. Simeone B, Maggio E, Schirone L, et al. Chronic thromboembolic pulmonary hypertension: the diagnostic assessment. *Front Cardiovasc Med.* 2024;11:1439402. [Crossref]
26. Inohara T, Kawakami T, Kataoka M, et al. Lesion morphological classification by OCT to predict therapeutic efficacy after balloon pulmonary angioplasty in CTEPH. *Int J Cardiol.* 2015;197:23-25. [Crossref]
27. Ishiguro H, Kataoka M, Inami T, et al. Diversity of lesion morphology in CTEPH analyzed by OCT, pressure wire, and angiography. *JACC Cardiovasc Imaging.* 2016;9(3):324-325. [Crossref]
28. Chamié D, Bahl R, Maia J, et al. Can contrast injections cause or propagate coronary injuries? Insights from vessel and guiding catheter hemodynamics. *J Soc Cardiovasc Angiogr Interv.* 2024;3(12):102396. [Crossref]
29. Alabed S, Garg P, Johns CS, et al. Cardiac magnetic resonance in pulmonary hypertension-an update. *Curr Cardiovasc Imaging Rep.* 2020;13(12):30. [Crossref]
30. Rajaram S, Swift AJ, Telfer A, et al. 3D contrast-enhanced lung perfusion MRI is an effective screening tool for chronic thromboembolic pulmonary hypertension: results from the ASPIRE Registry. *Thorax.* 2013;68(7):677-678. [Crossref]
31. Swift AJ, Rajaram S, Marshall H, et al. Black blood MRI has diagnostic and prognostic value in the assessment of patients with pulmonary hypertension. *Eur Radiol.* 2012;22(3):695-702. [Crossref]
32. Bergmann LL, Ackman JB, Starekova J, et al. MR angiography of pulmonary vasculature. *Magn Reson Imaging Clin N Am.* 2023;31(3):475-491. [Crossref]
33. Kreitner KF, Kunz RP, Ley S, et al. Chronic thromboembolic pulmonary hypertension - assessment by magnetic resonance imaging. *Eur Radiol.* 2007;17(1):11-21. [Crossref]
34. Nikolaou K, Schoenberg SO, Attenberger U, et al. Pulmonary arterial hypertension: diagnosis with fast perfusion MR imaging and high-spatial-resolution MR angiography--preliminary experience. *Radiology.* 2005;236(2):694-703. [Crossref]
35. Knobloch G, Colgan T, Schiebler ML, et al. Comparison of gadolinium-enhanced and ferumoxytol-enhanced conventional and UTE-MRA for the depiction of the pulmonary vasculature. *Magn Reson Med.* 2019;82(5):1660-1670. [Crossref]
36. Maxien D, Dietrich O, Thieme SF, et al. Value of oxygen-enhanced MRI of the lungs in patients with pulmonary hypertension: a qualitative and quantitative approach. *J Magn Reson Imaging.* 2012;35(1):86-94. [Crossref]

37. Bookwalter CA, McDonald RJ, Packard AT, Little JT, McDonald JS, Watson RE. Contrast media in pregnant and lactating patients, from the *AJR* special series on contrast media. *AJR Am J Roentgenol.* 2025;224(4):e2431415. [Crossref]
38. Bissell MM, Raimondi F, Ait Ali L, et al. 4D Flow cardiovascular magnetic resonance consensus statement: 2023 update. *J Cardiovasc Magn Reson.* 2023;25(1):40. [Crossref]
39. Schäfer M, Barker AJ, Kheyfets V, et al. Helicity and vorticity of pulmonary arterial flow in patients with pulmonary hypertension: quantitative analysis of flow formations. *J Am Heart Assoc.* 2017;6(12):e007010. [Crossref]
40. Kheyfets VO, Schafer M, Podgorski CA, et al. 4D magnetic resonance flow imaging for estimating pulmonary vascular resistance in pulmonary hypertension. *J Magn Reson Imaging.* 2016;44(4):914-922. [Crossref]
41. Helderman F, Mauritz GJ, Andringa KE, Vonk-Noordegraaf A, Marcus JT. Early onset of retrograde flow in the main pulmonary artery is a characteristic of pulmonary arterial hypertension. *J Magn Reson Imaging.* 2011;33(6):1362-1368. [Crossref]
42. Moher Alsady T, Voskrebenzev A, Behrendt L, et al. Multicenter Standardization of Phase-Resolved Functional Lung MRI in patients with suspected chronic thromboembolic pulmonary hypertension. *J Magn Reson Imaging.* 2024;59(6):1953-1964. [Crossref]
43. Frille A, Steinhoff KG, Hesse S, et al. Thoracic [18F]fluorodeoxyglucose uptake measured by positron emission tomography/computed tomography in pulmonary hypertension. *Medicine (Baltimore).* 2016;95(25):e3976. [Crossref]
44. Xi XY, Gao W, Gong JN, et al. Value of ^{18}F-FDG PET/CT in differentiating malignancy of pulmonary artery from pulmonary thromboembolism: a cohort study and literature review. *Int J Cardiovasc Imaging.* 2019;35(7):1395-1403. [Crossref]
45. Hagan G, Gopalan D, Church C, et al. Isolated large vessel pulmonary vasculitis as a cause of chronic obstruction of the pulmonary arteries. *Pulm Circ.* 2011;1(3):425-429. [Crossref]
46. Derenoncourt PR, Felder GJ, Royal HD, et al. Ventilation-perfusion scan: a primer for practicing radiologists. *Radiographics.* 2021;41(7):2047-2070. [Crossref]
47. Le Roux PY, Robin P, Tromeur C, et al. Ventilation/perfusion SPECT for the diagnosis of pulmonary embolism: a systematic review. *J Thromb Haemost.* 2020;18(11):2910-2920. [Crossref]
48. Johns CS, Swift AJ, Rajaram S, et al. Lung perfusion: MRI vs. SPECT for screening in suspected chronic thromboembolic pulmonary hypertension. *J Magn Reson Imaging.* 2017;46(6):1693-1697. [Crossref]
49. Bartnik A, Pepke-Zaba J, Bunclark K, et al. Cardiac MRI in the assessment of chronic thromboembolic pulmonary hypertension and response to treatment. *Thorax.* 2023;79(1):90-97. [Crossref]
50. Hahn LD, Papamatheakis DG, Fernandes TM, et al. Multidisciplinary approach to chronic thromboembolic pulmonary hypertension: role of radiologists. *Radiographics.* 2023;43(2):e220078. [Crossref]
51. Pamukçu E, Kaya MO. Meta-analysis of real-world clinical practice to assess the effectiveness of riociguat in treating chronic thromboembolic pulmonary hypertension. *J Clin Hypertens (Greenwich).* 2025;27(2):e70015. [Crossref]
52. Madani MM. Pulmonary endarterectomy for chronic thromboembolic pulmonary hypertension: state-of-the-art 2020. *Pulm Circ.* 2021;11(2):20458940211007372. [Crossref]
53. Delcroix M, Lang I, Pepke-Zaba J, et al. Long-term outcome of patients with chronic thromboembolic pulmonary hypertension: results from an international prospective registry. *Circulation.* 2016;133(9):859-871. [Crossref]
54. Roik M, Wretowski D, Łabyk A, et al. Refined balloon pulmonary angioplasty driven by combined assessment of intra-arterial anatomy and physiology--Multimodal approach to treated lesions in patients with non-operable distal chronic thromboembolic pulmonary hypertension--Technique, safety and efficacy of 50 consecutive angioplasties. *Int J Cardiol.* 2016;203:228-235. [Crossref]
55. Jain N, Sheikh MA, Bajaj D, et al. Periprocedural complications with balloon pulmonary angioplasty: analysis of global studies. *JACC Cardiovasc Interv.* 2023;16(8):976-983. [Crossref]
56. Pienn M, Gertz RJ, Gerhardt F, et al. CT-derived lung vessel morphology correlates with prognostic markers in precapillary pulmonary hypertension. *J Heart Lung Transplant.* 2024;43(1):54-65. [Crossref]
57. Rahaghi FN, Argemí G, Nardelli P, et al. Pulmonary vascular density: comparison of findings on computed tomography imaging with histology. *Eur Respir J.* 2019;54(2):1900370. [Crossref]
58. Tsubata H, Nakanishi N, Itatani K, et al. Pulmonary artery blood flow dynamics in chronic thromboembolic pulmonary hypertension. *Sci Rep.* 2023;13(1):6490. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263541

CONTRAST MEDIA

LETTER TO THE EDITOR

Reply: Radiologists' tendency to collaborate with referring physicians in managing contrast media-related risk factors

Burak Öztürk[1]
Özgür Karabıyık[2]

[1]Ünye State Hospital, Clinic of Radiology, Ordu, Türkiye

[2]Erciyes University Faculty of Medicine, Department of Radiology, Kayseri, Türkiye

Dear Editor,

We appreciate the authors' interest in our study evaluating the collaborative tendencies of radiologists in managing contrast media-related risk factors.[1] Their thoughtful remarks give us a welcome opportunity to clarify the methodological details they highlighted.

First, regarding the validation of the modified Control Preferences Scale (CPS), we acknowledge that we did not perform formal psychometric testing on the translated and adapted version before starting the investigation. Rather than altering the scale's fundamental construct, our modification primarily involved contextualizing the scenarios specifically for contrast media risk management.[2] The fact that the vast majority of participants successfully completed the scale suggests the core concept of decision-making control remained intact. However, we certainly agree that future large-scale studies would benefit from a properly validated psychometric assessment.

Second, the authors raised a question regarding the six excluded participants. The CPS requires participants to rank their preferences in a consistent, logical order.[2] If a participant's choices contradict each other—for example, choosing both the completely active and entirely passive cards as their top preferences—the permutation is invalid. We excluded these responses as a strict methodological requirement of the CPS design to ensure logical data reliability, not as a subjective clinical choice. Because this exclusion was purely procedural, we did not perform a separate demographic sub-analysis on this group.

Third, we agree that face-to-face administration by a single researcher carries a potential risk of social desirability bias. We opted for this method because having an interviewer present kept the survey process consistent. It also meant we could answer any immediate questions about the modified cards. Without this direct interaction, it would have been nearly impossible to get enough responses from doctors dealing with the fast-paced daily schedule of a city hospital.

Finally, the comments regarding workplace culture and inter-specialty dynamics are highly relevant. Radiologists often end up taking a passive role simply because of the historical divide and ingrained stereotypes between diagnostic and clinical specialties. This kind of behavior is not unique to us; it has been documented in various interprofessional settings.[3] Our research primarily considered systemic and infrastructural hurdles, but the psychosocial environment is clearly equally important. Exploring these social barriers through qualitative studies would be a positive next step for the field. We are very grateful to the authors for sharing their thoughts and helping to advance this conversation.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Corresponding author: Burak Öztürk

E-mail: dr.burak61@gmail.com

Received 02 March 2026; accepted 03 March 2026.

Epub: 11.03.2026

Publication date: 04.05.2026

DOI: 10.4274/dir.2026.263541

You may cite this article as: Öztürk B, Karabıyık Ö. Reply: Radiologists' tendency to collaborate with referring physicians in managing contrast media-related risk factors. *Diagn Interv Radiol*. 2026;32(3):304-305.

References

1. Öztürk B, Karabıyık Ö. Radiologists tendency to collaborate with referring physicians in managing contrast media-related risk factors. *Diagn Interv Radiol*. 2026;32(1):65-70. [Crossref]
2. Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. Can J Nurs Res. 1997 Fall;29(3):21-43. [Crossref]
3. Lewis C. The impact of interprofessional incivility on medical performance, service and patient care: a systematic review. *Future Healthc J*. 2023;10(1):69-77. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253549

EMERGENCY RADIOLOGY

ORIGINAL ARTICLE

Analysis of the baseline multiphase computed tomographic angiography findings to predict clinical outcomes in patients with middle cerebral artery M1 occlusion treated with mechanical thrombectomy

Esra Çıvgın[1]
Hasan Bayındır[2]

[1]Ankara Bilkent City Hospital, Department of Radiology, Ankara, Türkiye

[2]Ankara Bilkent City Hospital, Department of Neurology, Ankara, Türkiye

PURPOSE

We aimed to evaluate the predictive ability of baseline multiphase computed tomographic angiography (mCTA) findings and the time from symptom onset to imaging in predicting functional outcomes in patients with middle cerebral artery (MCA) M1 occlusion treated with mechanical thrombectomy (MT).

METHODS

A total of 70 patients were evaluated retrospectively. The time between the onset of symptoms and imaging, thrombus density, estimated thrombus length, the Alberta Stroke Program Early CT Score (ASPECTS) on non-contrast CT, collateral circulation (CC), actual thrombus length, and clot burden score were assessed on mCTA images. Patients with a 90-day modified Rankin scale score of 0–2 were categorized as having good outcomes, whereas the others were categorized as having poor outcomes. The mCTA findings of patients with good and poor outcomes were compared, and binary logistic regression analysis was performed to identify independent predictors that could affect clinical outcomes.

RESULTS

The estimated thrombus length, the ASPECTS, thrombus density, clot burden score, and CC grade were not significantly different between patients with good and poor outcomes. The actual thrombus length was shorter in patients with good outcomes than in those with poor outcomes (15.9 mm versus 21.5 mm, $P = 0.001$). Binary logistic regression analysis revealed that actual thrombus length [$P = 0.005$, odds ratio (OR): 0.754, 95% confidence interval (CI): 0.61–0.92] and thrombus density ($P = 0.022$, OR: 1.167, 95% CI: 1.02–1.33) were independent variables for a good outcome. The optimal cut-off value for actual thrombus length was 18.7 mm (area under the curve, 0.74; 95% CI: 0.62–0.86; $P = 0.001$) to predict good outcomes.

CONCLUSION

Higher thrombus density and actual thrombus length shorter than 18.7 mm were associated with good clinical outcomes. However, no significant correlation was found between clinical outcomes and the ASPECTS, CC degree, or clot burden scores.

CLINICAL SIGNIFICANCE

Thrombus length and density are associated with the clinical outcome of patients with MCA M1 occlusion treated with MT who have distal collateral filling sufficient to depict thrombus margins in mCTA.

KEYWORDS

CT angiography, ischemic stroke, mechanical thrombectomy, prognosis, thrombosis

Corresponding author: Esra Çıvgın

E-mail: esrayurduseven@gmail.com

Received 11 June 2025; revision requested 06 August 2025; last revision received 03 October 2025; accepted 21 October 2025.

Epub: 19.11.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253549

You may cite this article as: Çıvgın E, Bayındır H. Analysis of the baseline multiphase computed tomographic angiography findings to predict clinical outcomes in patients with middle cerebral artery M1 occlusion treated with mechanical thrombectomy. *Diagn Interv Radiol*. 2026;32(3):306-314.

Stroke is the second leading cause of death, according to 2022 World Health Organization data.[1] The global prevalence of stroke was 93.8 million in 2021, and 69.9 million of those cases were ischemic.[2,3] Strokes secondary to large-vessel occlusions constitute approximately one-third of all acute ischemic stroke cases; however, they have a larger infarct size than those caused by non-large-vessel occlusions.[4,5] Furthermore, large-vessel occlusions cause 60% of dependency and more than 90% of mortality in all cases.[6]

Neuroimaging is fundamental in acute stroke management, as it demonstrates stroke type (hemorrhagic or ischemic), identifies vessel occlusion, and assesses the affected and salvageable brain areas.[7] The American Heart Association/American Stroke Association 2019 update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke recommends that all patients with suspected acute stroke undergo non-contrast computed tomography (NCCT) or diffusion-weighted imaging for initial evaluation as quickly as possible. The guidelines also recommend non-invasive vascular imaging with computed tomographic angiography (CTA) or magnetic resonance angiography for patients who meet the criteria for mechanical thrombectomy (MT).[8] The ultimate goal of neuroimaging in patients with acute ischemic stroke is to identify those who may safely and effectively benefit from appropriate treatment options, such as intravenous thrombolytic therapy and/or MT.[7-9]

Many studies have been conducted on the utility of neuroimaging findings and scores to predict outcomes in patients with acute stroke.[10-18] These studies use one or more imaging modalities, such as magnetic resonance imaging (MRI), NCCT, single-phase CTA (sCTA), multiphase CTA (mCTA), CT perfusion, or digital subtraction angiography. Different studies have analyzed the possible effect of individual or combined imaging findings and thrombus characteristics on prognosis.[10-18] In some studies, the patient population was heterogeneous in terms of management (intravenous thrombolytics and/or MT or none).[17,18]

The Alberta Stroke Program Early CT Score (ASPECTS) is a 10-point topographic CT scan score developed to predict clinical outcomes in patients with middle cerebral artery (MCA) stroke. It is calculated by subtracting 1 point from the total of 10 points for each region (cortical M1–6 and insula, subcortical lentiform nucleus, internal capsule, and caudate) where early ischemic changes (intraparenchymal hypoattenuation and focal swelling) are detected.[19]

We aim to evaluate the ability of mCTA findings, including the ASPECTS, estimated and actual thrombus length, thrombus density, clot burden score, collateral circulation (CC) status at first admission to the emergency department, and time from symptom onset to imaging, to predict functional outcomes in a selected group of patients who underwent MT for MCA M1 occlusion.

Main points

- Multi-phase computed tomographic angiography is an increasingly important imaging modality in the initial evaluation and management of patients with acute ischemic stroke.
- Thrombus density reflects its structure; erythrocyte-rich thrombus shows higher Hounsfield units than platelet-rich thrombus.
- Thrombus characteristics influence the clinical outcome in patients with acute ischemic stroke.
- The higher thrombus density and shorter thrombus length than 18.7 mm were associated with good clinical outcomes of patients.

Methods

Study design and patient inclusion and exclusion criteria

This study was conducted in accordance with the tenets of the Declaration of Helsinki. This retrospective study was approved by the Ethics Committee of the Ankara Bilkent City Hospital, (approval number: E1-23-4231, date: 1.11.2023). Written informed consent could not be obtained due to the retrospective nature of the study. The medical records of patients who underwent MT between March 2022 and November 2023 were evaluated. Patients with MCA M1 segment occlusion, symptom onset of less than 6 hours, and a standard local mCTA protocol adjusted according to the method described by Menon et al.[20] were included in the study. The mCTA protocol developed by Menon et al.[20] was applied to patients over 18 years of age who presented to the emergency department with stroke-related symptoms within 12 hours of symptom onset. Patients diagnosed at another health center and referred to our hospital for the interventional procedure; those with missing or noisy CT images; undetected actual thrombus length (due to curved vessels and/or poor CC status); occlusion or considerable stenosis of the contralateral MCA; thrombosis of the internal carotid artery (ICA); pre-stroke modified Rankin scale (mRS) score greater than 2; and/or patients younger than 18 years were excluded from the study.

A total of 127 patients who underwent MT between March 2022 and October 2023 were evaluated, and 70 patients who met the inclusion criteria were included in the study (Figure 1).

Imaging procedures

CT scans were obtained using two devices with 128-detector and 64-detector systems (GE Revolution EVO 128 and 64 Slice CT Scanner, GE Medical Systems, Milwaukee, WI, USA). The standard acute stroke CTA protocol, identical for both 128- and 64-detector devices at our institution, included NCCT with 2.5 mm section thickness acquired using the following parameters: 120 kV, 350 mA, 20 mm collimation, 0.6 s/rotation, and pitch 0.531. The mCTA included three phases, each 8 seconds apart. Using the bolus-tracking technique, CTA was acquired by injection of 40–45 mL of contrast material (350 mg/mL iodine), followed by a 20 mL saline chaser via an 18–20 G cannula in the upper limb at a flow rate of 4 mL/s. The arterial phase was obtained from the aortic arch to the vertex using the following parameters: 100 kV, 300 mA, 0.6 s/rotation, 0.984 mm/rotation table speed, and 1.25 mm section thickness. Early and late venous phases were obtained sequentially from the skull base to the vertex, 8 seconds apart from the previous phase, using the same parameters. Scans were reformatted to a 0.6 mm thickness and 10 mm maximum intensity projection (MIP) images. Further image processing and analysis were performed at a remote workstation.

Image analysis

An emergency radiologist with more than 10 years of experience, blinded to clinical findings, analyzed the images on a remote workstation. The density of the thrombus was measured on NCCT images. In 54 of the 70 patients, the thrombus could be differentiated on NCCT, and the estimated thrombus length was measured. In 16 patients, the estimated thrombus length could not be measured because the thrombus was not depicted in the MCA on NCCT images. Patients' ASPECTS were evaluated on narrow window (window width: 40, window level: 30) NCCT images. The mCTA

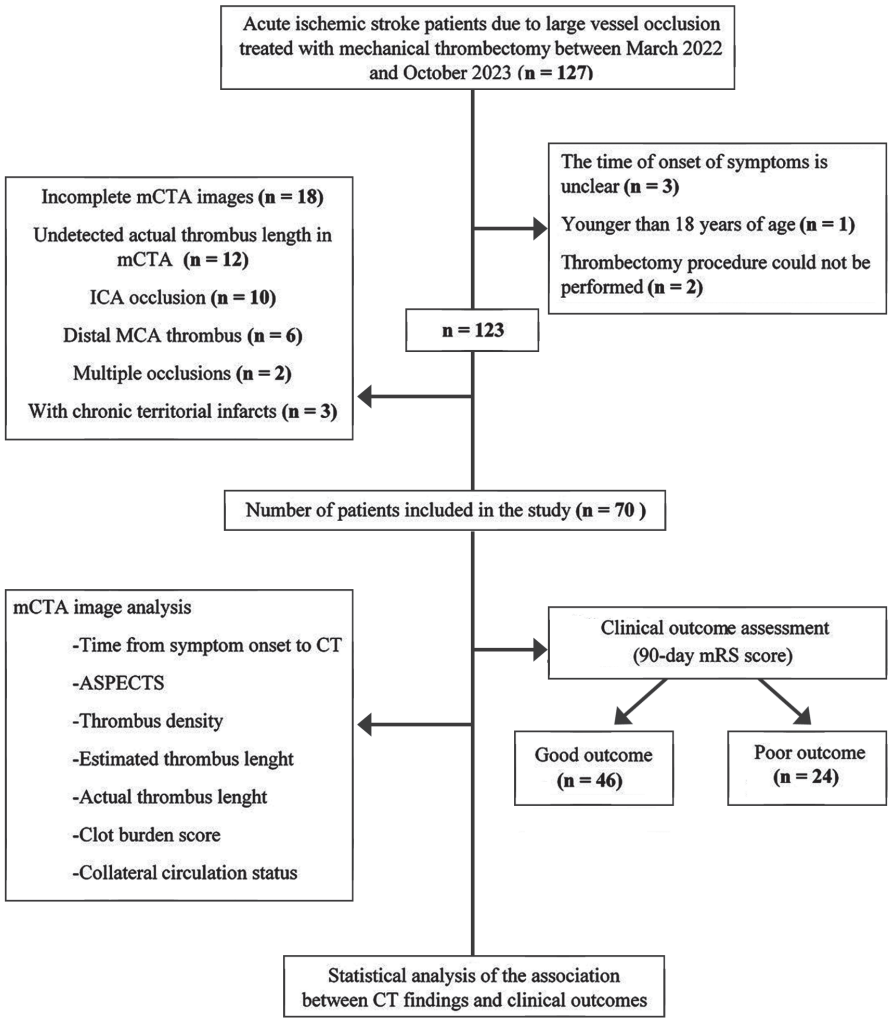


Figure 1. Flowchart of the study. mCTA, multiphase computed tomography angiography; ICA, internal carotid artery; MCA, middle cerebral artery; mRS, modified Rankin scale; ASPECTS, the Alberta Stroke Program Early CT Score.

images were used to measure the actual thrombus length (Figure 2c). Occlusions in curved vessels were calculated on multiplanar reformation images, and when the thrombus extended into branches of the MCA, the longest thrombus length was measured. Based on the pial arterial filling score within the symptomatic ischemic zone, CC status was evaluated on MIP images.[20] Patients were classified as having poor (grade 0–3) or good (grade 4–5) CC. The clot burden score was calculated, and a score of ≤5 was classified as a high thrombus burden.[10]

Assessment of clinical data: patient outcomes

A single neurologist documented patients' demographic data and 90-day mRS scores using electronic medical records. A good clinical outcome was defined as a 90-day mRS score of 0–2, and a poor clinical outcome as an mRS score of 3–6.

Statistical analysis

The demographic, clinical, and radiological characteristics of the patients included in the study were summarized using descriptive statistics. For continuous variables, mean ± standard deviation or median (minimum–maximum) values were used, whereas for categorical variables, frequencies (percentages) were reported.

For intergroup comparisons, the Mann–Whitney U test was applied to continuous variables that did not follow a normal distribution, the chi-square test was used for categorical variables, and for cross-tables with expected value issues, the Fisher exact test was applied for 2 × 2 tables, whereas the Fisher–Freeman–Halton test was used for k × k tables. The threshold for statistical significance was set at $P < 0.05$.

Binary logistic regression analysis was performed to identify independent predictors that could influence clinical outcomes. The dependent variable was defined as an mRS score of 0–2 at 90 days (indicating a good clinical outcome). Independent variables included the time from symptom onset to CTA, the ASPECTS, estimated and actual thrombus length, thrombus density, clot burden score, and patterns of CC. The logistic regression analysis results were reported as β ± standard error, odds ratio (OR), and 95% confidence interval (CI). Model fit was assessed using Cox and Snell R^2 and Nagelkerke R^2 coefficients.

The cut-off value was obtained by applying receiver operating characteristic analysis for the effective numerical independent variables according to the model results. The statistical significance level for regression analysis was set at α = 0.05. All statistical analyses were conducted using the Statistical Package for the Social Sciences (IBM SPSS Statistics for Windows, Version 30.0, IBM Corp., Armonk, NY, USA).

Result

Patient characteristics

The mean age of the patients was 68.56 ± 13.99 years, and the number of female (n = 35) and male (n = 35) patients was similar. Successful thrombectomy (thrombolysis in cerebral infarction 2b–3) was achieved in 66 (94.2%) patients. Of the 70 patients, 46 (65.7%) had a good clinical outcome and 24 (34.2%) had a poor clinical outcome. The 90-day mortality rate was 21.4% (n = 15). The mean age of patients was slightly, but not significantly, lower in those with good clinical outcomes than in those with poor clinical outcomes (63.04 ± 13.81 years versus 69.54 ± 14.57 years, $P = 0.590$). There was no significant difference in gender distribution between the two outcome groups [good clinical outcome: women (n = 24), men (n = 22); poor clinical outcome: women (n = 11), men (n = 13); $P = 0.615$]. The comparison of imaging findings between patients with good and poor clinical outcomes is shown in Table 1. The estimated thrombus length was slightly, but not significantly, shorter in patients with good clinical outcomes than in those with poor outcomes (12.8 mm vs. 15.0 mm; $P = 0.359$). Although a poor CC pattern was observed more frequently in patients with poor clinical outcomes than in those with good outcomes, this difference was not statistically significant ($P = 0.232$). A similar pattern was observed in CC subgroups evaluated using the Fisher–Freeman–Halton test ($P = 0.091$). The ASPECTS, thrombus density, and clot burden score were not significantly dif-

ferent between patients with good and poor clinical outcomes. Figures 2 and 3 show the imaging findings of two different patients with good and poor clinical outcomes, respectively. The mean actual thrombus length of the study population was 17.2 (5–36) mm. The actual thrombus length was shorter in patients with good clinical outcomes than in those with poor outcomes (15.9 mm vs. 21.5 mm; $P = 0.001$) (Figure 4).

The binary logistic regression analysis of the prediction of good clinical outcome by time from onset of symptoms to imaging and baseline mCTA findings in patients with MCA M1 occlusion treated with MT is shown in Table 2. With this model, 37.6% of the variation in the dependent variable was explained according to the Cox and Snell calculation, and 51.8% according to the Nagelkerke calculation. The actual thrombus length [$P = 0.005$; OR: 0.754 (95% CI: 0.61–0.92)] and thrombus density [$P = 0.022$; OR: 1.167 (95% CI: 1.02–1.33)] were predictors of good clinical outcomes in patients with MCA M1 occlusion treated with MT. The optimal cut-off value for actual thrombus length was 18.7 mm, with a sensitivity of 72.5% and specificity of 61.8% [area under the curve: 0.74 ± 0.06 (95% CI: 0.62–0.86); $P = 0.001$] (Figure 5).

Discussion

In this study, we found that actual thrombus length and density measured on mCTA images were useful in predicting clinical outcomes in patients with MCA M1 occlusion treated with MT.

The NCCT is used to assess the presence of intracranial hemorrhage, well-established hypodense ischemia (ASPECTS), and hyperdense vessel signs in patients with suspected acute ischemic stroke, and it also allows measurement of thrombus density, which reflects thrombus composition. Borst et al.[11] evaluated the value of thrombus CT characteristics, including thrombus density, in 199 patients with acute ischemic stroke and found that relative thrombus density on CTA was an independent predictor of functional outcome in multivariable analysis [adjusted common OR of 1.21 per 10% (95% CI: 1.02–1.43); $P = 0.029$].

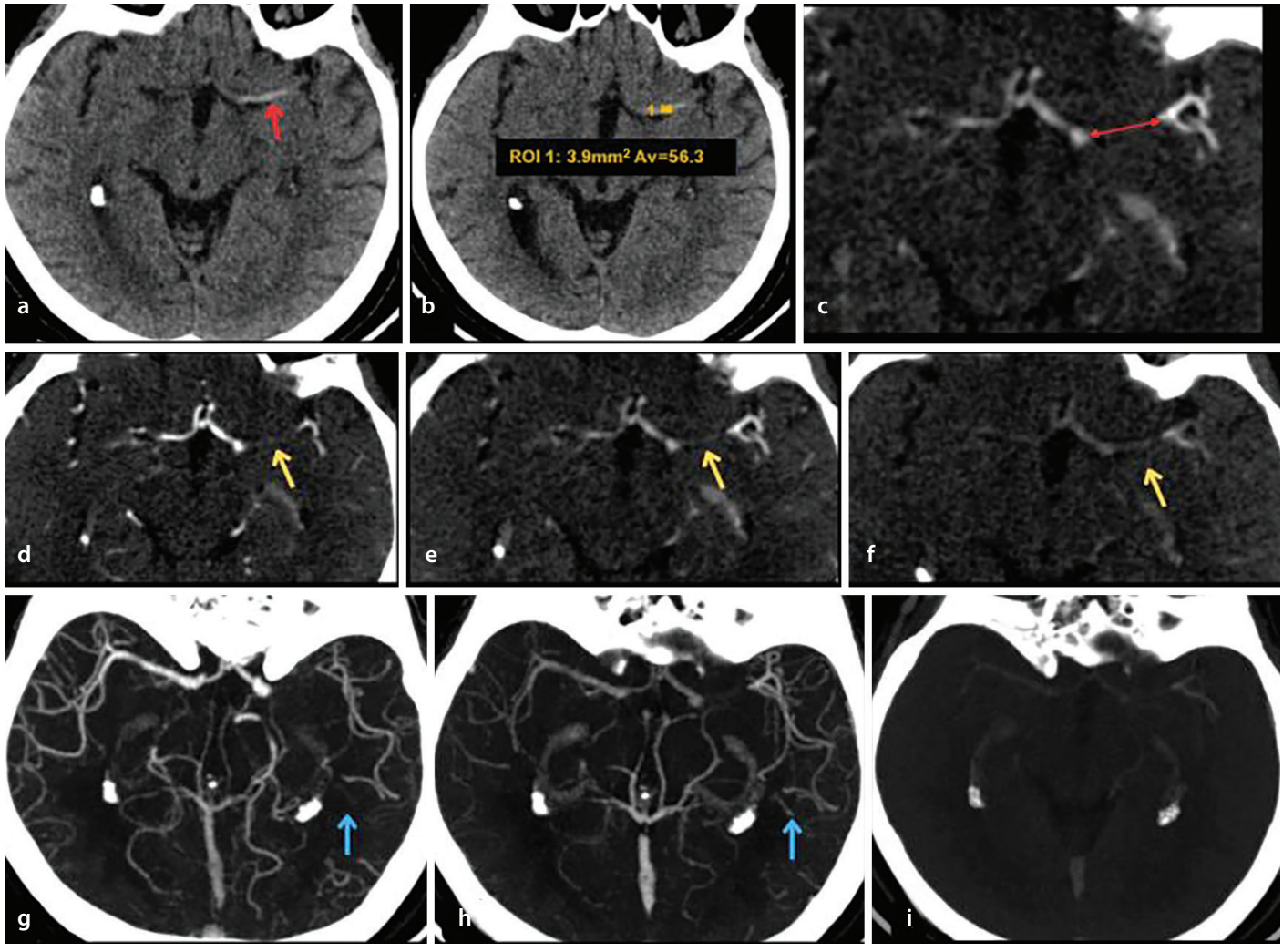


Figure 2. Non-contrast CT **(a, b)** and mCTA images **(c-i)** of a 68-year-old woman with acute ischemic stroke. The time from symptom onset to mCTA was 170 minutes. **(a)** NCCT shows a hyperdense MCA sign (*red arrow*); **(b)** thrombus density was calculated as 56 Hounsfield units (*yellow point*); **(c)** thrombus margins (*red arrow*) were best seen on the early venous phase image, and the actual thrombus length measured 15 mm (not shown). **(d)** Arterial phase, **(e)** early venous phase, and **(f)** late venous phase axial mCTA images show the thrombus (yellow arrows) in the MCA M1 segment. The thrombus was located in the proximal and distal parts of the MCA M1 segment, and the clot burden score was calculated as 10 – 2 – 2 = 6 points. MIP images of **(g)** arterial, **(h)** early venous, and **(i)** late venous phase mCTA images show grade 5 collateral circulation filling (blue arrows) in the peripheral vessels distal to the thrombus. Successful recanalization was achieved with mechanical thrombectomy, and the patient had a good outcome (90-day modified Rankin scale score = 1). NCCT, non-contrast computed tomography; mCTA, multiphase computed tomographic angiography; MIP, maximum intensity projection; CT, computed tomography; MCA, middle cerebral artery.

In a study of 141 patients with acute ischemia undergoing intra-arterial therapy, thrombus density was measured on 4.8 mm thick NCCT images, and no significant correlation was found between thrombus density and functional outcome.[12] In the present study, a significant positive correlation [$P = 0.022$; OR = 1.16 (95% CI: 1.02–1.33)] was found between thrombus density and good clinical outcomes, but there was no statistically significant difference in thrombus density between patients with good and poor outcomes. Thrombus density reflects its structure: erythrocyte-rich thrombi show higher Hounsfield units than platelet-rich thrombi and appear as a hyperdense MCA sign on NCCT.[21] Thrombus composition affects successful recanalization in thrombolytic therapy and thrombectomy,[22] and thus the prognosis of patients.

In previous studies of acute ischemic stroke, patients with a time from symptom onset to recanalization of >300 minutes had better clinical outcomes when the ASPECTS on baseline CT was >7 compared with those with an ASPECTS of ≤7, and MT showed no beneficial impact on the clinical outcomes of patients with a score of ≤4, regardless of time to recanalization.[13,14] Patients with a baseline ASPECTS of ≤7 have a higher probability of good outcomes when reperfusion is achieved within <300 minutes of symptom onset.[13] The current study included patients with a symptom onset of <300 minutes, and the ASPECTS was precisely calculated in narrow window settings (window width: 40, window level: 30) on 2.5 mm thick NCCT images; no statistically significant difference was found between the ASPECTS values of patients with good and poor clinical outcomes. This may have been due to the point reduction in the ASPECTS, regardless of the proportion of each anatomical region affected by ischemia, as all ASPECTS regions may have different predictive abilities for patients' functional outcomes.[23] Though not the subject of this study, parenchymal volumes and localizations affected by ischemia could vary greatly among patients with the same ASPECTS, leading to different prognoses. Broocks et al.[24] reported a retrospective analysis of 100 patients with large vessel occlusion treated with endovascular therapy, showing that even patients with an ASPECTS of ≤5 could have better functional outcomes if they were younger than 73 years and had good collateral status. In our study, patients with poor outcomes were older than those with good outcomes (63.04 ± 13.81 years versus 69.54 ± 14.57 years, $P = 0.590$), and the rate of poor collaterals was higher in the poor outcome group, though not statistically significant (good outcome poor collaterals: 23.9% vs. poor outcome poor collaterals: 37.5%, $P = 0.232$).

mCTA is an increasingly important imaging modality in the initial evaluation of patients with acute ischemic stroke.[25,26] It is useful for detecting large vessel occlusions and showing distal arterial occlusion (delayed vessel sign),[27] and it allows the assessment of CC and the measurement of thrombus length.[9,20,28] In patients with poor CC, arterial phase CTA images may overestimate the actual thrombus length. Delayed phase contrast-enhanced CT images are useful for accurately measuring thrombus length.[28] To overcome this limitation, we excluded patients in whom the actual thrombus length could not be measured because of poor CC. Therefore, we had no patients with grade 0 collateral, and the number of patients with poor CC (20 out of 70) was lower than those with good CC (50 out of 70) in the study group.

Table 1. Comparison of imaging findings between patients with good outcomes and poor outcomes

Imaging findings		All median (min–max) or n (%)	Good clinical outcome median (min–max) or n (%)	Poor clinical outcome median (min–max) n (%)	*P*
Time from symptom onset to CT		111 (26–350)	120 (26–258)	108 (35–350)	0.924[β]
Non-contrast computed tomography					
ASPECTS		9 (4–10)	9 (5–10)	8.5 (4–10)	0.454[β]
>7		47 (67.1%)	32 (69.6%)	15 (62.5%)	0.598[α]
≤7		23 (32.9%)	14 (30.4)	9 (37.5%)	
Thrombus density (HU)		55.0 (37–66)	57.0 (37–66)	53.0 (40–65)	0.053[β]
Estimated thrombus length (mm)		13.0 (5–29)	12.85 (5–29)	15.0 (8–25)	0.359[β]
Multiphase computed tomography angiography					
Actual thrombus length (mm)		17.25 (5–36)	15.9 (5–26)	21.5 (11–36)	0.001[β] *
Clot burden score		6.0 (4–9)	6.0 (4–9)	6.0 (4–9)	0.325[β]
>5		58 (82.9%)	40 (87.0%)	18 (75.0%)	0.316[α]
≤5		12 (17.1%)	6 (13.0%)	6 (25.0)	
Collateral circulation	1	2 (2.9%)	0 (0%)	2 (8.3%)	0.091[θ]
	2	11 (15.7%)	5 (10.9%)	6 (25.0%)	
	3	7 (10%)	6 (13%)	1 (4.2%)	
	4	39 (55.7%)	26 (56.5%)	13 (54.2%)	
	5	11 (15.7%)	9 (19.6%)	2 (18.3%)	
	Poor collateral	20 (28.6%)	11 (23.9%)	9 (37.5%)	
Collateral circulation	Good collateral	50 (71.4%)	35 (76.1%)	15 (62.5%)	0.232[α]

*$P < 0.05$; [α], Fisher's exact test *P* value; β, Mann–Whitney U test *P* value; [θ], Fisher–Freeman–Halton test *P* value. HU, Hounsfield unit; mm, Millimeter; CT, computed tomography; min-max, minimum-maximum; ASPECTS, Alberta Stroke Program Early CT Score

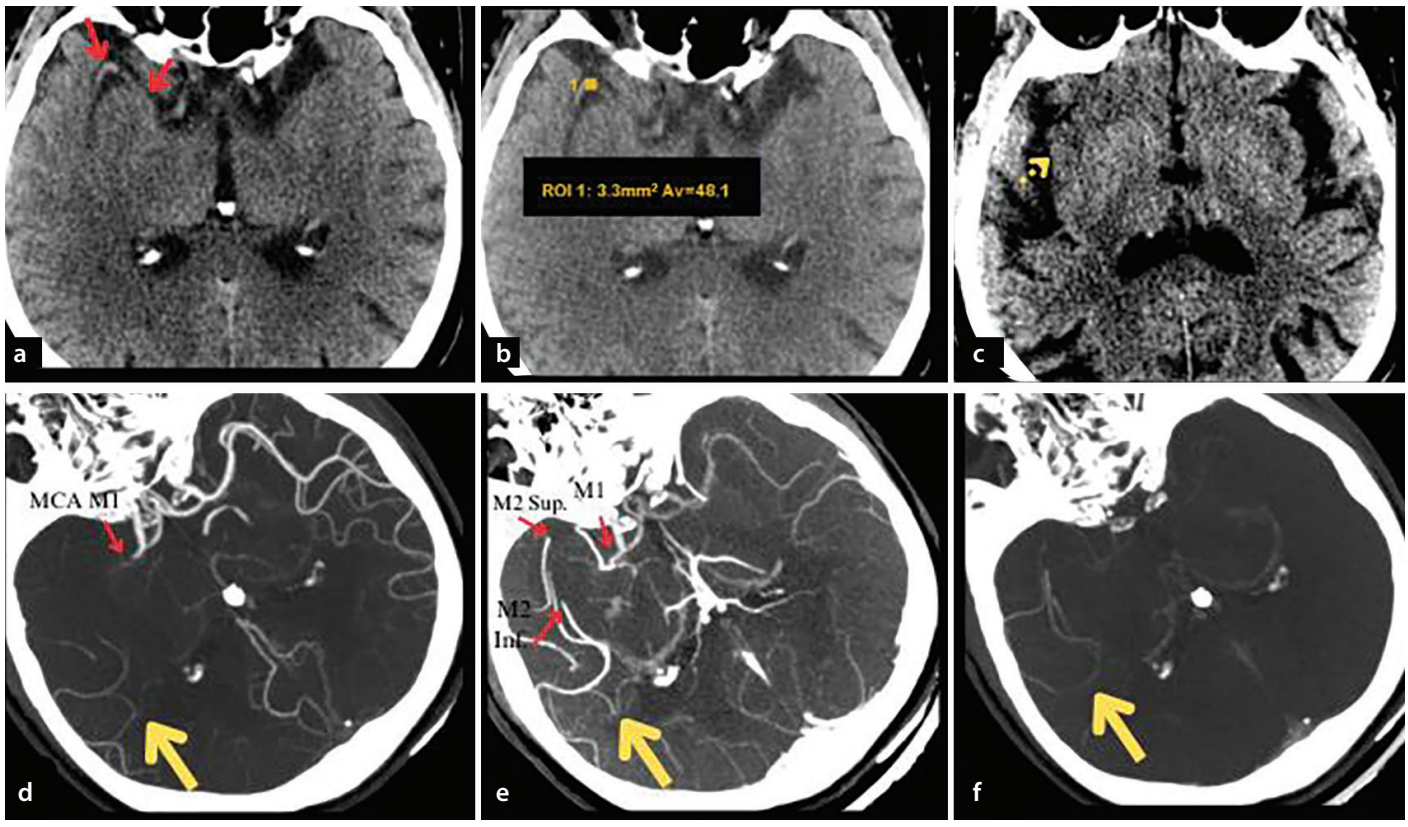


Figure 3. Non-contrast CT and mCTA images **(a-f)** of a 70-year-old man with acute ischemic stroke. The time from symptom onset to mCTA was 47 minutes. **(a)** The proximal M2 segment of the MCA was slightly hyperdense, but the M1 segment was isodense on NCCT (*red arrows*). **(b)** Thrombus density was calculated as 48 Hounsfield units (*yellow square*). **(c)** Axial NCCT image (width: 40; level: 30) at the level of the basal ganglia shows mild hypodensity (*dotted yellow arrow*) consistent with early ischemic changes in the right insula (ASPECTS: 9). **(d)** Arterial and **(e)** early venous phase MIP mCTA images show the thrombus extending from the proximal M1 segment to the superior (M2 sup) and inferior (M2 inf) divisions of the M2 segment in the MCA. The actual thrombus length, best delineated on early venous phase mCTA images (*red arrows*), measured 31 mm from the M1 segment to the M2 inferior division (*not shown*). The clot burden score was calculated as 10 – 2 – 2 – 1 – 1 = 4 points. **(d)** Arterial, **(e)** early venous, and **(f)** late venous phase MIP mCTA images show a one-phase delay in collateral filling (*yellow arrows*) in the peripheral vessels (grade 4 collateral circulation). Although successful recanalization was achieved with mechanical thrombectomy, the patient passed away after hospital discharge (90-day modified Rankin scale score = 6). NCCT, non-contrast computed tomography; mCTA, multiphase computed tomographic angiography; ASPECTS, Alberta Stroke Program Early CT Score; MIP, maximum intensity projection; mCTA, multiphase computed tomographic angiography; CT, computed tomography.

Table 2. Binary logistic regression analysis of the prediction of good clinical outcomes based on baseline multiphasic CTA findings and time from symptom onset to CTA

Independent variables	B ± SE	OR (95% CI)	*P*
Time from symptom onset to CTA	0.001 ± 0.006	1.001 (0.990–1.013)	0.847
Non-contrast computed tomography			
ASPECTS	0.340 ± 0.336	1.404 (0.727–2.713)	0.315
Estimated thrombus length	0.047 ± 0.080	1.048 (0.897–1.225)	0.575
Thrombus density	0.154 ± 0.067	1.167* (1.023–1.331)	0.022*
Multiphase computed tomography angiography			
Actual thrombus length	−0.282 ± 0.104	0.754* (0.615–.925)	0.005*
Clot burden score	−0.398 ± 0.457	0.671 (0.274–1.643)	0.398
Collateral circulation	−0.025 ± 0.442	0.975 (0.410–2.320)	0.207
Poor collateral (1)	0.340 ± 0.915	1.404 (0.234–8.432)	0.124
Poor to good (2)	−2.13 ± 41.690	0.118 (0.004–3.247)	0.306

$P < 0.05$; dependent variable, 90-day mRS (0–2); R^2, Cox and Snell: 0.376; Nagelkerke: 0.518; Hosmer–Lemeshow goodness-of-fit test, χ^2: 4.720; *$P > 0.05$. CTA, computed tomography angiography; ASPECTS, Alberta Stroke Program Early Computed Tomography Score; CI, confidence interval; OR, odds ratio

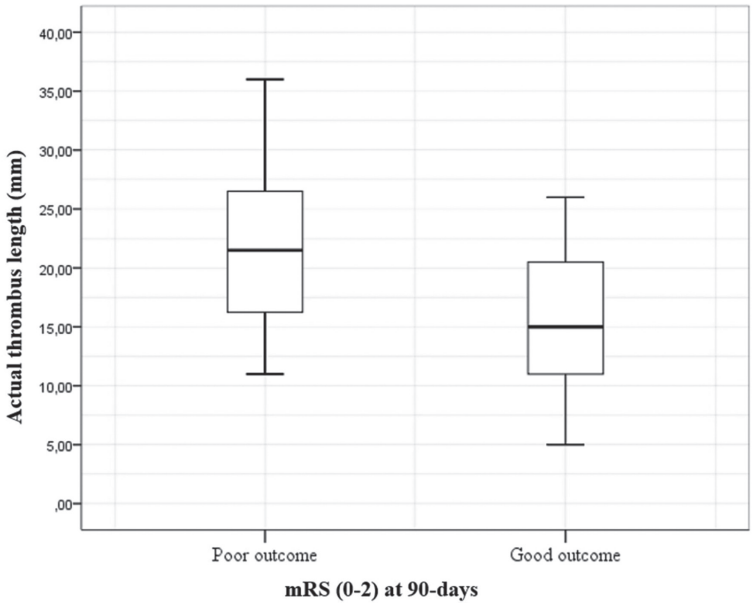


Figure 4. Comparison of the actual thrombus length between patients with poor and good clinical outcomes is shown using box plots. mRS, modified Rankin scale.

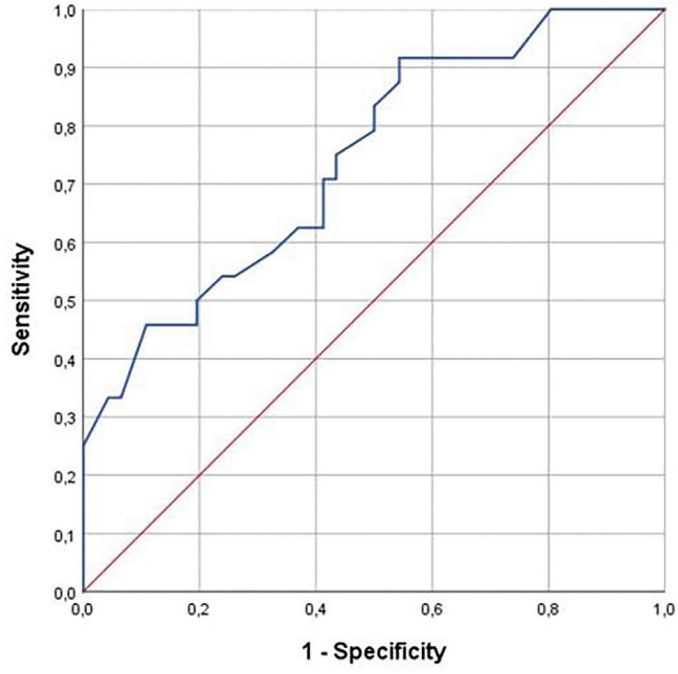


Figure 5. Receiver operating characteristic curve analysis of the actual thrombus length for predicting good clinical outcomes in patients with MCA occlusion treated with mechanical thrombectomy. The optimal cut-off value of thrombus length was 18.7 mm, with a sensitivity of 72.5% and specificity of 61.8% (area under the curve: 0.74 ± 0.06; 95% confidence interval: 0.62–0.86; $P = 0.001$). MCA, middle cerebral artery.

The probability of recanalization with intravenous thrombolysis is almost impossible when the thrombus length exceeds 8 mm,[29] and it remains controversial whether a longer thrombus affects recanalization in thrombectomy and procedural or post-procedural complications.[15,30] Borst et al.[11] measured thrombus length on NCCT images coregistered with CTA using software that calculated thrombus length marked on images and reported that a median thrombus length greater than 8 mm was associated with functional outcomes in patients with acute ischemic stroke ($P < 0.05$). Seker et al.[15] analyzed the impact of thrombus length on neurological outcomes in 72 patients with M1 occlusion treated with MT. They used at least two perpendicular planes to measure thrombus length in multiplanar reformatted MR angiography or CTA images, and when they were uncertain about the distal extension of the thrombus, they used susceptibility-weighted MR sequences or NCCT scans to estimate thrombus length. They found no association between thrombus length and the probability of a good clinical outcome (OR: 0.95; 95% CI: 0.84–1.03; $P = 0.176$).[15] Seker et al.[15] used MRI, CTA, and NCCT images to depict thrombus length, whereas in our study, we used mCTA images, which may be more accurate than sCTA and NCCT in depicting thrombus length. Spiotta et al.[12] calculated the estimated thrombus length on sCTA MIP images of 141 patients undergoing intra-arterial therapy for acute ischemia. By measuring the filling defect in the affected vessel, they found no significant associations between either thrombus length or functional outcome at 90 days.[12] The current study showed a significant negative correlation between actual thrombus length and good clinical outcome ($P = 0.005$; OR: 0.754; 95% CI: 0.61–0.92) and found that an actual thrombus length of 18.7 mm was the cut-off value for estimating good clinical outcome, with a sensitivity of 72.5% and a specificity of 61.8% [$P = 0.001$ (95% CI: 0.62–0.86)]. These discrepancies between studies may be because each researcher used a different method or modality to measure the estimated thrombus length.

Lee and Bang[16] reviewed recent reports and summarized the relationship between CC status and clinical outcomes in patients with acute ischemic stroke treated with thrombectomy. There are two studies in the literature estimating prognosis using the pial arterial filling score in mCTA, but unlike our study, the treatment methods in these studies were heterogeneous and did not include only patients treated with MT.[17,18] Drozdov et al.[17] found that the area under the curve of CC status (pial arterial filling score >3 vs. ≤3) in predicting a good outcome was 0.66 [OR: 4.11 (1.35–12.54); $P = 0.013$]. In the present study, no correlation was found between CC grade and clinical outcomes; this may be due to the smaller number of patients with poor collaterals in the study group for the reasons mentioned above. Larger series studies on mCTA collateral scoring in patient groups treated with the same method may clarify this issue.

Regarding the impact of the clot burden score on clinical outcomes, it is known that patients are considerably more likely to have good clinical outcomes and less likely to die as the clot burden score increases. Puetz et al.[10] demonstrated that high thrombus burden (clot burden score ≤5) was a predictor of mortality and functional outcome in patients with acute ischemic stroke. The present study found no association between clot burden score and clinical outcomes. This discrepancy might be due to the exclusion of patients with terminal ICA and/or anterior cerebral artery occlusion.

Our study has several limitations. It is a single-center retrospective study. To maintain homogeneity of the study group, we included only patients with MCA M1 occlusion who had sufficient CC to allow clear delineation of the thrombus edges, which resulted in a relatively small sample size. The small sample size may limit the reliability and generalizability of the results. The evaluation of CT images and measurements was performed by a single observer. We could not assess the effect of perfusion parameters on clinical outcomes, as perfusion CT is not routinely used in the management of patients with suspected acute stroke at our institution. Another limitation was the lack of venous flow evaluation.

In conclusion, higher thrombus density and thrombus lengths shorter than 18.7 mm were associated with good clinical outcomes in patients with MCA M1 occlusion treated with MT. However, no considerable correlation was found between clinical outcomes and the ASPECTS, CC degree, or clot burden score. Further multicenter studies with larger patient groups are needed to confirm these findings and to clarify the role of thrombus characteristics and other imaging features that may influence patient outcomes.

Acknowledgments

We are grateful to Assoc. Prof., M.D İnci Elif Erbahçeci Timur, Assoc. Prof., M.D. İzzet Selçuk Parlak, Assoc. Prof., M.D. Semra Duran, Prof., M.D. Bige Sayın, Assoc. Prof., M.D. Hatice Gül Hatipoğlu Çetin from the Department of Radiology, Ankara Bilkent City Hospital, and their invaluable support.

Footnotes

Conflict of Interest

The authors declared no conflicts of interest.

References

1. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *Int J Stroke*. 2022;17(1):18-29. Erratum in: *Int J Stroke*. 2022;17(4):478. [Crossref]
2. Institute for Health Metrics and Evaluation (IHME). Disease, injury, and risk factsheets: stroke-level cause [Internet]. Seattle (WA): IHME; 2021 [cited 2024 Nov 3]. [Crossref]
3. Institute for Health Metrics and Evaluation (IHME). Disease, injury, and risk factsheets: ischaemic stroke–level 4 cause [Internet]. Seattle (WA): IHME; 2021 [cited 2024 Nov 3]. [Crossref]
4. Laredo C, Zhao Y, Rudilosso S, et al. Prognostic significance of infarct size and location: the case of insular stroke. *Sci Rep*. 2018;8(1):9498. [Crossref]
5. Kodumuri N, Sebastian R, Davis C, et al. The association of insular stroke with lesion volume. *Neuroimage Clin*. 2016;11:41-45. [Crossref]
6. Malhotra K, Gornbein J, Saver JL. Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: a review. *Front Neurol*. 2017;8:651. [Crossref]
7. Abdalkader M, Siegler JE, Lee JS, et al. Neuroimaging of acute ischemic stroke: multimodal imaging approach for acute endovascular therapy. *J Stroke*. 2023;25(1):55-71. [Crossref]
8. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2019;50(12):e344-e418. Epub 2019 Oct 30. Erratum in: *Stroke*. 2019;50(12):e440-e441. [Crossref]
9. de Oliveira EP, Fiebach JB, Vagal A, Schaefer PW, Aviv RI. Controversies in imaging of patients with acute ischemic stroke: *AJR* expert panel narrative review. *AJR Am J Roentgenol*. 2021;217(5):1027-1037. [Crossref]
10. Puetz V, Dzialowski I, Hill MD, et al. Intracranial thrombus extent predicts clinical outcome, final infarct size and hemorrhagic transformation in ischemic stroke: the clot burden score. *Int J Stroke*. 2008;3(4):230-236. [Crossref]
11. Borst J, Berkhemer OA, Santos EMM, et al. Value of thrombus CT characteristics in patients with acute ischemic stroke. *AJNR Am J Neuroradiol*. 2017;38(9):1758-1764. [Crossref]
12. Spiotta AM, Vargas J, Hawk H, et al. Hounsfield unit value and clot length in the acutely occluded vessel and time required to achieve thrombectomy, complications and outcome. *J Neurointerv Surg*. 2014;6(6):423-427. [Crossref]
13. Goyal M, Menon BK, Coutts SB, et al. Effect of baseline CT scan appearance and time to recanalization on clinical outcomes in endovascular thrombectomy of acute ischemic strokes. *Stroke*. 2011;42(1):93-97. [Crossref]
14. Menon BK, Puetz V, Kochar P, Demchuk AM. ASPECTS and other neuroimaging scores in the triage and prediction of outcome in acute stroke patients. *Neuroimaging Clin N Am*. 2011;21(2):407-423. [Crossref]
15. Seker F, Pfaff J, Wolf M, et al. Impact of thrombus length on recanalization and clinical outcome following mechanical thrombectomy in acute ischemic stroke. *J Neurointerv Surg*. 2017;9(10):937-939. [Crossref]
16. Lee JS, Bang OY. Collateral status and outcomes after thrombectomy. *Transl Stroke Res*. 2023;14(1):22-37. [Crossref]
17. Drozdov AA, Arora M, Sheikhy A, Leon Guerrero CR, Taheri MR. Anterior ischemic stroke: analysis of the multivariable CT-based models for prediction of clinical outcome. *J Stroke Cerebrovasc Dis*. 2023;32(8):107242. [Crossref]
18. Lu SS, Zhang X, Xu XQ, et al. Comparison of CT angiography collaterals for predicting target perfusion profile and clinical outcome in patients with acute ischemic stroke. *Eur Radiol*. 2019;29(9):4922-4929. [Crossref]
19. Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JH, Hudon ME, et al. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet*. 2000;355(9216):1670–1674. [Crossref]
20. Menon BK, d'Esterre CD, Qazi EM, et al. Multiphase CT angiography: a new tool for the imaging triage of patients with acute ischemic stroke. *Radiology*. 2015;275(2):510-20. [Crossref]
21. Liebeskind DS, Sanossian N, Yong Whet al. CT and MRI early vessel signs reflect clot composition in acute stroke. Stroke. 2011;42(5):1237-1243. [Crossref]
22. Moftakhar P, English JD, Cooke DL, et al. Density of thrombus on admission CT predicts revascularization efficacy in large vessel occlusion acute ischemic stroke. *Stroke*. 2013;44(1):243-245. [Crossref]
23. Shin DH, Shin DJ, Kim JR. Do all ASPECT score regions have the same predictive power for functional outcomes? *J Stroke Cerebrovasc Dis*. 2020;29(2):104516. [Crossref]
24. Broocks G, Kniep H, Schramm P, et al. Patients with low Alberta stroke program early CT score (ASPECTS) but good collaterals benefit from endovascular recanalization. *J Neurointerv Surg*. 2020;12(8):747-752. [Crossref]

25. Dundamadappa S, Iyer K, Agrawal A, Choi DJ. Multiphase CT angiography: a useful technique in acute stroke imaging-collaterals and beyond. *AJNR Am J Neuroradiol*. 2021;42(2):221-227. [Crossref]

26. Busto G, Morotti A, Carlesi E, et al. Pivotal role of multiphase computed tomography angiography for collateral assessment in patients with acute ischemic stroke. *Radiol Med*. 2023;128(8):944-959. [Crossref]

27. Byrne D, Sugrue G, Stanley E, et al. Improved detection of anterior circulation occlusions: the "delayed vessel sign" on multiphase CT angiography. *AJNR Am J Neuroradiol*. 2017;38(10):1911-1916. [Crossref]

28. Mortimer AM, Little DH, Minhas KS, Walton ER, Renowden SA, Bradley MD. Thrombus length estimation in acute ischemic stroke: a potential role for delayed contrast enhanced CT. *J Neurointerv Surg.* 2014;6(3):244-248. [Crossref]

29. Riedel CH, Zimmermann P, Jensen-Kondering U, Stingele R, Deuschl G, Jansen O. The importance of size: successful recanalization by intravenous thrombolysis in acute anterior stroke depends on thrombus length. *Stroke.* 2011;42(6):1775-7. [Crossref]

30. Belachew NF, Dobrocky T, Meinel TR, et al. Risks of undersizing stent retriever length relative to thrombus length in patients with acute ischemic stroke. *AJNR Am J Neuroradiol.* 2021;42(12):2181-2187. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253311

HEAD AND NECK IMAGING

ORIGINAL ARTICLE

Vascularity assessment in Hashimoto's thyroiditis: a prospective comparative study with power Doppler and superb microvascular imaging

Tuba Selçuk Can[1]
Sevim Özdemir[1]
Türkan İkizceli[1]
Behice Kaniye Yılmaz[2]
Mehmet Akif Sarı[1]
Rüştü Türkay[1]
Özlem Doğan[3]

[1]University of Health Sciences Türkiye, İstanbul Haseki Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

[2]University of Health Sciences Türkiye, Kanuni Sultan Süleyman Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

[3]University of Health Sciences Türkiye, İstanbul Haseki Training and Research Hospital, Clinic of Endocrinology, İstanbul, Türkiye

Corresponding author: Tuba Selçuk Can

E-mail: drtubas@gmail.com

Received 19 March 2025; revision requested 04 May 2025; last revision received 28 June 2025; accepted 27 July 2025.

Epub: 27.08.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253311

PURPOSE

To quantitatively evaluate the vascularity of the thyroid parenchyma in patients diagnosed with Hashimoto's thyroiditis (HT) compared with healthy controls by using vascularity index (VI) through power Doppler (PD) and color superb microvascular imaging (cSMI) and to determine a threshold VI value to effectively differentiate patients with HT and hypothyroid HT.

METHODS

This prospective cross-sectional study involved 73 patients diagnosed with HT and 66 healthy controls. The diagnosis of HT was established based on clinical and laboratory findings. The total volume of the thyroid gland was measured, and the region of interest was drawn manually by delineating the gland boundaries for VI calculation on PD and cSMI. The mean VI for both lobes of the thyroid were computed for each participant. Statistical analyses were conducted using SPSS version 29.0, with receiver operating characteristic curve analysis employed to ascertain the optimal cSMI VI cut-off values for the diagnosis of HT and for patients with hypothyroid HT.

RESULTS

The analysis revealed no significant differences in the total thyroid volume between the HT group and the control group, or between the hypothyroid and euthyroid HT subgroups. The SMI VI values were recorded at 8.85 [interquartile range (IQR): 25%–75%, 6.55–12.6] for patients with HT and 8.40 (IQR: 25%–75%, 6.70–12.8) for the control group, indicating a statistically significant increase in the HT cohort ($P < 0.001$). Additionally, the PD VI values in patients with HT were significantly higher than in the control group ($P < 0.001$). A strong positive correlation was identified between thyroid-stimulating hormone levels and cSMI VI in patients with HT (rho = 0.739, $P < 0.001$), whereas the correlation with PD VI was found to be weak (rho = 0.346, $P < 0.001$). The optimal cut-off value for SMI VI was 6.75% for the general diagnosis of HT and 8.825% for patients with hypothyroid HT.

CONCLUSION

This study indicates that the optimal threshold values of 6.75% for the diagnosis of HT and 8.825% for patients with hypothyroid HT suggest that cSMI is an effective and promising diagnostic tool for detecting alterations in thyroid vascularization. Furthermore, there is a strong concordance among radiologists regarding the VI measurements.

CLINICAL SIGNIFICANCE

The SMI technique represents a promising diagnostic tool for the detection of subtle alterations in thyroid vascularization. The higher sensitivity of cSMI in comparison to PD positions it as an innovative and effective technology for the assessment of HT, offering valuable insights into disease activity and progression.

KEYWORDS

Hashimoto's thyroiditis, power Doppler, ultrasonography, color superb microvascular imaging, vascularity index

You may cite this article as: Selçuk Can T, Özdemir S, İkizceli T, et al. Vascularity assessment in Hashimoto's thyroiditis: a prospective comparative study with power Doppler and superb microvascular imaging. *Diagn Interv Radiol*. 2026;32(3):315-321.

Hashimoto's thyroiditis (HT) is an autoimmune disorder characterized by inflammation of the thyroid gland, which initially results in hyperthyroidism and subsequently progresses to hypothyroidism due to parenchymal degeneration.[1] The diagnosis of HT can be established through clinical observations, ultrasonographic findings, and the detection of circulating antithyroid autoantibodies that contribute to morphological changes.[2,3] Conventional ultrasound (US) and color Doppler imaging (CDI) typically reveal a heterogeneous echotexture characterized by lobulated contours, hypoechoic micronodules, and increased vascularity in the early stages, while diminished vascularity is observed in the chronic stages.[4,5] CDI, particularly power Doppler (PD), is essential for assessing the vascularity of the thyroid gland, which is often elevated in cases of HT. When combined with other sonographic characteristics, it enhances both the sensitivity and specificity of diagnosing diffuse thyroid pathologies, including HT. This comprehensive approach improves diagnostic accuracy in asymptomatic individuals.[6,7] However, CDI faces challenges in distinguishing genuine blood flow from motion artifacts. Superb microvascular imaging (SMI) is an advanced Doppler technique that employs adaptive algorithms to detect low-velocity blood flow while minimizing motion artifacts, thereby providing a clearer representation of true vascular perfusion. This technique can be utilized in two modalities: color SMI (cSMI) and monochrome SMI (mSMI). The cSMI modality generates color-coded Doppler signals superimposed on grayscale US images, whereas mSMI eliminates the background grayscale US data and exclusively displays the Doppler signals.[8] The vascularity index (VI), available in both PD and SMI, quantifies blood flow by calculating the ratio of colored pixels within a designated region of interest (ROI).[9,10]

This prospective cross-sectional study quantitatively evaluates the vascularity of the thyroid parenchyma in patients with HT and a control group by employing the VI in conjunction with PD and cSMI methodologies, thereby facilitating a comparative analysis of these techniques. Additionally, we establish a threshold VI value via cSMI to differentiate patients with HT from controls and to distinguish between hypothyroid and euthyroid cases of HT.

Main points

- This study compares power Doppler (PD) and color superb microvascular imaging (cSMI) techniques to assess thyroid parenchymal vascularity in patients with Hashimoto's thyroiditis (HT).
- The results indicate a significantly higher vascularity index (VI) in patients with HT compared with healthy controls, with cSMI demonstrating superior sensitivity over PD.
- A strong positive correlation was observed between thyroid-stimulating hormone levels and SMI VI, suggesting that SMI VI can be used to assess disease activity.
- The study highlights cSMI as a promising diagnostic tool for detecting subtle vascular changes in HT, with potential clinical applications for disease monitoring and management.

Methods

This study was approved by the Clinical Research Ethics Committee of İstanbul Haseki Training and Research Hospital (approval date: May 25, 2022; decision number: 73-2022). In this prospective cross-sectional investigation, the thyroid glands of 73 individuals diagnosed with HT and 66 healthy asymptomatic participants were comprehensively evaluated. The minimum required sample size to achieve a 95% confidence interval (CI) ($\alpha = 0.05$) and 90% power was determined to be 18 participants per group, resulting in a total of 54 participants, as calculated based on the study entitled "Vascularity Index for the Diagnosis of Autoimmune Thyroid Disease".[11]

Patients diagnosed with HT and monitored at the endocrinology outpatient clinic were included in this study (Figure 1). The diagnosis of HT was established through the detection of thyroid autoantibodies, specifically thyroid peroxidase and thyroglobulin antibodies (TgAbs), in conjunction with thyroid hormone levels, clinical manifestations, and ultrasonographic findings indicative of thyroiditis. Laboratory assessments conducted over a 1-week period were excluded from the analysis. Medical personnel who consented to participate in the investigation during their routine hospital assessments, who exhibited no indications of autoimmune disorders and exhibited normal laboratory findings, were incorporated into the control cohort.

The reference values employed in University of Health Sciences Türkiye, İstanbul Haseki Training and Research Hospital for laboratory parameters are delineated as follows: free thyroxine (T4) levels range from 0.70 to 1.74 ng/dL, thyroid-stimulating hormone (TSH) levels range from 0.35 to 4.5 mIU/L, TgAbs range from 0 to 115 IU/mL, thyroid peroxidase antibodies (TPOAbs) range from 0 to 34 IU/mL, and thyroglobulin levels range from 1.6 to 60 ng/mL. The cohort of patients diagnosed with HT was subsequently stratified into hypothyroid and euthyroid subgroups.

Ultrasonographic evaluations were performed utilizing the Canon Aplio 500 US system (Canon Medical Systems, Tokyo, Japan), equipped with a high-frequency probe (4–14 MHz). The examinations were conducted by two radiologists with 15 years (S.O.) and 10 years (T.S.C.) of relevant experience. Participants were positioned in a supine posture with their necks slightly extended. Initially, a grayscale US examination was conducted. Each thyroid lobe was assessed separately in both transverse and longitudinal planes.

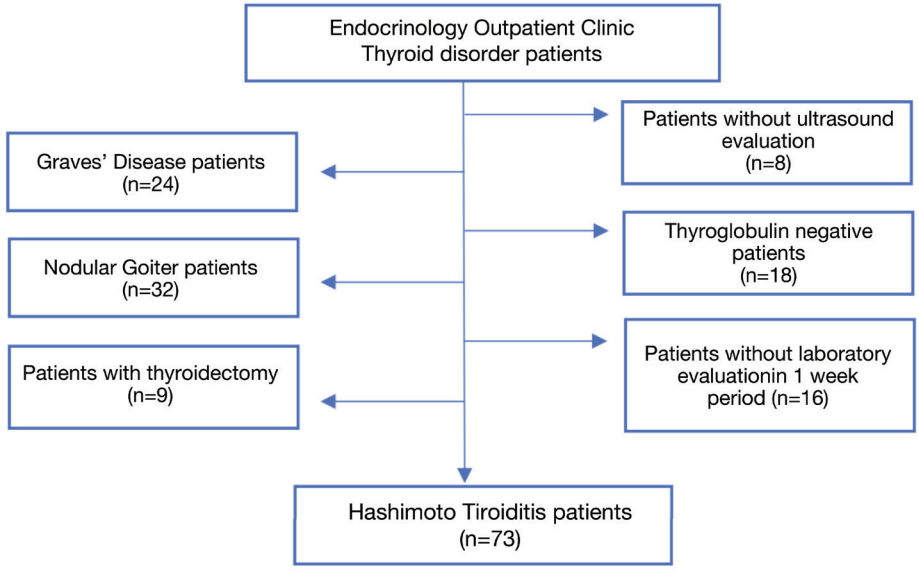


Figure 1. Flow diagram of the study.

The dimensions of the thyroid gland were measured across three planes for each lobe, and the thyroid volume for each lobe was calculated using the following formula: length × width × depth × 0.52. The total thyroid volume was determined by summing the volumes of both lobes while excluding the volume of the isthmus from the calculation. Furthermore, the structure and echogenicity of the thyroid parenchyma were evaluated in detail.

Following the grayscale examination, vascular imaging was performed in the longitudinal plane for both lobes using PD and cSMI techniques. The pulse repetition frequency was calibrated within a range of 150–80 Hz. During the PD and cSMI evaluations, participants were instructed to refrain from swallowing and to hold their breath for a duration of 5 seconds. After capturing the relevant images, the boundaries of the thyroid gland were manually delineated, and a ROI was established (Figure 2). The VI was calculated by quantifying the colored pixels using the device's algorithm and expressing this quantity as a proportion of the total pixel count within the ROI. This procedure was conducted independently three times for each lobe, and the mean of the three measurements was computed. To determine the overall thyroid VI, the average of the measurements from both lobes was calculated. All values were derived by averaging the VI results obtained from both radiologists.

Statistical analysis

The statistical analysis was conducted utilizing SPSS version 29.0 (IBM, Armonk, NY, USA). Categorical variables were represented as counts and percentages, and numerical variables were expressed as means ± standard deviations or medians with interquartile ranges (IQR), depending on the distribution characteristics of the data. To compare independent numerical variables between two distinct groups, either Student's t-test or the Mann–Whitney U test was employed, based on the normality assumption of the data distribution. Receiver operating characteristic (ROC) curve analysis was performed to determine the optimal cut-off value for the cSMI VI in distinguishing patients with HT from healthy controls, as well as in distinguishing hypothyroid from euthyroid HT. The data were stratified according to the threshold values established by the Youden index and the area under the curve (AUC) metrics. Subsequently, measures of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, and negative likelihood ratio were calculated. The study also assessed the concordance of findings among each radiologist, which was evaluated using the intraclass correlation coefficient (ICC). The interpretation of Kappa values was categorized as follows: 0.81–1.00 (very good), 0.61–0.80 (good), 0.41–0.60 (moderate), 0.21–0.40 (fair), and <0.20 (poor).

Results

In the current study, a cohort of 73 patients diagnosed with HT, aged between 11 and 78 years, was designated as the case group. A control group consisting of 66 healthy individuals, aged from 19 to 72 years, was also established. The median age of the 66 healthy controls, which included 43 women and 23 men, was determined to be 33 years (IQR: 25%–75%, 25.3–44). In contrast, the mean age of the 73 individuals in the HT group, comprising 63 women and 10 men, was calculated to be 38.8 ± 14.1 years. Statistical analysis revealed that the age differences between the control group and the case group were not statistically significant ($P = 0.187$, Mann–Whitney U test). Within the case group, the mean age of the 35 patients with hypothyroidism was recorded as 38.7 ± 12.8 years, whereas the mean age of the 38 patients with euthyroidism was 38.9 ± 15.3 years. The statistical evaluation indicated no significant age difference between the patients with hypothyroidism and those with euthyroidism within the HT cohort ($P = 0.950$, Student's t-test).

The differences in thyroid function assessments between the HT group and the control cohort were found to be statistically significant. Median TSH concentration was 3.90 mIU/L (IQR: 25%–75%, 1.73–6.33) in the HT group and 2.55 mIU/L (IQR: 25%–75%, 1.59–3.33) in the control group. In the HT group, median TSH concentration was 6.39 mIU/L (IQR: 25%–75%, 5.23–10.1) in the hypothyroid subgroup and 2.07 ± 1.08 mIU/L in the euthyroid subgroup. This showed that TSH levels were significantly elevated in the patients with HT compared with the control group, and were also significantly higher in the hypothyroid subgroup compared with the euthyroid subgroup ($P < 0.001$, Mann–Whitney U test). Median T4 levels were 11.6 ng/L (IQR: 25%–75%, 10.7–12.7) in the

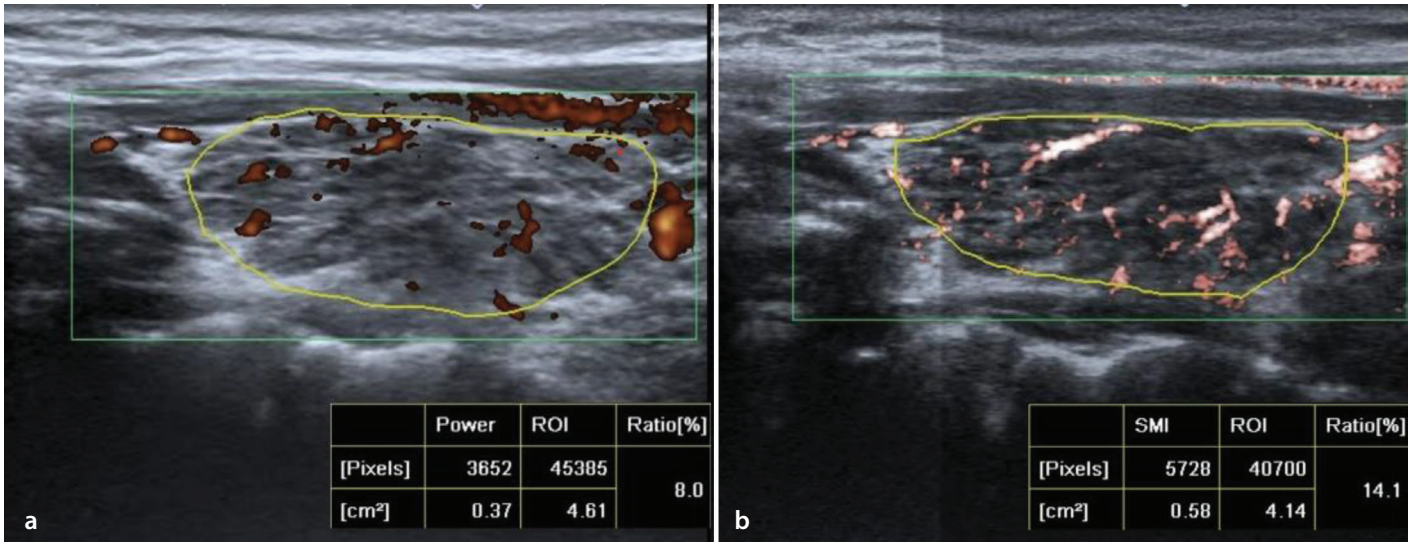


Figure 2. Quantitative vascularity index values were determined by manually tracing the contours of the thyroid gland structure in **(a)** power Doppler and **(b)** color superb microvascular imaging mode. The measurements were performed in the longitudinal planes of the right lobe. ROI, region of interest; SMI, superb microvascular imaging.

euthyroid group and 10.9 ng/L (IQR: 25%–75%, 9.14–11.6) in the hypothyroid group, showing notable reduction in the latter subgroup compared with the former ($P < 0.001$, Mann–Whitney U test). However, no statistically significant difference was observed in T3 levels.

There was no statistically significant difference observed between the HT group and the control group, or between the hypothyroid and euthyroid HT groups, regarding the volume of both thyroid lobes and the total thyroid volume, as determined using the Mann–Whitney U test (Table 1).

The SMI VI values for the right and left thyroid lobes in the control cohort were measured at 3.75% (IQR: 25%–75%, 2.82–5.16) and 3.97% ± 1.34%, respectively. In contrast, the HT cohort exhibited SMI VI values of 8.85% (IQR: 25%–75%, 6.55–12.6) for the right lobe and 8.40% (IQR: 25%–75%, 6.70–12.8) for the left lobe. Importantly, the SMI VI values for both lobes demonstrated a statistically significant increase in the HT cohort ($P < 0.001$, Mann–Whitney U test).

In the HT cohort, the SMI VI values for the right lobe were measured at 14.0% ± 4.53% for the hypothyroid subgroup and 6.77% ± 2.08% for the euthyroid subgroup. A statistically significant difference was observed, favoring the hypothyroid cohort ($P < 0.001$, Student's t-test). Regarding the left lobe, the median SMI VI value of 12.8% (IQR: 25%–75%, 10.8–14.1) in the hypothyroid cohort was significantly higher than the 6.78% (IQR: 25%–75%, 5.11–7.75) in the euthyroid cohort ($P < 0.001$, Mann–Whitney U test) (Table 2).

The PD VI metrics for the right and left lobes in patients diagnosed with HT were measured at 6.15% (IQR: 25%–75%, 3.65–8.05) and 6.30% (IQR: 25%–75%, 4.10–7.75), respectively. These values were found to be statistically significantly elevated when compared with the mean values of the corresponding lobes in the control cohort, which were 3.28% ± 1.35% and 3.35% ± 1.32%, respectively ($P < 0.001$, Mann–Whitney U test). Furthermore, the PD VI measurements obtained from the right and left lobes of the patients with hypothyroid HT were 7.35% (IQR: 25%–75%, 5.35–11) and 7.30 (IQR: 25%–75%, 5–8.50), respectively. These measurements were found to be significantly greater than those observed in the patients with euthyroid HT, which were 5.22% (IQR: 25%–75%, 3.26–6.86) and 4.80% (IQR: 25%–75%, 3.65–6.64), respectively, with statistical significance ($P = 0.002$, $P = 0.003$; Mann–Whitney U test) (Table 3).

When interobserver agreement was evaluated in VI values, the ICC value for cSMI was 0.994 [F (179,72) = 57.2, $P < 0.001$, 95% CI: 0.990–0.996] for the right lobe and 0.995 [F (214,72) = 57.7, $P < 0.001$, 95% CI: 0.992–0.997] for the left lobe. For PD VI, the ICC value was 0.995 [F (210,72) = 33.7, $P < 0.001$, 95% CI: 0.990–0.997] for the right lobe and 0.997 [F (405,72) = 21.9, $P < 0.001$, 95% CI: 0.994–0.998] for the left lobe. This shows a high level of agreement between the two radiologists regarding the SMI and PD VI measurements.

The results of the Spearman correlation analysis performed on a cohort of individuals diagnosed with HT revealed a moderate positive correlation between the right SMI VI and right PD VI (rho = 0.613, $P < 0.001$). Additionally, a similar moderate positive correlation was found between the left SMI VI and left PD VI (rho = 0.607, $P < 0.001$).

The results of the Spearman correlation analysis conducted on the control group revealed a strong positive correlation between right SMI VI and right PD VI (rho = 0.727, $P < 0.001$). Additionally, a robust positive correlation was identified between left SMI VI and left PD VI (rho = 0.825, $P < 0.001$).

When the correlation between TSH values and VI values was assessed in patients with HT, a strong positive correlation was found between TSH and SMI VI (rho = 0.739, $P < 0.001$). In contrast, a weak correlation was noted between TSH and PD VI values (rho = 0.346, $P < 0.001$).

The correlation between SMI VI and the antibodies, anti-TG and anti-TPO, was found to be weak, with Spearman's rho values of

Table 1. Thyroid gland volumes of Hashimoto's thyroiditis and control groups

Group	Right thyroid volume (mm^3) Median (IQR: 25%–75%)	Left thyroid volume (mm^3) Median (IQR: 25%–75%)	Total thyroid volume (mm^3)
Hashimoto's thyroiditis	43.6 (29.9–61.9)	43.7 (28.4–57.6)	88.8 (61.5–116)
Control	41.7 (31.8–51.3)	36.2 (30–48.4)	76.7 (60.5–94.8)
P value	0.464	0.280	0.440
Hashimoto's thyroiditis (euthyroid)	39.7 (27.2–64.8)	37.4 (26.7–56.7)	71.7 (58.2–116)
Hashimoto's thyroiditis (hypothyroid)	48.1 (35.4–60.9)	47.0 (29.8–57.3)	93.3 (64.2–115)
P value	0.359	0.365	0.353

IQR, interquartile range.

Table 2. Comparison of SMI VI values across groups

Group	SMI VI % (R) Radiologist 1	SMI VI % (R) Radiologist 2	SMI VI % (R)	SMI VI % (L) Radiologist 1	SMI VI % (L) Radiologist 2	SMI VI % (L)
Control	3.99 ± 1.5	3.91 ± 1.34	3.75 (2.82–5.16)	4.16 ± 1.54	3.79 ± 1.16	3.97 ± 1.34
Hashimoto's thyroiditis	8.80 (6.80–12.3)	9.00 (6.50–13.0)	8.85 (6.55–12.6)	8.30 (6.90–12.6)	8.60 (6.60–13.1)	8.40 (6.70–12.8)
P value	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
*Mann–Whitney U Test						
Hashimoto's thyroiditis (hypothyroid)	13.9 ± 4.47	14.2 ± 4.64	14.0 ± 4.53	12.6 (10.4–14.2)	13.1 (11.3–14.1)	12.8 (10.8–14.1)
Hashimoto's thyroiditis (euthyroid)	6.71 ± 2.14	6.84 ± 2.06	6.77 ± 2.08	7 (4.95–7.68)	6.75 (5.23–7.95)	6.78 (5.11–7.75)
P value	<0.001**	<0.001**	<0.001**	<0.001*	<0.001*	<0.001*

*Mann–Whitney U Test; **Student's t-test (Welch's t). SMI, superb microvascular imaging; VI, vascularity index.

Table 3. Comparison of PD VI values across groups

Group	PD VI % (R) Radiologist 1	PD VI % (R) Radiologist 2	PD VI % (R)	PD VI % (L) Radiologist 1	PD VI % (L) Radiologist 2	PD VI % (L)
Control	3.29 ± 1.48	3.28 ± 1.25	3.28 ± 1.35	3.30 ± 1.50	3.41 ± 1.17	3.35 ± 1.32
Hashimoto's thyroiditis	6.20 (3.60–8.20)	6.30 (3.80–8.10)	6.15 (3.65–8.05)	6.20 (4–7.80)	6.30 (4.10–7.80)	6.30 (4.10–7.75)
P value	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
*Mann–Whitney U Test						
Hashimoto's thyroiditis (hypothyroid)	7.10 (5.25–10.5)	7.30 (5.50–11.1)	7.35 (5.35–11)	7.20 (5–8.40)	7.40 (5–8.50)	7.30 (5–8.50)
Hashimoto's thyroiditis (euthyroid)	5.25 (3.20–6.68)	5.25 (3.30–7.05)	5.22 (3.26–6.86)	5.29 ± 2.35	5 (3.65–6.83)	4.80 (3.65–6.64)
P value	0.002*	0.003*	0.002*	0.004*	0.003*	0.003*

*Mann–Whitney U Test. PD, power Doppler; VI, vascularity index.

0.296 and 0.303, respectively. Similarly, the correlation between PD VI and these antibodies was also weak, with Spearman's rho values of 0.304 and 0.217.

The correlation between SMI and PD VI values and thyroid volumes in patients with HT was found to be weak, with Spearman's rho values of 0.254 and 0.268, respectively.

The AUC value, derived from the ROC curve, was calculated to differentiate patients diagnosed with HT based on varying threshold SMI VI values. The AUC was found to be 0.933, leading to the establishment of the threshold SMI VI value at 6.75%.

In alignment with the varying threshold values of the SMI VI, the AUC value derived from the ROC curve developed to differentiate between patients with hypothyroidism and those diagnosed with HT was determined to be 0.959. As a result, the SMI VI threshold value was established at 8.825%. The findings from the diagnostic value assessment conducted in accordance with these threshold values are presented in Table 4.

The AUC value, derived from the ROC curve constructed using varying threshold values of PD VI to distinguish patients diagnosed with HT was found to be 0.838. In contrast, the AUC for differentiating patients with hypothyroid HT from those with euthyroid HT was determined to be 0.714. However, the Youden's indexes for these analyses were 0–0.0455 and 0.311–0.353, respectively, indicating a low diagnostic performance.

Discussion

HT is recognized as the most prevalent autoimmune disorder affecting the thyroid gland, with an annual incidence rate ranging from 2 to 498 cases per 100,000 individuals.[12,13] The progression of this condition typically begins with a phase of hyperthyroidism, which is subsequently followed by hypothyroidism due to the degeneration of thyroid tissue.[14] US is a widely utilized imaging modality for the diagnostic evaluation and ongoing monitoring of HT.[3] In the assessment of vascular characteristics, SMI represents an innovative ultrasonographic technique that facilitates the detection of low-velocity blood flow.

In this study, we examined the vascular characteristics of the thyroid gland in patients diagnosed with HT utilizing both SMI and PD modalities. Our findings revealed that patients with HT exhibited significantly elevated VI values when compared with the control group. Additionally, in the HT cohort, the hypothyroid subgroup demonstrated higher VI values than the euthyroid subgroup. However, the diagnostic efficacy of PD for differentiating HT cases based on its VI threshold was found to be suboptimal. In contrast, the SMI method produced more favorable results, with a threshold VI of 6.75% for the detection of HT and 8.825% for the identification of hypothyroid HT.

Few studies have utilized the SMI technique in the context of HT, meaning our investigation makes a significant contribution to this specialized field.[15,16] Bayramoglu et al.[15] identified a VI cut-off value of 10.58% with an AUC of 0.794, which yielded a sensitivity of 67.1% and a specificity of 90% for diagnosing HT in a pediatric cohort consisting of 70 patients with HT and 30 controls. The reported sensitivity was 67.1%, specificity was 90%, PPV was 94%, NPV was 54%, and the overall diagnostic accuracy was determined to be 74%. In the present study, comparable sensitivity and specificity metrics were obtained for the SMI VI threshold of 6.75%, which was established alongside a superior AUC value. In addition to SMI VI, we additionally established a threshold value for PD VI. Nevertheless, our findings indicated that the threshold values derived from the ROC analysis displayed suboptimal diagnostic efficacy, as evidenced by exceedingly low Youden indices; consequently, these values were deemed insufficient for the differentiation of patients with HT from healthy controls, akin to SMI VI. Bayramoglu et al.[15] categorized patients based on the extent of glandular involvement as assessed by gray-scale US and noted a statistically significant difference in VI values between the two patient cohorts with HT. However, while they suggested that this finding was related to the degree of glandular involvement, they did not specify a cut-off value. In a larger cohort of pediatric patients with HT, Durmaz et al.[17] reported a cut-off value of 6.00% for the

Table 4. Diagnostic performance of cSMI VI

Group (cut-off value)	Sensitivity % (GA)	Specificity % (GA)	PPV % (GA)	NPV % (GA)	PLR (GA)	NLR (GA)	Accuracy % (GA)
Hashimoto's thyroiditis vs. Control (6.75)	75%, 34 (63, 86–84, 68)	100% (94, 56–100)	100%	78%, 57 (71, 06–84, 56)	–	0.25 (0.17–0.37)	87%, 05 (80, 31–92, 14)
Hypothyroid vs. Euthyroid (8.825)	34%, 29 (19, 13–52, 21)	100% (90, 75–100)	100%	62%, 3 (56, 53–67, 73)	–	0, 66 (0, 52–0, 84)	68%, 49 (56, 56–78, 87)

cSMI, color superb microvascular imaging; VI, vascularity index; GA, gray-scale area; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

average VI across all thyroid glands, achieving a sensitivity of 86.3% and a specificity of 82.2% for HT diagnosis, thereby indicating substantial diagnostic accuracy. Additionally, they established distinct cut-off values for various measurement planes: 6.350% for the VI of the right and left thyroid lobes in the longitudinal plane, 6.225% for the mean VI in the longitudinal plane, and 6.350% for the VI in the transverse plane. The threshold value identified in our research closely resembles that reported by Durmaz et al.[17] A significant factor contributing to the variability in VI values observed in both previous studies and our own may be the average age of the participants, which could influence VI values due to age-related changes in thyroid vascularization. Furthermore, Bayramoglu et al.[15] systematically excluded individuals with fibrotic thyroid glands, echogenic septa, and pronounced pseudonodular characteristics in their investigation. These factors have the potential to affect VI measurements, possibly leading to divergent cut-off values. Moreover, as noted by Durmaz et al.[17], the identification of varying cut-off values across different imaging planes may suggest that the measurement methodologies employed could contribute to the discrepancies observed in cut-off values. Such findings underscore the necessity for age-specific considerations in the assessment of thyroid vascularization using the VI, particularly when comparing results from different studies involving diverse populations.

In the present study, it was observed that the volumes of the thyroid gland were significantly elevated in patients diagnosed with HT, particularly within the hypothyroid subgroup, when compared with the euthyroid cohort. However, these findings did not achieve statistical significance. Similarly, the research conducted by Bayramoglu et al.[15] and Durmaz et al.[17] reported increased thyroid gland volumes in patients with HT relative to the control group. In the early stages of the pathology, an enlargement of the glandular structure may be noted, which can be attributed to increased vascularization. This enhancement in vascularity occurs as a result of the immunological response and the inflammatory processes taking place within the glandular tissue.[18]

Durmaz et al.[17] elucidated a positive relationship between VI values and the concentrations of TSH, TgAbs, and TPOAbs. However, these associations were characterized by a lack of strength. Additionally, Bayramoglu et al.[15] identified a significant fair positive correlation between the VI derived from SMI and serum TgAb concentrations, indicating that the VI exhibits an upward trend suggestive of enhanced vascularity within the thyroid gland. Furthermore, they observed a significant moderate positive correlation between the VI and TPOAb levels, which, being a more robust correlation compared with TgAb levels, implies that TPOAb concentrations are more closely associated with increased vascularity in the thyroid gland. A positive and statistically significant correlation between autoantibody concentrations and VI values suggests that the VI measurement of the thyroid gland may be associated with the extent of inflammation and could provide insights regarding disease activity. However, our investigation revealed a significant yet weak correlation between the SMI and PD VIs alongside the autoantibodies. The levels of serum antibodies may serve as indicators of the severity of thyroid gland inflammation in patients diagnosed with HT.[19] While TgAbs may be responsible for the early immune response, TPOAbs are associated with thyroid damage, representing a later immune response.[20] The significant yet modest correlation observed in our investigation, along with findings from other research, may be attributed to the variability inherent in the different stages of HT present within the studied patient cohorts. Bayramoglu et al.[15] found a statistically significant moderate correlation between VI values and serum TSH levels in their study, whereas Durmaz et al.[17] reported a statistically significant weak positive correlation. Significant correlations between TSH levels and thyroid blood flow have been documented in studies utilizing CDI techniques.[21-23] In contrast to previous studies that reported a weaker correlation, our study found a highly positive correlation between serum TSH levels and VI values (rho = 0.739, $P < 0.001$).

Given that the utilization of SMI within the framework of the adult HT demographic has not been explored or documented in the prevailing scholarly literature, we synthesized the conclusions drawn from our analysis in this particular context. The reproducibility of SMI VI and PD VI measurements executed by two separate operators was scrutinized, resulting in an interobserver concordance that can be described as almost perfect. The diagnostic performance of the cSMI VI in differentiating patients with HT from asymptomatic controls exhibited a substantial degree of precision, thereby offering significant contributions to the scholarly discourse, as the threshold VI value we established through SMI for individuals with HT is anticipated to enhance clinical management, particularly in the detection of aberrant blood flow while ensuring tissue uniformity.

Our investigation is subject to several limitations. The primary limitation is that the identification of HT was conducted through clinical evaluation rather than histopathological examination. Due to the absence of biopsy specimens, we were unable to assess and interpret the histopathological alterations and the extent of inflammation within the thyroid tissue. Additionally, a majority of the participants enrolled in the study were receiving pharmacological treatment, and the duration of the disease and the disease stage were not documented. A statistically significant disparity was observed in the TSH values between the hypothyroid and euthyroid cohorts within the HT population. Despite the majority of individuals in the HT group undergoing therapeutic interventions, the pronounced variation in TSH levels among these patients substantiates the hypothesis that the administered drug dosage was inadequate for those receiving treatment. Nevertheless, the absence of documentation regarding the treatment status, as well as the absence of the transient hyperthyroid phase of the patients, constitutes a notable limitation of this investigation. Moreover, a deficiency of information pertaining to autoantibody titers within the control group impedes the comprehensive exclusion of subclinical autoimmune thyroid pathology. In addition, our measurements of PD and VI were conducted within a single plane (longitudinal). Bayramoglu et al.[15] employed a single-plane (transverse) methodology for VI measurement similar to our approach; however, unlike our methodology, which involved calculating the mean of three measurements derived from the same segment (middle), they averaged three measurements taken from the upper, middle, and lower sections of the gland. While this may represent a limitation, Durmaz et al.[17] found no statistically significant variation in VI measurements across different planes, suggesting that a single measurement obtained from one plane can be utilized with a high degree of reliability. This finding reduces the need for multiple measurements and simplifies the process in clinical settings. Additional limitations inherent to our investigation include, first, the predominance of the female demographic within our cohort, which precluded the opportunity for subgroup analysis. While the predominance of female patients is anticipated in cases of HT, this gender disproportion remains a significant limitation. Second,

the vascular characteristics associated with various manifestations of HT, encompassing both atrophic and goitrogenic variants, were not assessed, and consequently, no subgroup analysis was conducted in this regard.

In conclusion, our study highlights the significant elevation of SMI VI values in patients with HT, particularly among those with hypothyroidism, in comparison to healthy controls. The identified optimal threshold values of 6.75% for the diagnosis of HT and 8.825% for hypothyroid HT suggest that SMI may serve as a promising diagnostic tool for detecting subtle alterations in thyroid vascularization. The enhanced sensitivity of SMI in relation to PD technology positions it as an innovative and effective method for the assessment of HT, providing valuable insights into disease activity and progression. Future research, particularly studies that incorporate histopathological data and multi-plane measurements, in addition to the therapeutic condition of patients, is essential to further refine the diagnostic capabilities of SMI and to explore its clinical applications.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. *N Engl J Med.* 2003;348(26):2646-2655. [Crossref]
2. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev.* 2014;13(4-5):391-397. [Crossref]
3. Li F, Sun W, Liu L, Meng Z, Su J. The application value of CDFI and SMI combined with serological markers in distinguishing benign and malignant thyroid nodules. *Clin Transl Oncol.* 2022;24(11):2200-2209. [Crossref]
4. Wu G, Zou D, Cai H, Liu Y. Ultrasonography in the diagnosis of Hashimoto's thyroiditis. *Front Biosci (Landmark Ed).* 2016;21(5):1006-1012. [Crossref]
5. Gürsel BE, Çağlar B, Özpar R, Saraydaroğlu Ö, Gökalp G, Topal NB. Multimodal ultrasonography evaluation in thyroid nodule characterization: what is the ideal algorithm? *Med Ultrason.* 2024;26(1):41-49. [Crossref]
6. Ceylan I, Yener S, Bayraktar F, Secil M. Roles of ultrasound and power Doppler ultrasound for diagnosis of Hashimoto thyroiditis in antithyroid marker-positive euthyroid subjects. *Quant Imaging Med Surg.* 2014;4(4):232-238. [Crossref]
7. Li W, Gao L, Du Y, et al. Ultrasound microflow patterns help in distinguishing malignant from benign thyroid nodules. *Cancer Imaging.* 2024;24(1):18. [Crossref]
8. Fu Z, Zhang J, Lu Y, et al. Clinical applications of superb microvascular imaging in the superficial tissues and organs: a systematic review. *Acad Radiol.* 2021;28(5):694-703. [Crossref]
9. Cappelli C, Pirola I, Gandossi E, et al. Ultrasound microvascular blood flow evaluation: a new tool for the management of thyroid nodule? *Int J Endocrinol.* 2019;2019:7874890. [Crossref]
10. Ates F, Topaloglu ÖF, Secilmis O, et al. Evaluation of the thyroid gland vascularity during pregnancy using 2-dimensional color superb microvascular imaging vascularization index technique. *Eur J Obstet Gynecol Reprod Biol.* 2024;296:316-320. [Crossref]
11. Banaka I, Kaltsas G, Antoniou S,et al. Prognostic value of vascularity index for the diagnosis of autoimmune thyroid disease. *JBR-BTR.* 2011;94(4):185-190. [Crossref]
12. Ralli M, Angeletti D, Fiore M, et al. Hashimoto's thyroiditis: an update on pathogenic mechanisms, diagnostic protocols, therapeutic strategies, and potential malignant transformation. Autoimmun Rev. 2020;19(10):102649. [Crossref]
13. Klubo-Gwiezdzinska J, Wartofsky L. Hashimoto thyroiditis: an evidence-based guide to etiology, diagnosis and treatment. *Pol Arch Intern Med.* 2022;132(3):16222. [Crossref]
14. Willms A, Bieler D, Wieler H, Willms D, Kaiser KP, Schwab R. Correlation between sonography and antibody activity in patients with Hashimoto thyroiditis. *J Ultrasound Med.* 2013;32(11):1979-1986. [Crossref]
15. Bayramoglu Z, Kandemirli SG, Caliskan E, et al. Assessment of paediatric Hashimoto's thyroiditis using superb microvascular imaging. *Clin Radiol.* 2018;73(12):1059. [Crossref]
16. Ozturk M, Durmaz MS, Ateş F, Kara T, Durmaz FG, Akyürek N. Assessment of thyroid gland vascularity with superb microvascular imaging in healthy children and its relationship with potential factors. *J Pediatr Endocrinol Metab.* 2020;33(9):1139-1145. [Crossref]
17. Durmaz MS, Akyürek N, Kara T, et al. Quantitative assessment of thyroid gland vascularization with vascularization index using color superb microvascular imaging in pediatric patients with hashimoto thyroiditis. *Ultrasound Q.* 2019;35(3):281-289. [Crossref]
18. Pedersen OM, Aardal NP, Larssen TB, Varhaug JE, Myking O, Vik-Mo H. The value of ultrasonography in predicting autoimmune thyroid disease. *Thyroid.* 2000;10(3):251-9. [Crossref]
19. Liu J, Zhang Y, Ji Y, Wan Q, Dun G. The value of shear wave elastography in diffuse thyroid disease. *Clin Imaging.* 2018;49:187-192. [Crossref]
20. Fukuhara T, Matsuda E, Izawa S, Fujiwara K, Kitano H. Utility of shear wave elastography for diagnosing chronic autoimmune thyroiditis. *J Thyroid Res.* 2015;2015:164548. [Crossref]
21. Anderson L, Middleton WD, Teefey SA, et al. Hashimoto thyroiditis: part 1, sonographic analysis of the nodular form of Hashimoto thyroiditis. *AJR Am J Roentgenol.* 2010;195(1):208-215. [Crossref]
22. Sarikaya B, Demirbilek H, Akata D, Kandemir N. The role of the resistive index in Hashimoto's thyroiditis: a sonographic pilot study in children. *Clinics (Sao Paulo).* 2012;67(11):1253-1257. [Crossref]
23. Tang K, Liu M, Zhu Y, Zhang M, Niu C. The clinical application of ultrasonography with superb microvascular imaging-a review. *J Clin Ultrasound.* 2022;50(5):721-732. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253416

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Preoperative arterial and venous embolization of heterotopic ossification of the hip: a case-control study

Thomas Le Tat[1,2]
Fabien Cale[3]
Mostafa El Hajjam[1]
François Genêt[4]
Robert-Yves Carlier[1]
Raphaël Amar[1]

[1]Raymond Poincaré Hospital, Clinic of Radiology, Garches, France

[2]Percy Army Training Hospital, Clinic of Radiology, Clamart, France

[3]Raymond-Poincaré Hospital, Clinic of Orthopedic Surgery, Garches, France

[4]Raymond Poincaré Hospital, Clinic of Physical and Rehabilitation Medicine, Garches, France

PURPOSE

Heterotopic ossification (HO) is an abnormal bone mass in soft tissue, often complicating spinal cord or brain injuries with paralysis. When ossification limits joint amplitudes and becomes symptomatic, surgical excision may be necessary, although it carries a high risk of hemorrhage. This study evaluates the role of preoperative arterial embolization on peri-operative bleeding.

METHODS

A retrospective case-control study was conducted on patients undergoing hip HO resection between September 2019 and April 2024. Only anteromedial or circumferential HO were included. Data on blood loss, transfusion requirements, duration of surgery, and length of hospital stay were analyzed. Embolization was performed with microspheres 500–700 µm and/or coils. Statistical significance was determined using the Mann–Whitney U test ($P < 0.05$).

RESULTS

A total of 18 resections of HO and 9 preoperative embolizations were performed. Embolization was technically successful in all cases, with an average of 1.2 embolized arteries per patient. Venous embolization was performed in three specific cases. The mean surgical time was 127 min. The estimated mean blood loss was 1.789 mL, with no significant reduction with embolization ($P = 0.25$). However, embolized patients had a significantly shorter hospital stay (6.2 vs. 8 days, $P = 0.03$). One complication (arterial thrombosis downstream of the puncture site) was reported and successfully treated.

CONCLUSION

Preoperative embolization may improve recovery by shortening hospital stay; however, its impact on surgical bleeding remains unclear. Further studies are needed to refine embolization strategies and evaluate long-term outcomes, including recurrence rates.

CLINICAL SIGNIFICANCE

Preoperative embolization is a promising adjunct to complex HO hip resections, with a favorable safety profile that justifies its consideration in multidisciplinary surgical planning.

KEYWORDS

Heterotopic ossification, preoperative arterial embolization, hip surgery, interventional radiology, surgical outcomes

Corresponding author: Thomas Le Tat

E-mail: letat.thomas@yahoo.fr

Received 06 May 2025; revision requested 08 June 2025; accepted 16 July 2025.

Epub:16.12.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253416

You may cite this article as: Le Tat T, Cale F, El Hajjam M, Genêtd F, Carlier R-Y, Amar R. Preoperative arterial and venous embolization of heterotopic ossification of the hip: a case-control study. *Diagn Interv Radiol*. 2026;32(3):322-327.

Heterotopic ossification (HO) is a proliferation of heterotopic extra-articular calcifications.[1] It is a common complication following brain or spinal cord injury with paralysis, appearing within the first 6 months in nearly 20% of patients.[2]

When HO significantly impairs quality of life by limiting joint mobility and causing pain, surgery is often required.[3,4] Surgical resection, though effective, is associated with substantial intraoperative bleeding due to the hypervascular nature of HO, and the presence of multiple vessels feeds these ossifications. Although tranexamic acid can reduce bleeding in orthopedic surgery, its efficacy in HO resection remains limited.[5] One way of limiting peri-operative bleeding could be preoperative embolization. However, although preoperative embolization is well-documented in reducing bleeding in malignant bone tumors, its role in HO resection remains underexplored, with only a few cases reported.[6]

The primary objective of this study is to evaluate the efficacy of pre-operative arterial embolization in reducing peri-operative bleeding during HO resection. Secondary objectives include assessing the safety of the procedure and its impact on operative time and hospital stay.

Methods

Study design

This retrospective case-control study included 18 surgical resections of symptomatic HO of the hip between September 2019 and April 2024. The inclusion criteria were age > 18 years and symptomatic circumferential or antero-medial HO requiring anterior surgical resection (Figure 1). HO was considered symptomatic when it caused functional discomfort or persistent pain or when signs of neuronal or vascular compression appeared.

Main points

- Preoperative embolization is feasible and technically successful in all cases of high-risk hip heterotopic ossification.
- Hospital stay was significantly shortened in embolized patients compared with those who were non-embolized.
- Venous embolization has been performed in selected cases and requires further studies to assess its efficacy.

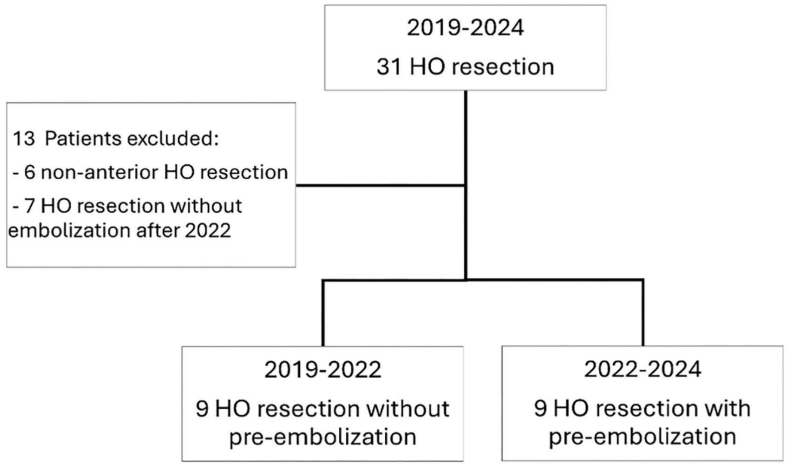


Figure 1. Flowchart. HO, Heterotopic ossification.

Nine non-embolized cases came from a previous study in which patients underwent surgery between September 2019 and January 2022.[7] From January 2022 onward, a multidisciplinary meeting including the surgical team and the interventional radiology teams took place before any resection of HO considered by the surgeon to be at high risk of bleeding to decide if preoperative embolization was required for the patient. High hemorrhagic risk was defined by the presence of major vessels embedded within the HO mass, making surgical exposure and potential ligation in case of bleeding particularly difficult. Nine preoperative arterial embolization operations were performed between January 2022 and April 2024; 7 patients underwent surgical resection of circumferential or antero-medial HO without preoperative embolization after January 2022 but were not included in this study, as they were considered by the surgeon to be at moderate risk of bleeding.

The following data were collected: age, sex, weight, height, body mass index, etiology of HO, size of HO, time from surgery to hospital discharge, preoperative and postoperative hemoglobinemia, blood transfusion details, and operative time. For embolized patients, the following data were also collected: procedure length (time from preparation of the patient's groin until the time of the last image acquired), dose area product, fluoroscopy time, and number of embolized vessels. Blood mass was estimated at 70 mL/kg. Blood loss was calculated using Gross' formula, based on the difference between preoperative hemoglobinemia and the lowest postoperative hemoglobinemia during hospitalization.[8]

Prior to embolization and surgery, computed tomography (CT) angiography was performed in all patients to plan the embolization and surgical procedure [SOMATOM Definition AS (Siemens Healthineers, Erlangen, Germany)]. A biphasic CT scan was performed, with a first injection of 120 mL of iodinated contrast medium (Iomeron 400 mg/mL, Bracco Imaging) at a speed of 1.5 mL/s immediately followed by a second injection of 80 mL at a speed of 3 mL/s. The CT scan was triggered 135 s after the start of the first injection. Multi-planar reconstructions were produced, along with three-dimensional images showing bone and the vascular tree. In a single acquisition, this protocol enables the radiologist and surgeon to locate the vessels in the surgical pathway easily and to plan embolization.[9]

Arterial and venous embolization and the surgical procedure

All procedures were performed under local anesthesia in an angiographic unit (Azurion©, PHILIPS, Amsterdam, Netherlands) by two interventional radiologists with 8 (RA) and 25 (EHM) years' experience in interventional radiology. After percutaneous introduction of a 5 Fr sheath into the femoral artery contralateral to ossification, a 5 Fr Cobra 2 catheter was used to cross-over and catheterize the deep femoral artery. Selective catheterization was then performed with a microcatheter (2.4 Fr and 2.7 Fr Progreat©, Terumo, Tokyo, Japan; 2.4 Fr Maestro©, Merit Medical, South Jordan, USA). Microspheres 500–700 µm (Embogold©, Merit Medical, South Jordan, USA) were then injected until the tumor blush disappeared, followed by coils (Nester© CookMedical, Bloomington,

USA; Mreye© CookMedical, Bloomington, USA). Technical success was defined as complete occlusion of the feeding vessels in the surgical pathway. As the surgical approach was anterior, only feeding vessels located anterior to the HO were considered on the surgical path.

Venous embolization was requested by the surgeon when one of the main branches of the femoral veins he planned to clamp was surrounded by HO, as these veins present a particular surgical challenge due to their proximity to ossified tissue, making them difficult to isolate and clamp during dissection. Venous embolization was intended to facilitate surgical exposure and potentially minimize intraoperative bleeding. Venous embolization was performed under local anesthesia by the same interventional radiologists. After percutaneous introduction of a 5 Fr sheath into the femoral vein contralateral to ossification, a 5 Fr Cobra 2 catheter was used to cross-over and catheterize the lateral circumflex vein. Embolization was then performed with coils.

Surgery was scheduled shortly after embolization, usually the following day, before any new vascularization occurred. Surgery was performed via an anterior approach similar to an extended Hueter approach, extending proximally along the iliac crest to the medial gluteal tuberosity and distally to the junction between the middle and upper thirds of the thigh.[4,10]

The French Ethics Committee for Medical Imaging Research (CERIM) approved this retrospective study and waived the requirement for written informed consent (approval number: CRM-2501-462, date: 23.03.2025).

Statistical analysis

Due to the small size of the study, the non-parametric Mann–Whitney U test was used for quantitative values. A P value of < 0.05 was defined as statistically significant. Statistical analyses were performed using R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Of the 18 cases of HO included, 16 were neurogenic HO and 2 were related to prolonged hospitalization in an intensive care unit (Table 1); 9 pre-operative embolizations were performed (Figure 2). All embolizations were technically successful, defined as embolization of vessels in the surgical tract, with an average of 1.2 arteries embolized (Table 2). HO was generally supplied by branches of the medial and lateral circumflex femoral arteries as well as by perforating branches of the deep femoral artery. Embolization most frequently involved branches of the lateral circumflex femoral artery, with additional vessels targeted when they were located within the planned surgical field (Figure 3). Three patients also benefited from embolization of the lateral circumflex vein, which was located on the operative path (Figure 4). The estimated mean blood loss was 1.789 mL, with no significant difference between embolized and non-embolized patients ($P = 0.25$). Six patients (33%) required transfusion during their hospital stay. The mean operative time was 127 min, with no significant difference between groups. The mean length of hospital stay after surgical resection was 7.1 days and was significantly shorter in embolized patients than in non-embolized patients: 6.2 days vs. 8 days ($P < 0.05$).

One complication was reported in the embolization group, classified as Grade 3 according to the Cardiovascular and Interventional Radiological Society of Europe classification system.[11] This involved the thrombosis of a distal artery downstream of the puncture site, which was effectively treated with thrombolytics. However, surgery had to be postponed for 2 months.

Discussion

This study evaluated the efficacy and safety of preoperative embolization on peri-operative bleeding prior to resection of HO. Only patients presenting with circumferential or anteromedial HO were included in this study, as these patterns are associated with the highest risk of bleeding.[12]

Preoperative embolization was associated with a significantly shorter hospital stay compared with patients without embolization. However, no significant differences were observed in terms of blood loss or blood transfusion requirements. This differs from the case-control study by Papalexis et al.,[6] who found a reduction in transfusion requirements and hemoglobin loss after pre-embolization with cyanoacrylate glue. In our study, the mean number of arteries embolized was 1.2. The interventional radiologist endeavored to embolize only vessels located in the surgical pathway, thereby limiting

Table 1. Continuous data are presented as the mean and standard deviation. Categorical data are shown as the number and percentage

Baseline characteristics	Embolization n = 9	No embolization n = 9	Total n = 18	*P* value
Age (years)	41 (13.7)	51.7 (12.1)	46.3 (13.7)	
Male, n (%)	8 (88.9%)	8 (88.9%)	16 (88.9%)	
Body mass index	23 (3.9)	24.3 (4.1)	23.7 (4)	
Etiology				
- Brain injury, n (%)	3 (33.3%)	3 (33.3%)	6 (33.3%)	
- Spinal injury, n (%)	5 (55.6%)	5 (55.6%)	10 (55.6%)	
- Prolonged stay in intensive care, n (%)	1 (11.1%)	1 (11.1%)	2 (11.1%)	
Pre-operative hemoglobinemia (g/dL)	13.6 (1.1)	13.3 (1.3)	13.5 (1.2)	
Estimated blood volume (mL)	5,024 (876)	5,616 (1219)	5,320 (1,074)	
Heterotopic ossification volume (mL)	757 (471)	570 (305)	664 (397)	
Post operative follow-up				
Operative time (min)	115 (38)	139 (38)	127 (39)	0.31
Hospital stay duration (days)	6.2 (3.1)	8 (3.9)	7.1 (3.5)	0.03
Estimated blood loss (mL)	2,097 (1,150)	1,497 (683)	1,796 (968)	0.25
Units of blood transfused	1.2 (2)	0.9 (1.5)	1.1 (1.7)	0.87

Table 2. Continuous data are presented as the mean and standard deviation	
Endovascular procedures	
Number of embolized arteries	1.2 (0.8)
Procedure length (min)	81 (32)
Fluoroscopy time (min)	20 (8)
Dose area product (Gy.cm^2)	63 (33)

the risk of osteonecrosis in healthy bone. In the study by Papalexis et al.,[6] the operator embolized an average of 2.6 arteries and > 90% of the arteries feeding the ossification. Thus, intraoperative bleeding also depends on the complexity of the surgical procedure, which is difficult to quantify. The retrospective case-control design of our study, with selection based on perceived bleeding risk, introduces a strong selection bias, which may explain why we did not achieve a significant reduction in bleeding. Other possible explanations are the use of particles rather than glue and the limited number of arteries embolized. Future studies could determine the best strategy.

The embolization procedure demonstrated a favorable safety profile, with only one complication observed across all cases, related to arterial puncture. None of the other cases reported in the literature had complications related to embolization.[13-17]

We only studied patients operated on via the anterior approach, in whom the risk of bleeding is highest.[12] However, given the safety profile of embolization, it would be interesting for further studies to examine the effect of preoperative embolization for other surgical approaches. Nevertheless, preoperative embolization of the HO has the same disadvantages and risks as other preoperative embolizations, including non-targeted embolization, risk of vascular complications, and increased overall cost.

The three cases of venous embolization in this study are the first published in the literature for this indication. Although this study does not assess the efficacy of this embolization, this technique could perhaps be of interest in special anatomical circumstances, notably when large veins are embedded in the ossification, as they are difficult to manage by the surgeon and can lead to life-threatening bleeding.[18,19]

Unfortunately, we were unable to carry out a long-term follow-up of our patients, as half of them were referred to us for surgery and, after the 1st year, conducted their follow-up elsewhere. It would be interesting in the future to determine whether embolization can limit the risk of symptomatic recurrence of HO.

This study has several limitations, including its retrospective design with non-random allocation and the small sample size. Thus, it remains possible that HO management improved during the study and that the observed effect on hospital stay is due solely to this, as this is a before-and-after study. Moreover, the length of hospital stay is not the best criterion in open-label studies, as medical staff may be encouraged to discharge patients, knowing they have been embolized.

In conclusion, preoperative embolization appears to be a safe adjunct to HO resection, significantly reducing hospital stay. However, its impact on surgical bleeding remains unclear. Further research is needed to refine embolization techniques and evaluate long-term outcomes, including recurrence rates.

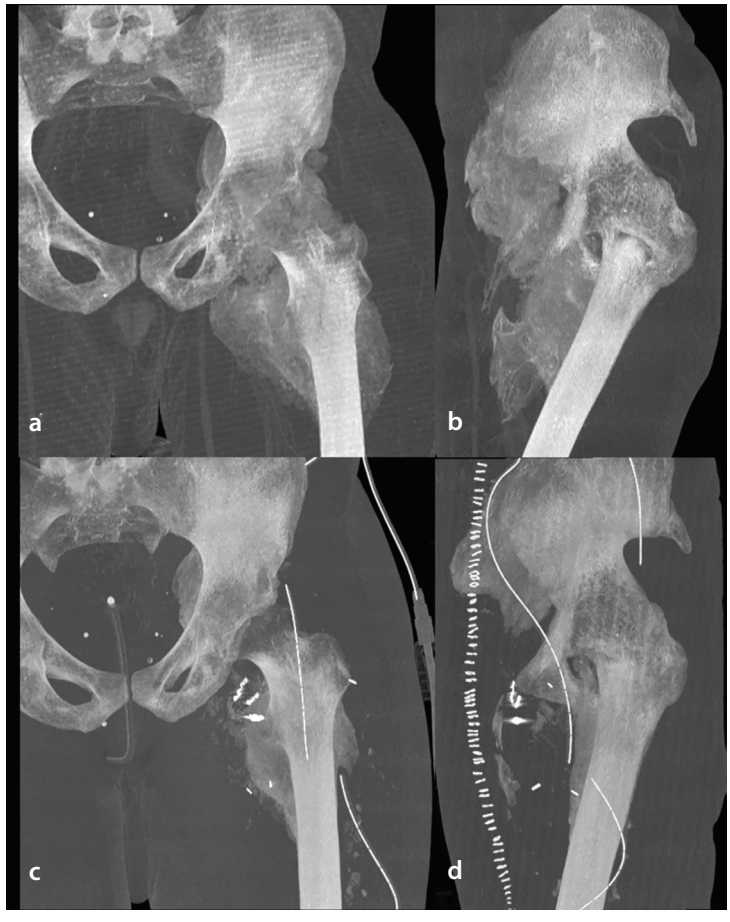


Figure 2. Pre-operative (**a, b**) and postoperative (**c, d**) computed tomography scan of the left hip in the maximum-intensity projection.

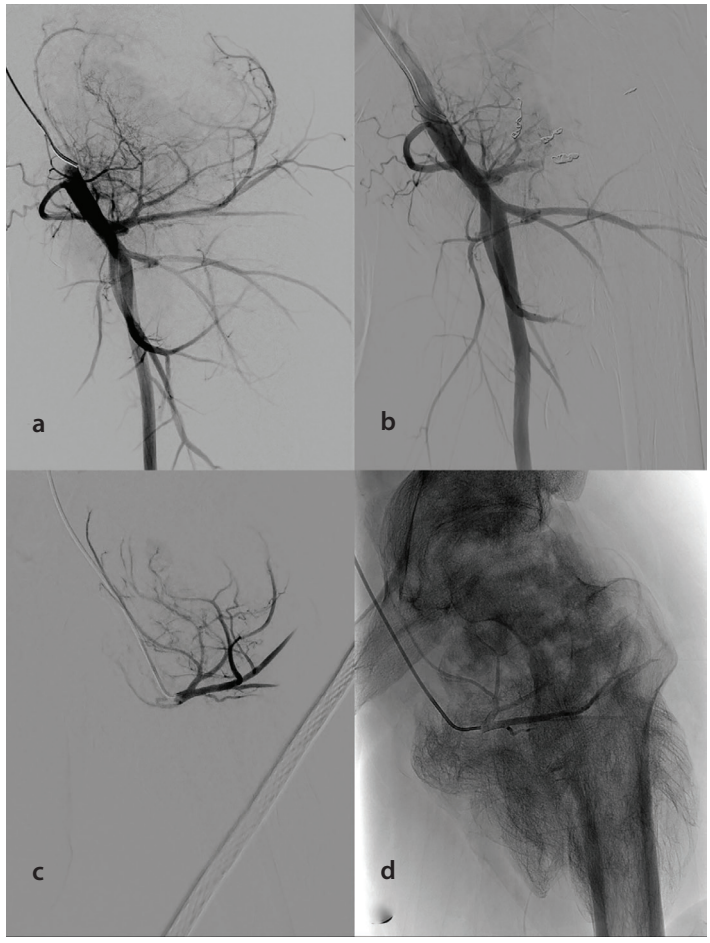


Figure 3. Angiographic views of the left hip (same patient as Figure 1). (**a**) Angiography of the deep femoral artery, showing pseudotumoral hypervascularization originating from the lateral femoral circumflex artery. (**b**) Angiography of the deep femoral artery after embolization of branches of the lateral femoral circumflex artery, showing reduced hypervascularization and the persistence of pseudotumoral blush from branches of the medial femoral circumflex artery. These arterial branches were not embolized, as they were not on the operative path. (**c, d**) Selective angiography of the lateral circumflex femoral artery prior to embolization.

Figure 4. Phlebography of the left femoral vein. (**a**) Pre-embolization phlebography showing the lateral circumflex vein. (**b**) Post-embolization.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Dejerine A, Ceillier A. Para-ostéo-arthropathies des Paraplégiques par lésion médullaire: Étude clinique et radiographique. *Ann Méd*. 1918:497-5.
2. Wittenberg RH, Peschke U, Bötel U. Heterotopic ossification after spinal cord injury. Epidemiology and risk factors. *J Bone Joint Surg Br*. 1992;74(2):215-218. [Crossref]
3. Ranganathan K, Loder S, Agarwal S, et al. Heterotopic ossification: basic-science principles and clinical correlates. *J Bone Joint Surg Am*. 2015;97(13):1101-1111. [Crossref]
4. Denormandie P, de l'Escalopier N, Gatin L, Grelier A, Genêt F. Resection of neurogenic heterotopic ossification (NHO) of the hip. *Orthop Traumatol Surg Res*. 2018;104(1 Suppl):121-127. [Crossref]
5. Kim C, Park SSH, Davey JR. Tranexamic acid for the prevention and management of orthopedic surgical hemorrhage: current evidence. *J Blood Med*. 2015;6:239-244. [Crossref]
6. Papalexis N, Peta G, Errani C, Miceli M, Facchini G. Preoperative arterial embolization for heterotopic ossification of the hip. *J Vasc Interv Radiol*. 2023;34(4):608-612. [Crossref]
7. Amar R, Salga M, Tordjman M, et al. Neurogenic heterotopic ossification of the hip: magnetic resonance imaging versus computed tomography for pre-surgical assessment. *Eur J Radiol*. 2024;175:111466. [Crossref]
8. Gross JB. Estimating allowable blood loss: corrected for dilution. *Anesthesiology*. 1983;58(3):277-280. [Crossref]
9. Law-Ye B, Hangard C, Felter A, et al. Pre-surgical CT-assessment of neurogenic myositis ossificans of the hip and risk factors of recurrence: a series of 101 consecutive patients. *BMC Musculoskelet Disord*. 2016;17(1):433. [Crossref]
10. Genêt F, Jourdan C, Schnitzler A, et al. Troublesome heterotopic ossification after central nervous system damage: a survey of 570 surgeries. *Plos One*. 2011;6(1):e16632. [Crossref]
11. Filippiadis DK, Binkert C, Pellerin O, Hoffmann RT, Krajina A, Pereira PL. Cirse quality assurance document and standards for classification of complications: the cirse classification system. *Cardiovasc Intervent Radiol*. 2017;40(8):1141-1146. [Crossref]
12. Mavrogenis AF, Guerra G, Staals EL, Bianchi G, Ruggieri P. A classification method for neurogenic heterotopic ossification of the hip. *J Orthop Traumatol*. 2012;13(2):69-78. [Crossref]
13. Appelt EA, Kenkel JM, Ballard JR, Lopez JA, Anthony T, Castillo T. Preoperative embolization of heterotopic ossification for the treatment of a recalcitrant pressure sore. *Plast Reconstr Surg*. 2005;116(4):50e-53e. [Crossref]
14. Vogl TJ, Wolff JD, Balzer J, Skripitz R. Preoperative arterial embolization in heterotopic ossification: a case report. *Eur Radiol*. 2001;11(6):962-964. [Crossref]
15. Kim JH, Park C, Son SM, et al. Preoperative arterial embolization of heterotopic ossification around the hip joint. *Yeungnam Univ J Med*. 2018;35(1):130-134. [Crossref]
16. Zielinski E, Chiang BJL, Satpathy J. The role of preoperative vascular imaging and embolisation for the surgical resection of bilateral hip heterotopic ossification. *BMJ Case Rep*. 2019;12(8):e230964. [Crossref]
17. Igei T, Nakasone S, Ishihara M, Onaga M, Nishida K. Embolization followed by resection of the heterotopic hip joint ossification with spinal cord injury. *J Orthop Sci*. 2024;29(1):454-457. [Crossref]
18. Garland DE, Hanscom DA, Keenan MA, Smith C, Moore T. Resection of heterotopic ossification in the adult with head trauma. *J Bone Joint Surg Am*. 1985;67(8):1261-1269. [Crossref]
19. Alshameeri Z, Bajekal R, Varty K, Khanduja V. Iatrogenic vascular injuries during arthroplasty of the hip. *Bone Joint J*. 2015;97-B(11):1447-1455. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253590

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Diagnostic performance of magnetic resonance imaging-targeted biopsy for PI-RADS ≥3 peripheral zone lesions in multiparametric prostate magnetic resonance imaging: correlation with clinically significant prostate cancer

Serdar Aslan[1],
Emrah Sülün[1],
Ertuğrul Çakır[1],
Ural Oğuz[2],
Tümay Bekçi[1]

[1]Giresun University Faculty of Medicine, Department of Radiology, Giresun, Türkiye

[2]Giresun University Faculty of Medicine, Department of Urology, Giresun, Türkiye

Corresponding author: Serdar Aslan

E-mail: serdaraslan28@hotmail.com

Received 31 July 2025; revision requested 08 September 2025; last revision received 13 October 2025; accepted 15 October 2025.

Epub: 03.11.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253590

PURPOSE

To evaluate magnetic resonance imaging (MRI)-targeted biopsy (MRI-TB) performance in detecting clinically significant prostate cancer (csPCa) with a Prostate Imaging Reporting and Data System (PI-RADS) score of ≥3 peripheral zone (PZ) lesions using multiparametric MRI (mpMRI)-histopathology correlation.

METHODS

This retrospective study included 141 patients with 187 PZ lesions who underwent mpMRI followed by both MRI-TB and transrectal ultrasound-guided systematic biopsy (SB) between December 2021 and December 2024. All mpMRI scans were evaluated by a board-certified experienced radiologist in accordance with the PI-RADS version 2.1 criteria. The csPCa detection rates of SB, MRI-TB, and combined biopsy (CB) were compared. Statistical analyses included McNemar's test, Fisher's exact test, and the Mann–Whitney U test. A *P* value <0.05 was considered statistically significant.

RESULTS

Among the 141 patients (187 PI-RADS ≥3 PZ lesions), patients with csPCa exhibited significantly higher prostate-specific antigen (PSA) levels (15.3 vs. 8.2 ng/mL; $P = 0.02$), lower prostate volume (52.4 vs. 78.6 mL; $P < 0.001$), and three-fold higher PSA density (PSAD) (0.30 vs. 0.10 ng/mL/mL; $P < 0.001$) than non-csPCa cases. Notably, PSAD > 0.15 ng/mL/mL occurred in 78% of patients with csPCa vs. 18% in non-csPCa cases ($P < 0.001$). Moreover, MRI-TB detected significantly more csPCa than SB (17.7% vs. 10.7% of lesions; $P < 0.001$), with maximal advantage in PI-RADS 4 lesions (20.7% vs. 10.9%; $P = 0.004$). By contrast, CB did not significantly increase csPCa detection over MRI-TB alone (19.8% vs. 17.7%; $P = 0.125$). Chronic prostatitis (CP) (34.0% of benign cases) confounded PI-RADS specificity.

CONCLUSION

For csPCa detection in PI-RADS ≥3 PZ lesions, particularly PI-RADS 4, MRI-TB outperforms SB. For PI-RADS 5, SB and MRI-TB showed equivalent efficacy. However, MRI-TB alone suffices for PI-RADS ≥4 lesions or PSAD >0.15 ng/mL/mL, whereas CB remains preferable for PI-RADS 3. The high CP prevalence underscores the need for adjunctive biomarkers to improve specificity.

CLINICAL SIGNIFICANCE

MRI-TB optimizes csPCa detection for PI-RADS ≥4 PZ lesions, reducing reliance on SBs. A PSAD threshold >0.15 ng/mL/mL effectively stratifies biopsy necessity, and high CP prevalence (34% of benign cases) underscores the need for adjunct biomarkers to improve specificity in PI-RADS 3–4 lesions.

KEYWORDS

Magnetic resonance imaging-targeted biopsy, Prostate Imaging Reporting and Data System, peripheral zone, magnetic resonance imaging

You may cite this article as: Aslan S, Sülün E, Çakır E, Oğuz U, Bekçi T. Diagnostic performance of magnetic resonance imaging-targeted biopsy for PI-RADS ≥3 peripheral zone lesions in multiparametric prostate magnetic resonance imaging: correlation with clinically significant prostate cancer. *Diagn Interv Radiol*. 2026;32(3):328-337.

Prostate cancer (PCa) ranks as the second most frequently diagnosed malignancy in men worldwide and the seventh leading cause of cancer-related death.[1] Approximately 95% of PCas are adenocarcinomas arising from glandular epithelial cells, predominantly in the peripheral zone (PZ), whereas approximately 25% originate in the transitional zone (TZ); rare cases involve neuroendocrine, basal cell, or mesenchymal tumors.[2,3]

The incidence of PCa increases substantially with age, predominantly affecting men over 65 years, with family history serving as a well-established risk factor, particularly when first-degree relatives are diagnosed before the age of 65.[1,4] Early-stage PCa often remains asymptomatic but may present with metastatic symptoms in advanced disease, emphasizing the importance of early detection to reduce morbidity and mortality.

The prostate-specific antigen (PSA) test remains the primary screening tool, although its specificity is limited by frequent elevation in benign conditions, such as benign prostatic hyperplasia (BPH) and prostatitis.[5] PSA density (PSAD), calculated by dividing the PSA level by the prostate volume, enhances diagnostic accuracy when combined with age, percent-free PSA, and family history.[6,7]

Main points

- Magnetic resonance imaging-targeted biopsy (MRI-TB) outperforms systematic biopsy (SB) in detecting clinically significant prostate cancer (csPCa) with a Prostate Imaging Reporting and Data System (PI-RADS) score ≥3 peripheral zone lesions (17.7% vs. 10.7%; $P < 0.001$), with maximal advantage in PI-RADS 4 lesions (20.7% vs. 10.9%; $P = 0.004$).
- Chronic prostatitis confounds PI-RADS specificity, present in 34% of benign cases and mimicking csPCa on multiparametric MRI, particularly in PI-RADS 3–4 lesions.
- Prostate-specific antigen density (PSAD) >0.15 ng/mL/mL is a robust predictor of csPCa (78% sensitivity) and reduces unnecessary biopsies in equivocal lesions.
- Combined biopsy did not significantly increase csPCa detection over MRI-TB alone (19.8% vs. 17.7%; $P = 0.125$), supporting MRI-TB as a first-line intervention for PI-RADS ≥4 or high PSAD.
- SB remains viable for PI-RADS 5 lesions (equivalent csPCa detection to MRI-TB: 18.5% each), optimizing resource use in high-volume settings.

Transrectal ultrasound (TRUS)-guided systematic biopsy (SB), typically involving 12-core sampling, is the standard diagnostic technique. However, this method has limited sensitivity, with a cancer detection rate of 27%–40% and a risk of missing up to 25% of clinically significant PCa (csPCa).[8,9] Recent advances in multiparametric magnetic resonance imaging (mpMRI) have substantially improved csPCa detection through high-resolution anatomical and functional imaging. MRI-targeted biopsy (MRI-TB), fusing mpMRI findings with real-time United States (US) to guide precise sampling, demonstrates superior csPCa detection over SB while reducing diagnosis of clinically insignificant PCa (ciPCa).[10,11] Prostate Imaging Reporting and Data System (PI-RADS) version 2.1 scoring standardizes lesion characterization, with scores ≥3 indicating biopsy-eligible risk.[12]

Although the principles of MRI-TB superiority and PSAD utility are recognized, their application to individual PI-RADS categories in the PZ—a region prone to both cancer and confounding inflammation—remains inadequately defined. The aim of this study is to (1) determine the differential performance of SB, MRI-TB, and combined biopsy (CB) for each PI-RADS category; (2) quantify the category-specific prevalence and impact of chronic prostatitis (CP); and (3) validate a PSAD threshold (>0.15 ng/mL/mL) in a PZ cohort enriched with CP. We hypothesize that the diagnostic advantage of MRI-TB varies by PI-RADS category and that PSAD can effectively stratify biopsy necessity.

Methods

This single-center retrospective study received approval from the institutional ethics committee of Giresun Training and Research Hospital (ethics committee approval: KAEK-55, decision number: 08, date: 13.03.2023), with waived informed consent due to its retrospective design.

Study population

Between December 2021 and December 2024, patients who underwent mpMRI using a 1.5-Tesla MRI scanner (Magnetom Aera, Siemens Medical Solutions, Erlangen, Germany) and had lesions initially localized to the PZ with PI-RADS version 2.1 ≥3 were identified through the Picture Archiving and Communication System. A total of 223 patients underwent MRI-TB during this period, and their histopathological data were collected. The inclusion criteria were as follows: (a) a final radiological confirmation of strictly PZ-localized lesions (PI-RADS version 2.1), (b) diagnostic-quality mpMRI, and (c) both MRI-TB and 12-core TRUS-guided SB within 3 months of imaging. The exclusion criteria included the following: (1) lesions extending beyond PZ boundaries (exclusively TZ or PZ–TZ overlap) (n = 36), (2) history of previous prostate biopsy/prostate surgery (n = 19), (3) prior PCa diagnosis (n = 9), (4) inadequate mpMRI quality or missing sequences (n = 8), (5) repeat MRI-TB (n = 4), (6) absence of SB (n = 3), and (7) failure to undergo MRI-TB within 3 months of mpMRI (n = 3). TZ lesions were excluded due to higher csPCa prevalence in the PZ (70%–80%), distinct PI-RADS criteria [PZ: diffusion-weighted imaging (DWI) dominant vs. TZ: T2-weighted imaging (T2WI) dominant], and frequent BPH overlap.[2,3] The final cohort comprised 141 patients (187 PZ lesions). Figure 1 summarizes patient selection.

Image acquisition

All mpMRI examinations were performed on a 1.5T MRI system (Siemens Medical Solutions). Images were obtained with the patient in the supine position using a 16-channel pelvic phased-array coil, without an endorectal coil, in accordance with PI-RADS version 2.1 recommendations for patient comfort and feasibility at 1.5T.[12] The imaging protocol included axial/sagittal/coronal T2WI, axial T1-weighted imaging (T1WI), DWI (b-values: 0, 800, 1400 s/mm^2) with apparent diffusion coefficient (ADC) maps, and dynamic contrast-enhanced (DCE) sequences. For DCE, gadolinium-based contrast (0.1 mmoL/kg) was injected intravenously at 2–3 mL/s, followed by high-resolution three-dimensional axial T1WI every 7s for 240–300 s (slice thickness ≤3 mm). Table 1 outlines the technical parameters of the MRI sequences.

Image analysis

A single board-certified radiologist with 13 years of prostate MRI experience, blinded to histopathology, evaluated all the mpMRI scans. The interpreting radiologist had 10 years of dedicated prostate MRI experience at the commencement of the study enrollment period (December 2021). The PI-RADS version 2.1 scoring was performed prospectively as part of the initial clinical interpretation prior to biopsy. The PZ lesions were scored in accordance with the PI-RADS version 2.1 criteria, which integrate T2WI, DWI, and DCE findings for standardized characterization.[12]

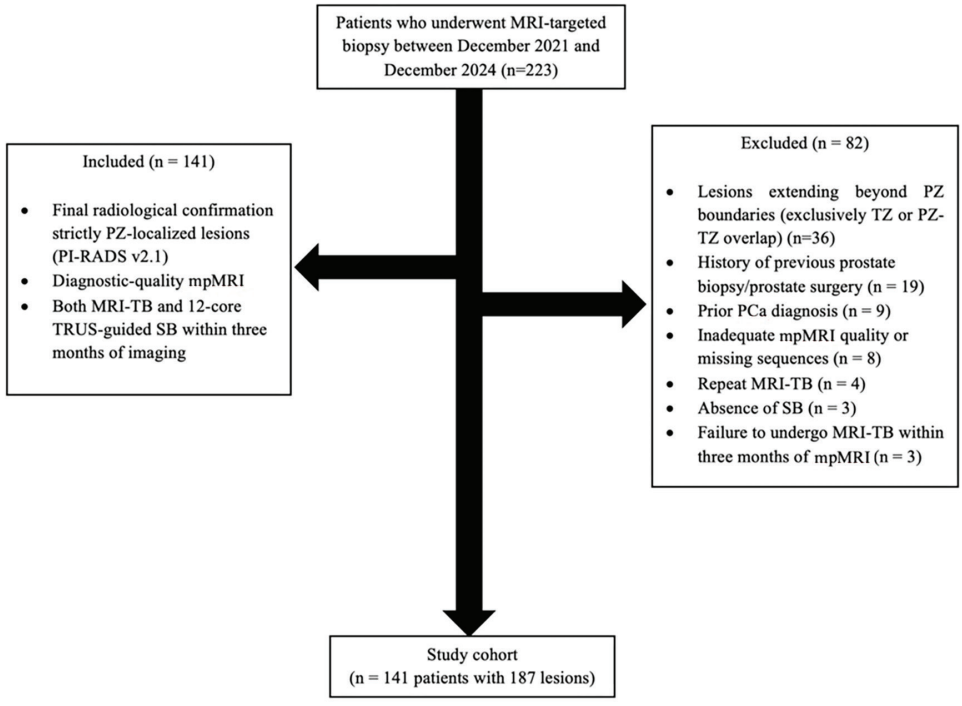


Figure 1. Flowchart of patient selection. PZ, peripheral zone; PI-RADS, Prostate Imaging Reporting and Data System; MRI-TB, magnetic resonance imaging-targeted biopsy; TRUS, transrectal ultrasound; SB, systematic biopsy; TZ, transitional zone; PCa, prostate cancer; mpMRI, multiparametric magnetic resonance imaging.

Table 1. Technical parameters for multiparametric magnetic resonance imaging acquisition

	Axial T2W	Sagittal T2W	Coronal T2W	DWI†	Axial T1W	DCE*
Sequence	TSE	TSE	TSE	EPI	GRE	3D GRE
Fat supression	No	No	No	No	Yes	Yes
TE (ms)	6.000	7.000	6.200	4.200	580	4.46
TR (ms)	108	108	108	82	13	1.72
FOV (mm)	200 × 200	200 × 200	200 × 200	260 × 260	200 × 200	260 × 260
Matrix	275 × 320	266 × 320	298 × 320	112 × 112	256 × 256	154 × 192
Slice thickness (mm)	3.5	3.5	3.5	4	3.5	3.5

†Acquired at b-values: 0, 800, 1, 400 s/mm^2 with ADC mapping.
*Dynamic acquisition: temporal resolution 7s, duration 240–300 s; gadolinium dose: 0.1 mmoL/kg at 2–3 mL/s.
ADC, apparent diffusion coefficient; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; DWI, diffusion-weighted imaging; EPI, echo-planar imaging; FOV, field of view; GRE, gradient recalled echo; TSE, turbo spin echo; TE, echo time; TR, repetition time; T2W, T2-weighted; T1W, T1-weighted; 3D, three-dimensional.

Prostate volume and prostate-specific antigen measurement

Prostate volume was calculated from axial and sagittal T2WI using the ellipsoid formula ($\pi/6 \times AP \times$ transverse $\times$ craniocaudal diameter), as recommended by PI-RADS version 2.1.[12] All PSA values were obtained from serum samples collected within 3 months prior to the mpMRI examination, ensuring contemporaneity with the imaging findings.[13]

Magnetic resonance imaging-targeted and 12-core transrectal ultrasound-guided systematic biopsy procedure

All MRI-TB procedures were performed by two experienced radiologists (each with Two radiologists performed MRI-TB using a dedicated MRI-TRUS fusion platform (RS85 Prestige; Samsung Medison, Seoul, South Korea) and software (NavigoR v2.1; Samsung Medison, Seoul, South Korea), consistent with established technical standards.[14] Pre-procedural steps included uploading mpMRI datasets for lesion segmentation, followed by intraprocedural dual registration combining sensor-based electromagnetic tracking and organ-based deformable registration. During target sampling, ≥2 cores per lesion were obtained under continuous US guidance with real-time needle trajectory visualization, maintaining <3 mm targeting accuracy. For each patient, target lesion size was measured as the maximal axial diameter on T2WI.

Concurrently, a standardized 12-core extended sextant biopsy was performed by urologists, blinded to the MRI-TB targets, under TRUS guidance, with cores systematically obtained from six anatomical sectors per prostatic lobe: (1) the apex (medial and lateral PZ; two cores), (2) mid-gland (medial and lateral PZ; two cores), and (3) base (medial and lateral PZ; two cores), using an 18-gauge spring-loaded biopsy needle to harvest tissue cores of 15–22 mm in length, ensuring comprehensive glandular sampling for histopathological correlation.

Digital rectal examination (DRE) was performed pre-biopsy by urologists, with abnormal DRE defined as palpable nodule or glandular asymmetry. All targeted lesions

received a median of 3 cores per lesion [interquartile range (IQR): 2–5], resulting in a median total of 15 biopsy cores per patient (IQR: 14–18) when combined with the 12 systematic cores.

Histopathological analysis

Biopsy cores were evaluated by a dedicated genitourinary pathologist with 12 years of experience in prostate histopathology, in accordance with the 2014 International Society of Urological Pathology (ISUP) guidelines, including Gleason scoring and tumor involvement per core. The percentage of core involvement was used to estimate overall tumor volume, which was integrated with the Gleason score and other findings to generate a final histopathologic diagnosis for each patient. The ISUP grade groups (GGs) correspond to Gleason scores: grade 1 (3 + 3 = 6), grade 2 (3 + 4 = 7), and grade 3 (4 + 3 = 7).[15] PCa was defined as a Gleason score of ≥6 (3 + 3) (equivalent to ISUP GG ≥1); csPCa was defined as a Gleason score of ≥7 (e.g., 3 + 4 = 7, ISUP GG ≥2), an estimated tumor volume of ≥0.5 cm^3, or extraprostatic extension; ciPCa was defined as a Gleason score of 6 (3 + 3) (ISUP GG 1); and CP was diagnosed histopathologically based on the presence of a chronic inflammatory cell infiltrate within the prostatic stroma.

Reference standard and definition of outcomes

For the purpose of calculating detection rates and comparing the performance of MRI-TB and SB, the CB result was used as the reference standard, as it represents the most comprehensive histopathological assessment available for each lesion.[14,16] Although radical prostatectomy is the gold standard, its use would introduce selection bias by including only surgical candidates. We acknowledge that CB may underestimate the true prevalence of cancer due to the potential for sampling error inherent in any biopsy method; however, it represents the best available benchmark for the comparative assessment of biopsy yields in a clinical setting. A lesion was considered truly positive for csPCa if it was detected through CB (ISUP GG ≥2). The detection rate of each method was calculated against this standard. Cases where CB detected csPCa that was missed by MRI-TB or SB were considered false negatives for the respective method. Lesions with a positive MRI (i.e., assigned a PI-RADS score ≥3) but a negative CB result (benign or ciPCa) were considered false-positive MRI findings.

Statistical analysis

Statistical analyses were performed using SPSS software, version 26 (IBM Corporation, Armonk, NY, USA). Normality of data distribution was evaluated using visual methods (histograms and probability plots) and analytical tests (Kolmogorov–Smirnov and Shapiro–Wilk tests). Comparisons between two continuous variables were performed using the Mann–Whitney U test, and categorical variables were compared using the McNemar test and Fisher's exact test. A P value <0.05 was considered statistically significant.

Results

Patient and lesion characteristics

Patient demographics are summarized in Table 2. Key characteristics of the study lesions, stratified by PI-RADS score, are detailed in Table 3. Median lesion size was 14 mm (IQR: 10–18 mm). Patients underwent a median of 3 MRI-TB cores per lesion (IQR: 2–5) and 15 total cores per patient (IQR: 14–18). DRE was abnormal in 44 patients (31.2%). Histopathological assessment identified PCa in 73 patients (51.8%), stratified as clinically significant (csPCa, ISUP GG ≥2, n = 37, 26.2%) or clinically insignificant (ciPCa, ISUP GG 1, n = 36, 25.6%). Benign pathology was observed in 67 patients (47.5%), including CP in 48 cases (34.0%) (Table 4).

Overall biopsy performance

For overall PCa detection, CB demonstrated superior detection (42.8%, 80/187) to both MRI-TB (34.8%, 65/187; $P < 0.001$) and SB (26.7%, 50/187; $P < 0.001$). For csPCa (ISUP ≥2), MRI-TB (17.7%, 33/187) outperformed SB (10.7%, 20/187; $P < 0.001$), and CB (19.8%, 37/187) provided no significant additional benefit over MRI-TB alone ($P = 0.125$). Category-specific detection rates and csPCa yields across PI-RADS categories are detailed in Table 5.

Category-specific performance and International Society of Urological Pathology upgrades

Significant differences in diagnostic performance emerged across PI-RADS categories, with MRI-TB demonstrating category-specific advantages in both PCa detection and risk stratification (Table 5). Critically, among the 41 PI-RADS 3 lesions, MRI-TB identified four csPCas (ISUP GG ≥2) missed by SB,

Table 2. Demographic and clinical characteristics of the study cohort

	Value
Patients	141
Lesion of numbers	187
Age	
Mean ± SD	64.3 ± 7
Median (range)	64 (45–83)
PSA (ng/mL)	
Mean ± SD	10 ± 9.2
<4.0 ng/mL	16 (11.3%)
4.0–10.0 ng/mL	80 (56.7%)
>10.0 ng/mL	45 (32%)
Prostate volume (mL)	
Mean ± SD	72.17 ± 34.82
Median (range)	70 (13–196)
PSA density	
Mean ± SD	0.16 ± 0.17
>0.15 ng/mL/mL	48/141 (34%)*
Lesion characteristics	
Median size (mm)	14 (IQR: 10–18)
PI-RADS 3	41 (21.9%)
PI-RADS 4	92 (49.1%)
PI-RADS 5	54 (29%)
Biopsy parameters	
Median targeted cores per lesion	3 (IQR: 2–5)
Median total cores per patient	15 (IQR: 14–18)
Clinical findings	
Abnormal digital rectal examination	44 (31.2%)
Chronic prostatitis on histopathology	48 (34)

*PSAD >0.15 ng/mL/mL calculated for entire cohort.
IQR, interquartile range; PI-RADS, Prostate Imaging Reporting and Data System; PSA, prostate-specific antigen; SD, standard deviation.

yielding a csPCa detection rate of 9.8% (4/41) for MRI-TB compared with 0% (0/41) for SB. For the 92 PI-RADS 4 lesions, MRI-TB demonstrated superior csPCa detection to SB [20.7% (19/92) vs. 10.9% (10/92); $P = 0.004$]. When using CB as the reference standard, this resulted in 3 false negatives for MRI-TB vs. 12 false negatives for SB among PI-RADS 4 lesions. In lesions where both methods detected cancer (n = 17), MRI-TB upgraded the ISUP grade in 35.3% (6/17) of cases—comprising two upgrades from GG 1 to GG 2, three from GG 2 to GG 3, and one from GG 1 to GG 3—thereby altering clinical risk stratification (Figure 2). In the 54 PI-RADS 5 lesions, both methods showed equivalent csPCa detection [18.5% (10/54) each; $P = 1.000$]. Among lesions where both MRI-TB and SB detected PCa (n = 13), MRI-TB upgraded the ISUP grade in 7.7% of cases (1/13), specifically from GG 1 to GG 3 (Figure 3). Importantly, CP remained a prevalent confounding factor, observed in 22.0% (9/41) of PI-RADS 3, 22.8% (21/92) of PI-RADS 4, and 27.8% (15/54) of PI-RADS 5 lesions, and coexisted with PCa in 55.6% (5/9), 52.4% (11/21), and 33.3% (5/15) of these CP cases, respectively.

To further investigate the performance of PI-RADS 5 lesions, two post-hoc sensitivity analyses were performed. First, to address potential overscoring, we applied an additional quantitative ADC threshold of <750 $\mu m^2/s$ to define high-risk PI-RADS 5 lesions. This refined subgroup (n = 38) demonstrated a higher csPCa detection rate of 28.9% (11/38) by using CB. Second, to evaluate the impact of tumor volume, we considered high-volume Gleason score 6 (ISUP GG 1) disease (defined as ≥50% core involvement or ≥2 positive cores) as csPCa. This reclassification increased the csPCa detection rate for all PI-RADS 5 lesions from 20.4% (11/54) to 24.1% (13/54).

Clinical parameter correlations

Patients with csPCa exhibited significantly higher PSA levels (15.3 vs. 8.2 ng/mL; $P = 0.02$), smaller prostate volumes (52.4 vs. 78.6 mL; $P < 0.001$), and elevated PSAD (0.30 vs. 0.10 ng/mL/mL; $P < 0.001$). A PSAD threshold >0.15 ng/mL/mL exhibited 78% sensitivity for csPCa; this threshold was significantly more frequent in the csPCa group than in the non-csPCa group (78% vs. 18%; $P < 0.001$). No significant age difference existed between groups (Table 6).

Table 3. Characteristics of peripheral zone lesions stratified by PI-RADS score

Characteristic	PI-RADS 3 (n = 41)	PI-RADS 4 (n = 92)	PI-RADS 5 (n = 54)	*P* value
Lesion size (mm)				
Median (IQR)	10 (8–12)	13 (11–14)	17 (15–22)	<0.001
Chronic prostatitis, n (%)	9 (22.0)	21 (22.8)	15 (27.8)	0.698
PSA, ng/mL				
Median (IQR)	7.1 (5.2–9.8)	8.9 (6.0–12.0)	14.5 (9.8–21.0)	<0.001
Prostate volume, mL				
Median (IQR)	82 (65–105)	68 (50–88)	55 (40–75)	<0.001
PSAD, ng/mL/mL				
Median (IQR)	0.09 (0.06–0.12)	0.13 (0.09–0.19)	0.26 (0.17–0.40)	<0.001
PSAD >0.15, n (%)	7 (17.1)	32 (34.8)	32 (59.3)	<0.001

Statistically significant *P* values ($P < 0.05$) are shown in bold. IQR, interquartile range; PSA, prostate-specific antigen; PSAD, PSA density; PI-RADS, Prostate Imaging Reporting and Data System.

Table 4. Histopathological diagnosis of patients with PI-RADS ≥3 peripheral zone lesions

Histopathological diagnosis	Patients, n (%)
Prostate cancer	73 (51.8)
Clinically significant (ISUP GG ≥2)	37 (26.2)
Clinically insignificant (ISUP GG 1)	36 (25.6)
Atypical small acinar proliferation	1 (0.7)
Benign pathology	67 (47.5)
Chronic prostatitis Benign without prostatitis	48 (34) 19 (13.5)
Total	141 (100)

GG, grade group; ISUP, International Society of Urological Pathology; PI-RADS, Prostate Imaging Reporting and Data System.

Table 5. Comparative detection rates of prostate cancer subtypes by biopsy method across PI-RADS categories in peripheral zone lesions

PI-RADS score		Systematic biopsy (SB) (n, %)	MRI-targeted biopsy (TB) (n, %)	Combined biopsy (CB) (n, %)	SB vs. MRI-TB *P* value	SB vs. CB *P* value	MRI-TB vs. CB *P* value
3 (n = 41)	PCa	6 (41)	16 (39.1)	17 (41.5)	0.006	0.001	1.000
	csPCa	0	4 (9.8)	4 (9.8)	0.125	0.125	1.000
	ciPca	6 (41)	12 (29.3)	13 (31.7)	–	–	–
	Benign	35 (85.4)	25 (60.9)	24 (58.5)	–	–	–
4 (n = 92)	PCa	28 (30.5)	30 (32.7)	41 (44.6)	0.839	**<0.001**	**0.001**
	csPCa	10 (10.9)	19 (20.7)	22 (23.9)	**0.004**	**<0.001**	0.25
	ciPca	18 (19.6)	11 (12)	19 (20.7)	–	–	–
	Benign	64 (69.5)	62 (67.3)	51 (55.4)	–	–	–
5 (n = 54)	PCa	16 (29.6)	19 (35.2)	22 (40.8)	0.508	0.031	0.250
	csPCa	10 (18.5)	10 (18.5)	11 (20.4)	1.000	1.000	1.000
	ciPca	6 (11.1)	9 (16.7)	11 (20.4)	–	–	–
	Benign	38 (70.4)	35 (64.8)	32 (59.2)	–	–	–

Comparisons were performed only for csPCa detection rates between biopsy methods. '–' indicates no statistical comparison performed. Statistically significant *P* values ($P < 0.05$) are shown in bold.
PI-RADS, Prostate Imaging Reporting and Data System; PCa, prostate cancer; csPCa, clinically significant prostate cancer; ciPCa, clinically insignificant prostate cancer.

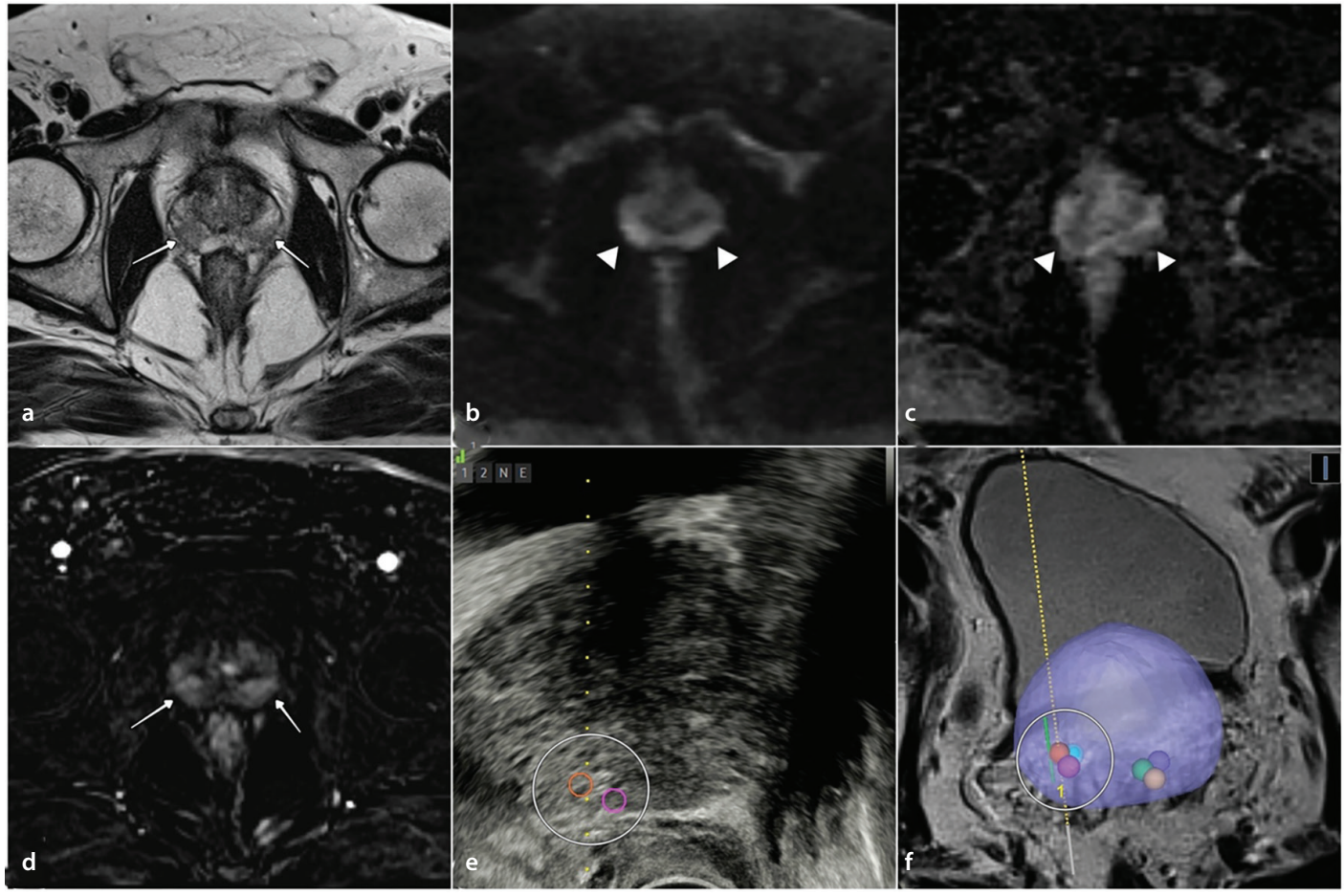


Figure 2. Multiparametric magnetic resonance imaging (MRI) and MRI-targeted biopsy (MRI-TB) findings in a 67-year-old man with bilateral PI-RADS category 4 peripheral zone (PZ) lesions. **(a)** Axial T2-weighted image demonstrates heterogeneous hypointense foci in bilateral PZs (arrows). **(b)** High b-value diffusion-weighted imaging (b = 1.400 s/mm^2) shows corresponding hyperintense signal. **(c)** Apparent diffusion coefficient map confirms diffusion restriction with hypointense signal (arrowheads). **(d)** Dynamic contrast-enhanced subtraction image reveals early arterial phase enhancement in both lesions (arrows). **(e)** Transrectal ultrasound with electromagnetic needle tracking during fusion biopsy of right PZ lesion (ring). **(f)** MRI-TB software overlay displaying co-registered bilateral targets with planned needle trajectory (ring). Histopathologic diagnosis: MRI-TB confirmed Gleason score 3 + 4 = 7 (ISUP GG 2) adenocarcinoma in both lesions. PI-RADS, Prostate Imaging Reporting and Data System; GG, grade group; ISUP, International Society of Urological Pathology.

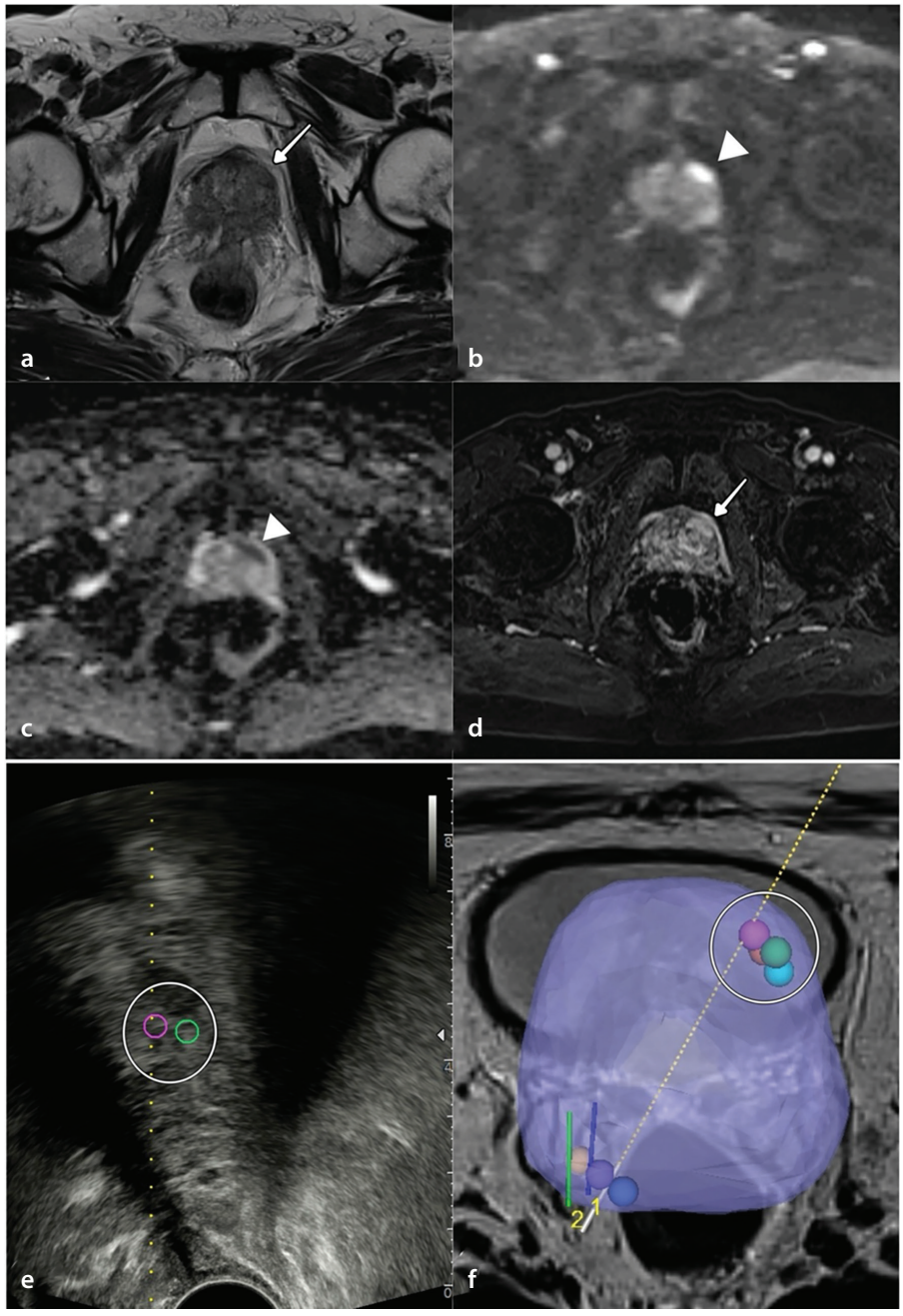


Figure 3. Multiparametric magnetic resonance imaging (MRI) and MRI-targeted biopsy (MRI-TB) of a PI-RADS category 5 peripheral zone lesion in a 75-year-old man. **(a)** Axial T2-weighted image shows a distinct hypointense focus in the left anterior peripheral zone (PZ) (arrow). **(b)** High b-value diffusion-weighted imaging (b = 1.400 s/mm^2) demonstrates marked hyperintensity. **(c)** Apparent diffusion coefficient map reveals corresponding hypointensity indicating diffusion restriction (arrowheads). **(d)** Dynamic contrast-enhanced subtraction image displays early intense arterial enhancement (arrow). **(e)** Transrectal ultrasound with electromagnetic tracking during biopsy of the left PZ target (ring). **(f)** MRI-TB fusion software overlay depicting needle trajectory registration (ring). Histopathologic diagnosis: MRI-TB confirmed Gleason score 4 + 4 = 8 (ISUP GG 4) adenocarcinoma. PI-RADS, Prostate Imaging Reporting and Data System; GG, grade group; ISUP, International Society of Urological Pathology.

Table 6. Comparative analysis of clinical parameters in patients with and without clinically significant prostate cancer (csPCa)

	csPCa (–) (n = 104, %)	csPCa (+) (n = 37, %)	*P* value
PSA (ng/mL)	8.2 ± 5.1	15.3 ± 14.2	**0.02**
Prostate volume (mL)	78.6 ± 33.7	52.4 ± 25.8	**<0.001**
PSA density (PSAD) **PSAD >0.15 ng/mL/mL**	0.10 ± 0.05 19 (18)	0.30 ± 0.25 29 (78.4)	**<0.001** **<0.001**
Age	63.8 ± 6.5	65.9 ± 8.1	0.185

Data presented as mean ± standard deviation or n (%). Statistically significant *P* values (*P* < 0.05) are shown in bold. PSA, prostate-specific antigen.

Discussion

Our study establishes that MRI-TB significantly enhances the detection of csPCa in PI-RADS ≥3 PZ lesions compared with SB, with maximal advantage observed in PI-RADS 4 lesions, where csPCa detection doubled (20.7% vs. 10.9%; $P = 0.004$). Crucially, our study addresses three critical gaps in PZ-specific diagnosis: 1) quantifying CP as a primary confounder (identified in 34.0% of benign cases), 2) demonstrating category-dependent biopsy performance across PI-RADS 3/4/5 subgroups, and 3) validating PSAD >0.15 ng/mL/mL as a robust predictor of csPCa (78% sensitivity). These findings refine biopsy pathways for the PCa-prone PZ—where 70%–80% of malignancies originate—through rigorous correlation of mpMRI features with histopathological outcomes.[2,3]

The principal novelty of our study lies in providing granular, category-specific data that refine the application of MRI-TB and PSAD for PZ lesions. We demonstrate that the diagnostic advantage of MRI-TB is maximal for PI-RADS 4 lesions, shows equipoise with SB for PI-RADS 5, and remains critical for detecting csPCa in PI-RADS 3. Furthermore, we quantitatively demonstrate that CP is a pervasive confounder not only in indeterminate (PI-RADS 3) but also in highly suspicious (PI-RADS 4–5) lesions, where it frequently coexists with cancer, thus presenting a profound diagnostic challenge. This precise stratification, combined with the validation of a PSAD >0.15 ng/mL/mL threshold in this CP-enriched cohort, provides a novel, actionable framework for personalizing biopsy decisions.

In our study, the CB approach demonstrated superior overall PCa detection (42.8% of lesions) to both MRI-TB (34.8%, $P < 0.001$) and SB (26.7%, $P < 0.001$), aligning with findings by Ahdoot et al.[14] For csPCa, MRI-TB significantly outperformed SB (17.7% vs. 10.7% of lesions; $P < 0.001$), corroborating the reported 38% vs. 26% detection advantage for MRI-TB in the study by Kasivisvanathan et al.[16] Our csPCa detection rates (SB: 10.7%; MRI-TB: 17.7%) were lower than those in high-prevalence cohorts, attributable to the following: (1) exclusion of TZ lesions with higher csPCa risk; (2) high CP prevalence (34% of benign cases), reducing specificity; and (3) a lower-risk cohort (only 32% with PSA >10 ng/mL vs. 45%–60% in other studies).[2,3,16-19] Crucially, CB provided no significant additional benefit over MRI-TB alone for csPCa detection (19.8% vs. 17.7%; $P = 0.125$), suggesting targeted sampling may suffice as a primary diagnostic method when supplemented with validated risk-stratification tools. Additionally, a meta-analysis by Schoots et al.,[9] including 16 studies and 1,926 patients, demonstrated that MRI-TB was 20% more effective than SB in detecting csPCa ($P < 0.05$), although no significant difference was noted in overall PCa detection between the two modalities. Our findings thus reinforce the superior csPCa detection of MRI-TB over SB and further confirm CB's advantage in overall PCa detection compared with SB alone.

The management of PI-RADS 3 lesions remains challenging due to their indeterminate nature. In our cohort, although the difference in csPCa detection between MRI-TB and SB in PI-RADS 3 lesions did not reach statistical significance, MRI-TB identified csPCa cases missed by SB. This highlights MRI-TB's added diagnostic value, even in lower-risk lesions, and supports its role in the diagnostic pathway for PI-RADS 3 cases. The low overall csPCa yield (9.8%) in PI-RADS 3 lesions—coupled with a 22% CP rate—further underscores the limitations of relying solely on mpMRI for indeterminate lesions. Thus, adjunctive biomarkers (e.g., PCA3, SelectMDx) or serial PSAD monitoring should be integrated to optimize risk stratification and reduce unnecessary biopsies.[17,18]

In de Braekt et al.[19] reported equivalent csPCa detection between SB and MRI-TB for PI-RADS 4 lesions ($P > 0.05$). This contrasts sharply with our findings, where MRI-TB demonstrated significantly superior csPCa detection over SB (20.7% vs. 10.9%; $P = 0.004$). The discrepancy may be attributed to key methodological differences—our cohort exclusively comprised patients who were biopsy naïve, and MRI-TB dominated csPCa detection within CB-positive cases. Crucially, the limited diagnostic advantage of MRI-TB over SB in our cohort also stems from prevalent CP (34% of benign cases), which mimics csPCa on mpMRI.

Our study reveals a critical nuance in PI-RADS 4 lesions; the limited diagnostic advantage of MRI-TB over SB (csPCa detection: 20.7% vs. 10.9%; $P = 0.004$) likely stems from prevalent CP, which constituted 34% of benign cases. This inflammatory condition mimics csPCa on mpMRI through characteristic DWI restriction and early DCE enhancement patterns, posing a major diagnostic challenge.[16,19] For such lesions, adjunctive PSAD (>0.15 ng/mL/mL) or non-focal morphology assessment are recommended to enhance specificity and reduce unnecessary biopsies.

The 26.9% prevalence of CP in our cohort (34.0% of benign cases) aligns with contemporary studies of PI-RADS ≥3 populations, confirming this reflects expected diagnostic challenges rather than cohort irregularity.[17–19] Patients with elevated PSA levels (mean: 10.03 ng/mL) and suspicious mpMRI lesions are inherently enriched for inflammatory conditions that mimic csPCa through shared imaging features—particularly diffusion restriction and early enhancement patterns on DCE sequences.[20] Furthermore, the exclusion of patients with prior biopsies likely contributed to the observed prevalence of detection of subclinical CP, as untreated inflammation accumulates over time.[1,5] Notably, the 31.2% abnormal DRE rate (44/141) correlates with both CP prevalence (26.9%) and contemporary MRI-TB studies, suggesting palpable glandular irregularities may reflect underlying inflammatory changes that confound mpMRI interpretation.[21,22]

For PI-RADS 5 lesions, the equivalent csPCa detection rates between SB and MRI-TB (18.5% each; $P = 1.000$) indicate comparable efficacy for large, conspicuous lesions (median size: 17 mm, IQR: 15–22 mm). However, our overall csPCa detection rate for PI-RADS 5 lesions (20.4% by CB) is lower than the very high rates (60%–90%) often reported in the literature.[16,19,21] This critical discrepancy can be attributed to several factors inherent to our study cohort: (1) the exclusion of TZ cancers, which frequently present as aggressive PI-RADS 5 tumors; (2) the high prevalence of CP (27.8%); and (3) nearly 40% of lesions having a PSAD below the 0.15 ng/mL/mL threshold, suggesting a less aggressive profile. Therefore, although PI-RADS 5 remains a high-risk category, its predictive value is not absolute and is substantially influenced by the underlying prevalence of confounding inflammation and the specific risk profile of the patient population. This aligns with the findings of In de Braekt et al.[19] but contrasts with studies reporting MRI-TB superiority.[21,22] In our cohort, 27.8% of PI-RADS 5 lesions showed imaging features consistent with CP, complicating lesion characterization. Technical heterogeneity in MRI acquisition, interpretation variability, operator experience, or biopsy targeting accuracy may underlie these discrepancies. Consequently, for morphologically overt PI-RADS 5 PZ lesions (≥15 mm), systematic sampling may achieve comparable csPCa detection to targeted ap-

proaches, potentially optimizing resource utilization in high-volume settings. Nevertheless, the 27.8% CP prevalence remains a critical confounder for mpMRI specificity.[20]

To address the potential for overscoring and refine the predictive value of PI-RADS 5 lesions, two post-hoc sensitivity analyses were performed. Our post-hoc sensitivity analyses offer potential pathways to refine the predictive value of PI-RADS 5 lesions. The application of a quantitative ADC threshold (<750 μm²/s) successfully identified a subgroup with a higher csPCa yield (28.9%), suggesting that incorporating quantitative metrics can help distinguish the highest-risk lesions within this category. Furthermore, considering high-volume Gleason 6 disease as clinically significant increased the overall csPCa detection rate to 24.1%, highlighting that a subset of these tumors may have greater clinical relevance than traditionally assigned. These exploratory analyses emphasize that beyond the qualitative PI-RADS 5 score, adjunctive quantitative and volumetric parameters could enhance risk stratification and clinical decision-making.

Regarding differential csPCa detection across PI-RADS categories, our data revealed two key patterns. First, for PI-RADS 4 lesions, CB detected significantly more csPCa than SB alone (23.9% vs. 10.9%; $P < 0.001$) and marginally more than MRI-TB alone (23.9% vs. 20.7%; $P = 0.25$). This three-case increment (22 vs. 19 lesions) reflects CB's capacity to sample both MRI-suspicious foci and occult csPCa outside targets. Importantly, contemporary evidence indicates that SB detects csPCa in approximately 10% of cases within MRI-normal contralateral lobes, reinforcing SB's role in comprehensive sampling despite negative mpMRI findings.[23] Prior studies confirm that 10%–15% of csPCa is MRI invisible but is detected by random sampling, particularly in glands with heterogeneous backgrounds such as CP.[16,19] Second, in PI-RADS 5 lesions, equivalent csPCa detection by SB and MRI-TB (18.5% each; $P = 1.000$) was observed. This equivalence likely stems from their conspicuous size (median 17 mm, IQR: 15–22 mm), enabling comparable sampling efficacy for both methods. Notably, 27.8% of these lesions exhibited concurrent CP, potentially contributing to false-positive MRI interpretations.[20]

Our findings validate PSAD as a pivotal discriminator for csPCa, with 78% of patients with csPCa exhibiting PSAD > 0.15 ng/mL/mL versus 18% in non-csPCa cases ($P < 0.001$). This aligns with prior evidence that PSAD enhances specificity in PI-RADS 4 lesions, where CP mimics malignancy.[17,20] The inverse correlation between prostate volume and csPCa risk (csPCa: 52.4 vs. non-csPCa: 78.6 mL; $P < 0.001$) further supports volumetric assessment in biopsy decisions. Although PSA was elevated in csPCa (15.3 vs. 8.2 ng/mL; $P = 0.02$), its overlap with inflammatory conditions (e.g., CP) limits standalone utility. We advocate integrating PSAD >0.15 ng/mL/mL into MRI-TB workflows to avoid unnecessary biopsies in equivocal PI-RADS 3–4 lesions.[17,24]

Although age is a well-established risk factor for PCa,[1,25] our study found no statistically significant difference between the cancer and non-cancer groups (median: 65 vs. 63.5 years; $P > 0.05$). This may reflect cohort-specific characteristics—nearly half the patients (49.7%) were aged 61–70 years, with balanced distribution between groups. Potential explanations include the relatively narrow age range, limited sample size for subgroup analyses, or selection bias from exclusively including PI-RADS ≥3 cases. Nevertheless, the established role of age in PCa risk stratification remains unchallenged in broader populations.[26]

Our study has several limitations requiring acknowledgment. First, its retrospective, single-center design and exclusive focus on PZ lesions may introduce selection bias, partially explaining lower csPCa rates than in studies including TZ cancers or high-PSA cohorts.[2,3,15,18] Second, the use of CB as the reference standard (as detailed under Methods) may affect accuracy estimates for MRI-invisible cancers, although this is a recognized limitation in biopsy comparison studies.[13–15,25] Additionally, tumor volume was estimated from core biopsy specimens rather than measured from radical prostatectomy specimens. Although this is a standard methodology for pre-treatment risk stratification, it remains an estimation subject to sampling error. Third, we did not stratify lesions by specific PZ location; thus, the advantage of MRI-TB over SB may be underestimated for under-sampled regions. The absence of this location-specific analysis limits the granularity of our conclusions regarding the differential advantage of MRI-TB. Future prospective studies designed to include such detailed anatomical mapping are warranted to provide more specific guidance. Fourth, reliance on a single radiologist for PI-RADS scoring precludes assessment of inter-observer variability. Fifth, although PI-RADS version 2.1 recommends 3T MRI, diagnostic accuracy for PI-RADS ≥3 PZ lesions is comparable at 1.5T with modern sequences; however, exclusive use of a pelvic phased-array coil (without endorectal coil) could reduce spatial resolution for sub-centimeter lesions—a limitation mitigated by our cohort's median lesion size.[13] Sixth, performing SB after MRI-TB may introduce hemorrhage-related sampling bias. Seventh, the transrectal biopsy approach, although appropriate for posteriorly located PZ lesions, contrasts current guideline recommendations favoring transperineal methods to reduce sepsis risk.[13,27] Finally, the high CP prevalence remains a key confounder for PI-RADS specificity, underscoring the need for future integration of quantitative imaging biomarkers (e.g., ADC histogram analysis) to improve discrimination.[18,21]

In conclusion, our study provides a refined, evidence-based algorithm for prostate biopsy in patients with PZ lesions, moving beyond broad principles to deliver category-specific guidance. We establish that the diagnostic advantage of MRI-TB is not uniform but is maximized for PI-RADS 4 lesions, whereas it is equivalent to SB for large, conspicuous PI-RADS 5 lesions. The high prevalence of CP (34% of benign cases) is a major confounder across all categories, frequently mimicking csPCa and reducing mpMRI specificity. Furthermore, we validate a PSAD threshold of >0.15 ng/mL/mL as a pivotal tool for risk stratification. These findings support using MRI-TB alone for PI-RADS ≥4 lesions or in patients with elevated PSAD, reserving CB for equivocal PI-RADS 3–4 cases. The high rate of inflammatory mimics underscores the critical need for integrating adjunctive biomarkers to improve specificity in the future.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;74(1):12-49. [Crossref]
2. Yacoub JH, Oto A. MR imaging of prostate zonal anatomy. *Radiol Clin North Am.* 2018;56(2):197-209. [Crossref]
3. Lee CH, Akin-Olugbade O, Kirschenbaum A. Overview of prostate anatomy, histology, and pathology. *Endocrinol Metab Clin North Am.* 2011;40(3):565-575, viii-ix. [Crossref]
4. Madersbacher S, Alcaraz A, Emberton M, et al. The influence of family history on prostate cancer risk: implications for clinical management. *BJU Int.* 2011;107(5):716-721. [Crossref]

5. Roddam AW, Duffy MJ, Hamdy FC, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/mL: systematic review and meta-analysis. *Eur Urol.* 2005;48(3):386-99; discussion 398-399. [Crossref]
6. Huang Y, Li ZZ, Huang YL, Song HJ, Wang YJ. Value of free/total prostate-specific antigen (f/t PSA) ratios for prostate cancer detection in patients with total serum prostate-specific antigen between 4 and 10 ng/mL: a meta-analysis. *Medicine (Baltimore).* 2018;97(13):e0249. [Crossref]
7. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. *J Urol.* 1992;147(3 Pt 2):815-816. [Crossref]
8. Omer A, Lamb AD. Optimizing prostate biopsy techniques. *Curr Opin Urol.* 2019;29(6):578-586. [Crossref]
9. Schoots IG, Roobol MJ, Nieboer D, Bangma CH, Steyerberg EW, Hunink MG. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. *Eur Urol.* 2015;68(3):438-450. [Crossref]
10. Brown AM, Elbuluk O, Mertan F, et al. Recent advances in image-guided targeted prostate biopsy. Abdom Imaging. 2015;40(6):1788-1799. [Crossref]
11. Gaziev G, Wadhwa K, Barrett T, et al. Defining the learning curve for multiparametric magnetic resonance imaging (MRI) of the prostate using MRI-transrectal ultrasonography (TRUS) fusion-guided transperineal prostate biopsies as a validation tool. *BJU Int.* 2016;117(1):80-86. [Crossref]
12. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol.* 2019;76(3):340-351. [Crossref]
13. Cornford P, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer-2024 Update. Part I: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol.* 2024;86(2):148-163. [Crossref]
14. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. *N Engl J Med.* 2020;382(10):917-928. [Crossref]
15. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol.* 2016;40(2):244-252. [Crossref]
16. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med.* 2018;378(19):1767-1777. [Crossref]
17. Görtz M, Radtke JP, Hatiboglu G, et al. The value of prostate-specific antigen density for prostate imaging-reporting and data system 3 lesions on multiparametric magnetic resonance imaging: a strategy to avoid unnecessary prostate biopsies. *Eur Urol Focus.* 2021;7(2):325-331. [Crossref]
18. Brizmohun Appayya M, Sidhu HS, Dikaios N, et al. Characterizing indeterminate (Likert-score 3/5) peripheral zone prostate lesions with PSA density, PI-RADS scoring and qualitative descriptors on multiparametric MRI. *Br J Radiol.* 2018;91(1083):20170645. [Crossref]
19. In de Braekt T, van Rooij SBT, Daniels-Gooszen AW, et al. Accuracy of MRI-ultrasound fusion-guided and systematic biopsy of the prostate. *Br J Radiol.* 2024;97(1158):1132-1138. [Crossref]
20. Ullrich T, Arsov C, Quentin M, et al. Analysis of PI-RADS 4 cases: management recommendations for negatively biopsied patients. *Eur J Radiol.* 2019;113:1-6. [Crossref]
21. Liu Y, Wang S, Xu G, et al. Effectiveness and accuracy of MRI-ultrasound fusion targeted biopsy based on PI-RADS v2.1 category in transition/peripheral zone of the prostate. *J Magn Reson Imaging.* 2023;58(3):709-717. [Crossref]
22. Nativ O, Shefler A, Bejar J, et al. Performance of standard systematic biopsy versus MRI/TRUS fusion biopsy using the Navigo® system in contemporary cohort. *Urol Oncol.* 2024;42(5):159.e1-159.e7. [Crossref]
23. Pepe P, Pepe L, Fiorentino V, Curduman M, Fraggetta F. Multiparametric MRI targeted prostate biopsy: when omit systematic biopsy? *Arch Ital Urol Androl.* 2024;96(4):12992. [Crossref]
24. Distler FA, Radtke JP, Bonekamp D, et al. The value of PSA density in combination with PI-RADS™ for the accuracy of prostate cancer prediction. *J Urol.* 2017;198(3):575-582. [Crossref]
25. Zhou CK, Check DP, Lortet-Tieulent J, et al. Prostate cancer incidence in 43 populations worldwide: an analysis of time trends overall and by age group. *Int J Cancer.* 2016;138(6):1388-400. [Crossref]
26. Fiorentino V, Pepe L, Zuccalà V, et al. Gleason score down and upgrading at radical prostatectomy in targeted vs. systematic prostate biopsy: findings from an institutional cohort. *Pathol Res Pract.* 2025;271:156040. [Crossref]
27. Greer MD, Brown AM, Shih JH, et al. Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: a multireader study. *J Magn Reson Imaging.* 2017;45(2):579-585. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253601

INTERVENTIONAL RADIOLOGY

TECHNICAL NOTE

Minimally invasive management of a post-cholecystectomy bile leak using a microvascular plug

Mohammed Misbahuddin-Leis[1-3]
Andreas R. R. Weiß[4]
Yazan Amin[4]
Robert Grützmann[4]
Axel Schmid[1]

[1]Friedrich-Alexander-University of Erlangen-Nuremberg, Institute of Radiology, University Hospital Erlangen, Erlangen, Germany

[2]Heidelberg University Faculty of Medicine, Heidelberg, Germany

[3]Sana Klinikum Hof GmbH, Department of Diagnostic and Interventional Radiology, Hof, Germany

[4]University Hospital Erlangen, Department of Surgery, Erlangen, Germany

ABSTRACT

Bouveret syndrome is a rare form of gallstone ileus in which a gallstone causes gastric outlet obstruction via a cholecystoenteric fistula. We present the case of an elderly patient who underwent emergency surgery due to clinical signs of gastric outlet obstruction. Intraoperatively, a large gallstone was identified and removed from the duodenum. A subtotal cholecystectomy was performed, during which the identification of the course of the cystic duct was limited due to considerable inflammatory changes. In the early postoperative period, bile-stained output from the surgical drain raised suspicion of a leak from the cystic duct stump. Given the patient's clinical status and the complexity of the local anatomy with severely inflamed and fibrotic tissues in the hepatoduodenal ligament, the decision was made to proceed with percutaneous treatment. A percutaneous transhepatic biliary drainage was established, and contrast extravasation from the cystic duct remnant was confirmed via fluoroscopic cholangiography. Through the same access, a microvascular plug was successfully deployed into the cystic duct, achieving immediate cessation of bile leakage. The patient's clinical condition improved markedly, and no further intervention was required. This case demonstrates the successful off-label use of polytetrafluoroethylene-covered vascular occlusion devices in managing biliary complications.

KEYWORDS

Cystic duct stump, bile leak, cholecystectomy complication, interventional radiology, microvascular plug, biliary embolization, iatrogenic injury

A leak from the cystic duct stump is an uncommon complication following cholecystectomy, with a reported incidence ranging from 0.3% to 2%, varying depending on the surgical context and patient factors.[1] Interventional radiology has played an increasingly major role in managing biliary leaks, particularly when conventional surgical or endoscopic options are limited. Techniques such as percutaneous drainage or embolization have been documented in the literature for controlling bile leaks. However, the use of microvascular plugs (MVPs), devices traditionally employed in vascular interventions, has not yet been explored in the context of biliary tract occlusion.

This note describes a case of cystic duct stump leakage following open cholecystectomy in the context of Bouveret syndrome that was successfully treated with percutaneous embolization using an MVP.

Methods

A 71-year-old Caucasian male patient presented to the emergency department with multiple episodes of vomiting (up to six times daily) and considerable oral intolerance over 3 days. Clinically, a distended abdomen with diffuse tenderness was noted on presentation, without signs of peritonism. Laboratory investigations on admission showed elevated inflammatory and hepatobiliary parameters. Written informed consent was obtained from the patient for publication of this case and accompanying images.

Corresponding author: Mohammed Misbahuddin-Leis

E-mail: misbahuddin.mohammed@outlook.com

Received 19 August 2025; revision requested 24 August 2025; accepted 07 September 2025.

Epub: 19.11.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253601

You may cite this article as: Misbahuddin-Leis M, Weiß ARR, Amin YA, Grützmann R, Schmid A. Minimally invasive management of a post-cholecystectomy bile leak using a microvascular plug. *Diagn Interv Radiol*. 2026;32(3):338-341.

Contrast-enhanced multiphase multidetector computed tomography of the abdomen (Somatom X.ceed, Siemens Healthineers, Forchheim, Germany) revealed a markedly distended stomach with upstream gastric outlet obstruction. A large, hyperdense intraluminal gallstone measuring approximately 4 cm was visualized in the descending duodenum. Additionally, the gallbladder appeared atrophic, with evidence of a cholecystoduodenal fistula. These findings were consistent with Bouveret syndrome (see Figure 1a, b).

Following radiological suspicion of gastric outlet obstruction due to a gallstone, the patient underwent emergency exploratory laparotomy. Intraoperatively, a large gallstone was identified and extracted from the duodenum. The cholecystoduodenal fistula was closed primarily with sutures and an omental patch, and a subtotal cholecystectomy was performed, leaving the fibrotic back wall of the gallbladder *in situ*, which was coagulated with bipolar forceps. The fragile and fibrotic cystic duct stump was oversewn with a 4/0 polydioxanone suture. At the end of the operation, white gauze was temporarily placed in the gallbladder fossa and on the cystic duct stump, showing no evidence of a bile leak.

However, on the first postoperative day, approximately 300 mL of bile-stained fluid was observed in the intra-abdominal Robinson drain over 24 hours, raising clinical suspicion of a cystic duct stump leak. Therefore, the interventional radiology team was consulted for further evaluation and image-guided treatment. On the following day (postoperative day 2), an ultrasound-guided puncture of the right biliary system was performed with a 22G needle (Neff Percutaneous Access Set™, Cook Medical, Bloomington, IN, USA) under local anesthesia and intravenous sedation using midazolam and piritramide in our angio suite. A 0.018-inch guidewire facilitated the placement of a 6F sheath, followed by biliary system recanalization using a 0.035-inch guidewire (Guide Wire M, Radiofocus™, Terumo Corporation, Tokyo, Japan) and a 4Fr/65 cm angle catheter (Glidecath®, Terumo). A cholangiogram demonstrated contrast extravasation originating from the cystic duct stump, thereby confirming the suspected leak. A 4F RIM catheter (Tempo™, Cordis, Miami Lakes, FL, USA) was positioned at the ostium of the cystic duct. The cystic duct was then accessed with a 2.6F microcatheter (Master Parkway HF, Asahi Intecc Co., Ltd., Aichi, Japan), and a 0.014-inch microwire (Transend™, Boston Scientific, Marlborough, MA, USA). An MVP (MVP-5Q, Medtronic, Minneapolis, MN, USA) designed for a target vessel diameter of 3.0–5.0 mm was successfully advanced and deployed through the microcatheter into the cystic duct to achieve occlusion. The sheath was removed, and the puncture site was dilated for percutaneous transhepatic cholangiodrainage (PTCD) insertion. An 8.5F biliary drainage catheter (ReSolve®, Merit Medical Systems, South Jordan, UT, USA) was positioned into the duodenum, and cholangiography confirmed its correct placement. A follow-up fluoroscopy demonstrated immediate and complete cessation of contrast leakage, confirming the technical success of the intervention (see Figure 2a, b).

Results

The clinical course was favorable, with complete cessation of bile leakage via the Robinson drain the following day and normalization of liver and cholestatic enzymes, so that on postoperative day 3, the Robinson drain was removed. On day 11, the PTCD was exchanged for an 8F Münchner drain (PerkuBil®, Peter Pflugbeil GmbH, Zorneding, Germany).

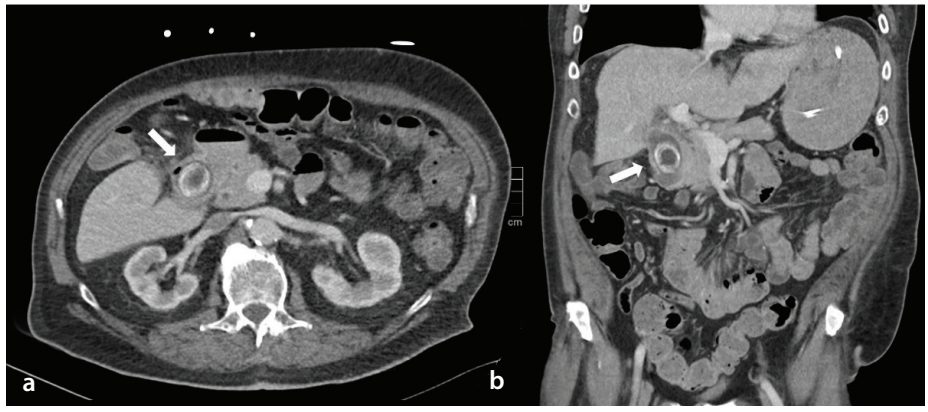


Figure 1. Axial **(a)** and coronal **(b)** contrast-enhanced computed tomography images show a markedly distended stomach due to gastric outlet obstruction. A large hyperdense gallstone is identified in the descending duodenum (Bouveret syndrome, white arrow). The gallbladder appears atrophic, with surrounding inflammatory changes suggesting chronic cholecystitis and a possible cholecystoenteric fistula.

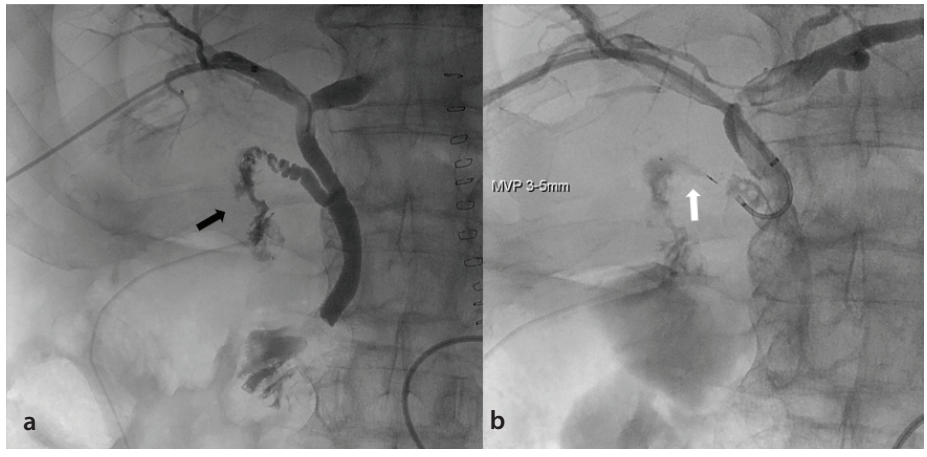


Figure 2. Percutaneous transhepatic cholangiography before **(a)** and after **(b)** deployment of the microvascular plug (MVP) into the cystic duct.

a. Contrast extravasation from the cystic duct stump confirms active bile leakage (black arrow). **b.** Successful placement of the MVP with immediate cessation of contrast leakage, indicating complete occlusion (white arrow).

Main points

- Cystic duct stump leakage is a challenging postoperative complication, particularly after a difficult cholecystectomy with severe inflammation.
- A partially polytetrafluoroethylene-covered microvascular plug can achieve immediate and complete occlusion without relying on thrombus formation.
- This minimally invasive approach offers a safe and effective alternative when standard endoscopic or surgical treatments are not feasible.

Cholangiography performed during the second intervention showed no evidence of persistent bile leakage. The patient tolerated a gradual reintroduction of oral intake and was discharged in a stable condition on postoperative day 19. At the time of discharge, inflammatory and cholestatic parameters had substantially improved. Cholangiography performed on postoperative day 31 still showed sufficient occlusion of the cystic duct, and the Münchner drain was removed (see Figure 3). No further bile leakage or postoperative complications were observed.

Discussion

Bouveret syndrome is a challenging variant of gallstone ileus, characterized by gastric outlet obstruction resulting from the impaction of a gallstone in the duodenum via a cholecystoduodenal fistula, contributing to a reported mortality rate ranging from 12% to 30%.[2] The standard surgical approach typically involves removing the obstructing duodenal gallstone, closing the cholecystoduodenal fistula, and cholecystectomy. In cases where severe inflammation or fibrosis makes complete cholecystectomy hazardous, a subtotal cholecystectomy is often chosen to minimize the risk of injury to the common bile duct and surrounding structures. Cystic duct stump insufficiency is a recognized complication following subtotal cholecystectomy, with incidence rates between 0.3% and 2%.[1] Traditionally, endoscopic retrograde cholangiopancreatography (ERCP) with sphincterotomy and stent placement has been the first-line treatment for bile leaks. However, in the early postoperative period, performing ERCP in the context of a recent duodenal repair carries an increased risk of compromising the suture line, as well as a higher risk of perforation and exacerbation of the leak.[3] In such scenarios, PTCD offers a viable alternative. The time required for the resolution of a leak from the cystic duct stump following PTCD treatment varies depending on the extent of the leak and the use of adjunctive embolization techniques. In most cases, minor bile leaks managed with PTCD alone resolve within 7–21 days as biliary pressure is reduced and the fistula tract seals spontaneously.[4] When embolic agents are used to directly occlude the cystic duct remnant, the time to leak resolution may be shortened considerably to a median of 10 days.[5] Embolic materials such as metallic coils and Amplatzer™ vascular plugs (AVPs) have been widely used, favored for their radiopacity, ease of delivery, and mechanical occlusion. However, their effectiveness can be limited in bile or lymphatic leaks, as both rely primarily on thrombus formation to achieve permanent occlusion. In cases with impaired coagulation or high-flow leakage, fluids may continue to pass through the coils. This represents a key advantage of the MVP, which provides immediate and complete mechanical occlusion without depending on the coagulation system.

Figure 3. Follow-up cholangiography on postoperative day 31 demonstrates sustained occlusion of the cystic duct stump after MVP deployment. No contrast extravasation is visible, confirming the long-term technical success of the intervention. MVP, microvascular plug.

In addition to coils, liquid embolic agents such as N-butyl cyanoacrylate (Histoacryl), Onyx, and absorbable gelatin sponges (Gelfoam) have also been used, either alone or in combination with coils, to manage biliary leaks. These agents are valued for their availability, ease of application, and, in the case of Gelfoam, low cost. However, liquid embolic agents carry a risk of non-target embolization, particularly in the biliary system, where unintended occlusion of adjacent ducts may lead to serious complications. Furthermore, temporary agents such as Gelfoam may be absorbed over time, potentially resulting in a recurrent leak.[6] The use of an MVP in this context represents a novel off-label application of a device traditionally employed in vascular embolization procedures. MVPs are self-expanding, detachable devices designed to occlude blood vessels with precision and minimal migration risk (see Figure 4).[7]

A key advantage in the biliary system is that the MVP can be delivered through a 0.027-inch microcatheter, enabling access to small or tortuous segments such as the cystic duct—an area that would be technically challenging to reach with larger devices such as the AVP. Moreover, the partial polytetrafluoroethylene cover of the MVP may offer

Figure 4. The MVP is an embolic device composed of two key components: a proximal PTFE covering (black star) and an electrochemically resorbable detachment point (black arrow) (product image used with the kind permission of Medtronic). PTFE, partial polytetrafluoroethylene; MVP, microvascular plug.

a major advantage, as its occlusive capability does not rely on thrombus formation. Nevertheless, the deployment of an MVP in the biliary system, particularly for sealing cystic duct leaks, scarcely documented in the literature, making this case a pioneering example of such use. The decision to utilize an MVP was influenced by several factors. The cystic duct's size and orientation made it amenable to occlusion with an MVP, which can conform to varying ductal diameters and provide a secure seal. Unlike coil or liquid embolization, which may require multiple coils and carry a risk of coil or glue migration, the MVP's design allows for single-device occlusion with a potentially lower risk profile.[8] In our case, cholangiography after MVP deployment confirmed sufficient occlusion of the cystic duct, indicating immediate technical success. Complete cessation of bile

leakage via the Robinson drain occurred almost immediately after MVP deployment, clearly outperforming the expected timeframe for resolution when using PTCD alone or in combination with other embolic agents. The successful use of an MVP to treat a leak from the cystic duct stump in this case highlights a potential new avenue in the management of this challenging complication.

In conclusion the use of a percutaneous transhepatic approach combined with targeted embolization using an MVP enabled the effective management of a cystic duct stump leak following cholecystectomy.

Acknowledgments

The authors would like to thank Dr. Farah Farooqui (Raha Medical Center, Al-Khor, Qatar) for her valuable assistance in the preparation of this manuscript.

Footnotes

Conflict of Interest

The authors declared that they have no conflict of interest.

References

1. Shaikh IA, Thomas H, Joga K, Amin AI, Daniel T. Post-cholecystectomy cystic duct stump leak: a preventable morbidity. *J Dig Dis*. 2009;10(3):207-212. [Crossref]
2. Gaillard F, Elfeky M, Saber M, et al. Bouveret syndrome [Internet]. Radiopaedia.org; 2025 Jun 1 [cited 2025 Nov 6]. [Crossref]
3. McVay TR, Adler DG. Bile duct injuries: multidisciplinary evaluation and treatment. *Pract Gastroenterol*. 2015;39(1). [Crossref]
4. Hassan AM, Mohammed AQ. Biliary leak after laparoscopic cholecystectomy: incidence and management. Egypt J Surg. 2021;40(2):685-694. [Crossref]
5. Nezami N, Jarmakani H, Arici M, et al. Selective trans-catheter coil embolization of cystic duct stump in post-cholecystectomy bile leak. *Dig Dis Sci*. 2019;64(11):3314-3320. [Crossref]
6. Abdel Aal AK, Jones DP, Caraway J, et al. Percutaneous embolization of cystic duct stump leak following failed endoscopic management. *Radiol Case Rep*. 2017;12(4):786-789. [Crossref]
7. Carrafiello G, Ierardi AM, Piacentino F, Cardim LN. Percutaneous transhepatic embolization of biliary leakage with N-butyl cyanoacrylate. *Indian J Radiol Imaging*. 2012;22(1):19-22. [Crossref]
8. Özgen A. Use of gelatin sponge to seal the biliary tract after percutaneous transhepatic biliary drainage in patients with liver transplants. *Diagn Interv Radiol*. 2024;30(5):325-327. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253555

MODALITY-BASED (US, CT, MRI, PET-CT) IMAGING

ORIGINAL ARTICLE

Automated calculation of slice-specific volume computed tomography dose index, water-equivalent diameter, and size-specific dose estimation for computed tomography scans

Supawitoo Sookpeng[1]
Rosario Lopez-Gonzalez[2]
Suwapim Chanlaor[3]
Boriphat Kadman[1]

[1]Naresuan University, Faculty of Allied Health Sciences, Department of Radiological Technology, Phitsanulok, Thailand

[2]University of Glasgow, Institute of Neurosciences and Psychology, Department of Clinical Physics and Bioengineering, Glasgow, United Kingdom

[3]Buddhachinaraj Phitsanulok Hospital, Department of Radiology, Phitsanulok, Thailand

Corresponding author: Supawitoo Sookpeng

E-mail: supawitoos@nu.ac.th

Received 18 July 2025; revision requested 01 September 2025; last revision received 05 October 2025; accepted 13 October 2025.

Epub: 03.11.2025

Publication date: 04.05.2026

DOI: 10.4274/dir.2025.253555

PURPOSE

To develop and validate an automated computational tool for calculating a slice-specific volume computed tomography (CT) dose index ($CTDI_{vol}$), a water-equivalent diameter (D_w), and size-specific dose estimates (SSDEs) from CT images, addressing limitations of conventional console-displayed values that provide only averaged values across scan regions.

METHODS

A custom ImageJ macro was developed based on methodologies proposed in American Association of Physicists in Medicine reports 220 and 293. The tool employs threshold-based body contour segmentation [−140 Hounsfield unit (HU)] to extract patient cross-sectional areas and calculates slice-specific D_w using mean CT numbers. Slice-specific $CTDI_{vol}$ values are estimated by normalizing scanner-displayed $CTDI_{vol}$ to individual slice exposure values from Digital Imaging and Communications in Medicine metadata. An SSDE was computed using appropriate correction factors for head and body examinations. Validation was performed using water phantoms, anthropomorphic phantoms, and clinical datasets from ≥30 patients. Two Siemens CT scanners were evaluated: SOMATOM go.Top®, with console-displayed values, and SOMATOM Force®, with Radimetrics software. Agreement was assessed using intraclass correlation coefficients (ICCs) and Bland–Altman analysis.

RESULTS

Water phantom validation demonstrated excellent accuracy, with differences of <2.3% for both D_w and SSDEs. The macro required approximately 30 seconds per examination to complete the analysis. Bland–Altman plots confirmed clinically acceptable mean differences. Importantly, the slice-specific approach revealed substantial intra-scan dose variations not captured by console-reported averages, particularly in the chest phantom, where SSDEs ranged from 5.77 to 23.68 mGy despite identical average values. For the clinical dataset, ICC (3,1) values for Scanner A indicated good to excellent agreement across both head and chest/abdomen examinations (head CT—$CTDI_{vol}$: 0.974, D_w: 0.893, SSDE: 0.965; chest/abdomen CT—$CTDI_{vol}$: 1.000, D_w: 0.994, SSDE: 0.989). By contrast, Scanner B demonstrated near-perfect agreement for head CT in $CTDI_{vol}$ (0.996) and SSDE (0.967) but poor agreement for D_w (0.267). For chest/abdomen CT, however, Scanner B showed consistently high agreement, with ICC values ranging from 0.884 to 1.000.

CONCLUSION

The developed ImageJ macro provides accurate, transparent, and low-cost open-source solution slice-specific CT dose estimation that correlates well with commercial systems while offering superior spatial resolution. This automated method overcomes the limitations of traditional dose reporting by providing detailed slice-by-slice dose variations, which are often overlooked in average summary values, allowing for more accurate and clinically meaningful dose assessments.

CLINICAL SIGNIFICANCE

This tool supports detailed dose evaluation across scan regions, helping optimize protocols and enhance radiation safety. Its slice-specific approach is especially useful in anatomically complex areas and research, offering clinicians more precise dose information to guide patient care.

KEYWORDS

Size-specific dose estimate, water-equivalent diameter, volume computed tomography dose index, ImageJ, dose calculation, patient-specific dosimetry

You may cite this article as: Sookpeng S, Lopez-Gonzalez R, Chanlaor S, Kadman B. Automated calculation of slice-specific volume computed tomography dose index, water-equivalent diameter, and size-specific dose estimation for computed tomography scans. *Diagn Interv Radiol*. 2026;32(3):342-350.

Computed tomography (CT) examinations contribute substantially to radiation exposure in the general population due to their relatively high radiation doses. Radiation doses to individual organs are associated with both deterministic effects, such as skin burns and epilation, and stochastic risks, including cancer induction and genetic mutations.[1-3] Accurate quantification of the radiation dose received by a patient undergoing a CT scan is essential for both radiation protection and clinical optimization. Traditionally, the CT radiation dose has been reported using the volume CT dose index ($CTDI_{vol}$) and the dose-length product (DLP), both of which are derived from output measurements in 16- and 32-cm cylindrical polymethyl methacrylate phantoms. However, $CTDI_{vol}$ is primarily dependent on exposure parameters (e.g., tube current and tube voltage) and does not account for patient size.[4-6] Given that larger patients receive relatively low radiation doses for the same $CTDI_{vol}$, size correction methods have been proposed to improve dose estimation accuracy.

To address the limitations of $CTDI_{vol}$, the American Association of Physicists in Medicine (AAPM) introduced the size-specific dose estimate (SSDE) in Report 204.[7] The SSDE adjusts $CTDI_{vol}$ based on patient size, providing a more individualized dose estimate at the center of the scanned region. The patient's physical dimensions, derived from CT images, are used in SSDE calculations. Initially, AAPM Report 204 relied on geometric size as a proxy for X-ray attenuation. However, since X-ray attenuation depends on tissue density and composition, different anatomical regions (e.g., thorax vs. abdomen) exhibit varying attenuation properties, even when their geometric sizes are identical. For instance, the thorax, being less dense than the abdomen, results in higher radiation exposure for the same $CTDI_{vol}$. To further refine SSDE calculations, AAPM Report 220[8] introduced the concept of the water-equivalent diameter (D_w), which represents the diameter of a cylindrical water volume with equivalent mean attenuation. This approach accounts for tissue composition and provides a more accurate, patient-specific dose estimation; D_w is derived from attenuation values in axial images along the z-axis. More recently, AAPM Report 293 extended the application of SSDEs to head CT examinations by incorporating region-specific correction factors.[9]

Several studies have suggested that SSDEs serve as a more reliable surrogate for organ-absorbed doses on a slice-by-slice basis.[10-12] In clinical practice, automatic tube current modulation (ATCM) is widely implemented in CT imaging, adjusting the tube current according to the attenuation level in the xy-plane and along the z-axis. Consequently, $CTDI_{vol}$ and SSDEs vary across slices throughout the scanned region.[13] CT manufacturers have begun displaying estimated SSDE values alongside other dosimetry quantities, but these features often incur additional costs. Furthermore, different CT manufacturers incorporate proprietary algorithms within their software to estimate D_w and SSDEs. Variations in computational methods across manufacturers may lead to discrepancies in reported D_w values, potentially affecting SSDE calculations and radiation dose assessments. In some manufacturers' software, the SSDE (geometrical) is displayed as an estimate of the radiation dose a patient receives from a CT scan. This estimate is calculated using the patient's dimensions, such as the effective diameter to derive a conversion factor from $CTDI_{vol}$, rather than using the D_w as the primary metric. Moreover, the SSDE value displayed by CT scanners is often a single value, representing either an average across the scanned region or a measurement from the middle slice. This approach limits the ability of users to assess the SSDE for individual slices, which may more accurately reflect localized radiation absorption. As a result, non-commercial automated solutions have gained popularity. Some studies have developed automated programs to calculate SSDEs; however, many of these lack direct comparison with established methods, making it difficult to confirm their validity and clinical reliability.[14-16] In this study, we developed and validated a user-friendly computational tool for estimating slice-specific D_w, $CTDI_{vol}$, and SSDEs, based on methodologies outlined in AAPM reports. A custom ImageJ macro was created to perform automated, threshold-based body contour segmentation and extract slice-specific exposure values directly from Digital Imaging and Communications in Medicine (DICOM) metadata. The accuracy and clinical relevance of the method were evaluated through comparison with both commercial CT scanner outputs and dose monitoring software using phantom and patient datasets. This approach provides an accessible and transparent solution for patient-specific dose estimation, enabling detailed slice-by-slice assessment and offering greater spatial resolution than conventional scanner displays or commercial software systems.

Main points

- A custom ImageJ macro was developed to automatically calculate the slice-specific volume computed tomography (CT) dose index ($CTDI_{vol}$), water-equivalent diameter, and size-specific dose estimate (SSDE) from CT images.
- The macro exhibited excellent agreement with both scanner console values and commercial software (Radimetrics), especially for $CTDI_{vol}$ and SSDE.
- Slice-by-slice dose evaluation revealed dose variations that are not visible in conventional average-based reporting.
- This method allows detailed, patient-specific dose assessment and supports protocol optimization in clinical and research settings.
- The tool is low cost, open source, and suitable for institutions without access to commercial dose monitoring systems.

Methods

CT imaging was performed using both homogeneous circular phantoms and anthropomorphic phantoms representing the head, thorax, and abdomen. All phantom scans were conducted using the Siemens SOMATOM go.Top® scanner (Siemens Healthineers, Erlangen, Germany).

In addition, clinical CT datasets were retrospectively collected from at least 30 patients who underwent routine head, chest, and abdomen examinations in accordance with the hospital's standard imaging protocols. Two CT scanners were involved in data collection. For Scanner A (Siemens SOMATOM go.Top®), the $CTDI_{vol}$, D_w, and SSDE values were recorded directly from the scanner console. For Scanner B (Siemens SOMATOM Force®), the corresponding dose parameters were extracted using Radimetrics, a commercial dose monitoring software integrated with the scanner.

Development of ImageJ macro for slice-specific dose calculations

A custom macro was developed using ImageJ (version 1.54g, National Institutes of Health, Bethesda, MD, USA)[17] to estimate the SSDE from DICOM-format CT image stacks, in accordance with the guidelines provided by AAPM reports 220 and 293.

Data acquisition and preprocessing

CT image datasets were acquired from a Siemens SOMATOM scanner (syngo CT VA40A software) and SOMATOM force, and DICOM files were imported into ImageJ as image stacks, each representing a series of con-

tiguous axial slices from either head or body examinations. Prior to analysis, metadata, including the number of slices, slice location, and exposure (DICOM tag 0018,1152—"Exposure"), were extracted using the ImageJ DICOM header parser.

Body contour segmentation and area calculation

To estimate patient size (D_w), each slice underwent automatic body contour segmentation using a threshold-based method. A threshold of –140 Hounsfield unit (HU) was applied to segment the patient's body contour, effectively distinguishing body tissues from the surrounding air to enable consistent and reproducible D_w measurements. This threshold has been successfully applied in previous SSDE-related studies in CT examinations[18] and was selected to balance the exclusion of air while avoiding inclusion of the scanner couch. Although lower thresholds (–300 to –500 HU) have been employed in other pipelines, such as the CT contour (–383 HU) for abdomen–pelvic CT,[19] these have been reported to occasionally include couch structures, potentially reducing contour accuracy. After segmentation, the "Analyze Particles" function in ImageJ was used to detect regions of interest (ROIs), with size and circularity constraints (size: 5000–∞ pixels, circularity: 0.2–1.0) to exclude non-patient structures such as the scanner couch. The cross-sectional area (A) of each ROI was then calculated based on pixel spacing from the DICOM metadata.

Slice-specific water-equivalent diameter calculation

For each slice, the D_w was calculated from the segmented area and average CT number within the ROI, as follows:

$$D_w = 2\sqrt{\left[\frac{1}{1000}\overline{CT(x,y)_{ROI}} + 1\right]\frac{A_{ROI}}{\pi}} \quad \text{Eq.1}$$

where D_w is the water-equivalent diameter (cm), $\overline{CT(x,y)_{ROI}}$ is the average CT number within the area of interest, and is the area of the ROI (cm^2).

Size-specific dose estimate calculation

The slice-specific SSDE was calculated as follows:

$$SSDE = CTDI_{vol} \times f \quad \text{Eq.2}$$

The correction factor (*f*) was determined from the calculated D_w value, accounting for variations in patient size and scanner phantom type. Separate exponential functions were applied depending on whether $CTDI_{vol}$ was derived from a 16- or 32-cm calibration phantom, with the source equations taken from AAPM Report 293 for head examinations[9] and AAPM Report 220 for body examinations[8]:

$$f^{H16} = 1.9852e^{(-0.0486D_w)} \quad \text{Eq.3}$$
$$f^{B32} = 3.7055e^{(-0.0367D_w)} \quad \text{Eq.4}$$

where *H16* and *B32* nomenclature are used in the superscript of the conversion factor *f* when 16- or 32-cm CTDI phantoms were used for the head (*H*) or body (*B*) $CTDI_{vol}$ measurements and D_w is the water-equivalent diameter.

$$CTDI_{vol,slice} = \left(\frac{Exposure_{slice}}{Exposure_{avg}}\right) \times CTDI_{vol,avg} \quad \text{Eq.5}$$

The slice-specific $CTDI_{vol}$ ($CTDI_{vol,slice}$) was estimated by normalizing the scanner-displayed $CTDI_{vol}$ to the exposure value of each slice) ($CTDI_{vol,avg}$) The exposure value (tag 0018,1152) represents the tube current–time product in mAs, where $Exposure_{slice}$ refers to the tube current–time product for that particular slice and $Exposure_{avg}$ refers to the average tube current–time product across the entire scan.

Finally, SSDE per slice ($SSDE_{slice}$) was calculated as follows:

$$SSDE_{slice} = CTDI_{vol,slice} \times f_{corr} \quad \text{Eq.6}$$

where f_{corr} is the correction factor corresponding to the slice-specific D_w and scan region (head or body).

The macro generated an output table containing the following parameters for each slice: slice location, exposure, segmented area, mean pixel value, D_w, correction factor, $CTDI_{vol}$ per slice, and SSDE per slice. All calculations were performed in real time within the ImageJ environment and exported to CSV format for further statistical analysis.

Data comparison and statistical analysis

Statistical analyses were performed using Stata version 17 (StataCorp, College Station, TX, USA). Agreement between $CTDI_{vol}$, D_w, and SSDE values obtained from the custom ImageJ macro and those from the scanner console and Radimetrics software was assessed using intraclass correlation coefficients (ICCs) and Bland–Altman plots. Specifically, ICC (3,1), a two-way mixed-effects model for absolute agreement with single measurements, was employed. Interpretation of ICC values followed established guidelines: <0.5 = poor agreement, 0.5–0.75 = moderate agreement, 0.75–0.9 = good agreement, and >0.9 = excellent agreement.

Research ethics standards compliance

This retrospective study was approved by the Institutional Review Board (IRB) under expedited review (COA No. 081/2025) on March 18, 2025. All CT images were originally acquired for clinical purposes and were fully anonymized before being retrospectively analyzed, with no identifiable patient information included. As the CT images were already taken for routine medical care and later de-identified, the IRB waived the requirement for obtaining informed consent in accordance with ethical guidelines for retrospective studies.

Results

Development of the dose calculation macro

The study population is presented in Figure 1a. The workflow of the macro is illustrated in Figure 1b, showing sequential steps from image input, DICOM metadata extraction, ROI detection, dose calculations, and final data export.

An ImageJ macro was successfully developed to calculate slice-specific dose metrics from CT images. The tool automatically segments the patient contour on each slice using a –140 HU threshold, calculates D_w, and determines slice-specific $CTDI_{vol}$ and $CTDI_{vol,slice}$ based on DICOM exposure values (tag 0018,1152). The SSDE is then computed for each slice using the appropriate correction factors from AAPM reports. The macro completes analysis in approximately 30 seconds per examination, providing comprehensive output including the slice location, exposure, area, D_w, correction factor, $CTDI_{vol}$, and SSDE for each slice.

Validation of the macro with phantom studies

The accuracy of the macro-calculated D_w and SSDE values was validated against scanner-displayed measurements using both uniform water phantoms and anthropomorphic phantoms. The results for water phantoms of different sizes (16 and 19 cm inner diameter) are presented in Table 1, showing excellent agreement with differences of less than 2.3% for both D_w and SSDE measurements.

For anthropomorphic phantoms of the head, chest, and abdomen (Table 2), the macro demonstrated comparable accuracy. In the head phantom, both the middle-slice and average values closely matched the scanner output. The chest phantom showed

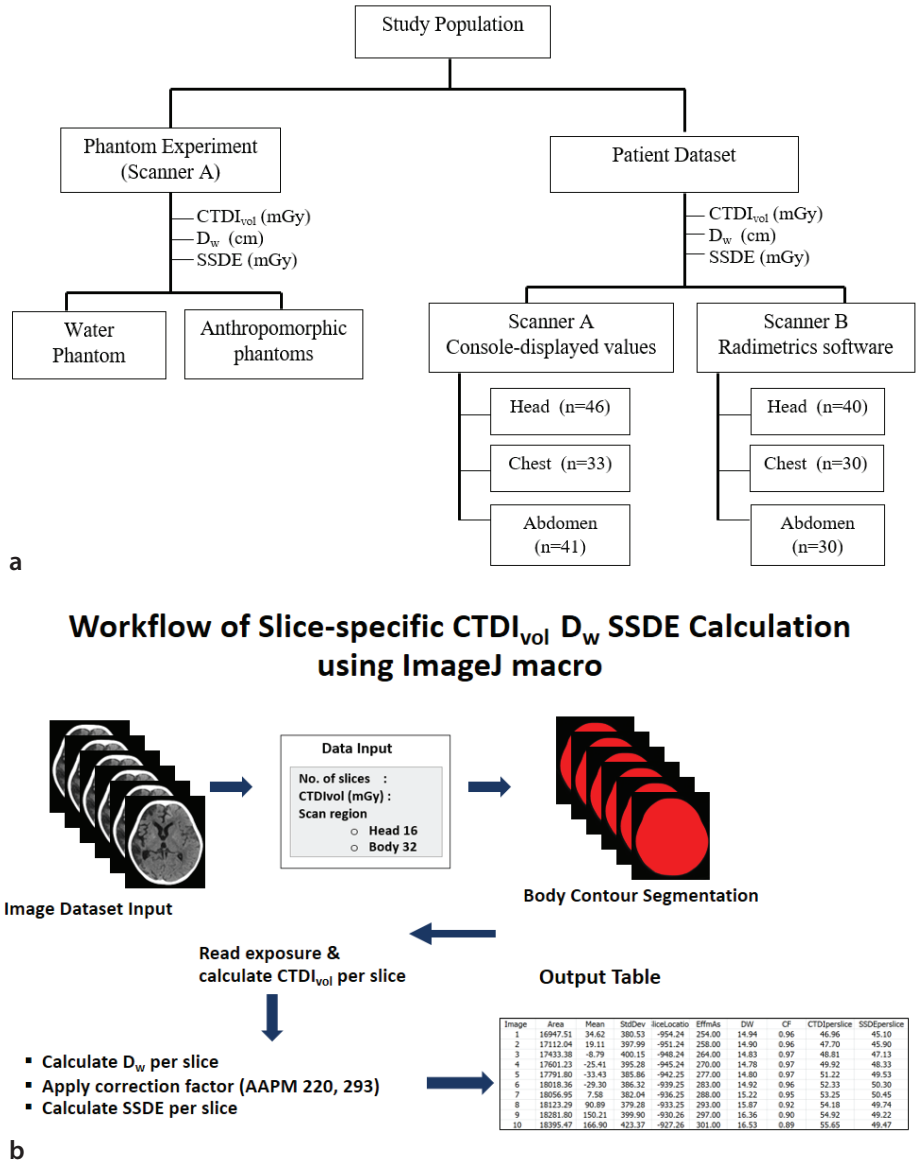


Figure 1. **(a)** The study population, and **(b)** the workflow for automated slice-specific calculation of the volume computed tomography dose index, water-equivalent diameter, and size-specific dose estimate using the ImageJ macro. $CTDI_{vol}$, volume computed tomography dose index; D_w, water-equivalent diameter; SSDE, size-specific dose estimate.

the most pronounced slice-to-slice variation in SSDE (range: 5.77–23.68 mGy) despite the average value (12.56 ± 6.99 mGy) being nearly identical to the scanner-reported value (12.5 mGy). This key variation highlights the importance of slice-specific dose assessment in anatomically heterogeneous regions. The abdomen phantom showed more consistent values across slices, with average measurements within 5% of scanner-displayed values.

These validation studies confirm that the macro provides accurate D_w and SSDE calculations that align well with scanner-reported values while offering the additional advantage of slice-specific analysis that reveals dose distribution patterns not captured by console-displayed averages.

Analysis of clinical patient scans

Scanner A

Agreement between the automated slice-specific dose calculations and the scanner-displayed values from Scanner A (Siemens SOMATOM go.Top®) was evaluated using Bland–Altman analysis and ICCs. The comparison included three dose parameters—$CTDI_{vol}$, D_w, and SSDE—assessed separately for head and chest/abdomen scans.

The Bland–Altman analysis shown in Table 3 and plots shown in Figure 2 demonstrate excellent overall agreement between the ImageJ macro and the scanner-reported values. For head scans, the mean differences between the macro-derived and scanner-reported values were 2.285 mGy for $CTDI_{vol}$, 0.649 cm for D_w, and −1.670 mGy for the SSDE, with 95% limits of agreement indicating acceptable variability. Similarly, for chest and abdomen scans, the mean differences were −0.001 mGy, −0.336 cm, and 0.284 mGy, respectively, within clinically acceptable ranges.

These findings were further supported by ICC analysis. The ICC (3,1) values for $CTDI_{vol}$, D_w, and the SSDE for head scans were 0.974, 0.893, and 0.965, respectively, indicating good to excellent agreement. For chest and abdomen scans, ICC (3,1) values were 1.000 for $CTDI_{vol}$, 0.994 for D_w, and 0.989 for the SSDE, also reflecting excellent agreement across methods.

Scanner B

The Bland–Altman analysis (Table 3 and Figure 3) for head CT showed a mean difference in $CTDI_{vol}$ of 0.067 mGy, with 95% limits of agreement from −0.653 to 0.787 mGy and an ICC (3,1) of 0.996. For the SSDE, the mean difference was −0.123 mGy (−1.911 to 1.665 mGy), with an ICC of 0.967, confirming near-perfect agreement between the ImageJ macro and Radimetrics for head scans on Scanner B. By contrast, D_w demonstrated poor agreement, with a mean difference of 1.250 cm (0.609 to 1.890 cm) and an ICC of 0.267. This discrepancy highlights a potential risk of misestimating patient size and dose if console-reported D_w is used alone; however, the SSDE values maintained near-perfect agreement (ICC: 0.967), underscoring its robustness as a clinically relevant dose metric.

For chest/abdomen CT, the mean difference in $CTDI_{vol}$ was −0.014 mGy, with limits of agreement ranging from −0.1 to 0.072 mGy and an ICC of 1.000. The D_w comparison showed a mean difference of −0.312 cm (−2.577 to 1.954 cm), with an ICC of 0.94. The SSDE values differed by a mean of −0.119 mGy (−4.670 to 4.432 mGy), and the ICC was 0.884.

A comparative analysis of D_w and SSDE values between the CT console output and the ImageJ macro for the two scanners across the head, chest, and abdomen regions is shown in Figure 4. For Scanner A, the macro slightly overestimated D_w in the head region [mean ± standard deviation (SD): 15.8 ± 1.4 vs. 15.2 ± 1.5 cm]; this was similar for Scanner B, with higher macro values (17.2 ± 0.6 vs. 16.0 ± 0.6 cm). In the chest and abdomen, the differences were minor, with both scanners showing close agreement between methods.

Table 1. Comparison of D_w and SSDE values between CT scanner output and automated calculation in water phantoms

Phantom	$CTDI_{vol}$ (mGy)	CT scanner		Automated calculations	
		D_w (cm)	SSDE (mGy)	D_w (cm)	SSDE (mGy)
Small water phantom*	37.5	16.5	33.4	16.75	33.00
Large water phantom**	37.5	19.5	28.9	19.95	28.24

Scan protocol, 120 kV; effective mAs, 410; field of view, 300 mm; pitch, 1.0; slice thickness, 3 mm; convolution kernel, Hr40; * D_{Eff} =16 cm and D_{Eff} includes outer shell = 17 cm; ** D_{Eff} =19 cm and D_{Eff} includes outer shell = 20 cm. CT, computed tomography; $CTDI_{vol}$, volume computed tomography dose index; D_w, water-equivalent diameter; SSDE, size-specific dose estimate.

Table 2. Comparison of D_w and SSDE from CT scanner display and automated calculations using the developed macro in head, chest, and abdomen phantoms

Phantom	D_w (cm)			$CTDI_{vol}$ (mGy)			SSDE (mGy)		
	CT scanner display	Automated calculations (MS)	Automated calculations (ASA)	CT scanner display	Automated calculations (MS)	Automated calculations (ASA)	CT scanner display	Automated calculations (MS)	Automated calculations (ASA)
Head	16.2	18.47	16.46 ± 2.75 (6.76 – 18.91)	50.6	60.16	55.41 ± 5.54 (37.48 – 60.53)	49.2	48.73	49.19 ± 1.44 (46.34 – 53.60)
Chest	22.3	22.38	22.29 ± 2.60 (13.65 – 25.22)	7.4	4.73	7.85 ± 4.69 (2.76 – 15.33)	12.5	7.70	12.56 ± 6.99 (5.77 – 23.68)
Abdomen	26.4	28.01	27.20 ± 1.55 (24.94 – 29.54)	7.09	7.22	7.15 ± 1.14 (5.53 – 8.74)	10.1	9.60	9.70 ± 1.13 (8.07 – 11.66)

Head protocol, 120 kV; effective mAs, 410; field of view (FOV), 250 mm; pitch, 0.55; slice thickness, 3 mm; convolution kernel, Hr40; SAFIRE strength level 3.
Chest protocol, 120 kV; effective mAs, 125; FOV, 400 mm; pitch, 0.8; slice thickness, 3 mm; convolution kernel, Hr40; SAFIRE strength level 3.
Abdomen protocol, 120 kV; effective mAs, 180; FOV, 400 mm; pitch, 0.8; slice thickness, 3 mm; convolution kernel, Hr40; SAFIRE strength level 3.
CT, computed tomography; $CTDI_{vol}$: volume computed tomography dose index, D_w, water-equivalent diameter; SSDE: size-specific dose estimate; MS, middle slice of scan length; ASA, all slices average.

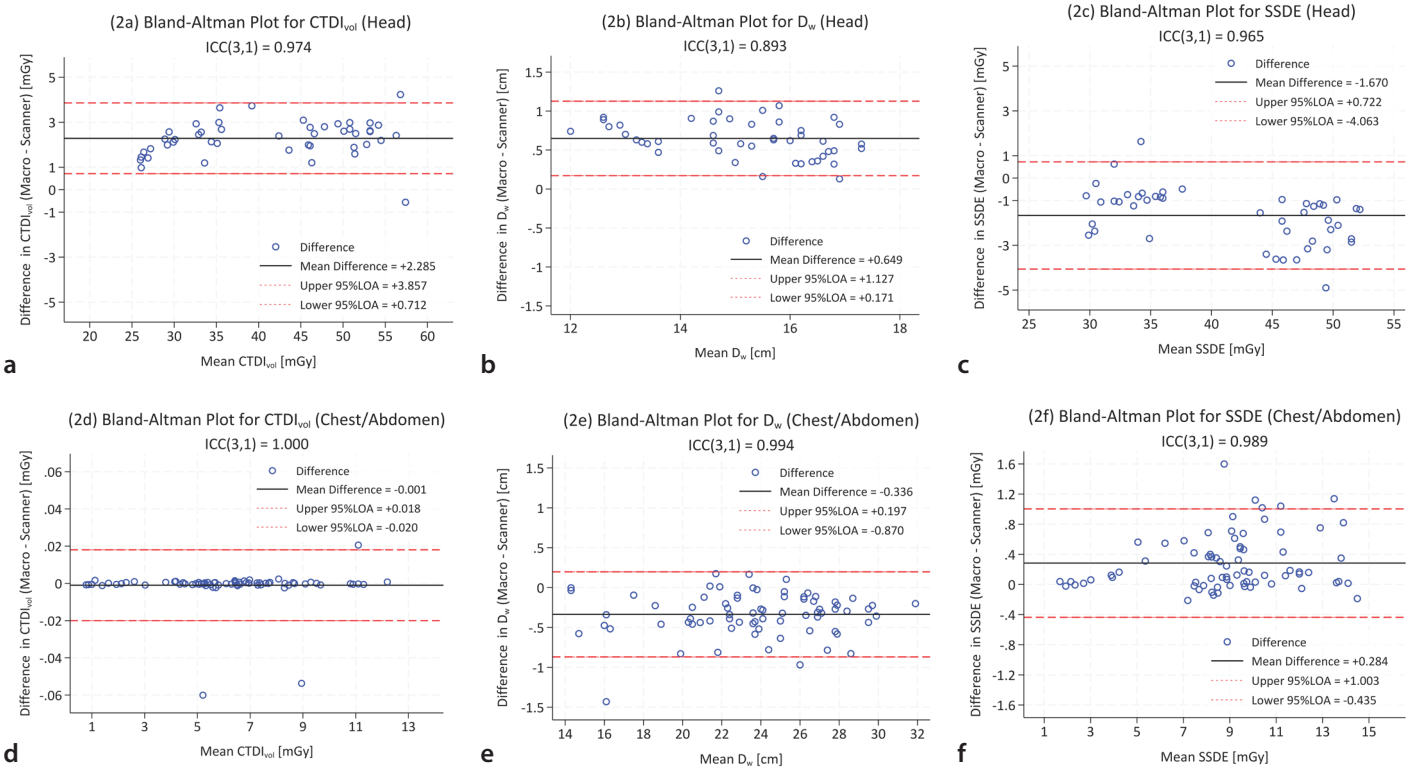


Figure 2. Bland–Altman plots comparing the dose parameters obtained from the ImageJ macro and the scanner console display (Scanner A) for **(a)** volume computed tomography dose index ($CTDI_{vol}$), **(b)** water-equivalent diameter (D_w), and **(c)** size-specific dose estimate (SSDE) in head computed tomography (CT) examinations and **(d)** $CTDI_{vol}$, **(e)** D_w, and **(f)** SSDE in chest/abdomen CT examinations. LOA, limits of agreement; ICC, intraclass correlation coefficient.

Table 3. Bland–Altman analysis and intraclass correlation coefficients (ICCs) comparing $CTDI_{vol}$, D_w, and SSDEs between the ImageJ macro and scanner console (Scanner A) and Radimetrics (Scanner B)

Scanner	CT exam		Bland & Altman			ICC (3,1)			
			Mean difference	Lower 95% LOA	Upper 95% LOA	Individual ICC	Lower 95% CI	Upper 95% CI	*P* value
A	Head (n = 46)	$CTDI_{vol}$ (mGy)	+2.285	+0.712	+3.857	0.974	0.084	0.995	<0.001
		D_w (cm)	+0.649	+0.171	+1.127	0.893	−0.022	0.976	<0.001
		SSDE (mGy)	−1.670	−4.063	+0.722	0.965	0.481	0.991	<0.001
	Chest/abdomen (n = 74)	$CTDI_{vol}$ (mGy)	−0.001	−0.020	+0.018	1.000	1.000	1.000	<0.001
		D_w (cm)	−0.336	−0.870	+0.197	0.994	0.906	0.998	<0.001
		SSDE (mGy)	+0.284	−0.435	+1.003	0.989	0.952	0.995	<0.001
B	Head (n = 40)	$CTDI_{vol}$ (mGy)	+0.067	−0.653	+0.787	0.996	0.992	0.998	<0.001
		D_w (cm)	+1.250	+0.609	+1.890	0.267	−0.030	0.658	<0.001
		SSDE (mGy)	−0.123	−1.911	+1.665	0.967	0.933	0.984	<0.001
	Chest/abdomen (n = 60)	$CTDI_{vol}$ (mGy)	−0.014	−0.100	+0.072	1.000	1.000	1.000	<0.001
		D_w (cm)	−0.312	−2.577	+1.954	0.940	0.899	0.964	<0.001
		SSDE (mGy)	−0.119	−4.670	+4.432	0.884	0.812	0.929	<0.001

CT, computed tomography; $CTDI_{vol}$, volume computed tomography dose index; D_w, water-equivalent diameter; SSDE, size-specific dose estimate; LOA, limits of agreement; CI, Confidence interval.

Macro-derived SSDE values closely matched those from the console. In head CT scans, Scanner A showed a mean SSDE value of 39.8 ± 7.5 (macro) and 41.4 ± 8.0 mGy (console), whereas Scanner B yielded 38.3 ± 3.6 (macro) and 38.5 ± 3.5 mGy (console). For the chest and abdomen, differences were minor for Scanner A; however, for Scanner B, slightly greater differences were observed: mean chest SSDEs of 13.5 ± 4.0 (macro) and 12.0 ± 3.3 mGy (console), and mean abdomen SSDEs of 16.7 ± 4.3 (macro) and 18.3 ± 4.6 mGy (console).

Larger variations were observed in the chest and abdomen for Scanner B than for Scanner A, with Bland–Altman analysis showing 95% limits of agreement from –4.670 to +4.432 mGy (Figure 3f) compared with –0.435 to +1.003 mGy for Scanner A (Figure 2f). Nevertheless, mean SSDE differences between the macro and console for Scanner B were minimal, with relative deviations of 0.52% for head CT and 0.66% for chest/abdomen CT, both well below the ± 10% threshold generally considered acceptable for clinical dose estimation.

Discussion

It is important to emphasize that $CTDI_{vol}$, DLP, and the SSDE are not intended to represent actual absorbed doses in individual patients; rather, these dose indicators are designed for quality assurance (QA), dose tracking, and protocol optimization, as recommended by the AAPM.

However, conventional metrics such as $CTDI_{vol}$ and DLP reflect only averaged values across the entire scan range, potentially masking substantial intra-scan dose variations. To address this limitation, we developed and validated an ImageJ macro capable of slice-specific dose estimation. By generating spatially resolved $CTDI_{vol}$, D_w, and SSDE values, the macro enhances the granularity of dose reporting and enables identification of within-scan dose fluctuations that may be clinically relevant.

Although DLP provides a cumulative measure of radiation output over the scanned region, it lacks spatial resolution. By contrast, the slice-specific $CTDI_{vol}$ values derived using our tool offer a detailed map of dose distribution along the z-axis. This capability allows for the identification of peak dose regions—commonly occurring at anatomical transitions such as the shoulders or skull base—and provides valuable data for tailoring protocols or conducting organ-level dosimetric studies.

Comparison between the developed macro and siemens console display

The ImageJ macro developed in this study provides patient-specific, slice-by-slice analysis of CT dose parameters, offering higher granularity than the Siemens console display. Although the console reports effective mAs, $CTDI_{vol}$, and SSDEs as modulation-weighted averages or values from a representative slice, these measurements do not capture anatomical variations throughout the scan range. By contrast, the macro calculates $CTDI_{vol}$ and SSDE per slice using exposure data extracted from DICOM metadata (tag 0018,1152) and cross-sectional area measurements (D_w) obtained through threshold-based segmentation.

This difference was particularly evident in head CT examinations, where substantial anatomical variation exists between the vertex and skull base. The slice-by-slice approach provided more detailed SSDEs, revealing dose variations that are not apparent in the averaged values shown on the scanner console. Notably, discrepancies of 15%–35% were observed at anatomical transition zones, such as the thoracoabdominal junction, where rapid changes in cross-sectional area occur.

Another key difference involves the handling of tube current modulation. The Siemens console provides only averaged effective mAs for the entire scan, whereas the macro captures actual slice-specific exposure values, reflecting real-time adjustments in tube current. This capability allows for more precise dose estimation, particularly for modern protocols that employ aggressive modulation techniques.

The macro also offers temporal resolution advantages by generating comprehensive dose profiles across the scan range, enabling identification of dose peaks and facilitating protocol optimization. Although console-displayed metrics remain useful for rapid dose feedback and protocol adjustments in clinical workflows, they typically provide only a single averaged value. By contrast, the macro yields high-resolution, slice-specific dose estimates that reveal otherwise obscured intra-scan dose variations not cap-

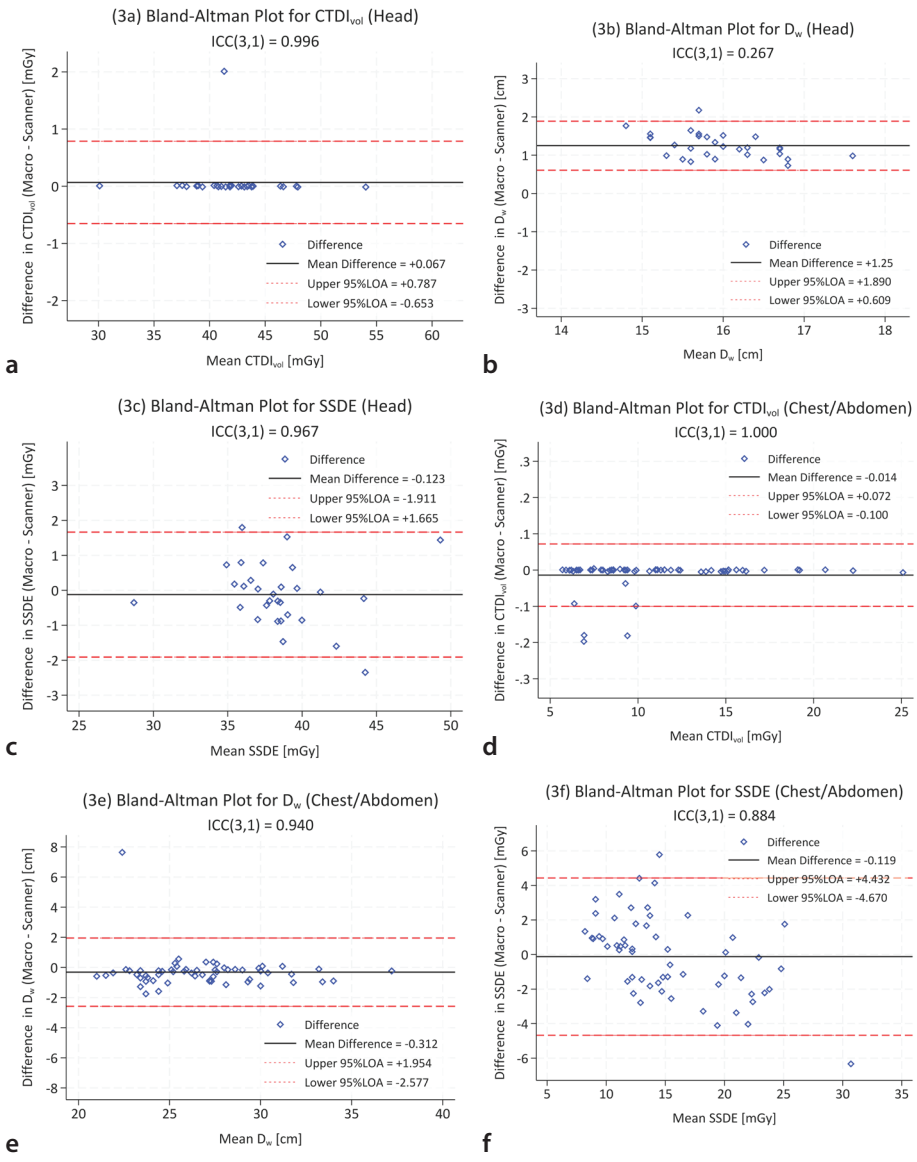


Figure 3. Bland–Altman plots comparing the dose parameters obtained from the ImageJ macro and Radimetrics (Scanner B) for **(a)** volume computed tomography dose index ($CTDI_{vol}$), **(b)** water-equivalent diameter (D_w), and **(c)** size-specific dose estimate (SSDE) in head computed tomography (CT) examinations and **(d)** $CTDI_{vol}$, **(e)** D_w, and **(f)** SSDE in chest/abdomen CT examinations. LOA, limits of agreement; ICC, intraclass correlation coefficient.

tured by conventional summary metrics, a feature particularly valuable in anatomically heterogeneous regions such as the thorax and pelvis.

Comparison between the developed imagej macro and commercial dose monitoring software (Radimetrics)

Considerable methodological and functional differences were observed between the developed ImageJ macro and the commercial dose monitoring software Radimetrics™ (Bayer Healthcare). Although both systems extract dose-relevant information from DICOM headers and estimate patient-specific dose metrics, the processes, assumptions, and applications differ substantially.

The observed differences in D_w agreement between anatomical regions can be attributed to distinct threshold segmentation approaches. Our ImageJ macro uses a fixed threshold of −140 HU; by contrast, although the exact threshold values used by commercial dose management systems such as Radimetrics are not publicly disclosed, several studies and indirect validations suggest that these systems commonly apply a body contour segmentation threshold in the range of approximately −300 to −500 HU. This threshold range is designed to avoid including excessive air (below −500 HU) while preventing the exclusion of soft tissue (above −300 HU).

The anatomical differences substantially influence the impact of threshold selection on D_w calculation. In head CT, the relatively uniform soft tissue density and distinct bone–air interfaces make D_w particularly sensitive to segmentation. For example, a threshold of −140 HU (as used in our macro) may inadvertently exclude bone edges or include small air pockets, leading to substantial changes in mean CT numbers within the ROI and consequently affecting D_w calculations, in accordance with AAPM 220 methodology. Similarly, lower thresholds within the range applied by commercial systems (−300 to −500 HU, e.g., −400 HU) have also been reported to occasionally include couch structures. These threshold-related factors explain the poor agreement (ICC = 0.267) observed for head D_w measurements.

Conversely, in chest and abdomen CT, the more heterogeneous tissue composition and larger soft tissue volumes make the calculations less sensitive to threshold variations. The predominance of soft tissue in these regions provides more stable mean CT numbers, resulting in improved D_w agreement (ICC = 0.940) despite different threshold approaches.

Quantitative comparison between the two systems using Bland–Altman plots and ICC analysis reflects these methodological differences. For head CT examinations, although $CTDI_{vol}$ and SSDEs showed near-perfect agreement with ICC values of 0.996 and 0.967, respectively, D_w demonstrated poor agreement (IC: 0.267), with a mean difference of 1.25 cm. This substantial discrepancy likely stems from the different threshold segmentation approaches affecting the CT number-based calculations more substantially in the relatively homogeneous head anatomy. Nevertheless, SSDEs remained robust despite this variability in D_w, as the exponential correction factor $f(D_w)$ changes only gradually in the head diameter range, limiting the propagation of D_w errors into SSDEs. Together with the near-perfect agreement in $CTDI_{vol}$, this explains the consistently high SSDE agreement observed for head CT examinations.

For chest/abdomen scans, the agreement was more consistent across all parameters: $CTDI_{vol}$ showed excellent agreement, with an ICC of 1.00 and mean difference of −0.014 mGy; D_w comparison yielded good agreement (ICC: 0.940), with a mean difference of −0.312 cm; and SSDEs showed excellent agreement (ICC: 0.884), with a mean difference of −0.119 mGy. This suggests that the heterogeneous tissue composition in these regions provides more robust results despite methodological differences.

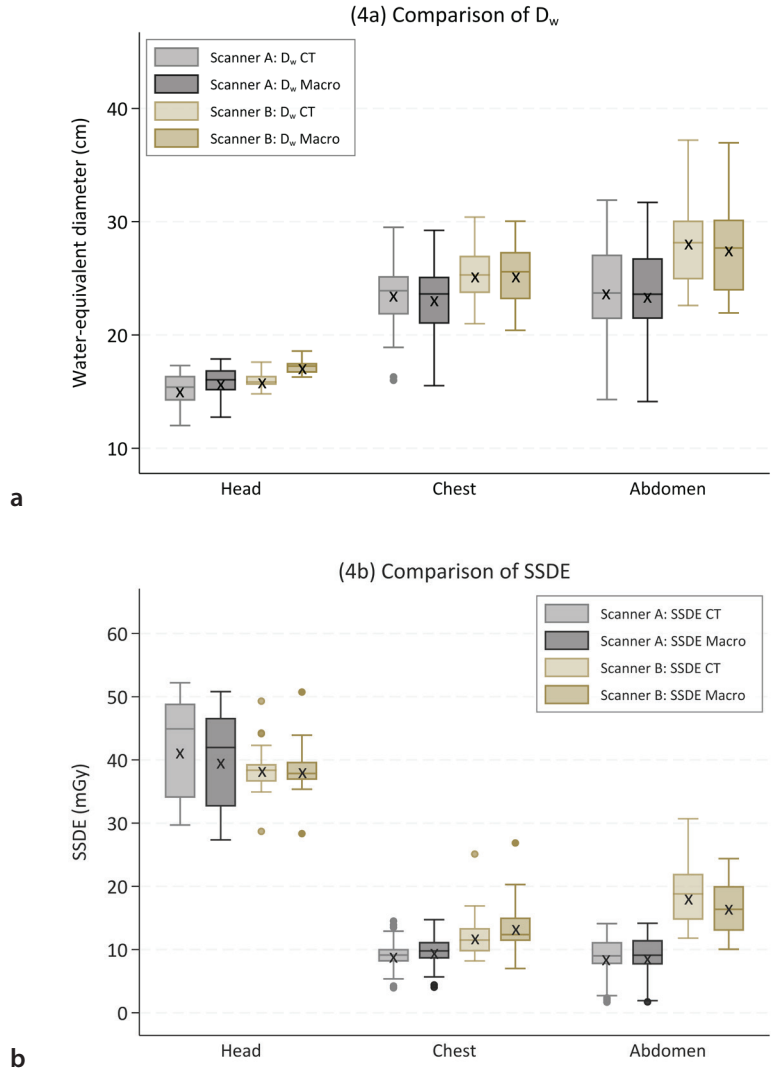


Figure 4. Box plot comparison of **(a)** the water-equivalent diameter and **(b)** size-specific dose estimate obtained from the ImageJ macro, computed tomography scanner console, and Radimetrics across anatomical regions and scanners. Boxes represent the interquartile range, center lines indicate the median, whiskers extend to the data range, and crosses (×) represent the mean values. CT, computed tomography; D_w, water-equivalent diameter; SSDE, size-specific dose estimate.

The macro's threshold-based segmentation was validated using phantoms and yielded results within 5% of physical dimensions while requiring fewer computational resources. However, challenges such as low-contrast boundaries (e.g., at air–bone interfaces) can reduce accuracy in specific regions such as the head. These findings highlight the importance of understanding the underlying algorithmic differences when comparing dose monitoring systems. The threshold-dependent nature of D_w calculations, particularly in anatomically uniform regions such as the head, suggests that standardization of segmentation approaches may be necessary for consistent inter-system comparisons.

From a practical standpoint, Radimetrics excels in enterprise-wide integration. It automatically processes all CT examinations without user interaction and provides compliance dashboards and centralized QA tools. By contrast, the ImageJ macro requires manual execution, input selection, and parameter adjustment. Although this limits automation, it offers unmatched flexibility for protocol-specific analysis, research customization, and pediatric dose assessment. In terms of cost, the macro presents a clear advantage. Commercial systems such as Radimetrics require considerable investment in licensing, infrastructure, and ongoing support. The macro, built entirely on open-source tools, is a cost-effective alternative for institutions with research goals, limited budgets, or custom analysis needs.

Practical considerations for clinical implementation

Significant differences in D_w and SSDEs were observed in chest and abdomen scans, likely due to variations in field-of-view (FOV) settings. In some cases, the FOV may not fully cover the patient's body cross section, particularly at the lateral margins, leading to underestimation of body size in image-based calculations. This effect may be more pronounced in larger or misaligned patients.

Despite being statistically significant, these differences are generally within ± 10% and may have limited clinical impact, as key organs typically remain within the scan range. To reduce such discrepancies, consistent patient positioning and standardized FOV protocols are essential. Regular QA processes should validate dose calculation accuracy across various scanning conditions. Further studies in broader patient groups, including pediatric and atypical body types, are recommended to ensure the robustness of SSDEs.

Clinical implication of slice-specific dose estimates for protocol optimization

The granularity provided by slice-specific SSDEs offers critical insight into dose heterogeneity along the scan length. For instance, anatomical regions such as the shoulders or skull base often exhibit elevated SSDE values due to increased patient attenuation, which can trigger ATCM to deliver higher output. Conversely, areas such as the lungs or extremities may receive lower doses. Identifying these "higher" or "lower" dose sections allows radiology teams to fine-tune scan protocols accordingly.

For high-dose regions, protocol adjustments may involve reducing the maximum tube current limit, modifying ATCM curves, adjusting the noise index, or narrowing z-axis coverage if clinically acceptable. By contrast, persistently low-dose areas—especially those critical to diagnosis—may benefit from increased minimum mA thresholds, longer rotation times, or customized reconstruction parameters to preserve image quality. Such region-specific refinement, when guided by slice-level SSDE data and supported by institutional or national diagnostic reference levels, enables a more personalized approach to CT protocol design. Ultimately, this facilitates a balance between diagnostic performance and radiation safety.

This study has several limitations. The use of a fixed threshold (–140 HU) may not be optimal for all anatomical regions, particularly in low-contrast boundaries such as air–bone interfaces in head CT, where D_w accuracy may be affected. Future work could explore adaptive or data-driven thresholding approaches to improve segmentation performance across diverse anatomies, thereby enhancing the robustness and clinical applicability of the tool. Validation was limited to Siemens scanners and Radimetrics software; testing with other vendors (e.g., GE, Philips, Canon) is needed to assess broader applicability. The clinical dataset included only 30 patients, restricting variability in body size, pathology, and scan protocols. Inter-observer variability in manual parameter settings was not evaluated, which may affect reproducibility. Additionally, the macro currently lacks automated quality checks and standardized reporting, limiting its readiness for routine clinical integration. Moreover, we did not compare scanner-reported DLP values with the cumulative $CTDI_{vol}$ from the macro, as the precise irradiation length used by the scanner was not available in the DICOM metadata. Since DLP calculation depends on irradiation length, which may differ from the reconstructed scan length due to overscanning or collimation, direct comparisons would be inaccurate or misleading. Therefore, instead of focusing on DLP comparison, our study emphasizes the added value of slice-specific $CTDI_{vol}$ analysis, which reveals spatial variations in dose that global metrics such as DLP or average $CTDI_{vol}$ cannot capture. Future development should address automation, segmentation robustness, cross-vendor compatibility, and streamlined clinical implementation.

In conclusion, the developed ImageJ macro enables automated, slice-specific calculation of $CTDI_{vol}$, D_w, and SSDEs with high agreement with scanner and commercial software outputs. Its ability to reveal intra-scan dose variations offers improved slice-by-slice information for dose assessment and supports protocol optimization in clinical and research settings.

Acknowledgments

The authors would like to express their sincere gratitude to Kamphaeng Phet Hospital and Bumrungrad International Hospital for their valuable contribution in providing clinical data essential for this research. Their cooperation and support were instrumental in the successful completion of this study. The authors also gratefully acknowledge Buddhachinaraj Hospital, Phitsanulok, for providing access to Stata version 17 software used in the statistical analysis.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Bos D, Guberina N, Zensen S, Opitz M, Forsting M, Wetter A. Radiation exposure in computed tomography. *Dtsch Arztebl Int*. 2023;120(9):135-141. [Crossref]
2. Burtt JJ, Thompson PA, Lafrenie RM. Non-targeted effects and radiation-induced carcinogenesis: a review. *J Radiol Prot*. 2016;36(1):R23-35. [Crossref]
3. Hedgire S, Ghoshhajra B, Kalra M. Dose optimization in cardiac CT. *Phys Med*. 2017;41:97-103. [Crossref]
4. Huda W, Mettler FA. Volume CT dose index and dose-length product displayed during CT: what good are they? *Radiology*. 2011;258(1):236-242. [Crossref]
5. Durand DJ, Mahesh M. Understanding CT dose display. *J Am Coll Radiol*. 2012;9(9):669-671. [Crossref]
6. McCollough CH, Leng S, Yu L, Cody DD, Boone JM, McNitt-Gray MF. CT dose index and patient dose: they are not the same thing. *Radiology*. 2011;259(2):311-316. [Crossref]
7. American Association of Physicists in Medicine (AAPM). Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations. In: AAPM Report No. 204. College Park, MD: American Association of Physicists in Medicine; 2011. [Crossref]
8. American Association of Physicists in Medicine (AAPM). Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT. In: AAPM Report No. 220. College Park, MD: American Association of Physicists in Medicine; 2014. [Crossref]
9. American Association of Physicists in Medicine (AAPM). AAPM task group 293: Size-specific dose estimates (SSDE) for body CT examinations based on effective diameter and water-equivalent diameter. In: AAPM Report No. 293. College Park, MD: American Association of Physicists in Medicine; 2023. [Crossref]
10. Abdulkadir MK, Osman ND, Achuthan A, et al. A segmentation-based automated calculation of patient size and size-specific dose estimates in pediatric computed tomography scans. *J Med Phys*. 2024;49(3):456-463. [Crossref]
11. Moore BM, Brady SL, Mirro AE, Kaufman RA. Size-specific dose estimate (SSDE) provides a simple method to calculate organ dose for pediatric CT examinations. *Med Phys*. 2014;41(7):071917. [Crossref]
12. McMillan K, Bostani M, Cagnon C, Zankl M, Sepahdari AR, McNitt-Gray M. Size-specific, scanner-independent organ dose estimates in contiguous axial and helical head CT examinations. *Med Phys*. 2014;41(12):121909. [Crossref]
13. Sookpeng S, Martin CJ, Krisanachinda A. Design and use of a phantom for testing and comparing the performance of computed tomography automatic tube current modulation systems. *J Radiol Prot*. 2020;40(3):753-773. [Crossref]
14. Anam C, Haryanto F, Widita R, Arif I, Dougherty G. Automated calculation of water-equivalent diameter (DW) based on AAPM task group 220. *J Appl Clin Med Phys*. 2016;17(4):320-333. [Crossref]
15. Anam C, Mahdani FR, Dewi WK, et al. An improved method for automated calculation of the water-equivalent diameter for estimating size-specific dose in CT. *J Appl Clin Med Phys*. 2021;22(9):313-323. [Crossref]
16. Payne S, Badawy M. Comparison of average water equivalent diameter values between CTContour and vendor-specific estimates in CT dosimetry. *Phys Med*. 2023;114:103142. [Crossref]
17. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671-675. [Crossref]
18. Sookpeng S, Martin CJ, López-González MR. Simplified approach to estimation of organ absorbed doses for patients undergoing abdomen and pelvis CT examination. *J Radiol Prot*. 2021;41(4). [Crossref]
19. Pace E, Caruana CJ, Bosmans H, Cortis K, D'Anastasi M, Valentino G. CTContour: an open-source Python pipeline for automatic contouring and calculation of mean SSDE along the abdomino-pelvic region for CT images; validation on fifteen systems. *Phys Med*. 2022;103:190-198. [Crossref]