EISSN 1305-3612

# DIR

## Diagnostic and Interventional Radiology

TSR 1924
TURKISH SOCIETY OF RADIOLOGY

dirjournal.org

VOLUME 32
ISSUE 4
July 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:** July 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 4

July 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/).*

Diagnostic and Interventional Radiology

Volume 32 • Issue 4

July 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.2026.261305

EDITORIAL



# The final biological component: AI and radiology's mechanistic drift

Şükrü Mehmet Ertürk
Editor-in-Chief

İstanbul University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

Artificial intelligence (AI) in radiology is often described as an external technological force: algorithms that detect nodules, triage hemorrhage, quantify liver fat, or classify mammographic abnormalities. The prevailing debate therefore centres on replacement versus augmentation. Yet this framing may obscure a quieter and potentially more consequential transformation. The greater risk may not be that AI replaces radiologists from outside the profession, but that radiologists gradually adapt themselves—and their professional environment—toward forms of practice increasingly aligned with the operational logic of machine systems. Importantly, this process did not begin with deep learning or the recent AI boom. It has been underway for decades, driven by productivity metrics, standardisation protocols, and medicolegal pressures. AI then accelerates and deepens an existing drift rather than initiating it.

Radiology has always occupied a complex position between technical analysis and clinical interpretation. Imaging interpretation involves pattern recognition, but it also requires contextual integration: weighing findings against clinical history, communicating uncertainty, integrating temporal change, and understanding the implications of a report for a particular patient. As Osler[1] observed more broadly in medicine, scientific generalisation must ultimately return to the care of singular individuals.

The concern developed here is therefore not directed against AI itself. Many AI systems demonstrably improve efficiency, reduce repetitive workload, and may enhance diagnostic performance in selected domains. Nor is standardisation inherently problematic; structured reporting, evidence-based imaging pathways, and protocol harmonisation have often improved consistency, communication, and patient safety. Rather, the concern is that existing institutional pressures—including productivity metrics, medicolegal anxieties, and workflow optimisation—may increasingly privilege those dimensions of radiological work that are easiest to standardise, quantify, and computationally integrate. This editorial refers to this process as *mechanistic drift*: the gradual reorientation of professional identity toward machine-compatible modes of practice. The phenomenon predates contemporary AI and reflects broader industrial tendencies within modern medicine. AI amplifies these tendencies not because the technology is intrinsically dehumanising, but because computational systems naturally reward tasks that are measurable, reproducible, and scalable.

## Mechanistic drift

Contemporary radiology operates under substantial structural pressure. Relative value unit–based reimbursement systems, productivity metrics, turnaround-time expectations, and protocol standardisation have improved throughput and consistency across many healthcare systems. These developments have undeniable practical value. Yet they may also produce secondary epistemic effects by privileging aspects of practice that are easiest to count and operationalise.

Activities such as rapid image interpretation, standardised reporting, and protocol adherence are readily measurable. Other dimensions of radiological expertise—including reflective deliberation, extensive clinical correlation, communication with referring physicians, or nuanced contextual reasoning—are more difficult to quantify within institutional systems.

Epub: 13.05.2026
Publication date: 01.07.2026
DOI: 10.4274/dir.2026.261305

You may cite this article as: Ertürk ŞM. The final biological component: AI and radiology's mechanistic drift. *Diagn Interv Radiol*. 2026;32(4):351-353.

Over time, this imbalance may subtly influence what becomes recognised as professional excellence. Crucially, this imbalance existed long before AI entered clinical radiology. The introduction of AI merely intensifies an already established tendency: if a task can be measured, it can be optimised; if it can be optimised, it can be automated.

Clinical decision support systems illustrate this tension. Such systems were developed to improve appropriateness and reduce unnecessary imaging rather than to create "cookbook medicine." Nevertheless, implementation within high-volume environments may unintentionally encourage guideline adherence to function as a surrogate marker for quality itself.[2] Similarly, AI systems designed for detection support may gradually shape expectations regarding efficiency, concordance, and acceptable interpretive variability.

The concern is therefore not that radiologists use AI tools, but that radiological cognition itself may increasingly adapt toward machine-compatible priorities. Pattern recognition, protocol recall, rapid triage, and workflow acceleration are valuable capacities. However, when these capacities become dominant markers of competence, other forms of judgment risk becoming institutionally secondary. Human beings, not algorithms, are the ones who progressively redesign their own professional territory to resemble the conditions under which machines excel.

Research on automation bias demonstrates that clinicians may overaccept automated recommendations, particularly under conditions of cognitive load and workflow pressure.[3,4] Existing literature does not establish that AI erodes professional identity directly. However, it does suggest that workflow environments can influence how human judgment is exercised, deferred, or verified. The present concern is therefore sociotechnical rather than deterministic: institutional systems may progressively incentivise narrower forms of practice without any explicit intention to do so, and this tendency has been reinforced by successive waves of computational technology—not only by contemporary AI.

## H-knowledge and I-knowledge

To clarify the epistemic stakes of mechanistic drift, this editorial proposes a heuristic distinction between two poles of professional knowledge. These categories are analytic rather than binary; most radiological practice contains elements of both.

*H-knowledge* (Human/Relational knowledge) refers to forms of understanding grounded in contextual interpretation, embodied experience, and moral answerability. In radiology, H-knowledge includes integrating imaging findings with the patient's clinical narrative, recognising subtle contextual discordances, communicating uncertainty responsibly, and appreciating the downstream consequences of interpretation. It includes tacit dimensions of expertise that are difficult to formalise completely within algorithmic systems. This conception draws partly from phenomenological traditions associated with Merleau-Ponty,[5] in which perception is not treated as passive data acquisition but as embodied and situated interpretation. H-knowledge also resembles *phronesis*, or practical wisdom: the capacity to integrate technical knowledge with situational judgment under conditions of uncertainty.[6]

*I-knowledge* (Information/Digital knowledge) refers to forms of knowledge structured for abstraction, scalability, standardisation, and computational manipulation. In radiology, this includes labelled imaging datasets, probabilistic classification systems, standardised reporting structures, and algorithmically optimised workflows. I-knowledge is extraordinarily powerful precisely because it permits reproducibility, scalability, and pattern detection beyond unaided human capacity.

The distinction should not be interpreted as a moral hierarchy. Human cognition itself depends heavily on abstraction and probabilistic reasoning, while advanced computational systems increasingly incorporate multimodal contextual information. The concern is therefore not abstraction per se, but the possibility that institutional systems may progressively privilege I-oriented competencies while rendering H-oriented competencies less visible, less rewarded, or less cultivable. If that occurs, professional identity may gradually shift toward narrower definitions of expertise centred primarily on efficiency, concordance, and standardisation. Such a shift would not eliminate human radiologists, but it could reduce the practical space within which reflective judgment and contextual interpretation operate.

## Answerability

The concept most vulnerable within mechanistic drift may be *answerability*. By answerability, this essay refers not simply to legal accountability or procedural compliance, but to the lived moral exposure associated with diagnostic judgment.

Radiological reports guide surgery, chemotherapy, surveillance, reassurance, and prognostic expectation. A missed finding or overcalled abnormality may alter the course of a patient's life. Human clinicians experience these consequences psychologically and ethically even when they satisfy formal procedural requirements. This differs fundamentally from computational systems, which may participate causally in clinical outcomes but do not themselves experience moral consequence.

The distinction between accountability and answerability is therefore important. Accountability may be satisfied through compliance with guidelines or institutional workflow. Answerability involves a deeper phenomenological dimension: the recognition that one's judgment has affected another person. A radiologist may follow every procedural expectation correctly yet still experience profound distress after a missed diagnosis. Conversely, a technically compliant system may generate harmful outcomes without any lived experience of responsibility.

The "second victim" phenomenon illustrates this dimension of medical practice.[7] Radiologists who miss consequential findings may experience guilt, anxiety, shame, and long-lasting professional self-doubt. Such reactions are painful but also reveal that diagnostic interpretation remains *morally inhabited* rather than merely procedurally executed.

At the same time, the desire to minimise personal exposure to error may paradoxically reinforce mechanistic practice. Under medicolegal pressure, radiologists may increasingly rely on protocol adherence, guideline conformity, or algorithmic concordance as forms of defensive protection. This tendency does not arise from moral failure but from understandable adaptation to institutional environments that reward standardisation and penalise deviation. Preserving answerability does not require rejecting technological assistance. Rather, it requires maintaining active interpretive engagement. Brief but meaningful practices—reviewing the clinical history carefully, comparing prior examinations, communicating uncertainty directly, or discussing unexpected findings with referring clinicians—help preserve the relational grounding of radiological work even within technologically advanced systems.

### Structural incentives and professional adaptation

Mechanistic drift should not be understood primarily as an individual cognitive failure. It is better interpreted as an adaptive response to broader structural incentives embedded within healthcare systems. Moreover, these incentives have operated for decades, long before the recent proliferation of AI.

Reimbursement models often reward volume more directly than reflective consultation. Medicolegal frameworks may discourage deviation from standardised pathways even when contextual judgment suggests nuance. Commercial AI systems are frequently evaluated according to measurable performance metrics such as sensitivity, specificity, throughput improvement, or workflow acceleration rather than according to their effects on professional reasoning.

Institutionally, standardised error may also be perceived differently from individual deviation. A radiologist who follows an accepted algorithmic recommendation may be viewed as participating in a system-level failure, whereas independent deviation from protocol may attract greater scrutiny even when clinically justified. Over time, such asymmetries may subtly shape professional behaviour.

Yet it would be mistaken to frame this process as inevitable technological determinism. Professional cultures retain agency in deciding what forms of expertise they choose to value, reward, and teach. The future role of radiologists will depend not only on algorithmic capability, but also on how institutions define clinical excellence itself.

### Preserving radiological judgment

Preserving radiological judgment does not require resisting AI wholesale. The challenge is to integrate computational systems in ways that support rather than displace active interpretation.

For practicing radiologists, one practical strategy is to formulate an independent preliminary assessment before reviewing algorithmic output whenever feasible. Explicit documentation of why a guideline or AI recommendation was overridden may also help preserve reflective reasoning rather than passive concordance.

For educators, training environments should ensure that residents continue developing unaided interpretive skills and contextual clinical reasoning. Cases in which AI systems fail, oversimplify, or misclassify findings should be discussed not merely as technical errors, but as opportunities to examine the limits of computational abstraction.

For institutions, workflow architecture matters. Systems could be designed to encourage active preliminary interpretation before AI output becomes visible. Time protected for discrepancy review, multidisciplinary discussion, and reflective quality forums may help preserve the moral and interpretive dimensions of radiological practice.[8] Importantly, such measures should not romanticise individual intuition at the expense of evidence-based medicine, but rather maintain a productive balance between computational support and professional judgment.

The goal is therefore not to preserve radiology as an artisanal resistance movement against technology. It is to ensure that efficiency and standardisation do not become the sole visible markers of expertise.

Radiology's mechanistic evolution has produced extraordinary gains in speed, consistency, accessibility, and diagnostic capability. AI will almost certainly deepen many of these advances. Yet when optimisation becomes the dominant organising principle of professional identity, radiology risks narrowing its own conception of expertise.

The central question is not whether AI will eliminate radiologists entirely. Human clinicians remain indispensable for contextual interpretation, communication, procedural care, multidisciplinary integration, and ethical responsibility. The more important question is whether radiological authority may gradually compress into increasingly standardised and machine-compatible forms of practice. Because this compression is driven not by AI invading human territory, but by humans rendering their own territory ever more AI-like—a process well underway before the current wave of AI—the profession retains the capacity to reverse or moderate it.

Radiology is not defined solely by image classification accuracy. At its best, it remains an interpretive discipline in which technical perception is integrated with contextual understanding and moral responsibility. If these dimensions become secondary to workflow optimisation and algorithmic concordance, the radiologist risks becoming the *final biological component* within the imaging system: a human presence retained less for independent judgment than for procedural ratification, legitimising decisions increasingly shaped elsewhere.

The future of radiology therefore depends not only on what AI can do, but on what forms of judgment the profession continues to cultivate, reward, and defend.


### Acknowledgement

During the preparation of this work, the author used Gemini 3 Flash (Google) to refine language, improve clarity and concision, and reformat references. The author subsequently reviewed, edited, and takes full responsibility for the final content of the article, including all interpretations, arguments, and conclusions.


## References


1. Osler W. Aequanimitas: with other addresses to medical students, nurses and practitioners of medicine. London: H.K. Lewis; 1906. [Crossref]
2. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):1458-1465. [Crossref]
3. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121-127. [Crossref]
4. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc*. 2017;24(2):423-431. [Crossref]
5. Merleau-Ponty M. Phenomenology of perception. Smith C, translator. London: Routledge; 1962. [Crossref]
6. Montgomery K. How doctors think: clinical judgment and the practice of medicine. Oxford: Oxford University Press; 2006. [Crossref]
7. Wu AW. Medical error: the second victim. The doctor who makes the mistake needs help too. *BMJ*. 2000;320(7237):726-727. [Crossref]
8. Lown BA, Manning CF. The Schwartz Center Rounds: evaluation of an interdisciplinary approach to enhancing patient-centered communication, teamwork, and provider support. *Acad Med*. 2010;85(6):1073-1081. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253655

ABDOMINAL IMAGING

ORIGINAL ARTICLE



# Multi-organ non-contrast computed tomography radiomics model to predict hepatic encephalopathy in patients with cirrhosis and hepatorenal failure

Jin-ming Cao[1,2]
Ming-ya Zhang[3]
Xue-mei Ding[2]
Hai-ying Zhou[4]
Xiao-ming Zhang[1,4*]
Tian-wu Chen[5*]

[1]The First Clinical College of Jinan University, Department of Radiology, Guangzhou, China

[2]The Second Clinical Medical School of North Sichuan Medical College/Nanchong Central Hospital, Department of Radiology, Nanchong, China

[3]Nanjing University, Department of Computer Science, Xianlin Campus, Nanjing, China

[4]Medical Imaging Key Laboratory of Sichuan Province, Department of Radiology, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

[5]The Second Affiliated Hospital of Chongqing Medical University, Department of Radiology, Chongqing, China

***Contributed equally to this work.**

**Handling editor:** Burak Koçak

**Corresponding author:** Tian-wu Chen

**E-mail:** tianwuchen_nsmc@163.com

Received 17 September 2025; revision requested 21 October 2025; last revision received 27 November 2025; accepted 05 December 2025.

Epub: 11.12.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253655

**PURPOSE**

To develop and validate a model by incorporating abdominal multi-organ non-contrast computed tomography (CT) radiomics and clinical features to predict the feasibility of hepatic encephalopathy (HE) occurrence in patients with cirrhosis and hepatorenal failure.

**METHODS**

In total, 351 consecutive patients with cirrhosis and hepatorenal failure undergoing non-contrast abdominal CT scans at Centers 1 and 2 were enrolled. Patients from Center 1 were randomly allocated to training (n = 191) and internal test (n = 81) groups, and those from Center 2 were assigned to the external test group (n = 79). The nnU-Net framework was used for automated three-dimensional (3D) segmentation of abdominal organs—the liver, spleen, portal and splenic vein, inferior vena cava, esophagogastric junction, stomach, liver-adjacent small bowel, and colon. Segmented multi-organ radiomics features were extracted using 3D Slicer, with R software used for feature processing and model construction. Model performance in predicting HE occurrence was evaluated using receiver operating characteristic (ROC) analysis in the training, internal test, and external test cohorts. Decision curve analysis (DCA) was used to evaluate clinical utility. The SHapley Additive exPlanations (SHAP) tool was used to provide a basis for model interpretability analysis.

**RESULTS**

In total, 351 patients (mean age, 61.3 ± 10.7 years; 231 men) were enrolled in this study. Esophageal variceal bleeding, peritonitis, and ascites were independent clinical predictors of HE. Twenty discriminative radiomics features, selected from the abovementioned multi-organs through intraclass correlation coefficient and least absolute shrinkage and selection operator analysis, were used to construct the radiomics model. The integrated model, incorporating both radiomics and clinical features, obtained higher areas under the ROC curve than the radiomics and clinical models in the training (0.87 vs. 0.83 vs. 0.68), internal test (0.85 vs. 0.81 vs. 0.66), and external test (0.83 vs. 0.78 vs. 0.72) cohorts, as evidenced by favorable integrated discrimination improvement values ($P < 0.05$ for all). The integrated model demonstrated superior clinical utility in DCA. Moreover, SHAP feature contribution analysis revealed that the top five features in terms of contribution were all extracted from the digestive tract.

**CONCLUSION**

The integrated model can effectively predict HE occurrence in patients with cirrhosis and hepatorenal failure.

**CLINICAL SIGNIFICANCE**

This novel model, developed by integrating abdominal multi-organ non-contrast CT radiomics and clinical features, demonstrates robust performance in predicting the occurrence of cirrhosis-related HE in patients with cirrhosis and hepatorenal failure. It thus provides a valuable tool for clinical decision-making, facilitating the prevention of this complication.

**KEYWORDS**

Liver cirrhosis, hepatic encephalopathy, computed tomography, radiomics, hepatorenal function

You may cite this article as: Cao JM, Zhang MY, Ding XM, Zhou HY, Zhang XM, Chen TW. Multi-organ non-contrast computed tomography radiomics model to predict hepatic encephalopathy in patients with cirrhosis and hepatorenal failure. *Diagn Interv Radiol*. 2026;32(4):354-362.

Cirrhosis is highly prevalent globally and is a primary cause of morbidity and mortality in patients with chronic liver disease, accounting for 2.4% of global deaths in 2019.[1-3] With disease progression, cirrhosis often leads to hepatic decompensation, which in turn gives rise to severe complications such as hepatic encephalopathy (HE), esophageal variceal bleeding (EVB), and hepatorenal failure.[1-5] As a major complication of cirrhosis, HE affects more than 45% of patients with cirrhosis, causing substantial cognitive, psychomotor, and psychiatric impairments.[2] Its symptoms progress from subclinical or subtle psychomotor changes to confusion, somnolence, coma, or even death.[6,7] Notably, HE onset marks a critical turning point in cirrhosis progression, signifying the transition from the compensated to the decompensated stage. Consequently, patients with cirrhosis and HE have a considerably shorter median survival time, underscoring the urgency of early intervention.[6,8,9]

Radiomics-based machine learning involves extracting vast amounts of high-dimensional mineable data from radiological images, applying advanced characterization algorithms, and deriving imaging features.[10] This technology has been increasingly used in digital medical image analysis, supporting tasks such as automatic organ segmentation on computed tomography (CT) images.[11] By analyzing large datasets, machine learning equips computers with capabilities for pattern recognition, decision-making, and unbiased comprehensive image feature analysis, thereby facilitating holistic assessment of disease factors and potentially improving clinical outcome prediction accuracy.[2,12,13] Radiomics-based prediction models for HE risk assessment, treatment efficacy evaluation, and prognosis analysis are well established.[14-16] However, prior studies have been limited to single-center data, single-organ analysis, and reliance on contrast-enhanced CT.

**Main points**

- Hepatic encephalopathy (HE) is a severe complication in patients with cirrhosis; early clinical intervention can reduce the occurrence of this complication.
- Predictive models based on contrast-enhanced computed tomography (CT) radiomic features lack clinical utility for patients with hepatorenal failure.
- The integrated model, combining abdominal multi-organ non-contrast CT radiomics and clinical features, effectively predicts HE occurrence in patients with cirrhosis and hepatorenal failure.

Notably, patients with cirrhosis and hepatorenal failure are often unable to undergo contrast-enhanced CT, as underlying hepatic and renal impairment compromises the metabolism of contrast agents. Such impairment not only hinders contrast agent clearance but also further deteriorates hepatic and renal function. This critical limitation thus renders prior prediction models clinically unfeasible. Accordingly, this study aimed to develop and validate a multi-center model on the basis of radiomics features extracted from multiple organs on non-contrast abdominal CT scans to comprehensively predict HE occurrence in patients with cirrhosis and hepatorenal failure, thereby facilitating early clinical intervention.

## Methods

### Patients

The study protocol was approved by the Institutional Review Board of Nanchong Central Hospital [2024 (149)], with informed consent waived due to the study's retrospective nature. It was conducted in accordance with the Declaration of Helsinki and relevant ethical regulations.

Between February 2021 and December 2023, we consecutively collected the data of 411 patients with cirrhosis at two medical centers (Centers 1 and 2) in accordance with the following inclusion criteria: (1) cirrhosis was confirmed with laboratory examinations, radiological manifestations, and clinical findings according to the European Association for the Study of the Liver (EASL) and American Association for the Study of Liver Diseases clinical practice guidelines;[17,18] (2) all patients underwent biochemical workup and non-contrast abdominal CT scans within 1 week after initial hospitalization, and the quality of the CT data was satisfactory for image analysis according to the 5-point scale image-quality scoring system;[19] (3) patients did not experience HE before or during the initial hospitalization; and (4) the presence of hepatorenal failure. Hepatorenal failure was defined as concurrent abnormalities in liver and renal function parameters, which were confirmed via laboratory testing. Specifically, it included severe liver failure (total bilirubin ≥ 171 μmol/L and international normalized ratio ≥ 1.5) and renal insufficiency [estimated glomerular filtration rate (eGFR) < 45 mL/min/1.73 $m^2$], in which an eGFR < 30 mL/min/1.73 $m^2$ was defined as severe renal failure.[20] Other markers, including alanine transaminase, albumin, creatinine, and urine protein, were also used to confirm the presence of hepatorenal failure.

The exclusion criteria were as follows: (1) patients with intrahepatic or splenic space-occupying lesions (n = 31); (2) patients with a history of abdominal surgical procedures (n = 11); (3) patients complicated with other nervous system diseases (n = 5); (4) patients with incomplete clinical data (n = 3); and (5) patients lost to follow-up (n = 10). Based on the inclusion and exclusion criteria, 351 participants were enrolled in the study.

Patient follow-up was conducted in accordance with the following protocol. Monthly telephone assessment after discharge using cognitive assessment scales was conducted during the 2-year follow-up period, focusing on HE-related symptoms such as impaired calculation ability, memory deficits, or abnormal limb signs.[4,17] Patients with suspected HE or onset of related symptoms were advised to be admitted for definitive diagnosis, confirmed in accordance with EASL guidelines based on consistent neurological symptoms or test results. Follow-up ceased upon HE confirmation or after 2 years if undiagnosed, with final follow-up in May 2025.

Of these participants, 191 and 81 patients from Center 1 were randomly assigned to training and internal test groups, respectively, and 79 patients from Center 2 were allocated to the external test group. In both the training and test cohorts, patients with or without HE were divided into HE or non-HE subgroups, respectively, according to the follow-up results. Patient recruitment is illustrated in Figure 1.

### Computed tomography techniques

All the patients underwent non-contrast abdominal CT using a second-generation dual-source scanner (SOMATOM Definition Flash; Siemens Healthineers, Erlangen, Germany) in Center 1 and a 120-slice multi-detector spiral CT scanner (uCT710; United Imaging Healthcare Co., Ltd., Shanghai, China) in Center 2. The CT scanning parameters were a peak voltage of 140 kV, a tube current of 240 mA, detector collimation of 128 × 0.125 mm, a pitch of 1.75, slice thickness of 1.0 mm, and a matrix of 512 × 512 mm. All patients were required to fast for 6–8 hours prior to CT scanning and drink 800–1,000 mL of water 20 minutes before the procedure to distend the gastrointestinal tract. Each patient lay supine on the scanning bed and maintained breath-hold during the non-contrast abdominal CT acquisition. The

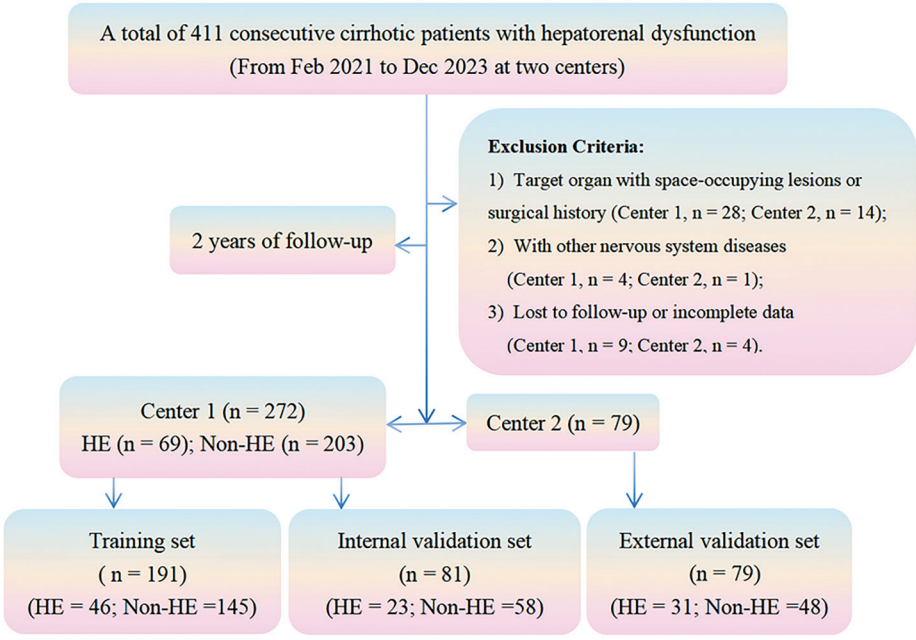



**Figure 1.** Patient selection flowchart. HE, hepatic encephalopathy.

scanning coverage extended from the dome of the diaphragm to the pubic symphysis. Subsequently, the CT data were transferred to the picture archiving and communication system (PACS), and all image data were saved in Digital Imaging and Communications in Medicine (DICOM) format for subsequent image analysis.

## Segmentation of multi-organ on computed tomography imaging

The CT data from both centers were extracted from the PACS system following image quality assessment. These DICOM files were then converted to NIfTI format (.nii.gz), and images were adjusted to standard abdominal window settings [width: 400 Hounsfield unit (HU); level: 50 HU] to facilitate subsequent segmentation. Preprocessed images were homogenized via resampling with an isotropic resolution of 1×1×1 $mm^3$ and via Z-score normalization to reduce systematic dataset variations, ensuring reliability for subsequent structural segmentation, feature extraction, and model construction. After completing CT image quality assessment and matching the images to clinical records, three-dimensional (3D) Slicer was used for additional preprocessing, including format validation by checking NIfTI compatibility for nnU-Net recognition, and visualization checks via multiplanar reconstruction to exclude poorly covered images. The preprocessed data were fed into the nnU-Net framework, an open-source deep learning tool developed by the German Cancer Research Center and implemented in Python, for automated segmentation of abdominal anatomical structures in non-contrast CT images using "TotalSegmentator".[21] The liver, spleen, portal and splenic vein, inferior vena cava, esophagogastric junction, stomach, liver-adjacent small bowel, and colon were selected as the target structures for subsequent radiomics feature extraction (Figure 2a-f). The selection of these nine organs was hypothesis driven, based on their clear links to HE pathogenesis in our target population. The liver is key for ammonia metabolism; the spleen and portal/splenic veins reflect portal hypertension; the stomach, esophagogastric junction, duodenum, and liver-adjacent bowel/colon relate to the gut–brain axis; and the inferior vena cava reflects systemic venous return changes in patients with cirrhosis and hepatorenal failure.[2,6,7]

## Evaluation of segmentation result accuracy

To validate the accuracy of automated multi-organ segmentation, all targeted anatomical structures segmented by nnU-Net underwent qualitative and quantitative assessments. For qualitative evaluation, two radiologists (first and third authors with 13 and 15 years of experience in abdominal imaging, respectively) reviewed the segmentation masks to assess alignment between automated segmentation boundaries and actual anatomical contours, with discrepancies resolved through consensus. For quantitative assessment, 20 randomly selected consecutive cases were semi-automatically segmented by the first author using 3D Slicer. The Dice similarity coefficient (DSC) was calculated to compare automated and semi-automated results, with a DSC ≥ 0.85 for all structures defined as the reliability criterion for subsequent feature extraction.[22]

## Computed tomography radiomics feature extraction

After image segmentation and accuracy assessment, CT radiomics features were extracted from the validated volume of interest (VOI) using the "Pyradiomics" extension in 3D Slicer. A total of 1,107 radiomics features were initially extracted from the nine target organs, categorized into four major types: (1) morphological features (e.g., surface area, volume, compactness), reflecting organ size and shape; (2) first-order statistical features (e.g., mean, median, skewness, kurtosis), reflecting the distribution of CT attenuation values within the VOI; (3) texture features from gray-level co-occurrence matrices (GLCMs), gray-level run-length matrices (GLRLMs), and gray-level size-zone matrices (GLSZMs) (e.g., contrast, correlation, energy, entropy), capturing spatial patterns of pixel intensity variations; and (4) higher-order features derived from wavelet transforms, quantifying texture information across multiple scales.

## Consistency analysis for radiomics feature selection

To ensure the stability and reliability of radiomics features, intraclass correlation coefficient (ICC) analysis was performed on 20 randomly selected cases. For these cases, radiomics features were extracted from both the initial automatic segmentation results via nnU-Net and the semi-automated segmentation results generated by the first author. ICC analysis was then used to compare the two feature sets, with those demonstrating an ICC > 0.8 retained as indicators of good reproducibility for subsequent analyses.[17] This approach enhanced the stability of features, mitigated potential biases from variability in image acquisition and segmentation, and laid a robust foundation for subsequent feature optimization.

## Feature selection and model construction

After ICC analysis, R software (version 4.5.1) was used to implement a rigorous pipeline encompassing feature optimization and selection, model construction, and performance evaluation. The packages included "rms," "glmnet," "caret," "pROC," "dplyr," and "ggplot2."

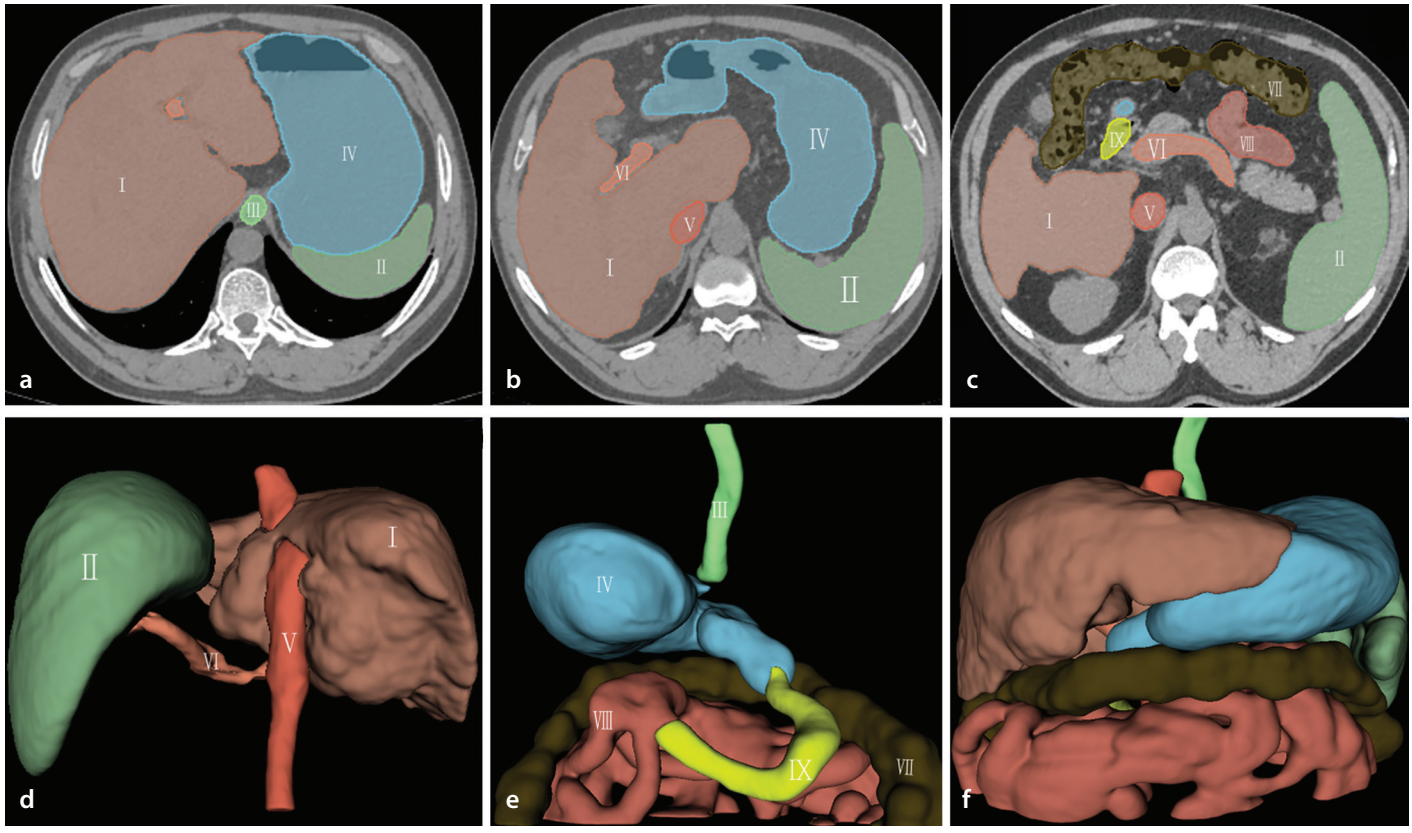



**Figure 2.** Example of automatic segmentation of multiple abdominal target organs. Images **(a-c)** show contour delineations of three two-dimensional cross sections. Images **(d-f)** depict three-dimensional visualizations of the segmented organs. I, liver; II, spleen, III, esophagogastric junction;s IV, stomach; V, inferior vena cava; VI, portal vein and splenic vein; VII, colon; VIII, small bowel; IX, duodenum.

The dataset from Center 1 was first split into training and internal test sets at a 7:3 ratio. Quantitative features in the training set were standardized using the Z-score method, with the same transformation parameters applied to the test set to avoid data leakage. Sample weights were calculated for the training set to address class imbalance. For features in the training set, Pearson correlation matrices with Holm–Bonferroni correction were computed to identify feature pairs with $|r| > 0.8$ (adjusted $P < 0.05$). Next, the variance inflation factor (VIF) was used to quantify multicollinearity (with VIF > 5 indicating significant multicollinearity). The specific feature that exhibits the highest VIF was explicitly identified and prioritized for removal to mitigate multicollinearity. Subsequently, a 10-fold cross-validated least absolute shrinkage and selection operator (LASSO) analysis was applied to the preprocessed training set using the binomial family with L1 regularization. The optimal regularization parameter (λ) was determined by minimizing the cross-validation error (λ-min). Feature stability was evaluated through bootstrapping (100 iterations), retaining only variables with Cohen's d > 0.5 selected in >70% of resamples to ensure robustness. A radiomics score (RS) was calculated for each case in the training, internal test, and external test sets using the predict function [predict(model, newdata = feature_data, type = "link")] in R software, and this function output the model's linear predictor. The calculation uses the following formula: $RS = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n$, where β represents regression coefficients and X represents radiomics features. Subgroup differences in the RS between patients with and without HE were compared.

Additionally, univariate analysis for clinical features was used to identify clinically important factors between the HE and non-HE groups. For this analysis, categorical factors, including sex, EVB, peritonitis, and ascites, were assessed using the chi-square test, and the continuous factor (age) was evaluated with the independent samples t-test.

Based on insights from previous radiomics and clinical feature analyses, three predictive models were developed using multivariate logistic regression—a clinical model, a radiomics model, and an integrated model—by combining radiomics features with the clinical factors. All three models were trained through 10-fold cross-validation in the training set, with coefficients optimized using maximum likelihood estimation, and predictive performance of the three models was evaluated in both the internal and external test sets. To interpret the predictive contribution of radiomics features, SHapley Additive exPlanations (SHAP) analysis was conducted. Mean absolute SHAP values were computed to quantify the importance of each radiomics feature, whereas SHAP beeswarm plots and feature importance bar plots were employed to visualize the impact of individual features and their contributions across samples.

## Statistical analysis

The performance of the models was evaluated using receiver operating characteristic (ROC) curve analysis and area under the ROC curves (AUCs), with the ROC test used to compare model performance. Based on the predicted classifications of the three models, confusion matrices were derived, and accuracy, sensitivity, and specificity were calculated for all datasets. Additionally, Brier scores were used to assess the calibration of predicted probabilities for each model, and integrated discrimination improvement (IDI) was employed to compare their discriminative ability.[23] DeLong's test was used to evaluate model overfitting across datasets.

Violin plots were used to visually depict the distributional differences in the RS between groups with and without HE. For baseline characteristics comparisons, normally distributed continuous variables were analyzed using the independent t-test; otherwise, the Mann–Whitney U test was used. Categorical variables were analyzed using the chi-square test or Fisher's exact test, as appropriate. A two-sided $P < 0.05$ was considered statistically significant.

## Results

### Baseline characteristics

The development set included 272 patients (173 men, 99 women), aged 61.2 ± 10.5 years (mean ± standard deviation), with an age range of 22–86 years. Of these, 69 (25.4%) had HE and 203 (74.6%) did not. The external test set comprised 79 patients (58 men, 21 women), aged 61.6 ± 11.1 years, with an age range of 34–87 years, including 31 (39.2%) with HE and 48 (60.8%) without. No significant differences were observed between sets ($P > 0.05$ for all). Among the clinical features, ascites, peritonitis, and EVB were identified as independent predictors of HE in the development and test sets ($P < 0.05$ for all) (Table 1).

### Radiomics feature extractions and selections

Across all target organs (the liver, spleen, portal and splenic veins, inferior vena cava, esophagogastric junction, stomach, and liver-adjacent small bowel and colon), the median DSC value between automated and semi-automated segmentations was 0.91 (ranging from 0.87 to 0.95), indicating the accuracy of automated multi-organ segmentation. Based on ICC analysis, 119 initial features with an ICC ≤ 0.8 were excluded, and the remaining 988 features with an ICC > 0.8 were retained for subsequent radiomics feature selection. To address multicollinearity, the feature with the highest VIF was removed, followed by 10-fold cross-validated LASSO logistic regression and correlation analysis. Eventually, a total of 20 discriminative radiomics features were selected (Figure 3a, b).

For these 20 selected features, categories included 3 first-order statistical features, 13 texture features (including 4 from GLCMs, 1 from GLRLMs, 3 from neighboring gray-tone difference matrices, 2 from gray-level dependence matrices, and 3 from GLSZMs), and 4 shape features. Based on the distribution across organs, the features were derived as follows: two from the spleen, one from the liver, one from the stomach, one from the

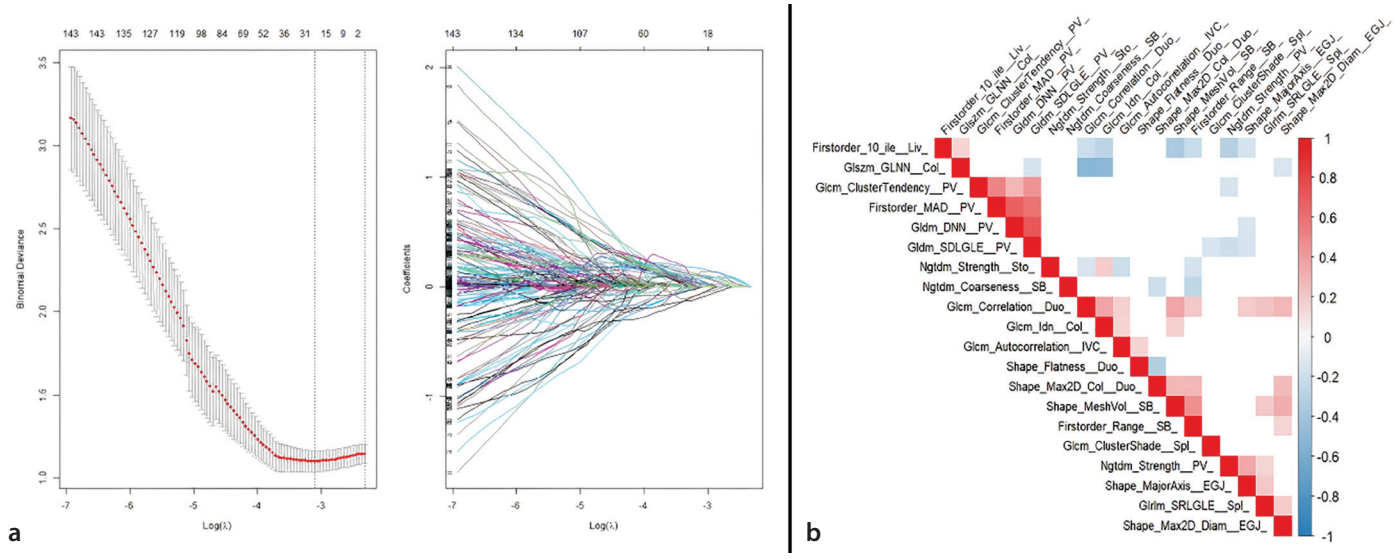



**Figure 3.** Feature selection using least absolute shrinkage and selection operator. Image **(a)** shows 10-fold cross-validation used for parameter tuning to obtain the best lambda, which is subsequently used for feature selection. Image **(b)** shows the correlation of selected features without multicollinearity in a heatmap. Spl, spleen; Liv, liver; Sto, stomach; IVC, inferior vena cava; PV, portal vein; EGJ, esophagogastric junction; SB, small bowel; Duo, duodenum; Col, colon.

**Table 1.** Patient characteristics in the development and external test sets

| Characteristics | Description | Development set (n = 272) | | | Test set (n = 79) | | |
|---|---|---|---|---|---|---|---|
| | | HE (n = 69) | Non-HE (n = 203) | *P* value | HE (n = 31) | Non-HE (n = 48) | *P* value |
| Gender, n (%) | Women | 25 (36.2%) | 74 (36.5%) | 1.0 | 8 (25.8%) | 13 (27.1%) | 1.0 |
| | Men | 44 (63.8%) | 129 (63.5%) | | 23 (74.2%) | 35 (72.9%) | |
| Age, years | Mean ± SD | 59.9 ± 11.2 | 61.7 ± 10.3 | 0.23 | 62.35 ± 11.11 | 61.17 ± 11.25 | 0.65 |
| EVB, n (%) | Y | 30 (43.5%) | 58 (28.6%) | 0.03 | 16 (51.6%) | 9 (18.8%) | 0.005 |
| | N | 39 (56.5%) | 145 (71.4%) | | 15 (48.4%) | 39 (81.2%) | |
| Ascites, n (%) | Y | 40 (58.0%) | 66 (32.5%) | <0.001 | 14 (45.2%) | 7 (14.6%) | 0.006 |
| | N | 29 (42.0%) | 137 (67.5%) | | 17 (54.8%) | 41 (85.4%) | |
| Peritonitis, n (%) | Y | 55 (79.7%) | 118 (58.1%) | 0.002 | 22 (71.0%) | 21 (43.8%) | 0.03 |
| | N | 14 (20.3%) | 85 (41.9%) | | 27 (56.2%) | 27 (56.2%) | |

HE, hepatic encephalopathy; SD, standard deviation; EVB, esophageal variceal bleeding; Y, yes; N, no.

inferior vena cava, four from the portal and splenic veins, two from the esophagogastric junction, three from the liver-adjacent small bowel, three from the liver-adjacent colon, and three from the duodenum.

Violin plots revealed distinct distributions of RSs between the HE and non-HE subgroups across the three cohorts. Patients with HE exhibited significantly higher RSs than those without HE ($P < 0.001$ for all), indicating that the RS effectively discriminated between both subgroups (Figure 4a-c).

### Performance of predictive models

Three predictive models were constructed: a radiomics model (incorporating selected radiomics features), a clinical model (incorporating selected clinical features), and an integrated model (combining all selected radiomics and clinical features). The performance and stability of all three models were evaluated using AUCs, accuracy, sensitivity, specificity, and Brier scores. The integrated model demonstrated superior clinical utility through decision curve analysis (DCA), revealing a higher net benefit at a range of 0.2–0.8 clinically relevant threshold probabilities (Figure 5a-d). Additionally, the integrated model showed significantly better predictive accuracy than the radiomics model, as evidenced by the IDI values, with the differences reaching statistical significance ($P < 0.05$) (Table 2). Moreover, SHAP feature contribution analysis revealed that the top five features in terms of contribution were extracted from the digestive tract (stomach, duodenum, liver-adjacent small bowel, and colon), including two texture features and three shape features (Figure 6a, b).

### Overfitting assessment for radiomics and integrated models

Overfitting was assessed by comparing model performance across datasets using DeLong's test (Table 3). For both models, pairwise comparisons between the training and internal and external test sets showed no statistically significant differences in AUCs ($P > 0.05$ for all), verifying that the models had no substantial overfitting risk.

## Discussion

This study developed and tested three predictive models for HE prediction in patients with cirrhosis and hepatorenal failure using multi-organ radiomics from non-contrast abdominal CT and clinical features. The integrated model, by combining multi-organ radiomics and clinical features, showed robust performance across the training, internal test, and external test cohorts, with AUCs of 0.87, 0.85, and 0.83, respectively; the absence of significant overfitting risk further strengthened its potential for clinical utility. Within these models, the RS demonstrated strong discriminative power for HE, with significantly higher values across cohorts. Additionally, three clinical features (ascites, EVB, and peritonitis) were identified as independent HE predictors, and their integration with radiomics features demonstrated superior performance, which strongly reinforced the merit of combining radiomics with clinical insights. This aligns with established mechanisms: ascites indicates decompensated cirrhosis, a critical transition from compensated liver dysfunction to overt failure of hepatic synthetic and metabolic capacity;[24] peritonitis triggers systemic inflammation, disrupting hepatic clearance of neurotoxins;[25] and variceal bleeding increases the ammonia burden by releasing nitrogen-rich blood into the gut, where bacterial metabolism further raises toxic ammonia levels.[26]

Notably, the radiomics features in this study were derived from nine abdominal organs, encompassing three categories: first-order statistical, texture, and shape features. This multi-organ strategy offers a distinct advantage over conventional radiomics studies, which often focus on isolated organs and thus overlook systemic pathological changes in cirrhosis.[2,4] By contrast, our multiple-organ coverage comprehensively captures multi-system dysfunction, giving it a critical advantage for evaluating complex disorders such as HE. Beyond features from well-studied organs (liver, spleen, and portal vein), 11 gastrointestinal radiomics features from the stomach, esophagogastric junction, liver-adjacent small bowel, and colon add unique value. These features not only highlight associations between portal hypertensive gastroenteropathy and decompensated cirrhosis but also provide radiomics-based evidence supporting the gut–brain axis and gut microbiota theories in HE pathogenesis. Specifically, the top five contributing radiomics features identified by SHAP extracted from the stomach, duodenum, liver-adjacent small bowel, and colon indicate that our observations align closely with the previously proposed gut–brain axis theory, corroborating its critical association with HE development; these features are insufficiently represented in radiomics studies limited to hepatic or splenic features.[2,27,28] More importantly, our logistic regression-based model offers superior interpretability, allowing clinicians to clearly quantify how multi-organ radiomics features contribute to HE risk predictions, which is critical for clinical trust and model adoption.

Critically, we employed non-contrast CT, a deliberate choice tailored to patients with cirrhosis and hepatorenal failure. Contrast agents are known to exacerbate renal dysfunction in this population,[20] yet many existing HE prediction models rely on contrast-enhanced CT,[2] restricting their applicability. Our use of non-contrast CT avoids this risk, leverages routine clinical data, and reduces costs, facilitating its integration into

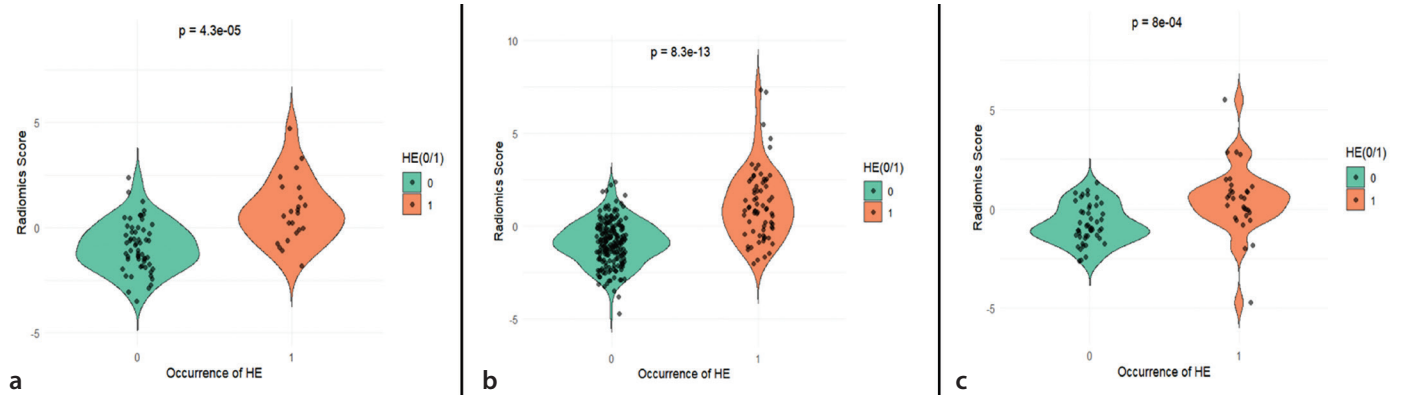



**Figure 4.** Violin plots show the distribution of radiomics scores. Training set **(a)**, internal test set **(b)**, and external test set **(c)**. HE, hepatic encephalopathy; 1, with hepatic encephalopathy; 0, without hepatic encephalopathy.

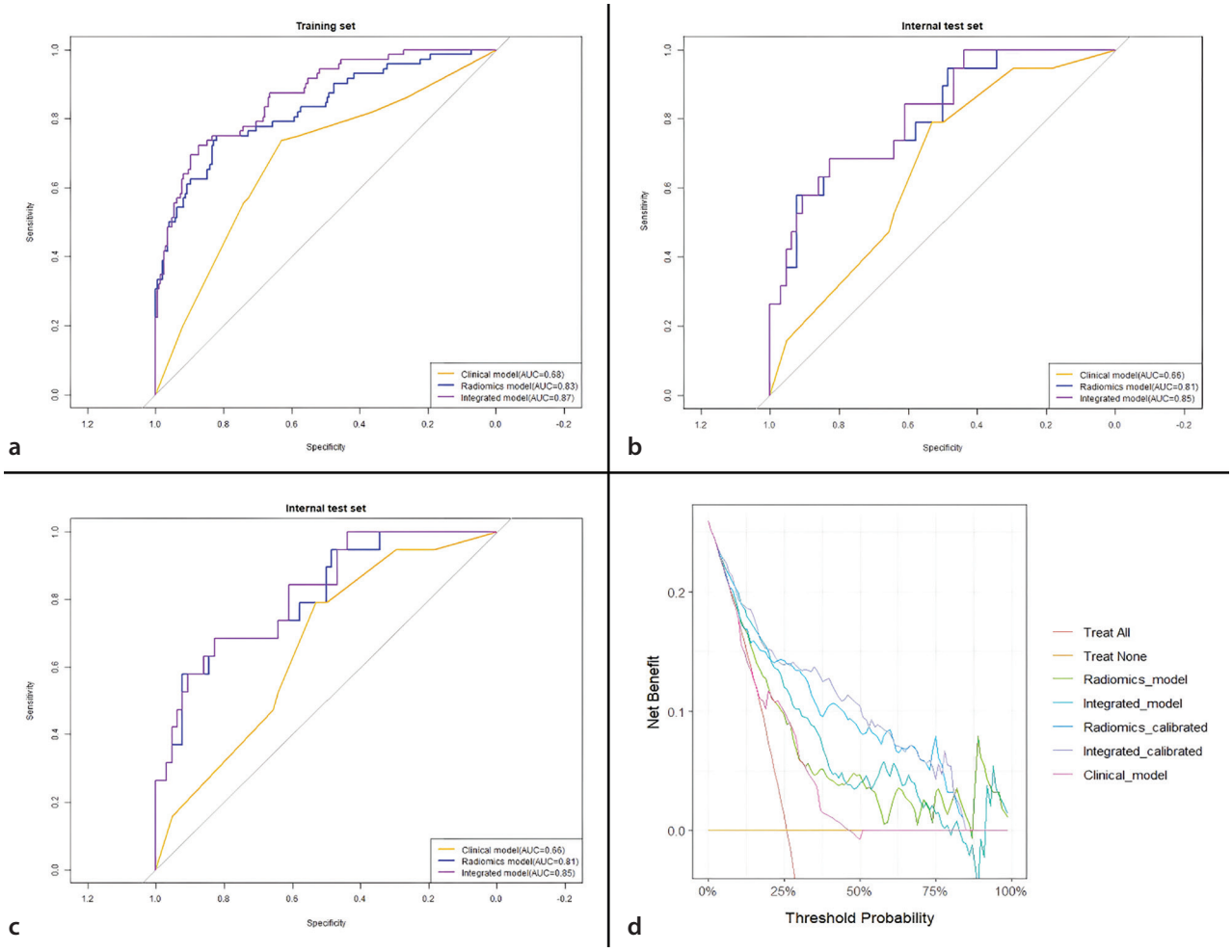



**Figure 5.** Receiver operating characteristic curves comparing the integrated clinical and radiomics models in the training set **(a)**, internal test set **(b)**, and external test set **(c)**. Decision curve analysis depicts net benefit curves to compare models across threshold probabilities, highlighting the integrated model's advantage **(d)**.

**Table 2.** Prediction performance of models in each dataset

| Model categories | Metrics | Training set | Internal test set | External test set |
|---|---|---|---|---|
| Clinical model | AUC (95% CI) | 0.68 (0.61–0.76) | 0.66 (0.54–0.79) | 0.72 (0.61–0.84) |
| | Acc | 0.66 | 0.59 | 0.72 |
| | Sen | 0.74 | 0.79 | 0.61 |
| | Spe | 0.63 | 0.53 | 0.79 |
| | Brier score | 0.1749 | 0.1648 | 0.1979 |
| Radiomics model | AUC (95% CI) | 0.83 (0.77–0.89) | 0.81 (0.71–0.92) | 0.78 (0.66–0.89) |
| | Acc | 0.85 | 0.75 | 0.72 |
| | Sen | 0.62 | 0.74 | 0.87 |
| | Spe | 0.92 | 0.76 | 0.63 |
| | Brier score | 0.1284 | 0.1498 | 0.1815 |
| Integrated model | AUC (95% CI) | 0.87 (0.82–0.91) | 0.85 (0.76–0.94) | 0.83 (0.74–0.93) |
| | Acc | 0.84 | 0.74 | 0.79 |
| | Sen | 0.64 | 0.91 | 0.77 |
| | Spe | 0.91 | 0.67 | 0.79 |
| | Brier score | 0.1212 | 0.1397 | 0.1569 |
| IDI (*P* value) | IM vs. CM | < 0.001 | < 0.001 | < 0.001 |
| | IM vs. RM | < 0.001 | 0.043 | < 0.001 |

AUC, area under the receiver operating characteristic curve; CI, confidence interval; IDI, integrated discrimination improvement; Acc, accuracy; Sen, sensitivity; Spe, specificity; IM, integrated model; CM, clinical model; RM, radiomics model.

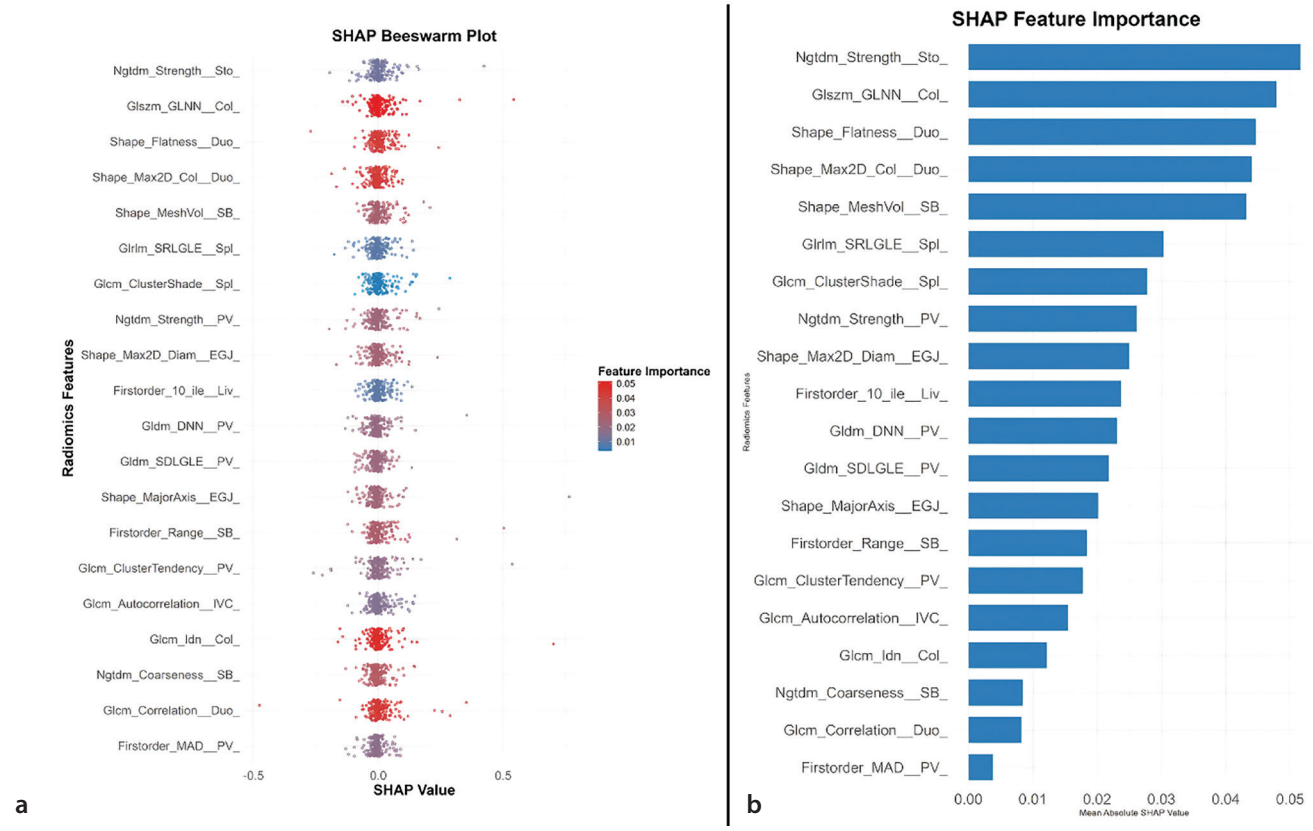



**Figure 6.** SHapley Additive exPlanations (SHAP) beeswarm plot with radiomics features on the y-axis (ordered by descending importance), SHAP values on the x-axis (denoting prediction impact), color-coded importance (red = high, blue = low), and dots representing each feature's per-sample contribution **(a)**. SHAP feature importance bar plot ranking radiomics features by their mean absolute SHAP values in descending order to illustrate their impact on model predictions **(b)**. Spl, spleen; Liv, liver; Sto, stomach; IVC, inferior vena cava; PV, portal vein; EGJ, esophagogastric junction; SB, small bowel; Duo, duodenum; Col, colon.

**Table 3.** Overfitting assessment across datasets using DeLong's test

| Model | Comparison group | AUC (TR-set) | AUC (TE-set) | *P* value | Overfitting (Y or N) |
|---|---|---|---|---|---|
| Clinical model | T-set vs. IT-set | 0.68 | 0.66 | 0.80 | N |
| | T-set vs. ET-set | 0.68 | 0.72 | 0.60 | N |
| Radiomics model | T-set vs. IT-set | 0.83 | 0.81 | 0.81 | N |
| | T-set vs. ET-set | 0.83 | 0.78 | 0.41 | N |
| Integrated model | T-set vs. IT-set | 0.87 | 0.85 | 0.74 | N |
| | T-set vs. ET-set | 0.87 | 0.83 | 0.56 | N |

TR-set, training set; IT-set, internal test set; ET-set, external test set; TE-set, test set; AUC, area under receiver operating characteristic curve, Y, yes; N, no.

practice, particularly in resource-limited settings. Notably, a prior study reported a weighted random forest model with an AUC of 0.82 for HE prediction in general patients with cirrhosis.[29] By contrast, our integrated model achieves comparable or superior performance, with an AUC ranging from 0.83 to 0.87, specifically for patients with cirrhosis and hepatorenal failure in which contrast-enhanced CT is contraindicated. This underscores our model's advantage in both filling a clinical gap and maintaining robust accuracy. These methodological strengths are further supported by our DCA results. Specifically, the integrated calibrated model delivered superior net benefit at a 0.2–0.8 threshold, which matches the clinical intervention time-point for HE prevention in patients with cirrhosis. Compared with the clinical and radiomics model, this advantage addresses the unmet need for early HE risk stratification, helping clinicians distinguish between low-risk and medium-risk patients to balance the benefits with overtreatment.

This study has several limitations. First, cirrhosis was not stratified by etiology; however, our integrated model had broad potential applications despite HE resulting from any causes. We will design cohort studies stratified by different etiologies to enhance the model's generalizability in future research. Second, although automatic segmentation via nnU-Net achieved high DSC values, its reliability may be decreased in cases of severe anatomical distortion (e.g., massive ascites or organ displacement). Follow-up studies could integrate manual correction for complex cases or explore more advanced segmentation algorithms to improve feature stability. Third, air in the gastrointestinal lumen may still introduce residual bias. Subsequent studies could use water filling or artificial intelligence-based air segmentation to further reduce this interference, ensuring more accurate extraction of radiomics features. Fourth, differences in CT scanners may introduce systematic biases in radiomics feature extraction despite standardization through Z-score normalization. Future studies should include device-specific calibration or multi-center harmonization protocols to miti-

gate this issue. In conclusion, extending the follow-up period or evaluating longitudinal changes in radiomics features may provide new insights into HE progression; we plan to investigate this in subsequent studies.

In conclusion, our combined model, integrating multi-organ radiomics features on abdominal non-contrast CT and important clinical factors, demonstrated robust performance in the prediction of HE secondary to cirrhosis in patients with hepatorenal failure. This model could provide valuable evidence for clinical decision-making regarding potential HE in this population.

### Footnotes

### Conflict of interest disclosure

The authors declared that there is no conflict of interest.

## References


1. Ginès P, Krag A, Abraldes JG, Solà E, Fabrellas N, Kamath PS. Liver cirrhosis. *Lancet*. 2021;398(10308):1359-1376. [Crossref]
2. Cao JM, Yang JQ, Ming ZQ, et al. A radiomics model of liver CT to predict risk of hepatic encephalopathy secondary to hepatitis B related cirrhosis. *Eur J Radiol*. 2020;130:109201. [Crossref]
3. Huang DQ, Terrault NA, Tacke F, et al. Global epidemiology of cirrhosis - aetiology, trends and predictions. *Nat Rev Gastroenterol Hepatol*. 2023;20(6):388-398. [Crossref]
4. Tan BG, Tang Z, Ou J, et al. An improved model based on quantitative features of right liver lobe, maximum varices, and portal vein system measured on magnetic resonance imaging to predict oesophagogastric variceal haemorrhage secondary to hepatitis B-related cirrhosis. *Quant Imaging Med Surg*. 2023;13(12):7741-7752. [Crossref]
5. Tapper EB, Ufere NN, Huang DQ, Loomba R. Review article: current and emerging therapies for the management of cirrhosis and its complications. *Aliment Pharmacol Ther*. 2022;55(9):1099-1115. [Crossref]
6. Louissaint J, Deutsch-Link S, Tapper EB. Changing epidemiology of cirrhosis and hepatic encephalopathy. *Clin Gastroenterol Hepatol*. 2022;20(8S):S1-S8. [Crossref]
7. Häussinger D, Dhiman RK, Felipo V, et al. Hepatic encephalopathy. *Nat Rev Dis Primers*. 2022;8(1):43. [Crossref]
8. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol*. 2006;44(1):217-231. [Crossref]
9. Tapper EB, Aberasturi D, Zhao Z, Hsu CY, Parikh ND. Outcomes after hepatic encephalopathy in population-based cohorts of patients with cirrhosis. *Aliment Pharmacol Ther*. 2020;51(12):1397-1405. [Crossref]
10. Mirestean CC, Pagute O, Buzea C, Iancu RI, Iancu DT. Radiomic machine learning and texture analysis - new horizons for head and neck oncology. *Maedica (Bucur)*. 2019;14(2):126-130. [Crossref]
11. Mai DVC, Drami I, Pring ET, et al. A systematic review of automated segmentation of 3D computed-tomography scans for volumetric body composition analysis. *J Cachexia Sarcopenia Muscle*. 2023;14(5):1973-1986. [Crossref]
12. Skrede OJ, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet*. 2020;395(10221):350-360. [Crossref]
13. Rakaee M, Tafavvoghi M, Ricciuti B, et al. Deep learning model for predicting immunotherapy response in advanced non-small cell lung cancer. *JAMA Oncol*. 2025;11(2):109-118. [Crossref]
14. Bo Z, Song J, He Q, et al. Application of artificial intelligence radiomics in the diagnosis, treatment, and prognosis of hepatocellular carcinoma. *Comput Biol Med*. 2024;173:108337. [Crossref]
15. Avanzo M, Wei L, Stancanello J, et al. Machine and deep learning methods for radiomics. *Med Phys*. 2020;47(5):e185-e202. [Crossref]
16. Chen B, Zhang R, Gan Y, Yang L, Li W. Development and clinical application of radiomics in lung cancer. *Radiat Oncol*. 2017;12(1):154. [Crossref]
17. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on the management of hepatic encephalopathy. *J Hepatol*. 2022;77(3):807-824. [Crossref]
18. Karvellas CJ, Bajaj JS, Kamath PS, et al. AASLD Practice guidance on acute-on-chronic liver failure and the management of critically ill patients with cirrhosis. *Hepatology*. 2024;79(6):1463-1502. [Crossref]
19. Chen TW, Yang ZG, Dong ZH, et al. Whole tumour first-pass perfusion using a low-dose method with 64-section multidetector row computed tomography in oesophageal squamous cell carcinoma. *Eur J Radiol*. 2011;80(2):284-291. [Crossref]
20. Lu S, Robyak K, Zhu Y. The CKD-EPI 2021 equation and other creatinine-based race-independent eGFR equations in chronic kidney disease diagnosis and staging. J Appl Lab Med. 2023;8(5):952-961. [Crossref]
21. 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]
22. Häntze H, Xu L, Mertens CJ, et al. Segmenting whole-body MRI and CT for multiorgan anatomic structure delineation. *Radiol Artif Intell*. 2025;7(6):e240777. [Crossref]
23. Huang S, Liang X, Lou K, et al. Comparing habitat, radiomics, and fusion models for predicting micropapillary/solid components in stage I lung adenocarcinoma. *Acad Radiol*. 2025;32(10):6307-6319. [Crossref]
24. Gracia-Sancho J, Marrone G, Fernández-Iglesias A. Hepatic microcirculation and mechanisms of portal hypertension. *Nat Rev Gastroenterol Hepatol*. 2019;16(4):221-234. [Crossref]
25. Lima LCD, Miranda AS, Ferreira RN, Rachid MA, Simões E Silva AC. Hepatic encephalopathy: lessons from preclinical studies. *World J Hepatol*. 2019;11(2):173-185. [Crossref]
26. Jindal A, Jagdish RK. Sarcopenia: ammonia metabolism and hepatic encephalopathy. *Clin Mol Hepatol*. 2019;25(3):270-279. [Crossref]
27. Agirman G, Yu KB, Hsiao EY. Signaling inflammation across the gut-brain axis. *Science*. 2021;374(6571):1087-1092. [Crossref]
28. Fried DE, Watson RE, Robson SC, Gulbransen BD. Ammonia modifies enteric neuromuscular transmission through glial γ-aminobutyric acid signaling. *Am J Physiol Gastrointest Liver Physiol*. 2017;313(6):G570-G580. [Crossref]
29. Yang H, Li X, Cao H, et al. Using machine learning methods to predict hepatic encephalopathy in cirrhotic patients with unbalanced data. *Comput Methods Programs Biomed*. 2021;211:106420. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253582

ABDOMINAL IMAGING

LETTER TO THE EDITOR



# Hepatic alveolar echinococcosis: a great tumor mimicker

Diğdem Kuru Öz
Ayşe Erden

Ankara University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

**To the editor,**

We read with great interest the recent article titled "Tumor-like conditions that mimic liver tumors" by Stanietzky et al.[1] which comprehensively reviewed non-neoplastic hepatic lesions that can resemble liver malignancies on imaging. We would like to contribute to this important discussion by drawing attention to another critical mimicker of hepatic tumors: alveolar echinococcosis (AE). In this context, we aim to present a case from our institution that demonstrated magnetic resonance imaging (MRI) findings mimicking a hepatic mass. A 35-year-old female patient was referred following the detection of a liver mass on ultrasonography due to abdominal pain. Dynamic contrast-enhanced liver MRI performed at our hospital revealed a 7-cm hypovascular lesion in segments 2/4a of the liver. The lesion appeared hypointense on T1-weighted images and mildly hyperintense on T2-weighted images. On T2-weighted sequences, punctate hyperintense foci were observed within the lesion. Diffusion-weighted imaging demonstrated peripheral ring-like diffusion restriction, and the apparent diffusion coefficient map showed a hypointense rim limited to the periphery. Except for septum-like internal structures, no significant contrast enhancement was noted in the lesion (Figure 1). Prospectively, the lesion was interpreted as an unspecified hepatic mass with concerning features. The mass-like appearance, peripheral diffusion restriction, and hypovascular enhancement pattern suggested intrahepatic cholangiocarcinoma as the primary differential diagnosis, with metastatic disease also considered despite no known primary malignancy. The case was discussed in a multidisciplinary tumor board meeting, and surgical resection was planned for both diagnostic and therapeutic purposes, as percutaneous biopsy was not performed due to concerns regarding potential tumor seeding. Preoperative imaging evaluation suggested a localized lesion confined to the left hepatic lobe without evidence of lymphadenopathy or distant metastases (clinical stage T1–2 N0 M0, assuming intrahepatic cholangiocarcinoma). Given the resectable nature of the lesion and the patient's good performance status, upfront surgical resection was indicated according to standard management guidelines for early-stage intrahepatic cholangiocarcinoma, as neoadjuvant therapy is not routinely recommended for resectable disease. The patient underwent left hepatic lobectomy. Macroscopic examination of the resected specimen revealed a tumor-like lesion measuring 8 × 7 × 4 cm, infiltrating the liver capsule, with a cream-yellow color and focal areas of hemorrhage, but without overt necrosis. Histopathological evaluation confirmed the diagnosis of AE. Upon retrospective review of the MRI, punctate hyperintense foci within the lesion—possibly corresponding to small vesicles—were noted on T2-weighted images. The term "alveolar" in AE refers to the presence of multiple vesicles resembling alveoli within the lesion.[2] Therefore, these small cystic components could have raised the suspicion of AE in the differential diagnosis of this mass-like lesion on MRI.

AE is a parasitic infection caused by *Echinococcus multilocularis* that primarily involves the liver. This disease may manifest as infiltrative hepatic masses with irregular borders, central necrosis, calcifications, and absence of significant contrast enhancement—features that can closely mimic cholangiocarcinoma or metastasis. Its tumor-like growth pattern and potential for local invasion and distant spread further complicate differentiation from true neoplastic entities.[2,3] In endemic areas, including parts of Central Europe and Türkiye, AE should be considered in the differential diagnosis of atypical hepatic lesions, particularly when calcifications and multiple small vesicular components are present.

**Handling editor:** İlkay İdilman

**Corresponding author:** Diğdem Kuru Öz

**E-mail:** digdem_k@hotmail.com

Received 25 July 2025; revision requested 01 September 2025; accepted 10 October 2025.

Epub: 17.11.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253582

You may cite this article as: Kuru Öz D, Erden A. Hepatic alveolar echinococcosis: a great tumor mimicker. *Diagn Interv Radiol*. 2026;32(4):363-364.

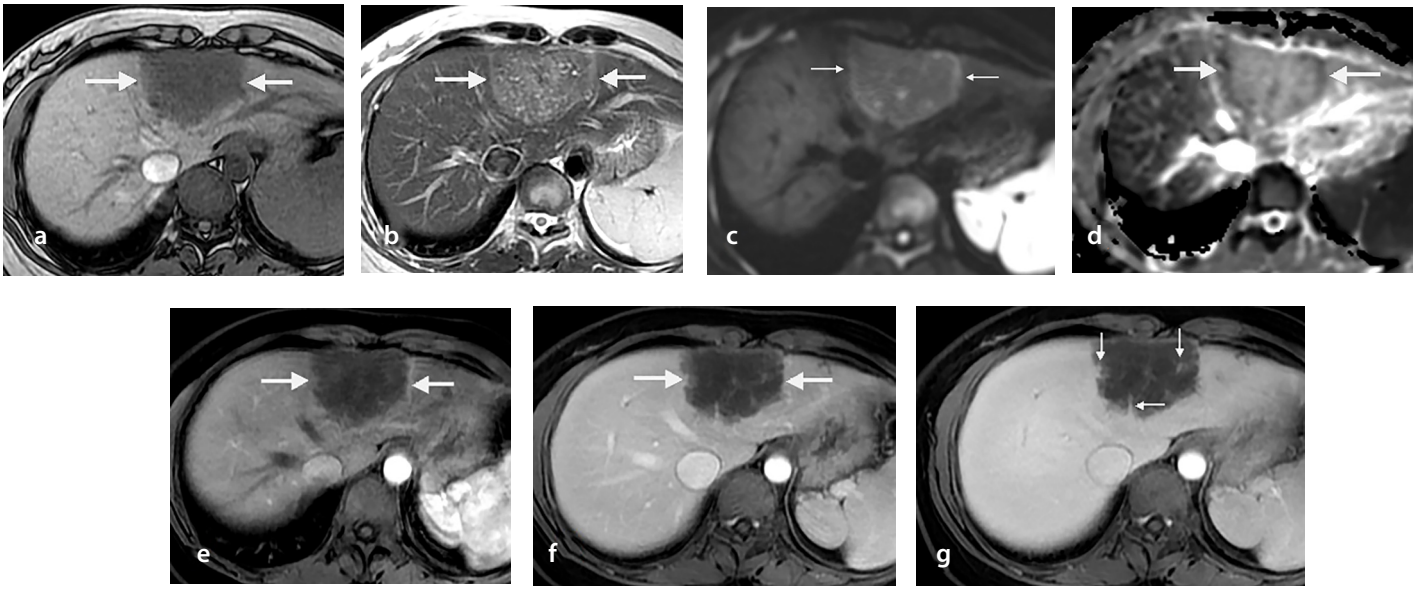



**Figure 1.** A 35-year-old woman with hepatic alveolar echinococcosis. On dynamic contrast-enhanced liver magnetic resonance imaging, a 7-cm mass-forming lesion (arrows in **a–f**) was observed in segments 2/4a. It appeared hypointense on T1-weighted images **(a)** and mildly hyperintense on T2-weighted images **(b)**. Note the punctate hyperintense foci within the lesion on T2-weighted images, possibly corresponding to small vesicles. Diffusion-weighted imaging **(c)** demonstrated peripheral ring-like diffusion restriction, and the apparent diffusion coefficient map **(d)** revealed a hypointense rim confined to the periphery. On post-contrast arterial **(e)**, venous **(f)**, and hepatobiliary **(g)** phases, no significant contrast enhancement was noted within the lesion, except for minimal enhancement along septum-like internal structures (small arrows in **g**).

Misdiagnosis may lead to inappropriate management, including unnecessary surgical resections or delayed antiparasitic treatment in inoperable patients.[3] We believe that recognizing AE as a hepatic tumor mimic and including it among radiologic differentials—especially in endemic regions—is vital for accurate diagnosis and appropriate therapy planning.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References

1. Stanietzky N, Salem AE, Elsayes KM, et al. Tumor-like conditions that mimic liver tumors. *Diagn Interv Radiol*. 2025;31(4):285-294. [Crossref]
2. Liu W, Delabrousse É, Blagosklonov O, et al. Innovation in hepatic alveolar echinococcosis imaging: best use of old tools, and necessary evaluation of new ones. *Parasite*. 2014;21:74. [Crossref]
3. Bulakçı M, Kartal MG, Yılmaz S, et al. Multimodality imaging in diagnosis and management of alveolar echinococcosis: an update. *Diagn Interv Radiol*. 2016;22(3):247-256. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253270

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE



# Comparative analysis of tumor and mesorectum radiomics in predicting neoadjuvant chemoradiotherapy response in locally advanced rectal cancer

Ali Cantürk[1]
Raif Can Yarol[2]
Ali Samet Tasak[3]
Hakan Gülmez[4]
Kenan Kadirli[1]
Tayfun Bişgin[5]
Berke Manoğlu[5]
Selman Sökmen[5]
İlhan Öztop[6]
İlknur Görken Bilkay[6]
Özgül Sağol[7]
Sülen Sarıoğlu[8]
Funda Barlık[2]

[1]University of Health Sciences Türkiye, Sultan 2. Abdülhamid Han Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

[2]Dokuz Eylül University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[3]Ministry of Health, Department of Radiology, Ağrı, Türkiye

[4]İzmir Demokrasi University Faculty of Medicine, Department of Family Medicine, İzmir, Türkiye

[5]Dokuz Eylül University Faculty of Medicine, Department of General Surgery, İzmir, Türkiye

[6]Dokuz Eylül University Faculty of Medicine, Department of Medical Oncology, İzmir, Türkiye

[7]Dokuz Eylül University Faculty of Medicine, Department of Pathology, İzmir, Türkiye

[8]Memorial Şişli Hospital, Clinic of Pathology, İstanbul, Türkiye

**Handling editor:** İlkay İdilman

**Corresponding author:** Ali Cantürk

**E-mail:** alicanturk.md@gmail.com

Received 09 February 2025; revision requested 09 March 2025; last revision received 29 June 2025; accepted 10 July 2025.

Epub: 12.08.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253270

**PURPOSE**

Neoadjuvant chemoradiotherapy (CRT) is known to increase sphincter preservation rates and decrease the risk of postoperative recurrence in patients with locally advanced rectal tumors. However, the response to CRT in patients with locally advanced rectal cancer (LARC) varies significantly. The objective of this study was to compare the performance of models based on radiomics features of the tumor alone, the mesorectum alone, and a combination of both in predicting tumor response to neoadjuvant CRT in LARC.

**METHODS**

This retrospective study included 101 patients with LARC. Patients were categorized as responders (modified Ryan score 0–1) and non-responders (modified Ryan score 2–3). Pre-CRT magnetic resonance imaging evaluations included tumor-T2 weighted imaging (T2WI), tumor-diffusion weighted imaging (DWI), tumor-apparent diffusion coefficient (ADC) maps, and mesorectum-T2WI. The first radiologist segmented the tumor and mesorectum from T2-weighted images, and the second radiologist performed tumor segmentation using DWI and ADC maps. Feature reproducibility was assessed by calculating the intraclass correlation coefficient (ICC) using a two-way mixed-effects model with absolute agreement for single measurements [ICC(3,1)]. Radiomic features with ICC values <0.60 were excluded from further analysis. Subsequently, the least absolute shrinkage and selection operator method was applied to select the most relevant radiomic features. The top five features with the highest coefficients were selected for model training. To address class imbalance between groups, the synthetic minority over-sampling technique was applied exclusively to the training folds during cross-validation. Thereafter, classification learner models were developed using 10-fold cross-validation to achieve the highest performance. The performance metrics of the final models, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC), were calculated to evaluate the classification performance.

**RESULTS**

Among the 101 patients, 36 were classified as responders and 65 as non-responders. A total of 25 radiomic features from the tumor and 20 from the mesorectum were found to be statistically significant ($P < 0.05$). The AUC values for predicting treatment response were 0.781 for the tumor-only model (random forest), 0.726 for the mesorectum-only model (logistic regression), and 0.837 for the combined model (logistic regression).

**CONCLUSION**

Radiomic features derived from both the tumor and mesorectum demonstrated complementary prognostic value in predicting treatment response. The inclusion of mesorectal features substantially improved model performance, with the combined model achieving the highest AUC value. These findings highlight the added predictive contribution of the mesorectum as a key peritumoral structure in radiomics-based assessment.

**CLINICAL SIGNIFICANCE**

Currently, the response of locally advanced rectal tumors to neoadjuvant therapy cannot be reliably predicted using conventional methods. Recently, the significance of the mesorectum in predicting treatment response has gained attention, although the number of studies focusing on this area remains limited. In our study, we performed radiomics analyses of both the tumor tissue and the mesorectum to predict neoadjuvant treatment response.

**KEYWORDS**

Artificial intelligence, locally advanced rectal cancer, magnetic resonance imaging, neoadjuvant treatment, radiomics

You may cite this article as: Cantürk A, Yarol RC, Tasak AS, et al. Comparative analysis of tumor and mesorectum radiomics in predicting neoadjuvant chemoradiotherapy response in locally advanced rectal cancer. *Diagn Interv Radiol*. 2026;32(4):365-375.

The standard imaging modality for locally advanced rectal cancer (LARC) is magnetic resonance imaging (MRI) to assess rectal wall invasion (T stage), evaluation of locoregional lymph nodes, macroscopic tumor invasion into the mesorectum, mesorectal fascia involvement, and extramural vascular invasion.[1,2] Neoadjuvant chemoradiotherapy (CRT) plays a crucial role in the management of LARC by not only increasing sphincter preservation rates but also facilitating organ preservation through non-operative strategies, such as the watch-and-wait approach, in carefully selected patients who achieve a complete clinical response. Furthermore, CRT has been shown to reduce the risk of postoperative recurrence significantly.[3,4] However, the response of patients with LARC to neoadjuvant CRT is variable. Neoadjuvant CRT results in tumor stage regression in 50% of patients, and pathologic complete response is observed in 15%–20% of patients.[5] Currently, the response of locally advanced rectal tumors to neoadjuvant therapy cannot be estimated by conventional methods. The prediction of tumor response to neoadjuvant treatment at the time of diagnosis can contribute to patient-specific tailoring of radiation doses and thus increase pathologic complete response and organ preservation rates.[6] Therefore, estimating the tumor's response to the neoadjuvant treatment is important for treatment management.

The influence of adipocytes on tumor pathogenesis has been intensively investigated in recent years. The molecular interaction between tumor cells and adipocytes has been associated with an increase in inflammatory markers and angiogenic factors, such as vascular endothelial growth factor (VEGF) and insulin-like growth factor 1 (IGF-1), that may locally and systematically provoke tumor growth and metastasis. The interaction between rectal cancer and mesorectal adipose tissue has been demonstrated to induce molecular alterations in adipocytes. These changes may lead to subtle MRI findings that are not readily detectable with conventional radiologic methods.[7,8] Some radiomics studies in the literature have evaluated peritumoral adipose tissue to predict clinical outcomes and prognosis. In breast tumors, evaluation of the peritumoral area has been proven to improve the differentiation between benign and malignant breast lesions.[9] Likewise, in non-small cell lung cancers, peritumoral lung parenchyma may also predict recurrence after surgery.[10]

In this study, we performed radiomics analyses of the tumor and mesorectum to predict the response to neoadjuvant CRT; a tumor-only model, mesorectum-only model, and combined tumor-mesorectum model were constructed.

**Main points**

- In this study, we developed machine-learning models to predict tumor response to neoadjuvant therapy using radiomics analysis of both the tumor and mesorectum. The area under the receiver operating characteristic values were 0.781 for the tumor-only model, 0.726 for the mesorectum-only model, and 0.837 for the combined tumor and mesorectum model.
- Molecular alterations in peritumoral adipocytes may induce subtle magnetic resonance imaging signal changes that are not visually apparent, highlighting the value of radiomics in quantitatively capturing these hidden imaging features.
- Radiomic-based assessment of the mesorectum underscores its added prognostic value in evaluating neoadjuvant treatment response, providing complementary insights beyond tumor-derived radiomic signatures.

## Methods

### Study participants

This study was approved by the Non-Interventional Research Ethics Committee of Dokuz Eylül University Hospital (approval number: 2023/33-18, date: August/2023). Due to the study's retrospective nature, the requirement for informed consent was waived. Details of patients with LARC who underwent neoadjuvant CRT followed by total mesorectal excision between March 2017 and May 2022 were retrieved from the hospital database. Patients who underwent rectal MRI before CRT were included in the study. The exclusion criteria were patients with MRI images with different parameters, pathologic evaluation performed outside the hospital, poor image quality, and patients who refused to be operated on. The patient accrual is summarized in Figure 1.

### Image acquisition

Examinations were performed on a 1.5-T MRI machine (Philips Achieva Release 1.8, Eindhoven, The Netherlands) with a pelvic phased-array coil. Turbo spin-echo T2-weighted images (T2WI) were acquired in the sagittal, para-axial (perpendicular to the long axis of the tumor), and para-coronal (parallel to the long axis of the tumor) planes using a repetition time (TR) of 4,500 ms, a field of view (FOV) of 180–220 mm, a matrix size of 256 × 512, a slice thickness of 3 mm, an intersection interval of 0.8 mm, and an echo train length of 16. Diffusion-weighted images (b: 0 and b: 1.000 s/mm$^2$) were acquired in the axial and sagittal planes with a single-shot echo-planar sequence using a 4.200/95 TR/echo time, 350–400 mm FOV, 90° bank angle, and 5-mm slice thickness. Apparent diffusion coefficient (ADC) maps were generated automatically by the software. Fat suppression techniques and contrast agents were not used. Scopolamine butyl bromide (20 mg) was injected intravenously 10 minutes before scanning to reduce intestinal motility.

### Protocol for neoadjuvant chemoradiotherapy

All patients received 45 gray (Gy) of pelvic radiotherapy before surgery. Subsequently, a boost of 5.4 Gy in three fractions was administered to the primary tumor. After the first and fifth weeks of radiotherapy, patients received 400 mg/m$^2$/day fluorouracil and 20 mg/m$^2$/day leucovorin for 3 days. Restaging MRI was performed approximately 6–8 weeks after completion of neoadjuvant CRT.

### Evaluation of the pathologic response to treatment

In this study, the modified Ryan scoring system was used as the gold standard (Table 1). The modified Ryan scoring has proven to be a reliable tool for classifying tumor regression due to its high reproducibility and inter-observer agreement.[11] It is based on the ratio of residual cancer cells to the fibrosis amount. In the modified Ryan scoring system, 0 points are given for complete response, and a score of 3 points indicates a poor response or no response to neoadjuvant treatment.

In the study, the patients were divided into two groups. Patients with a modified Ryan score of 0–1 were classified as responders to neoadjuvant treatment, and patients with a modified Ryan score of 2–3 were classified as non-responders to neoadjuvant treatment.

### Image interpretation–texture feature extraction

Data in Digital Imaging and Communications in Medicine format were transferred to a workstation and analyzed by dedicated software (LIFEx version 7.4, Inserm, Orsay, France).

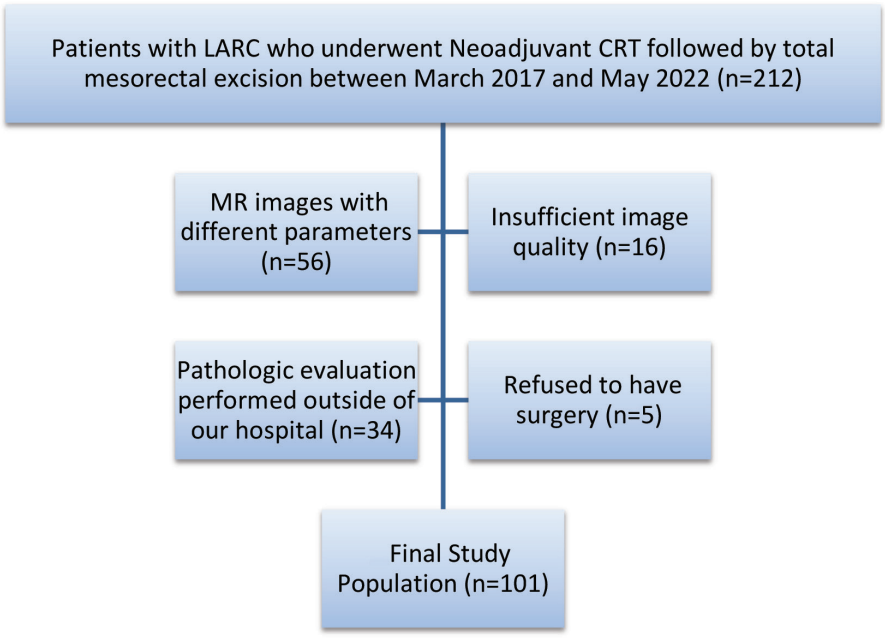



**Figure 1.** Flowchart of the study. LARC, locally advanced rectal cancer; CRT, chemoradiotherapy; MR, magnetic resonance.

**Table 1.** Modified Ryan scheme for tumor regression score

| Description | Tumor regression score |
|---|---|
| No viable cancer cells (complete response) | 0 |
| Single cells or rare small groups of cancer cells (near complete response) | 1 |
| Residual cancer with evident tumor regression, but more than single cells or rare small groups of cancer cells (partial response) | 2 |
| Extensive residual cancer with no evident tumor regression (poor or no response) | 3 |

Both tumor tissue and mesorectal adipose tissue were examined in this study. Tumor tissue and mesorectum were segmented separately from T2WI. In addition, tumor tissue was segmented using diffusion-weighted imaging (DWI).

Gy-level normalization and Gy-level discretization were performed to minimize the impact of differences in acquisition protocols on texture features and to generate a homogeneous dataset. For this reason, the voxel values of each lesion in three axes (x, y, z) were recorded, and the median values of these recorded data were obtained. These median values were then utilized as optimized parameters in the texture analysis of each lesion.[12] The intensity range was normalized using Z-scoring [mean ± 3 standard deviation (SD)]. Image intensities were discretized into 128 fixed bins.

In the study, the MRI images obtained at the time of diagnosis (pre-treatment MRI) were evaluated. Three radiologists with 5 years (AC), 4 years (RCY), and 33 years (FB) of experience in radiology evaluated the images of 10 patients together. The first radiologist (AC) performed a three-dimensional (3D) semi-automatic segmentation of the entire tumor (Figure 2a, b) and mesorectal adipose tissue (Figure 3) from the axial T2WI without fat suppression of all patients. Mesorectum segmentation was conducted from the point of attachment of the anterior peritoneal reflection to the rectal wall in the cranial section to the intersphincteric area in the caudal section. The second radiologist (RCY) performed a 3D semi-automatic segmentation of the entire tumor using DWI images (Figure 2c, d) and ADC mapping (Figure 2e, f) in the axial plane of all patients. The total number of radiomics features obtained was 17,978.

## Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 24.0 (IBM Corp., Armonk, NY, USA). The normality of numerical variables, such as age, was assessed using the Kolmogorov–Smirnov test. Correlation analyses between radiomic features were performed using the Spearman rank correlation coefficient, as the features did not follow a normal distribution. Continuous variables were expressed as mean ± SD, and differences in mean age between groups were analyzed using the independent samples t-test. Categorical variables, including sex, distance of extramural extension, and distance to the mesorectal fascia, were compared between groups using the Pearson chi-squared test, as all expected cell frequencies were ≥5. For the comparison of pretreatment, where expected cell counts were below the acceptable threshold, the Fisher–Freeman–Halton test was applied. A $P$ value of <0.05 was considered statistically significant.

## Feature selection and machine learning models

Radiomic analysis was conducted using LIFEx software to extract features from tumor and mesorectal segmentations. Prior to feature selection, all radiomic features were normalized using Z-score normalization. To ensure reproducibility, interobserver agreement was assessed on 20 randomly selected patients using (ICC)(3,1) (two-way mixed-effects model, absolute agreement, single measures). Features with ICC values of <0.60 were excluded from further analysis. Feature selection was performed using the least absolute shrinkage and selection operator (LASSO) regression method to reduce dimensionality and retain the most predictive features while minimizing the risk of overfitting. The top five features with the highest coefficients were selected for model training. To address class imbalance between groups, the synthetic minority over-sampling technique (SMOTE) was applied exclusively to the training folds during cross-validation to avoid data leakage (Figure 4). The extracted radiomic data were transferred to Python (version 3.9). Machine learning classifiers–including logistic regression, random forest, extreme gradient boosting (XGBoost), support vector machine (SVM) with radial basis function (RBF) kernel, and k-nearest neighbors (KNN)–were implemented using the scikit-learn and XGBoost libraries. Model performance was evaluated using 10-fold cross-validation. In each iteration, the dataset was split into 9 folds for training and 1 fold for testing, repeated 10 times to calculate average performance.[13,14] Evaluation metrics included accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC).

The methodological quality of this study was evaluated using the METhodological RadiomICs Score (METRICS), a standardized tool comprising 30 parameters that assess key aspects of radiomics research, including imaging acquisition, feature extraction, and model validation. The METRICS tool categorizes studies into quality ratings of very low (0%–20%), low (20%–40%), moderate (40%–60%), good (60%–80%), and excellent (80%–100%).[15]

## Results

In this study, a total of 101 patients [mean age 61.6 ± 13.59 years, 34 women (33.7%) and 67 men (66.3%)] with LARC were evaluated using high-resolution rectal MRI.

In the initial MRI, of 101 patients, 15.8% (n = 16) were staged as T2, 41.6% (n = 42) were staged as T3b, 20.8% (n = 21) were staged as T3c, 11.9% (n = 12) were staged as T3d, 6% (n = 6) were staged as T4a, and 4% (n = 4) were staged as T4b. In the MRI images obtained for re-staging after neoadjuvant CRT, 8% (n = 8) were in the T0 stage, 15% (n = 15) were in the T1 stage, 52.5% (n = 53) were in the T2 stage, 12% (n = 12) were in the T3b stage, 6% (n = 6) were in the T3c stage, 2% (n = 2) were in the T3d stage, 2% (n = 2) were in the T4a stage, and 3% (n = 3) were in the T4b stage.

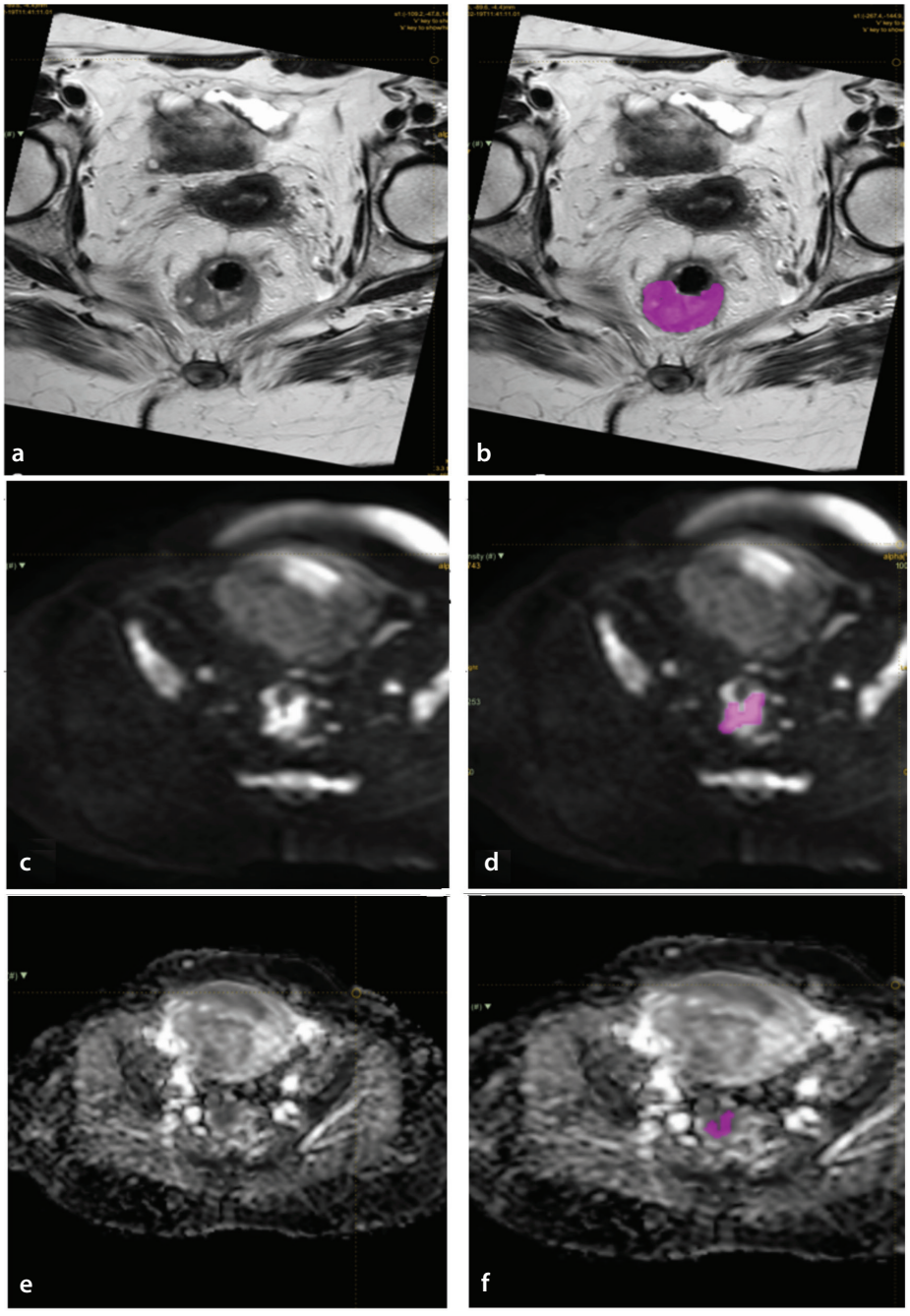



**Figure 2.** Tumor segmentation T2 weighted imaging **(a-b)**, tumor directed acyclic graph segmentation **(c-d)**, Tumor apparent diffusion coefficient map segmentation **(e-f)**.

The response to neoadjuvant treatment, according to the findings in the postoperative pathological material, was divided into groups by modified Ryan scoring. There were 21 patients (20%) with a Ryan score of 0, 15 patients (15%) with a modified Ryan score of 1, 50 patients (50%) with a score of 2, and 15 patients (15%) with a score of 3. Patients with modified Ryan scores of 0–1 were classified as responding, and patients with modified Ryan scores of 2–3 were classified as non-responding (Figure 5).

A total of 101 patients were included in the study, of whom 36 were classified as good responders (36%) and 65 as poor responders (64%). The mean age of good responders was 62 ± 12.5 years, and the mean age of poor responders was 65 ± 9.5 years. No statistically significant difference was observed between the mean age of patients who responded well and poorly to neoadjuvant treatment ($P = 0.115$). No significant correlation was identified between the T stage ($P = 0.196$), extramural extension (0.167), the proximity of the tumor to the mesorectal fascia ($P = 0.316$), and the neoadjuvant treatment response (Table 2).

A radiomic analysis was conducted on the tumor and mesorectum to predict the response of the neoadjuvant CRT.

### Prediction of treatment response

In the analyses performed to predict neoadjuvant CRT response, 25 radiomics features from the tumor (Table 3) and 20 radiomics features from the mesorectum (Table 3) were found to be significant ($P < 0.05$).

Radiomic features were extracted from the tumor region on T2WI and DWI MRI images to construct the tumor-only model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The random forest classifier achieved an accuracy of 69.2%, a precision of 70.2%, a recall of 66.7%, an F1-score of 68.4%, and an AUC of 0.781. The XGBoost model yielded an AUC of 0.737. The logistic regression, SVM (RBF kernel), and

KNN models resulted in AUCs of 0.714, 0.676, and 0.700, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the tumor-only model are illustrated in Figure 6a. The odds ratios (ORs) and 95% confidence intervals (CIs) from the logistic regression models are summarized in Table 5.

Radiomic features were extracted from the mesorectum on T2WI images to construct the mesorectum-only model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The logistic regression model achieved an accuracy of 66.7%, a precision of 66.1%, a recall of 68.3%, an F1-score of 67.2%, and an AUC of 0.726. The XGBoost and random forest models yielded AUCs of 0.708 and 0.700, respectively. The SVM (RBF kernel) and KNN models resulted in AUCs of 0.711 and 0.661, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the mesorectum-only model are illustrated in Figure 6b. The ORs and 95% CIs from the logistic regression models are summarized in Table 5.

Radiomic features extracted from both the tumor and mesorectum regions were combined to construct the combined model. The five most predictive parameters were selected using LASSO. Multiple machine learning models were constructed. The logistic regression model achieved an accuracy of 81%, a precision of 82.1%, a recall of 81.4%, an F1-score of 81.9%, and an AUC of 0.837. The random forest model yielded an AUC of 0.816. The AUCs for the XGBoost, SVM (RBF kernel), and KNN models were 0.789, 0.811, and 0.754, respectively. A detailed summary of the performance metrics for all classifiers in the models is presented in Table 4. The ROC curves for all five classifiers constructed in the combined model are illustrated in Figure 6c. The ORs and 95% CIs from the logistic regression models are summarized in Table 5.

Based on the METRICS assessment, the study achieved a score of 80.3%, classifying it as "excellent quality" (80 ≤ score ≤ 100%) (Appendix 1).

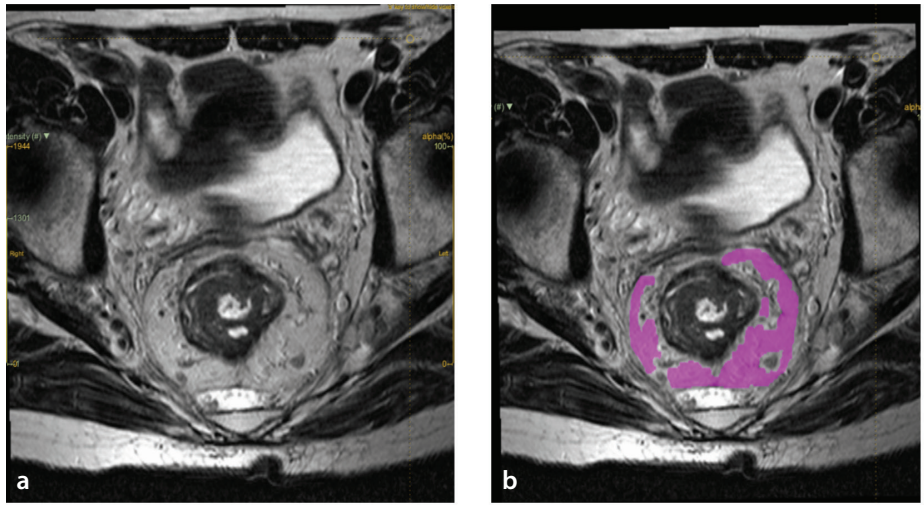



**Figure 3.** Mesorectum segmentation T2 weighted imaging **(a, b).**

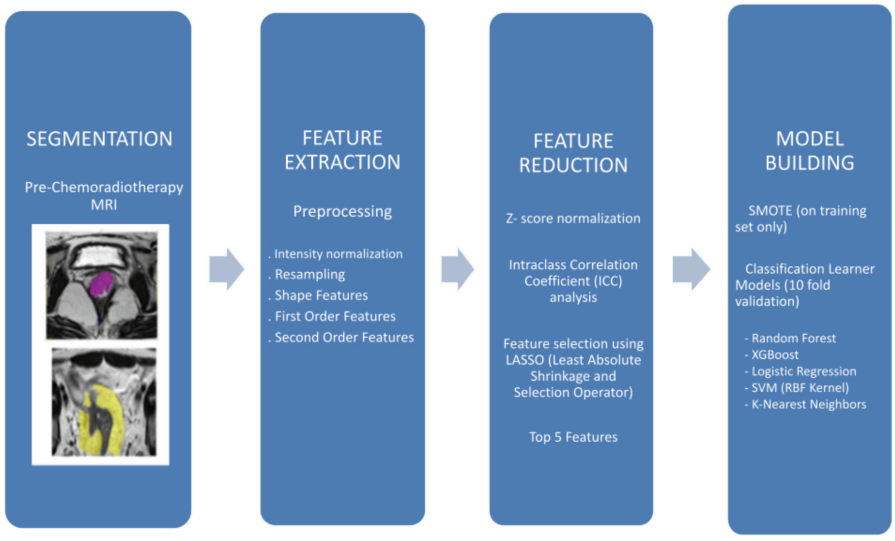



**Figure 4.** Pipeline for radiomic feature extraction and predictive model development in magnetic resonance imaging. MRI, magnetic resonance imaging.

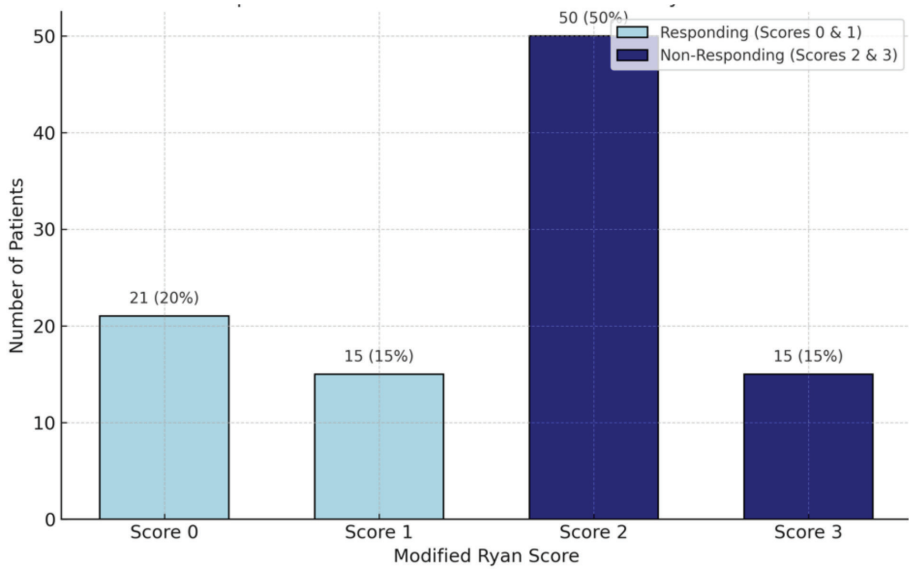



**Figure 5.** Distribution of patient responses to treatment based on modified Ryan score.

**Table 2.** Comparison of radiological and clinical parameters between responders and non-responders in rectal cancer treatment

| | Responding (modified Ryan score 0–1) n (%) | Non-responding (modified Ryan score 2–3) n (%) | *P* value |
|---|---|---|---|
| Sex | | | 0.137 |
| Male | 20 (29.9%) | 47 (70.1%) | |
| Female | 16 (47.1%) | 18 (52.9%) | |
| Pre-CRT T stage (MRI) | | | **0.201** |
| T2 | 8 (50%) | 8 (50%) | |
| T3b | 16 (38.1%) | 26 (61.9%) | |
| T3c | 9 (42.9%) | 12 (57.1%) | |
| T3d | 1 (8.33%) | 11 (91.7%) | |
| T4a | 1 (16.7%) | 5 (83.3%) | |
| T4b | 1 (25%) | 3 (75%) | |
| Distance of extramural extension | | | 0.167 |
| ≤5 mm | 22 (43.1%) | 29 (56.9%) | |
| >5 mm | 14 (28%) | 36 (72%) | |
| Distance to the mesorectal fascia | | | 0.316 |
| 0 mm | 9 (25.7%) | 26 (74.3%) | |
| 1–2 mm | 9 (40.9%) | 13 (59.1%) | |
| ≥3 mm | 18 (40.9%) | 26 (59.1%) | |

CRT, chemoradiotherapy; MRI, magnetic resonance imaging.

## Discussion

In this study, we constructed a series of machine-learning models to predict tumor response to neoadjuvant therapy by analyzing radiomic features extracted from the tumor, mesorectum, and their combination. The AUC values for the three segmentation approaches were as follows: 0.781 for the tumor-only model (random forest), 0.726 for the mesorectum-only model (logistic regression), and 0.837 for the combined model (logistic regression). This finding highlights the complementary value of the mesorectal compartment in radiomics modeling and its contribution to improving the performance of prediction models in LARC.

Personalized treatment protocols have become a prominent feature of clinical practice to minimize side effects, increase the frequency of organ-sparing surgery, and improve the clinical complete response rate in LARCs.[6,16,17] The prediction of CRT response has emerged as a valuable marker for guiding the development of personalized therapies. The potential of radiomics for predicting the response to LARC treatment has been the subject of numerous studies. In the majority of studies, radiomics models of tumor tissue were constructed from MRI obtained before and/or after CRT.[18-20]

Mesorectal adipocytes not only act as an anatomical barrier surrounding the tumor but also actively contribute to the tumor microenvironment. The dynamic crosstalk between tumor cells and adipocytes induces profound morphological and functional changes in adipocytes, altering the secretion of adipokines (e.g., leptin, adiponectin) and angiogenic factors (e.g., VEGF, IGF-1). These changes promote key biological processes, such as tumor progression, angiogenesis, and therapeutic and radiotherapy resistance.[21-25] Furthermore, molecular profile alterations within peritumoral adipocytes can lead to subtle MRI signal changes that may not be detectable through conventional visual assessment. This underscores the importance of radiomic analyses, which facilitate the extraction of hidden imaging data and provide a quantitative evaluation of subtle changes that would otherwise remain undetected.[26,27]

In our study, we aimed to detect changes at the cellular level by performing radiomics measurements from morphologically non-pathologic mesorectum, which did not include tumor deposits, extramural tumor extension, or lymph nodes. The mesorectum contains adipocytes whose molecular profiles are altered in response to tumor processes, as well as venous and lymphatic structures that facilitate the drainage of waste products from both the tumor and surrounding tissues. Recent literature suggests that this microenvironment harbors prognostic information comparable with the tumor itself.[7,8]

Relatively few MRI-based studies have incorporated mesorectal features into radiomics modeling. Shaish et al.[8] reported an AUC of 0.800 using both tumor and mesorectal features from pretreatment MRI in 132 patients. Jayaprakasam et al.[7] evaluated mesorectal features alone in a larger cohort of 236 patients and achieved an AUC of 0.890 for predicting pathological complete response. Kaval et al.[28] assessed tumor-only and combined models in 93 patients, reporting AUCs of 0.850 and 0.830, respectively. Although tumor segmentation yielded the highest AUC in that study, the addition of mesorectal features led to improved sensitivity (90%) and overall accuracy (79%), further supporting the complementary role of the mesorectum in response assessment.

Although variations in study design, sample size, and endpoints may account for differences in performance, our results remain consistent with the existing literature, highlighting the importance of including mesorectal features for more accurate prediction of treatment response.

Compared with our models, which relied solely on MRI-based tumor and mesorectal features, the computed tomography-based radiomics approach developed by Wang et al.[29] demonstrated lower predictive performance, with an AUC of 0.68 for identifying high-risk neoadjuvant rectal (NAR) scores. Notably, their analysis found mesorectal features to be more predictive than intratumoral features. In contrast, our results indicated that tumor-derived features contributed more strongly to treatment response prediction, suggesting that differences in imaging modality, feature representation, and endpoint definition (Ryan score vs. NAR) may explain the discrepancy. These findings support the utility of MRI-based radiomics as a more accurate and robust non-invasive tool for individualized response prediction.

**Table 3.** Diagnostic performance of tumor and mesorectum radiomics features for treatment response prediction

| T2WI-tumor radiomics features (treatment response) | Sensitivity | Specificity | PPV | NPV | AUC | *P* value |
|---|---|---|---|---|---|---|
| GLCM_Autocorrelation | 0.43 | 0.43 | 0.43 | 0.43 | 0.64 | 0.01 |
| INTENSITY-HISTOGRAM_IntensityHistogramMedian | 0.58 | 0.4 | 0.49 | 0.49 | 0.62 | 0.02 |
| INTENSITY-HISTOGRAM_IntensityHistogram50$^{th}$Percentile | 0.58 | 0.4 | 0.49 | 0.49 | 0.62 | 0.02 |
| INTENSITY-HISTOGRAM_IntensityHistogramSkewness | 0.53 | 0.53 | 0.53 | 0.53 | 0.6 | 0.02 |
| GLSZM_ZoneSizeNonUniformity | 0.58 | 0.58 | 0.58 | 0.58 | 0.6 | 0.02 |
| GLCM_SumAverage | 0.4 | 0.4 | 0.4 | 0.4 | 0.67 | 0.02 |
| GLCM_JointAverage | 0.4 | 0.4 | 0.4 | 0.4 | 0.67 | 0.02 |
| GLCM_ClusterShade | 0.55 | 0.55 | 0.55 | 0.55 | 0.6 | 0.03 |
| INTENSITY-HISTOGRAM_IntensityHistogram75$^{th}$Percentile | 0.48 | 0.48 | 0.48 | 0.48 | 0.59 | 0.04 |
| DWI-tumor radiomics features (treatment response) | | | | | | |
| INTENSITY-HISTOGRAM_MinimumHistogramGradient(IBSI:VQB3)[Intensity] | 0.42 | 0.38 | 0.4 | 0.4 | 0.64 | <0.01 |
| GLRLM_LongRunsEmphasis(IBSI:W4KF) | 0.6 | 0.6 | 0.6 | 0.6 | 0.63 | <0.01 |
| GLRLM_RunPercentage(IBSI:9ZK5) | 0.42 | 0.42 | 0.42 | 0.42 | 0.62 | <0.01 |
| GLSZM_ZonePercentage(IBSI:P30P) | 0.43 | 0.43 | 0.43 | 0.43 | 0.62 | <0.01 |
| GLRLM_ShortRunsEmphasis(IBSI:22OV) | 0.42 | 0.42 | 0.42 | 0.42 | 0.62 | <0.01 |
| GLSZM_LargeZoneEmphasis(IBSI:48P8) | 0.57 | 0.57 | 0.57 | 0.57 | 0.62 | 0.01 |
| NGTDM_Busyness(IBSI:NQ30) | 0.55 | 0.55 | 0.55 | 0.55 | 0.6 | 0.01 |
| GLRLM_GreyLevelNonUniformity(IBSI:R5YN) | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.02 |
| GLSZM_ZoneSizeVariance(IBSI:3NSA) | 0.58 | 0.58 | 0.58 | 0.58 | 0.62 | 0.02 |
| GLSZM_GreyLevelNonUniformity(IBSI:JNSA) | 0.57 | 0.57 | 0.57 | 0.57 | 0.58 | 0.02 |
| GLSZM_SmallZoneEmphasis(IBSI:5QRC) | 0.43 | 0.43 | 0.43 | 0.43 | 0.61 | 0.02 |
| GLSZM_NormalisedZoneSizeNonUniformity(IBSI:VB3A) | 0.43 | 0.43 | 0.43 | 0.43 | 0.61 | 0.03 |
| GLCM_InverseDifferenceMoment(IBSI:WF0Z) | 0.53 | 0.53 | 0.53 | 0.53 | 0.59 | 0.03 |
| MORPHOLOGICAL_voxelsCounting(IBSI:No)[#vx] | 0.55 | 0.55 | 0.55 | 0.55 | 0.56 | 0.04 |
| GLRLM_RunLengthNonUniformity(IBSI:W92Y) | 0.55 | 0.55 | 0.55 | 0.55 | 0.56 | 0.04 |
| INTENSITY-HISTOGRAM_MaximumHistogramGradient(IBSI:12CE)[Intensity] | 0.58 | 0.53 | 0.56 | 0.56 | 0.57 | 0.05 |
| T2WI-mesorectum radiomics features (treatment response) | | | | | | |
| T2M-MORPHOLOGICAL_SurfaceArea(IBSI:C0JK)[mm$^2$] | 0.63 | 0.63 | 0.63 | 0.63 | 0.69 | <0.01 |
| T2M-GLSZM_ZoneSizeNonUniformity(IBSI:4JP3) | 0.65 | 0.65 | 0.65 | 0.65 | 0.72 | <0.01 |
| MORPHOLOGICAL_Maximum3DDiameter(IBSI:L0JK)[mm] | 0.65 | 0.65 | 0.65 | 0.65 | 0.68 | <0.01 |
| INTENSITY-HISTOGRAM_RootMeanSquare(IBSI:No)[Intensity] | 0.33 | 0.33 | 0.33 | 0.33 | 0.72 | <0.01 |
| MORPHOLOGICAL_Sphericity(IBSI:QCFX)[] | 0.4 | 0.4 | 0.4 | 0.4 | 0.66 | <0.01 |
| MORPHOLOGICAL_SphereDiameter(IBSI:No)[mm] | 0.62 | 0.62 | 0.62 | 0.62 | 0.66 | <0.01 |
| MORPHOLOGICAL_Compactness1(IBSI:SKGS)[] | 0.4 | 0.4 | 0.4 | 0.4 | 0.66 | <0.01 |
| MORPHOLOGICAL_Asphericity(IBSI:25C7)[] | 0.58 | 0.58 | 0.58 | 0.58 | 0.65 | <0.01 |
| MORPHOLOGICAL_SphericalDisproportion(IBSI:KRCK)[] | 0.58 | 0.58 | 0.58 | 0.58 | 0.65 | <0.01 |
| MORPHOLOGICAL_Compactness2(IBSI:BQWJ)[] | 0.4 | 0.4 | 0.4 | 0.4 | 0.66 | <0.01 |
| MORPHOLOGICAL_Compacity(IBSI:No)[] | 0.58 | 0.58 | 0.58 | 0.58 | 0.65 | <0.01 |
| NGTDM_Strength(IBSI:1X9X) | 0.38 | 0.38 | 0.38 | 0.38 | 0.64 | <0.01 |
| NGTDM_Coarseness(IBSI:QCDE) | 0.38 | 0.38 | 0.38 | 0.38 | 0.63 | <0.01 |
| INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel(IBSI:RHQZ)[Intensity] | 0.42 | 0.38 | 0.5 | 0.39 | 0.62 | <0.01 |
| GLSZM_GreyLevelNonUniformity(IBSI:JNSA) | 0.62 | 0.62 | 0.62 | 0.62 | 0.67 | <0.01 |
| INTENSITY-HISTOGRAM_IntensityHistogramMinimumGreyLevel(IBSI:1PR8)[Intensity] | 0.49 | 0.44 | 0.5 | 0.42 | 0.57 | 0.01 |
| INTENSITY-BASED_IntensityBasedCoefficientOfVariation(IBSI:7TET)[] | 0.47 | 0.47 | 0.47 | 0.47 | 0.58 | 0.03 |
| NGTDM_Complexity(IBSI:HDEZ) | 0.62 | 0.62 | 0.62 | 0.62 | 0.64 | 0.04 |
| GLRLM_RunLengthNonUniformity(IBSI:W92Y) | 0.62 | 0.62 | 0.62 | 0.62 | 0.67 | 0.04 |
| MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5)[mm] | 0.42 | 0.42 | 0.42 | 0.42 | 0.59 | 0.05 |

T2WI, T2-weighted imaging; DWI, diffusion weighted imaging; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; NGTDM, neighborhood gray-tone difference matrix; GLSZM, gray-level size zone matrix; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

**Table 4.** Summary of classification performance metrics across tumor-only, mesorectum-only, and combined radiomic models

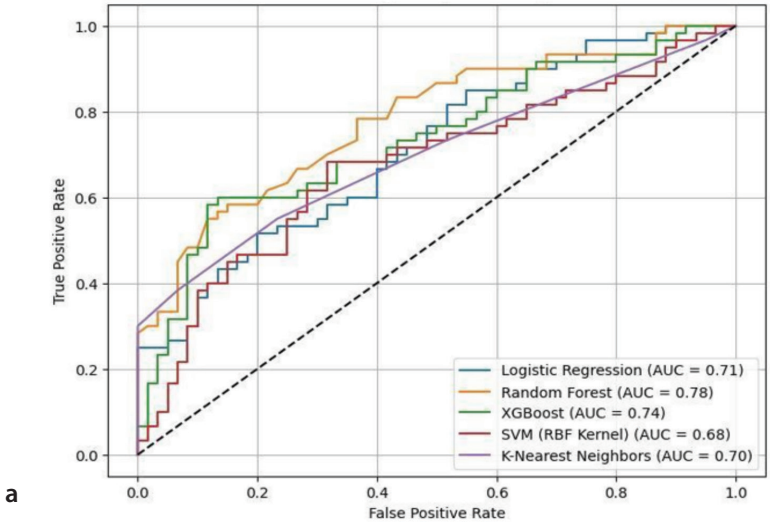
| Tumor-only model | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| Logistic regression | 0.617 | 0.621 | 0.600 | 0.610 | 0.714 |
| **Random forest** | **0.692** | **0.702** | **0.667** | **0.684** | **0.781** |
| XGBoost | 0.658 | 0.661 | 0.650 | 0.655 | 0.737 |
| SVM (RBF kernel) | 0.667 | 0.685 | 0.617 | 0.649 | 0.676 |
| K-nearest neighbors | 0.658 | 0.702 | 0.550 | 0.617 | 0.700 |
| **Mesorectum-only model** | | | | | |
| **Logistic regression** | **0.667** | **0.661** | **0.683** | **0.672** | **0.726** |
| Random forest | 0.575 | 0.579 | 0.550 | 0.564 | 0.700 |
| XGBoost | 0.658 | 0.661 | 0.650 | 0.655 | 0.708 |
| SVM (RBF kernel) | 0.658 | 0.686 | 0.583 | 0.631 | 0.711 |
| K-nearest neighbors | 0.575 | 0.574 | 0.583 | 0.579 | 0.661 |
| **Combined model** | | | | | |
| **Logistic regression** | **0.810** | **0.821** | **0.814** | **0.819** | **0.837** |
| Random forest | 0.750 | 0.768 | 0.717 | 0.741 | 0.816 |
| XGBoost | 0.683 | 0.704 | 0.633 | 0.667 | 0.789 |
| SVM (RBF kernel) | 0.767 | 0.767 | 0.767 | 0.767 | 0.811 |
| K-nearest neighbors | 0.717 | 0.724 | 0.700 | 0.712 | 0.754 |

AUC, Area under the curve; XGBoost, extreme gradient boosting; SVM (RBF kernel), support vector machine with radial basis function kernel.

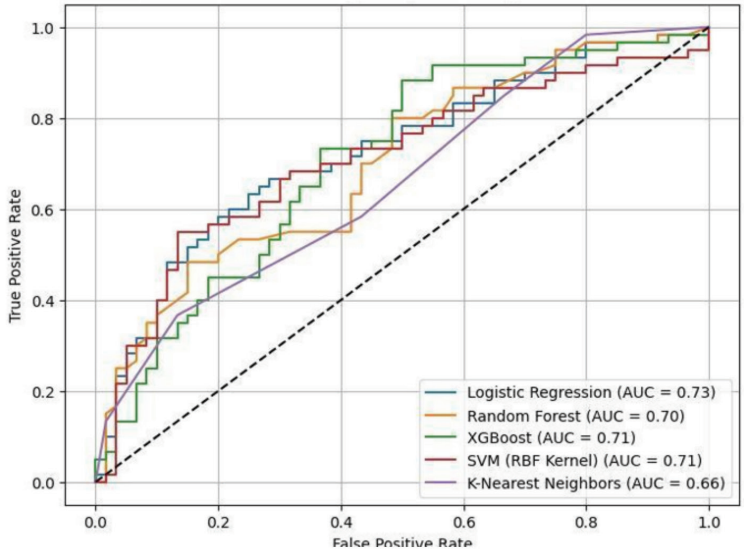
**Table 5.** Odds ratios and 95% confidence intervals of the logistic regression models constructed for the tumor-only, mesorectum-only, and combined model

| Tumor-only model | | | | |
|---|---|---|---|---|
| **Radiomic features** | *P* value | Odds ratio | 95% CI lower | 95% CI upper |
| T2T-GLCM_Autocorrelation (IBSI:QWB0) | **0.010** | 7.989 | 1.645 | 38.800 |
| D-GLCM_ClusterProminence (IBSI:AE86) | 0.222 | 1.351 | 0.833 | 2.190 |
| D-INTENSITY-HISTOGRAM_MinimumHistogramGradient (IBSI:VQB3) [Intensity] | **<0.001** | 2.968 | 1.571 | 5.610 |
| T2T-INTENSITY-HISTOGRAM_IntensityHistogramMode (IBSI:AMMC) [Intensity] | 0.587 | 0.855 | 0.485 | 1.510 |
| D-INTENSITY-HISTOGRAM_IntensityHistogramCoefficientOfVariation (IBSI:CWYJ) [Intensity] | **0.035** | 8.861 | 1.160 | 67.670 |
| **Mesorectum-only model** | | | | |
| **Radiomic features** | | | | |
| T2M-GLSZM_ZoneSizeNonUniformity (IBSI:4JP3) | **0.037** | 0.602 | 0.373 | 0.971 |
| T2M-MORPHOLOGICAL_SphericalDisproportion (IBSI:KRCK) | 0.515 | 0.491 | 0.058 | 4.187 |
| T2M-MORPHOLOGICAL_SurfaceArea (IBSI:C0JK) [$mm^2$] | 0.345 | 0.750 | 0.414 | 1.361 |
| T2M-INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel (IBSI:RHQZ) [Intensity] | 0.052 | 1.592 | 0.997 | 2.544 |
| T2M-MORPHOLOGICAL_Sphericity (IBSI:QCFX) | 0.847 | 0.811 | 0.096 | 6.834 |
| **Combined model** | | | | |
| **Radiomic features** | | | | |
| T2T-GLCM_Autocorrelation (IBSI:QWB0) | **0.014** | 9.819 | 1.577 | 61.139 |
| T2M-INTENSITY-HISTOGRAM_MinimumHistogramGradientGreyLevel (IBSI:RHQZ) [Intensity] | **0.017** | 2.069 | 1.142 | 3.747 |
| T2M-GLSZM_ZoneSizeNonUniformity (IBSI:4JP3) | **0.002** | 0.438 | 0.258 | 0.746 |
| T2M-NGTDM_Complexity (IBSI:HDEZ) | **0.009** | 0.461 | 0.257 | 0.827 |
| D-GLCM_InverseDifferenceMoment (IBSI:WF0Z) | **<0.001** | 0.313 | 0.163 | 0.602 |

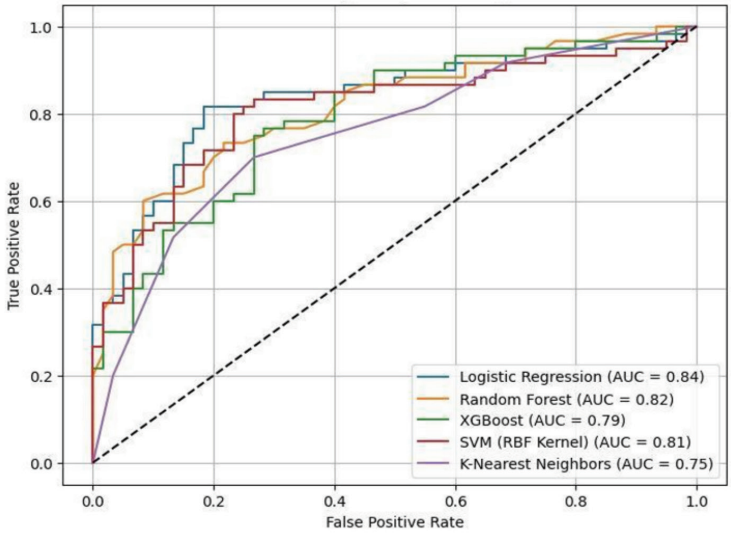
CI, confidence Interval; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; NGTDM, Neighborhood Gray-Tone difference matrix; GLSZM, gray-level size zone matrix.




a




b




c

**Figure 6.** Receiver operating characteristic curves of the five classifiers (10-fold cross-validation) **(a)** tumor-only model, **(b)** mesorectum-only model, **(c)** combined model.

Multiple models were developed to predict treatment response using radiomic features extracted from the tumor, mesorectal, and combined regions. Although classification performance varied across models, logistic regression, by providing ORs, enabled clinically meaningful interpretation across the three datasets.[30,31] The tumor-only logistic regression model was primarily driven by texture-and intensity-based features, reflecting intratumoral heterogeneity. In contrast, the mesorectum-only model included several morphological descriptors, though only a zone-based texture feature showed statistical significance. These findings indicate that mesorectal adipose tissue may reflect structural or spatial texture changes relevant to treatment response, even in the absence of pronounced intensity heterogeneity.

The combined logistic regression model demonstrated a more balanced and robust predictive performance than the individual models. All five features selected via LASSO contributed significantly to the model's performance. Notably, features indicative of tissue homogeneity, such as Gy-level autocorrelation and smooth intensity gradient transitions, were associated with favorable response, whereas heterogeneity-related features, including zone size non-uniformity and local textural complexity, were linked to poor response. These results support the hypothesis that radiomic heterogeneity reflects underlying biological disorganization or resistance, whereas homogeneity may indicate a more organized and treatment-sensitive tumor architecture. This interpretation aligns with existing literature. Tumor heterogeneity has been widely associated with treatment resistance and poor prognosis.[32,33]

According to our most predictive model (tumor + mesorectum), second-order radiomic features–particularly those derived from the Gy-level co-occurrence matrix (GLCM), Gy-level size zone matrix (GLSZM), and neighboring Gy tone difference matrix (NGTDM)–demonstrated the highest predictive value. These matrices assess tissue heterogeneity at different levels: GLCM captures local structural variability; GLSZM quantifies the size and uniformity of homogeneous zones; and NGTDM evaluates visual texture by comparing a central voxel to its neighbors. Supporting our findings, Shaish et al.[8] reported similar prognostic relevance of these features in evaluating response to neoadjuvant therapy. Additionally, Mazzei et al.[33] showed that changes in GLCM features before and after treatment in patients with gastric cancer correlated with response, em-

phasizing the potential of these features as imaging biomarkers.

In radiomics-based machine learning, model performance is strongly shaped by factors such as limited sample size, high feature dimensionality, multicollinearity, and class imbalance.[34] Our study reflects these challenges, as we analyzed 101 patients with rectal cancer using 17,978 radiomic features extracted from pretreatment MRI images. To mitigate the risk of overfitting and improve model generalizability, we applied LASSO-based feature selection and SMOTE-based class balancing. Among the tested algorithms, logistic regression with LASSO stood out by consistently providing robust and interpretable predictions, especially in the combined and mesorectum-only models, with strong AUC and F1 scores.[35] Ensemble methods, such as random forest and XGBoost, also performed well, reflecting their ability to model complex, non-linear relationships in high-dimensional data.[36,37] Notably, in the tumor-only model, random forest yielded the highest predictive performance, possibly due to its inherent ensemble structure, which reduces variance and captures localized, non-linear dependencies within tumor-derived radiomic features.

Conversely, distance-based algorithms, such as KNN and SVM, showed moderate but generally lower performance than other models. Their results may reflect methodological limitations, such as sensitivity to feature scaling, reduced robustness to noise, and a higher risk of overfitting in high-dimensional, low-sample-size contexts–issues that often require careful tuning and preprocessing to overcome.[38,39] Nevertheless, SVM yielded relatively strong performance in the combined model, suggesting that, when provided with sufficiently rich and diverse input features, distance-based algorithms may perform competitively despite their known limitations.

This study has several limitations, including its retrospective nature, single-center origin, and limited sample size. Although external validation was not feasible due to the small cohort, we employed 10-fold cross-validation to support model robustness. A hold-out set was not conducted due to the limited number of cases, and dividing the data into training and test sets would have resulted in information loss. Furthermore, one mucinous tumor was not excluded from our patient cohort. Images with different pixel and FOV sizes were registered in the picture archiving and communication system in our study. This limitation was overcome by utilizing techniques such as pixel size readjustment, normalization, and Gy-level discretization.

In conclusion, our study showed that combining radiomic features from both the tumor and mesorectum improves the prediction of response to neoadjuvant CRT in LARC. The combined model outperformed tumor-only and mesorectum-only models, achieving the highest AUC (0.837) and superior overall classification metrics. Incorporating mesorectal features resulted in a more balanced and more accurate model, highlighting the complementary role of the mesorectum in individualized response prediction. To enable the routine clinical application of these findings, further validation through large-scale, multicenter prospective studies is warranted.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

**Appendix 1:** https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/e22f6fa3-9eb3-4d20-9504-efa7b951fff7.pdf

## References


1. Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of rectal cancer: tumor staging, imaging techniques, and management. *Radiographics*. 2019;39:367-387. [Crossref]
2. Çelik H, Barlık F, Sökmen S, et al. Diagnostic performance of magnetic resonance imaging in preoperative local staging of rectal cancer after neoadjuvant chemoradiotherapy. *Diagn Interv Radiol*. 2023;29:219-227. [Crossref]
3. Chari RS, Tyler DS, Anscher MS, et al. Preoperative radiation and chemotherapy in the treatment of adenocarcinoma of the rectum. *Ann Surg*. 1995;221(6):786-787. [Crossref]
4. Roh MS, Colangelo LH, O'Connell MJ, et al. Preoperative multimodality therapy improves disease-free survival in patients with carcinoma of the rectum: NSABP R-03. *J Clin Oncol*. 2009;27(31):5124-5130. [Crossref]
5. Bonnen M, Crane C, Vauthey JN, et al. Long-term results using local excision after preoperative chemoradiation among selected T3 rectal cancer patients. *Int J Radiat Oncol Biol Phys*. 2004;60(4):1098-1105. [Crossref]
6. Burbach JP, den Harder AM, Intven M, van Vulpen M, Verkooijen HM, Reerink O. Impact of radiotherapy boost on pathological complete response in patients with locally advanced rectal cancer: a systematic review and meta-analysis. *Radiother Oncol*. 2014;113(1):1-9. [Crossref]
7. Jayaprakasam VS, Paroder V, Gibbs P, et al. MRI radiomics features of mesorectal fat can predict response to neoadjuvant chemoradiation therapy and tumor recurrence in patients with locally advanced rectal cancer. *Eur Radiol*. 2022;32(2):971-980. [Crossref]
8. Shaish H, Aukerman A, Vanguri R, et al. Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. *Eur Radiol*. 2020;30(11):6263-6273. [Crossref]
9. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res*. 2017;19(1):57. [Crossref]
10. Akinci D'Antonoli T, Farchione A, Lenkowicz J, et al. CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk. *Acad Radiol*. 2020;27:497-507. [Crossref]
11. Tiang T, Sidhu A, Williams D, Bui A. Impact of neoadjuvant chemotherapy interval on tumour regression grading for rectal cancer. *Int J Surg Res Pract*. 2022;9:137. [Crossref]
12. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging*. 2004;22(1):81-91. [Crossref]
13. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40-79. [Crossref]
14. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. London, *Springer*; 2013:176-177. [Crossref]
15. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. *Insights Imaging*. 2024;15(1):8. [Crossref]
16. Domingo-Boluda C, Dualde D, Taberner-Bonastre T, et al. Impact of dose-escalated chemoradiation on pathological complete response in patients with locally advanced rectal cancer. *Cancers*. 2024;16(18):3170. [Crossref]
17. Hoendervangers S, Burbach JPM, Lacle MM, et al. Pathological complete response following different neoadjuvant treatment strategies for locally advanced rectal cancer: a systematic

review and meta-analysis. *Ann Surg Oncol*. 2020;27:4319-4336. [Crossref]

18. Delli Pizzi A, Chiarelli AM, Chiacchiaretta P, et al. MRI-based clinical-radiomics model predicts tumor response before treatment in locally advanced rectal cancer. *Sci Rep*. 2021;11:5379. [Crossref]
19. Horvat N, Veeraraghavan H, Khan M, et al. MR Imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology*. 2018;287(3):833-843. [Crossref]
20. Miranda J, Tan GXV, Fernandes MC, et al. Rectal MRI radiomics for predicting pathological complete response: where we are. *Clin Imaging*. 2022;82:141-149. [Crossref]
21. Amor S, Iglesias-de la Cruz MC, Ferrero E, et al. Peritumoral adipose tissue as a source of inflammatory and angiogenic factors in colorectal cancer. *Int J Colorectal Dis*. 2016;31(2):365-375. [Crossref]
22. Neto NIP, Murari ASP, Oyama LM, et al. Peritumoural adipose tissue pro-inflammatory cytokines are associated with tumoural growth factors in cancer cachexia patients. *J Cachexia Sarcopenia Muscle*. 2018;9(6):1101-1108. [Crossref]
23. Cao Y. Adipocyte and lipid metabolism in cancer drug resistance. *J Clin Invest*. 2019;129(8):3006-3017. [Crossref]
24. Duong MN, Geneste A, Fallone F, Li X, Dumontet C, Muller C. The fat and the bad: mature adipocytes, key actors in tumor progression and resistance. *Oncotarget*. 2017;8(34):57622-57641. [Crossref]
25. Zoico E, Rizzatti V, Darra E, et al. Morphological and functional changes in the peritumoral adipose tissue of colorectal cancer patients. *Obesity (Silver Spring)*. 2017;25(Suppl 2):87-94. [Crossref]
26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [Crossref]
27. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. [Crossref]
28. Kaval G, Dagoglu Kartal MG, Azamat S, et al. Evaluating complete response prediction rates in locally advanced rectal cancer with different radiomics segmentation approaches. *Pathol Oncol Res*. 2024. [Crossref]
29. Wang F, Tan BF, Poh SS, et al. Predicting outcomes for locally advanced rectal cancer treated with neoadjuvant chemoradiation with CT-based radiomics. *Sci Rep*. 2022;12:6167. [Crossref]
30. Hua Y, Stead TS, George A, Ganti L. Clinical risk prediction with logistic regression: best practices, validation techniques, and applications in medical research. *Acad Med Surg*. 2025;3(1). [Crossref]
31. Petresc B, Lebovici A, Caraiani C, Feier DS, Graur F, Buruian MM. Pre-treatment T2-WI based radiomics features for prediction of locally advanced rectal cancer non-response to neoadjuvant chemoradiotherapy: a preliminary study. *Cancers*. 2020;12(7):1894. [Crossref]
32. Fu YC, Liang SB, Luo M, et al. Intratumoral heterogeneity and drug resistance in cancer. Cancer Cell Int. 2025;25:103. doi:10.1186/s12935-025-03734-w. [Crossref]
33. Mazzei MA, Di Giacomo L, Bagnacci G, et al. Delta-radiomics and response to neoadjuvant treatment in locally advanced gastric cancer-a multicenter study of GIRCG (Italian Research Group for Gastric Cancer). *Quant Imaging Med Surg*. 2021;11(6):2055-2066. [Crossref]
34. Koçak B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. Diagn Interv Radiol. 2022 Sep;28(5):450-462. [Crossref]
35. Zhu Y, Wei Y, Chen Z, et al. Different radiomics annotation methods comparison in rectal cancer characterisation and prognosis prediction: a two-centre study. Insights Imaging. 2024 Aug 26;15(1):211. [Crossref]
36. Zhou ZH. Ensemble Methods: Foundations and Algorithms. *CRC Press*; 2012. [Crossref]
37. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett*. 2016;382(1):110-117. [Crossref]
38. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11(70):2079-2107. [Crossref]
39. Radovanović M, Nanopoulos A, Ivanović M. Hubs in space: popular nearest neighbors in high-dimensional data. *J Mach Learn Res*. 2010;11:2487-2531. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253428

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE



# Comparison of the diagnostic performance of the artificial intelligence-based TIRADS algorithm with established classification systems for thyroid nodules

Abdilkadir Bozkuş[1]
Yeliz Başar[1]
Koray Güven[2]

[1]Acıbadem Maslak Hospital, Clinic of Radiology, İstanbul, Türkiye

[2]Acıbadem Mehmet Ali Aydınlar University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

**PURPOSE**

This study aimed to evaluate and compare the diagnostic performance of various Thyroid Imaging Reporting and Data Systems (TIRADS), with a particular focus on the artificial intelligence-based TIRADS (AI-TIRADS), in characterizing thyroid nodules.

**METHODS**

In this retrospective study conducted between April 2016 and May 2022, 1,322 thyroid nodules from 1,139 patients with confirmed cytopathological diagnoses were included. Each nodule was assessed using TIRADS classifications defined by the American College of Radiology (ACR-TIRADS), the American Thyroid Association (ATA-TIRADS), the European Thyroid Association (EU-TIRADS), the Korean Thyroid Association (K-TIRADS), and the AI-TIRADS. Three radiologists independently evaluated the ultrasound (US) characteristics of the nodules using all classification systems. Diagnostic performance was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value, and comparisons were made using the McNemar test.

**RESULTS**

Among the nodules, 846 (64%) were benign, 299 (22.6%) were of intermediate risk, and 147 (11.1%) were malignant. The AI-TIRADS demonstrated a PPV of 21.2% and a specificity of 53.6%, outperforming the other systems in specificity without compromising sensitivity. The specificities of the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS were 44.6%, 39.3%, 40.1%, and 40.1%, respectively (all pairwise comparisons with the AI-TIRADS: $P < 0.001$). The PPVs for the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS were 18.5%, 17.9%, 17.9%, and 17.4%, respectively (all pairwise comparisons with the AI-TIRADS, excluding the ACR-TIRADS: $P < 0.05$).

**CONCLUSION**

The AI-TIRADS shows promise in improving diagnostic specificity and reducing unnecessary biopsies in thyroid nodule assessment while maintaining high sensitivity. The findings suggest that the AI-TIRADS may enhance risk stratification, leading to better patient management. Additionally, the study found that the presence of multiple suspicious US features markedly increases the risk of malignancy, whereas isolated features do not substantially elevate the risk.

**CLINICAL SIGNIFICANCE**

The AI-TIRADS can enhance thyroid nodule risk stratification by improving diagnostic specificity and reducing unnecessary biopsies, potentially leading to more efficient patient management and better utilization of healthcare resources.

**KEYWORDS**

Thyroid Imaging Reporting and Data Systems, thyroid cancer, artificial intelligence, thyroid nodules, ultrasonography

**Handling editor:** Tuğba Akıncı D'Antonoli

**Corresponding author:** Koray Güven

**E-mail:** koray.guven@acibadem.com

Received 25 April 2025; revision requested 02 June 2025; accepted 27 July 2025.

Epub: 01.09.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253428

You may cite this article as: Bozkuş A, Başar Y, Güven K. Comparison of the diagnostic performance of the artificial intelligence-based TIRADS algorithm with established classification systems for thyroid nodules. *Diagn Interv Radiol*. 2026;32(4):376-381.

Thyroid nodules are a widespread clinical concern, detected in a substantial proportion of the general population through high-resolution ultrasound (US) examinations. The frequency of thyroid nodules detected in people during routine US screenings varies between 20% and 68%. This range is influenced by factors such as demographic traits and the precision of the imaging technology employed.[1,2] Although the majority of thyroid nodules are benign, approximately 5%–15% exhibit malignant potential, necessitating accurate risk stratification to guide clinical management.[3] The most important challenge in the evaluation of thyroid nodules is the distinction between benign and malignant lesions to determine the appropriate necessity for invasive investigations such as fine needle aspiration biopsy (FNAB).[4,5]

FNAB is a cornerstone diagnostic tool for thyroid nodules with high sensitivity and specificity in differentiating benign from malignant lesions. However, it is not without drawbacks, including being invasive, having the potential for nondiagnostic or indeterminate results, and causing patient discomfort.[6,7] Consequently, there is a pressing need for non-invasive, reliable methods to enhance the accuracy of thyroid nodule classification, thereby reducing unnecessary biopsies and associated healthcare costs.[8]

US remains the primary imaging modality for the evaluation of thyroid nodules due to its accessibility, lack of ionizing radiation, and ability to provide detailed anatomical and structural information.[9] These classification systems risk-stratify the probability of malignancy based on specific sonographic features such as echogenicity, composition, shape, margin, and the presence of calcifications.[10,11]

**Main points**

- The artificial intelligence-based Thyroid Imaging Reporting and Data System (AI-TIRADS) exhibited enhanced specificity in comparison with other recognized systems such as the American College of Radiology (ACR)-TIRADS, the American Thyroid Association TIRADS, the European TIRADS, and the Korean TIRADS, concurrently preserving elevated sensitivity.
- The AI-TIRADS notably reduced the incidence of avoidable fine needle aspiration biopsies to 40.8%, compared with 49% for the ACR-TIRADS and exceeding 51% for the other systems.
- The investigation validated that the existence of multiple concerning ultrasound characteristics is considerably correlated with an elevated risk of malignancy, whereas singular features have diminished predictive value, thereby underscoring the necessity for thorough risk evaluation.

Multiple Thyroid Imaging Reporting and Data Systems (TIRADS) have been proposed by different organizations, including the American College of Radiology (ACR-TIRADS),[12] the American Thyroid Association (ATA-TIRADS),[4] the European Thyroid Association (EU-TIRADS),[13] and the Korean Thyroid Association (K-TIRADS).[14] Despite their widespread adoption, variability exists among these systems in criteria weighting, risk categorization, and recommended management strategies, leading to inconsistencies in clinical practice.[15] This lack of consensus highlights the need for further refinement and the potential integration of advanced technologies to enhance diagnostic performance.

Artificial intelligence (AI) and machine learning have been transformative in medical imaging over the past decade, with the promise of improving traditional diagnostic practices.[16] The AI-based TIRADS (AI-TIRADS) leverage computational algorithms to analyze complex patterns in US data, aiming to improve the accuracy and consistency of thyroid nodule classification.[17] Preliminary studies suggest that the AI-TIRADS may exhibit higher specificity and reduced rates of unnecessary FNABs compared with the conventional TIRADS, without compromising sensitivity.[18,19] These advancements enable better thyroid nodule assessment, where AI-assisted decision-making can enhance clinical outcomes and optimize resource utilization.

Moreover, the integration of AI into the TIRADS addresses critical issues such as interobserver variability and subjective interpretation inherent in manual US evaluations.[20] By providing objective, reproducible assessments of nodular characteristics, the AI-TIRADS can standardize risk stratification across different healthcare settings and practitioners.[21] This is particularly pertinent in regions with limited access to specialized radiologists, where AI-driven tools can support primary care providers in making informed decisions.[22]

Despite the promising potential of the AI-TIRADS, comprehensive evaluations comparing its diagnostic performance against established TIRADS are limited. Furthermore, the impact of AI integration on clinical workflows, patient anxiety, and overall healthcare costs warrants thorough investigation.[23] Addressing these gaps is essential to validate the efficacy of the AI-TIRADS and facilitate its widespread adoption in routine clinical practice.

This study aims to compare and evaluate the diagnostic effectiveness of the AI-TIRADS with other currently established classification systems, such as the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, in characterizing thyroid nodules. By analyzing a large cohort of patients with confirmed cytopathological diagnoses, this research seeks to determine the efficacy of the AI-TIRADS in improving specificity, reducing unnecessary FNABs, and enhancing overall diagnostic accuracy.

## Study sample

This retrospective investigation was conducted with the approval of the local ethics committee, which also waived the requirement for informed consent due to the use of de-identified medical records (REDACTED). Approved by the Acıbadem University and Acıbadem Healthcare Institutions Medical Research Ethics Committee (ATADEK) on June 16, 2023, with decision number 2023-10/360. The study included adult individuals who underwent thyroid US assessments at a single tertiary care institution between April 2016 and May 2022. Initially, 1,322 thyroid nodules from 1,139 patients with confirmed cytopathological diagnoses were identified via the hospital information system and the picture archiving and communication system. All images were reviewed in Digital Imaging and Communications in Medicine format. Patients lacking complete cytopathological information were excluded, resulting in a final cohort comprising 1,110 patients with 1,292 thyroid nodules. Further details are presented as a flowchart in Figure 1. The study adhered to the Standards for Reporting of Diagnostic Accuracy guidelines to ensure integrity and transparency in the reporting process.[24]

All thyroid nodules were evaluated and classified according to five different TIRADS: the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, the K-TIRADS, and the AI-TIRADS algorithm.[4,12-14,17]

Wildman-Tobriner et al.[17] optimized the ACR-TIRADS classification using AI. Three board-certified radiologists with more than 5 years of experience in thyroid imaging independently reviewed all nodules' sonographic features, including echogenicity, composition, shape, margin, and calcifications. Dis-

agreements were resolved by consensus to prevent variation in classification. Independent review for ground truth verification was not performed, as inter-rater reliability testing was beyond the scope of the study.

Cytopathological diagnoses were categorized based on the Bethesda System for Reporting Thyroid Cytopathology.[25] Nodules were classified as benign (Bethesda category 2), indeterminate (Bethesda categories 3 and 4), or malignant (Bethesda categories 5 and 6). Nodules diagnosed as non-diagnostic or having inadequate material (Bethesda category 1) were excluded from the analysis. This classification enabled a standardized assessment of malignancy risk across the different TIRADS.

### Statistical analysis

Statistical analyses were performed using the R programming language (R Core Team, 2023) within the RStudio environment (RStudio Team, 2023). All analyses were two-tailed, and a *P* value of less than 0.05 was considered statistically significant.

Descriptive statistics were used to characterize the patient cohort. Continuous variables, such as age, were presented as mean ± standard deviation for normally distributed data or as median (minimum–maximum) for non-normally distributed data. Categorical variables were expressed as frequency and percentage.

Comparative evaluations among groups were conducted using appropriate statistical methods. Independent sample t-tests were applied to compare continuous variables between two groups, and analysis of variance alongside Tukey's post-hoc tests was used for comparisons among multiple groups. Categorical variables were analyzed using chi-square or Fisher's exact tests, as appropriate.

The performance of all TIRADS was assessed by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) to evaluate each system's accuracy in identifying malignant and benign nodules. Histopathology results with Bethesda scores of 2, 3, and 4 were considered negative, whereas scores of 5 and 6 were considered positive. For TIRADS classification, scores of 1, 2, and 3 were considered negative, and scores of 4 and 5 were considered positive. These metrics were computed using the caret package in R.[26] Additionally, confidence intervals for sensitivity, specificity, PPV, and NPV were calculated to assess the precision of these estimates. Differences in sensitivity and specificity between classification methods were evaluated using the McNemar test. Comparisons of PPV and NPV were conducted using the statistical package of Stock et al.[27]

## Results

In this study, 1,324 nodules detected in 1,139 patients were evaluated, and 32 nodules in 29 patients were excluded due to non-diagnostic or insufficient cytopathological material detected by FNAB. A total of 1,110 patients with 1,292 nodules confirmed by cytopathological diagnosis were included. Of the 1,292 nodules included, 846 (65.4%) were diagnosed as benign, 299 (23.1%) as intermediate risk, and 147 (11.3%) as malignant. The overall incidence of malignancy was approximately 11.5%. Example images with evaluations are shown in Figure 2.

Of the patients, 826 (72.5%) were women and 313 (27.5%) were men. Among the 947 nodules in women, 629 (66.4%) were benign, 217 (22.9%) were intermediate risk, and 101 (10.6%) were malignant. Of the 345 nodules in men, 216 (62.6%) were benign, 81 (23.4%) were intermediate risk, and 48 (13.9%) were malignant. A statistically significant difference was found between malignancy rates in men and women ($P < 0.001$).

The average age of the patients was 46.33 ± 11.95 years; for women, it was 46.18 ± 11.94 years, and for men, it was 47.37 ± 11.57 years. The average age of patients with benign nodules was 47.04 ± 11.62 years, whereas those with intermediate-risk nodules had an average age of 46.46 ± 11.90 years. Patients diagnosed with malignant nodules were younger, with an average age of 40.97 ± 11.93 years ($P < 0.001$). The lower average age of patients with malignant nodules was observed. Further distribution of patient characteristics is summarized in Table 1.

The diagnostic performance parameters of the AI-TIRADS classification are as follows: sensitivity, specificity, PPV, and NPV, calculated as 95.3%, 53.8%, 21.2%, and 98.8%, respectively. The statistical analysis of malignancy risk according to the AI-TIRADS guidelines is shown in Table 2.

The AI-TIRADS missed 7 cancer cases out of 149 (4.6%), the ACR-TIRADS missed 5 cancer cases out of 149 (3.3%), the ATA-TIRADS missed 3 cancer cases out of 146 (2%), the EU-TIRADS missed 3 cancer cases out of 149

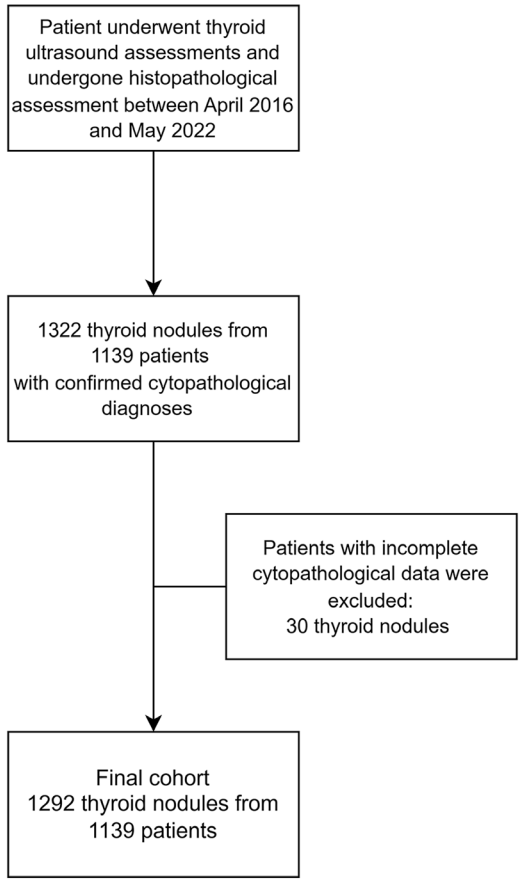



**Figure 1.** Flowchart of the study.

(2%), and the K-TIRADS missed 3 cancer cases out of 149 (2%).

The sensitivity and specificity for the AI-TIRADS were 94.6% and 53.6%, respectively. The AI-TIRADS showed no statistically significant difference from all other TIRADS in sensitivity but showed a statistically significant difference in specificity. In comparison, the ACR-TIRADS showed a sensitivity of 96.6% and a specificity of 44.6%. The ATA-TIRADS, the EU-TIRADS, and the K-TIRADS exhibited similar sensitivities of 97.9% but lower specificities ranging from 39.3% to 40.1%.

The PPV for the AI-TIRADS was 21.2%, which was statistically significantly higher than those of the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, which were 17.4%, 17.9%, and 17.9%, respectively. The PPV for the AI-TIRADS was not statistically significantly different from that of the ACR-TIRADS (18.5%). Unnecessary biopsy rates were interpreted based on PPVs (higher is better). The NPV for the AI-TIRADS was 98.8% and was not statistically significantly different from those of the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, which were 99%, 99.3%, 99.3%, and 99.3%, respectively. Examples of false-positive images are provided in Figure 3. The diagnostic performances of the classification systems used in our study are shown in Table 3.

## Discussion

This study compared the diagnostic accuracy of the AI-TIRADS with other classification systems, such as the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, for thyroid nodule characterization. The study revealed that the AI-TIRADS had a sensitivity of 94.6% and a specificity of 53.6%, surpassing the other systems in specificity while maintaining comparable sensitivity levels. This indicates that the AI-TIRADS can effectively reduce unnecessary FNABs without compromising the detection of malignant nodules.

The higher specificity of the AI-TIRADS aligns with recent studies suggesting that AI-enhanced algorithms improve risk stratification accuracy by minimizing false-positive results.[17,18] The reduction in unnecessary biopsies observed with the AI-TIRADS (PPV 21.2%) compared with the ACR-TIRADS (PPV 18.5%) and other systems (PPV <18%) underscores its potential to optimize clinical workflows and alleviate patient burden.[12]

Consistent with our findings, previous research has indicated that AI-based systems can enhance diagnostic performance by accurately analyzing complex US features beyond human capability.[28,29] Additionally, the lack of substantial differences in malignancy rates between genders and the association of multiple suspicious US features with increased cancer risk corroborate existing literature.[4,5]

The superior specificity of the AI-TIRADS (53.6%) compared with other systems such as the ACR-TIRADS (44.6%), the ATA-TIRADS (39.3%), the EU-TIRADS (40.1%), and the K-TIRADS (40.1%) is crucial in this context, as it suggests a greater ability to correctly classify benign nodules within this group, thereby reducing the rate of unnecessary FNABs. Wildman-Tobriner et al.[17] reported findings consistent with ours, highlighting the better specificity of the AI-TIRADS. When intermediate-risk lesions (Bethesda 3 and 4) were clas-

**Table 1.** Distribution of patient characteristics across different histopathological diagnoses

| Patient characteristics | Benign | | Malignant | *P* value |
|---|---|---|---|---|
| | Low-risk | Intermediate risk | | |
| Sex | - | - | - | <0.001 |
| Female | 629 (66.4%) | 217 (22.9%) | 101 (10.7%) | - |
| Male | 216 (62.6%) | 81 (23.5%) | 48 (13.9%) | - |
| Age | 47.04 ± 11.62 | 46.46 ± 11.90 | 40.97 ± 11.93 | <0.001 |

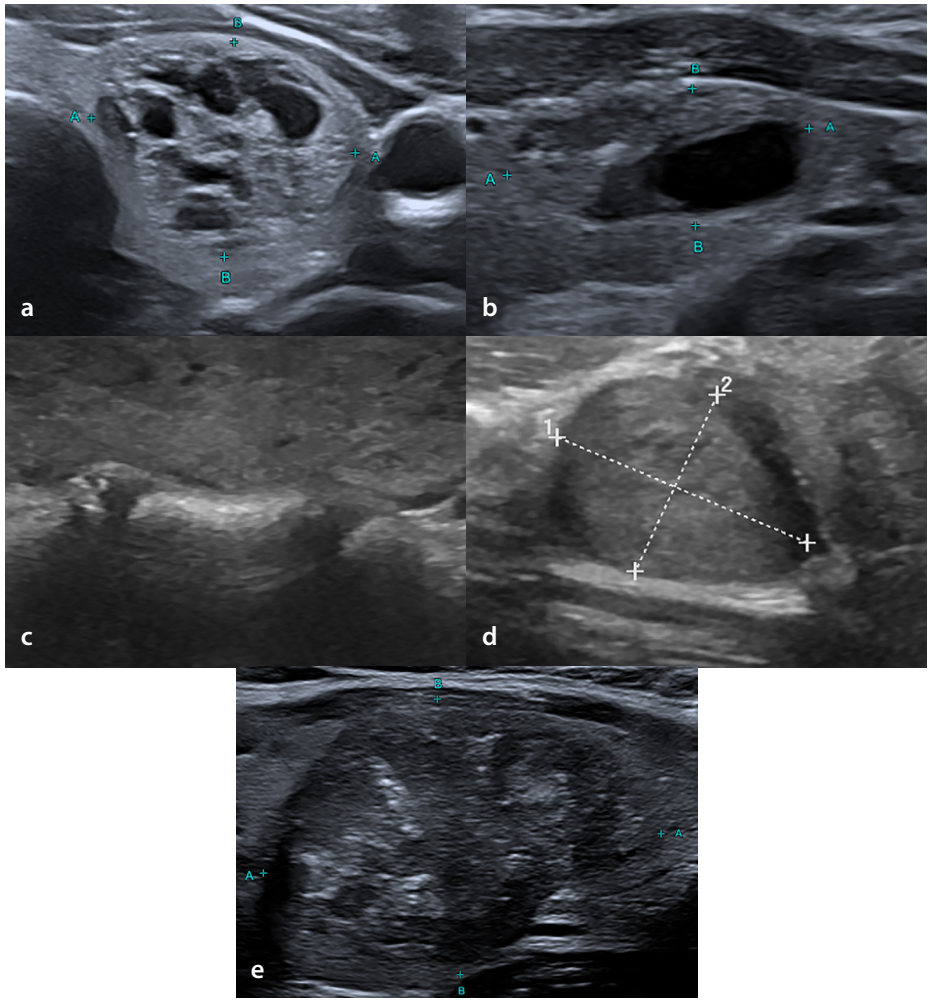



**Figure 2.** Examples of evaluated images. **(a)** AI-TIRADS 1, ACR-TIRADS 2, EU-TIRADS 2, K-TIRADS 2, and ATA-TIRADS 2 lesions. Histopathological evaluation was benign (Bethesda category 2). **(b)** AI-TIRADS 2, ACR-TIRADS 3, EU-TIRADS 2, K-TIRADS 2, and ATA-TIRADS 2 lesions. Histopathological evaluation was benign (Bethesda category 2). **(c)** AI-TIRADS 3, ACR-TIRADS 3, EU-TIRADS 3, K-TIRADS 3, and ATA-TIRADS 3 lesions. Histopathological evaluation was benign (Bethesda category 2). **(d)** AI-TIRADS 4, ACR-TIRADS 4, EU-TIRADS 4, K-TIRADS 4, and ATA-TIRADS 4 lesions. Histopathological evaluation was benign (Bethesda category 2). **(e)** AI-TIRADS 5, ACR-TIRADS 5, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was malignant (Bethesda category 6). AI-TIRADS, artificial intelligence-based Thyroid Imaging Reporting and Data System; ACR, American College of Radiology; EU, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association.

sified as positive, the AI-TIRADS demonstrated a PPV of 48%, which was higher than that of the other systems (43%–45%). Although other systems showed higher sensitivities in this analysis (78%–81% compared with 72% for the AI-TIRADS), they did so at the cost of lower precision and PPV. This trade-off underscores the clinical utility of the AI-TIRADS in the intermediate-risk category, where avoiding unnecessary biopsies for benign nodules is a primary goal.

**Table 2.** Distribution of the AI-TIRADS classifications

| Category | Benign | | Malignant (n = 149) | Malignancy risk | *P* value |
|---|---|---|---|---|---|
| | Low-risk (n = 845) | Intermediate risk (n = 298) | | | |
| AI-TIRADS 1 | 153 (18.1%) | 16 (5.3%) | 0 (0%) | 0% | <0.001 |
| AI-TIRADS 2 | 228 (27.0%) | 47 (15.8%) | 2 (1.3%) | 0.7% | - |
| AI-TIRADS 3 | 116 (13.7%) | 53 (17.8%) | 5 (3.3%) | 2.9% | - |
| AI-TIRADS 4 | 265 (31.3%) | 154 (51.7%) | 35 (22.8%) | 7.7% | - |
| AI-TIRADS 5 | 83 (9.9%) | 28 (9.4%) | 107 (72%) | 49% | - |

AI, Artificial intelligence; TIRADS, Thyroid Imaging Reporting and Data System.

**Table 3.** Diagnostic performance of the classification systems used

| Classifications System | Sensitivity | *P* value* | Precision | *P* value* | PPV | *P* value* | NPV | *P* value |
|---|---|---|---|---|---|---|---|---|
| AI-TIRADS | 95.3% | - | 53.8% | - | 21.2% | - | 98.8% | - |
| ACR-TIRADS | 96.6% | 0.39 | 44.6% | <0.001 | 18.5% | 0.07 | 99% | 0.62 |
| EU-TIRADS | 97.9% | 0.12 | 40.1% | <0.001 | 17.9% | 0.03 | 99.3% | 0.32 |
| K-TIRADS | 97.9% | 0.12 | 40.1% | <0.001 | 17.9% | 0.03 | 99.3% | 0.32 |
| ATA | 97.9% | 0.12 | 39.3% | <0.001 | 17.4% | 0.04 | 99.3% | 0.32 |

*Compared with the AI-TIRADS. AI, artificial intelligence; TIRADS, Thyroid Imaging Reporting and Data Systems; ACR, American College of Radiology; EU-TIRADS, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association; PPV, positive predictive value; NPV, negative predictive value.

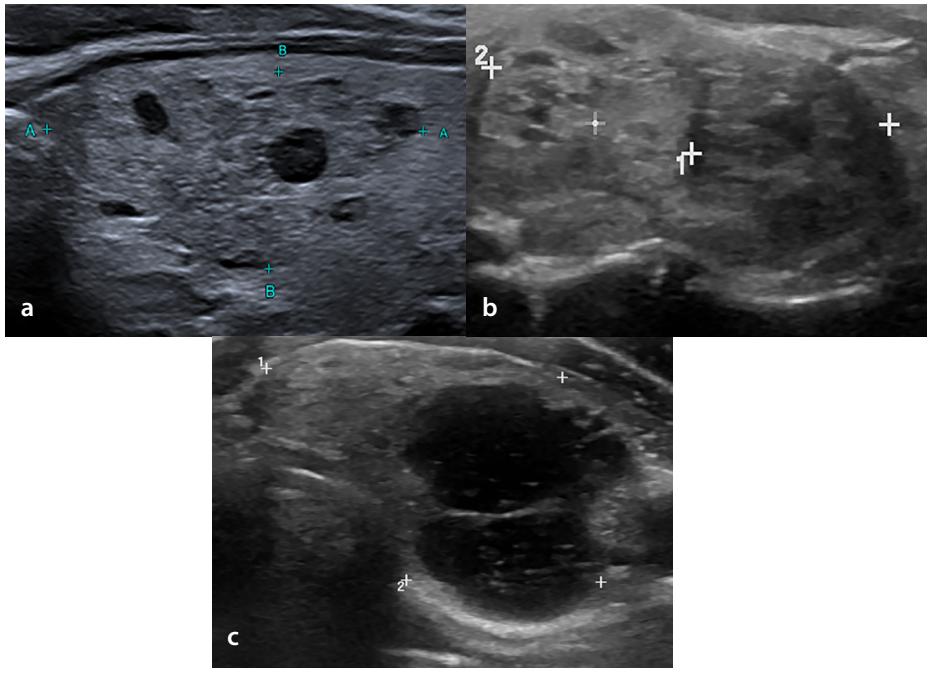



**Figure 3.** Examples of false-positive images. **(a)** AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. **(b)** AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. **(c)** AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. AI-TIRADS, artificial intelligence-based Thyroid Imaging Reporting and Data System; ACR, American College of Radiology; EU, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association.

However, this study has limitations, including its retrospective design, which may lead to selection bias and potentially restrict the generalizability of the findings. In particular, it must be acknowledged that patients with histopathology results of Bethesda score 1 and only biopsied patients were included in the study. Additionally, the lack of an interobserver reliability assessment might impact the consistency of nodule classification.[30] To confirm the effectiveness of the AI-TIRADS and examine its incorporation into standard clinical practice, future prospective studies involving larger and more diverse populations are needed.

Moreover, variability among different TI-RADS in criteria weighting and risk categorization highlights the need for standardized guidelines to ensure consistent application across healthcare settings.[31,32] The integration of AI into the TIRADS offers a promising solution to these challenges by providing objective and reproducible assessments, thereby enhancing diagnostic accuracy and reducing interobserver variability.[33]

The AI-TIRADS demonstrated improved specificity and a reduced unnecessary biopsy rate in thyroid nodule classification without sacrificing high sensitivity compared with other traditional TIRADS. The findings suggest that the AI-TIRADS can be utilized to enhance clinical decision-making, optimize resource utilization, and improve patient management in thyroid nodule assessment. Further prospective studies are required to confirm these findings and facilitate broader implementation of the AI-TIRADS in clinical practice.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest*. 2009;39(8):699-706. [CrossRef]
2. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med*. 1997;126(3):226-231. [CrossRef]

3. Frates MC, Benson CB, Doubilet PM, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab.* 2006;91(9):3411-3417. [CrossRef]
4. American Thyroid Association (ATA) Guidelines taskforce on thyroid nodules and differentiated thyroid cancer; Cooper DS, Doherty GM, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19(11):1167-1214. [CrossRef]
5. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical Guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *Endocr Pract*. 2010;16(3):468-475. [CrossRef]
6. Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. *Thyroid*. 2004;14(3):207-215. [CrossRef]
7. Weber AL, Randolph G, Aksoy FG. The thyroid and parathyroid glands. CT and MR imaging and correlation with pathology and clinical findings. *Radiol Clin North Am*. 2000;38(5):1105-1129. [CrossRef]
8. Burch HB. Evaluation and management of the solid thyroid nodule. *Endocrinol Metab Clin North Am.* 1995;24(4):663-710. [CrossRef]
9. Gharib H. Fine-needle aspiration biopsy of thyroid nodules: advantages, limitations, and effect. *Mayo Clin Proc.* 1994;69(1):44-49. [CrossRef]
10. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab.* 2009;94(5):1748-1751. [CrossRef]
11. Kwak JY, Han KH, Yoon JH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology*. 2011;260(3):892-899. [CrossRef]
12. Li X, Hou XJ, Du LY, et al. Virtual touch tissue imaging and quantification (VTIQ) combined with the American College of Radiology thyroid imaging reporting and data system (ACR TI-RADS) for malignancy risk stratification of thyroid nodules. *Clin Hemorheol Microcirc*. 2019;72(3):279-291. [CrossRef]
13. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association Guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur Thyroid J*. 2017;6(5):225-237. [CrossRef]
14. Ha EJ, Na DG, Baek JH. Korean thyroid imaging reporting and data system: current status, challenges, and future perspectives. *Korean J Radiol.* 2021;22(9):1569-1578. [CrossRef]
15. Grani G, Lamartina L, Ascoli V, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "right" TIRADS. *J Clin Endocrinol Metab*. 2019;104(1):95-102. [CrossRef]
16. Chen Y, Gao Z, He Y, et al. An artificial intelligence model based on ACR TI-RADS characteristics for US diagnosis of thyroid nodules. *Radiology*. 2022;303(3):613-619. [CrossRef]
17. Wildman-Tobriner B, Buda M, Hoang JK, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology*. 2019;292(1):112-119. [CrossRef]
18. Middleton WD, Teefey SA, Reading CC, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association Guidelines. *AJR Am J Roentgenol.* 2018;210(5):1148-1154. [CrossRef]
19. Lauria Pantano A, Maddaloni E, Briganti SI, et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. *Eur J Endocrinol*. 2018;178(6):595-603. [CrossRef]
20. Shah JP. Thyroid carcinoma: epidemiology, histology, and diagnosis. *Clin Adv Hematol Oncol*. 2015;13(4 Suppl 4):3-6. [CrossRef]
21. Borlea A, Borcan F, Sporea I, et al. TI-RADS diagnostic performance: which algorithm is superior and how elastography and 4D vascularity improve the malignancy risk assessment. *Diagnostics (Basel).* 2020;10(4):180. [CrossRef]
22. Nam-Goong IS, Kim HY, Gong G, et al. Ultrasonography-guided fine-needle aspiration of thyroid incidentaloma: correlation with pathological findings. *Clin Endocrinol (Oxf).* 2004;60(1):21-28. [CrossRef]
23. Ayvaz E, Kaplan K, Kuncan F, Ayvaz E, Türkoğlu H. Reducing operation costs of thyroid nodules using machine learning algorithms with thyroid nodules scoring systems. *Applied Sciences.* 2022;12(22):11559. [CrossRef]
24. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527. [CrossRef]
25. Acharya K, Shrivastav S, Triipathi P, et al. The Bethesda system for reporting thyroid cytopathology: validating at tribhuvan university teaching hospital. *Int Arch Otorhinolaryngol.* 2021;26(1):e097-e102. [CrossRef]
26. Kuhn M, Wing J, Weston S, et al. Caret: Classification and Regression Training. Published online March 21, 2023. Accessed September 15, 2024. [CrossRef]
27. Stock C, Hielscher T, Discacciati A. DTComPair: comparison of binary diagnostic tests in a paired study design. Published online August 20, 2023. Accessed January 2, 2024. [CrossRef]
28. Castellana M, Castellana C, Treglia G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *J Clin Endocrinol Metab*. 2020;105(5):dgz170. [CrossRef]
29. Zhang J, Zhang X, Meng Y, Chen Y. Contrast-enhanced ultrasound for the differential diagnosis of thyroid nodules: an updated meta-analysis with comprehensive heterogeneity analysis. *PLoS One*. 2020;15(4):e0231775. [CrossRef]
30. Hoang JK, Middleton WD, Farjat AE, et al. Reduction in thyroid nodule biopsies and improved accuracy with American College of Radiology thyroid imaging reporting and data system. *Radiology*. 2018;287(1):185-193. [CrossRef]
31. Cappelli C, Castellano M, Pirola I, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM*. 2007;100(1):29-35. [CrossRef]
32. Bakhshaee M, Davoudi Y, Mehrabi M, et al. Vascular pattern and spectral parameters of power Doppler ultrasound as predictors of malignancy risk in thyroid nodules. *Laryngoscope*. 2008;118(12):2182-2186. [CrossRef]
33. Cao CL, Li QL, Tong J, et al. Artificial intelligence in thyroid ultrasound. *Front Oncol*. 2023;13:1060702. [CrossRef]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253310

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE



# Automated detection and characterization of small cell lung cancer liver metastasis on computed tomography

Sophia Ty[1]
Fahmida Haque[1]
Parth Desai[2]
Nobuyuki Takahashi[3,4]
Usamah Chaudhary[1,5]
Peter L. Choyke[1]
Anish Thomas[4]
Barış Türkbey[1]
Stephanie A. Harmon[1]

[1]National Cancer Institute, Artificial Intelligence Resource, Maryland, USA

[2] Fox Chase Cancer Center at Temple University Hospital, Philadelphia, USA

[3]National Cancer Center Hospital East, Department of Medical Oncology, Kashiwa, Japan

[4]National Cancer Institute, Developmental Therapeutics Branch, Maryland, USA

[5]University of Texas Southwestern Medical Center, Texas, USA

**PURPOSE**

Small cell lung cancer (SCLC) is an aggressive disease with diverse phenotypes that reflect the heterogeneous expression of tumor-related genes. Recent studies have shown that neuroendocrine (NE) transcription factors may be used to classify SCLC tumors with distinct therapeutic responses. The liver is a common site of metastatic disease in SCLC and can drive a poor prognosis. Here, we present a computational approach to detect and characterize metastatic SCLC (mSCLC) liver lesions and their associated NE-related phenotype as a method to improve patient management.

**METHODS**

This study utilized computed tomography scans of patients with hepatic lesions from two data sources for segmentation and classification of liver disease: (1) a public dataset from patients of various cancer types (segmentation; n = 131) and (2) an institutional cohort of patients with SCLC (segmentation and classification; n = 86). We developed deep learning segmentation algorithms and compared their performance for automatically detecting liver lesions, evaluating the results with and without the inclusion of the SCLC cohort. Following segmentation in the SCLC cohort, radiomic features were extracted from the detected lesions, and least absolute shrinkage and selection operator regression was utilized to select features from a training cohort (80/20 split). Subsequently, we trained radiomics-based machine learning classifiers to stratify patients based on their NE tumor profile, defined as expression levels of a preselected gene set derived from bulk RNA sequencing or circulating free DNA chromatin immunoprecipitation sequencing.

**RESULTS**

Our liver lesion detection tool achieved lesion-based sensitivities of 66%–83% for the two datasets. In patients with mSCLC, the radiomics-based NE phenotype classifier distinguished patients as positive or negative for harboring NE-like liver metastasis phenotype with an area under the receiver operating characteristic curve of 0.73 and an F1 score of 0.88 in the testing cohort.

**CONCLUSION**

We demonstrate the potential of utilizing artificial intelligence (AI)-based platforms as clinical decision support systems, which could help clinicians determine treatment options for patients with SCLC based on their associated molecular tumor profile.

**CLINICAL SIGNIFICANCE**

Targeted therapy requires accurate molecular characterization of disease, which imaging and AI may aid in determining.

**KEYWORDS**

Computer vision, segmentation, neuroendocrine gene expression, radiomics, tumor classification, transcriptomics, molecular tumor profile

**Handling editor:** Tuğba Akıncı D'Antonoli

**Corresponding author:** Barış Türkbey

**E-mail:** turkbeyi@mail.nih.gov

Received 28 February 2025; revision requested 24 March 2025; last revision received 14 July 2025; accepted 11 September 2025.

Epub: 06.10.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253310

You may cite this article as: Ty S, Haque F, Desai P, et al. Automated detection and characterization of small cell lung cancer liver metastasis on computed tomography. *Diagn Interv Radiol*. 2026;32(4):382-390.

Small cell lung cancer (SCLC) is an aggressive form of lung cancer strongly associated with smoking and accounts for 13%–15% of all lung cancer cases.[1,2] Patients often present with advanced disease, resulting in a poor prognosis with a 5-year survival rate of 7%.[3] These outcomes reflect the challenges in clinical management of a recalcitrant cancer marked by the ubiquitous presence of TP53, RB1 loss-of-function events, and high chromosomal instability,[4,5] which drive rapid progression, widespread metastasis,[5,6] and treatment resistance following initial response to therapy.[7] Among many complications associated with cancer progression, SCLC typically leads to hepatic metastasis, which is seen in 21%–27% of patients at presentation and 69% at autopsy.[8] This makes the liver the most prevalent metastatic site after mediastinal lymph nodes—an important characteristic since liver metastasis is also an independent marker of poor prognosis.[9,10]

SCLC also demonstrates a high degree of heterogeneity, manifesting under various transcriptional subtypes. The classification of SCLC subtypes is defined by the expression levels of four transcription regulators, namely neuronal differentiation 1 (NEUROD1), achaete-scute family basic helix-loop-helix transcription factor 1 (ASCL1), POU class 2 homeobox3, and yes-associated protein (YAP1).[11] The relative expression of these regulators leads to heterogeneous neuroendocrine (NE) gene expression, which has therapeutic implications.[12,13] The SCLC tumors associated with relatively high NEUROD1 and ASCL1 expression are considered NE positive and demonstrate greater susceptibility to DNA-damaging agents;[14,15] non-NE SCLC tumors have greater POU2F2 and YAP1 expression and have been shown to possess better response to immunotherapy.[13,16-18]

Despite emerging insights into its molecular subtypes, SCLC is still currently treated as a homogenous disease. As we gain more insights into the molecular underpinnings of SCLC that drive tumor response to treatments, there is a need for clinical workflows that can stratify patients based on their tumor profile. Methods that identify subpopulations of patients with SCLC who are likely to benefit from specific targeted treatments without requiring additional invasive testing can offer physicians actionable insights, especially when treatment response status can be determined at the time of diagnosis. Furthermore, computational platforms that can accurately detect and characterize tumors offer practical utility in supporting physicians from diagnosis to treatment. They can automate critical tasks, integrate different types of medical data (e.g., radiology scans, biopsy findings, blood panel information), extract clinically relevant tumor characteristics, and consolidate medical information for health practitioners.

Within the past decade, artificial intelligence (AI) has been integrated into automating medical image processing tasks. More specifically, deep learning has shown promise in segmenting objects at different imaging scales, including tissue[19] and cellular[20] levels, for a variety of medical conditions, including lung cancer[21,22] and hepatic diseases.[23-25] Preceding the popularity of deep learning for medical image segmentation is the use of radiomics, a common research approach to describe tumors quantitatively, including their intensity, shape, and texture, which may be used as image-based biomarkers for downstream analysis and association studies. Radiomics has been adopted to address various clinical decision tasks, such as lesion classification[26,27] and treatment response prediction,[28,29] including applications in liver-associated malignancies.[30-32]

Given the implications of transcriptional subtypes to treatment response in SCLC and the association between SCLC hepatic metastasis and prognosis, we investigated whether NE status, as determined by tumor gene expression analysis, can be determined from computed tomography (CT) scans of confirmed SCLC metastasis in the liver. In this study, we present a two-step machine learning framework for automated detection and characterization of SCLC liver metastasis. We employ deep learning for three-dimensional (3D) segmentation of hepatic lesions followed by radiomics-based analysis to characterize and classify image scans by NE status, defined as high expression of a pre-selected gene set.

## Methods

A graphical summary of the study objective is shown in Figure 1.

**Main points**

- The liver is a frequent site of metastases in small cell lung cancer, and artificial intelligence helps identify and segment tumors.
- The computed tomography (CT) imaging characteristics of liver lesions have a moderate correlation with neuroendocrine transcription factors.
- An end-to-end machine learning pipeline may help characterize the molecular profile of liver lesions in CT.

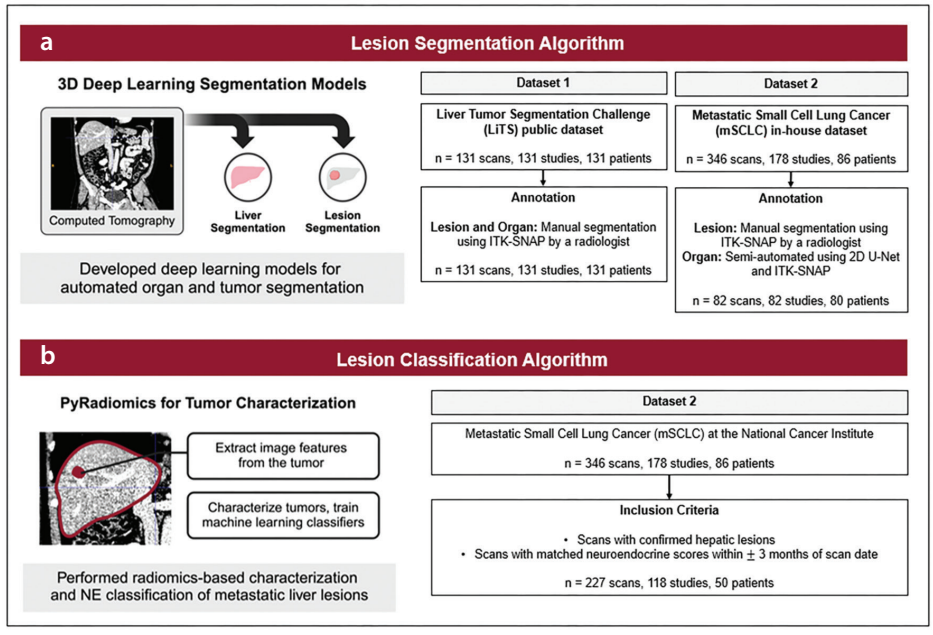



**Figure 1.** Graphical summary of the machine learning framework for automated detection and characterization of small cell lung cancer (SCLC) liver metastases on computed tomography. **(a)** Three-dimensional deep learning models were developed to simultaneously segment the liver and tumors within the region. **(b)** Radiomics-based characterization of metastatic SCLC liver tumors was performed, and subsequently supervised machine learning models were trained to classify patients' neuroendocrine (NE) status as NE positive or NE negative.

### Study population and data description

Two datasets were utilized in this study: (1) the liver tumor segmentation (LiTS) dataset, a publicly available dataset containing 131 CT scans; and (2) a retrospective cohort of patients with metastatic SCLC (mSCLC) underwent CT at the National Cancer Institute (Bethesda, MD, USA).

The LiTS dataset consists of multi-center scans of primary and secondary hepatic tumors. All scans were manually annotated by a radiologist (>3 years' experience) using the ITK-SNAP open source software platform to obtain liver and lesion labels. Annotations were confirmed by three additional radiologists; the most senior reader's findings were used in any labeling conflicts. This research study was conducted retrospectively using human participant data made available as open-access materials by Bilic et al.[25] This open-source cohort was used for training and development of a segmentation algorithm for the detection and segmentation of focal liver lesions on CT. This cohort was used exclusively for the segmentation task (Figure 1).

An initial query for the mSCLC dataset identified 88 patients diagnosed with SCLC and undergoing disease monitoring or treatment at the institution under one or more clinical protocols, including the following ClinicalTrials.gov identifiers: NCT02769962 (IRB 16-C-0107; 2016-05-09), NCT03554473 (IRB 18-C-0110; 2018-09-11), NCT02487095 (IRB 15-C-0150; 2015-07-30), NCT02484404 (IRB 15-C-0145; 2015-06-29), NCT02146170 (IRB 14-C-0105, 2014-05-28). Each protocol was approved by the local institutional review board, and written informed consent was obtained from all patients. From these patients, a total of 346 abdominal CT scans obtained during 178 CT sessions were identified for possible inclusion. This cohort was used for both the segmentation and classification tasks (Figure 1). Multiple series were included from each study date: for example, thick-slice and soft tissue thin-slice reconstructions to evaluate model robustness (Supplementary Table 1). Radiology reports were manually reviewed for each CT scan to confirm the presence or absence of hepatic lesions. From all available scans, a subset of 82 scans was manually reviewed by an expert radiologist (>15 years' experience), and liver lesions were segmented using ITK-SNAP. Liver organ annotations were obtained using a previously developed two-dimensional (2D) U-Net liver segmentation model[33] and were manually adjusted using ITK-SNAP. All annotated scans were used for the tumor segmentation task (Figure 1).

All patients in the mSCLC cohort underwent either tissue or blood sampling for bulk RNA or circulating free DNA (cfDNA) chromatin immunoprecipitation sequencing at multiple timepoints, corresponding with CT study dates (± 3 months). Expression profiles from sequencing data were used to classify patients broadly into NE-positive and NE-negative phenotype groups based on previously published methods.[13,34] Briefly, single-sample gene set enrichment analysis from a 50-gene signature panel was used to classify samples as NE (score >0) or non-NE (score <0), with a lower score in the non-NE group reflecting more confidence that the sample does not exhibit NE differentiation.[35] Strong correlation observed between cfDNA-derived and RNA-derived expression scores for NE phenotyping has been previously reported;[13,34-36] therefore, either reference standard was used for ground truth assignment in this cohort. The NE phenotype expression scores (range: −1,1) and classification (NE, non-NE) were recorded for use in this study (Table 1).

### Deep learning model development for tumor segmentation

Three deep learning algorithms were selected to build a hepatic lesion detection model: (1) a 3D U-Net, (2) a 3D SegResNet, and (3) a 3D nnU-Net. During initial model development and selection, each algorithm was trained solely using the LiTS dataset, and mSCLC data were used as an independent test set. For all training, data partitions were stratified at the patient level and are summarized in Table 2.

The U-Net and SegResNet models were built using the Medical Open Network for AI platform (version 1.3.0).[37] For these two network architectures, training was conducted with the following data pre-processing and augmentations: CTs were resampled to uniform spacing (0.5 mm × 0.5 mm × 1 mm), foreground cropping, variable CT windowing, and random cropping by labels with a sampling ratio of 4:1:3 for the background, liver, and lesion, with 12 samples taken per image. Each sample crop was of size 512 × 512 × 16. Both models were trained using an adaptive moment estimator to minimize lesion-level DICE loss, with a learning rate of 0.0001 for 1,500 epochs. The final model was selected based on the highest validation DICE reported during training. Inference was conducted using the sliding window technique.

The nnU-Net model, an auto-configuring semantic segmentation model, was implemented using the built-in *3dfullres* five-fold cross-validation.[38] Pre-processing configurations selected by the model included spacing (0.789 mm × 0.789 mm × 2 mm), patch size 80 × 80 × 60, and per-image z-score standardization.

For all three models, performance was evaluated in the test set using lesion-level DICE coefficients and tumor detection sensitivity on both the LiTS and mSCLC datasets.

Due to differences in the burden and imaging characteristics of the mSCLC cohort compared with the LiTS cohort, which may potentially impact generalizability, a final nnUNet model was trained from all LiTS training data along with a subset of mSCLC scans partitioned with an approximate training/test

**Table 1.** Summary of key characteristics of the mSCLC dataset describing the distribution of patient scans, lesions, and neuroendocrine status in the cohort

| mSCLC data characteristics | Quantity |
|---|---|
| Age (years) | 63 (23–82) |
| Sex: male, female | 45, 41 |
| Number of scans per patient | Median: 3, range: (1–14) |
| Number of lesions per scan | Median: 10, range: (1–148) |
| **Hepatic lesions** | |
| Present | 252 |
| Absent | 94 |
| **Final study cohort for classification†** | |
| **Neuroendocrine classification by sequencing-based assessment†** | |
| Positive | 172 |
| Negative | 55 |

†, reported for the final 227 series volumes for the classification task; mSCLC, metastatic small cell lung cancer.

split of 80%/20% scans. Inference of the final finetuned segmentation model was completed for all mSCLC scans for use in the classification model.

### Radiomics characterization and neuroendocrine phenotype classification

Each mSCLC scan was labeled based on matched gene expression-based NE score as NE positive (1) or NE negative (0). All CT studies with confirmed hepatic lesions that were matched to NE scores within ± 3 months of the imaging date were included in the NE phenotype classification. Scans were partitioned with an approximate training/test split of 80%/20% images using the same stratification applied during segmentation. Splits were determined at the patient level to avoid bias, resulting in 177 scans for training and 50 scans for testing (Table 2).

Liver lesion contours obtained from the final segmentation model were characterized using radiomics. Quantitative image features were extracted using PyRadiomics (v3.0.1) with a resampling pixel spacing of (1 mm, 1 mm, 1 mm) for the (x, y, z) voxel coordinates and default image standardization parameters. A total of 107 radiomic features were obtained, representing first-order statistics, shape (2D and 3D), gray level co-occurrence matrix, gray level run length matrix (glrlm), gray level size zone matrix (glszm), neighboring gray tone difference matrix, and gray level dependence matrix (gldm). The number of lesions per image was determined using connected-components-3D (v3.12.1) and served as an additional feature, resulting in a total of 108 features. From these, a subset of imaging characteristics correlated with NE phenotype was selected using least absolute shrinkage and selection operator (LASSO) regression (scikit-learn v1.2.2).

Radiomics-based NE phenotype classification was conducted using three machine learning models: (1) logistic regression (scikit-learn v1.2.2), (2) random forest (scikit-learn v1.2.2), and (3) XGBoost (v2.0.3). For all models, default parameters were used. Code and raw data for how these models were trained are available at https://github.com/NIH-MIP/mSCLC_Segmentation_Classification. All models incorporated the subset of imaging features selected using LASSO regression for binary classification of tumors as NE (1) or non-NE (0) phenotype.

Each model was trained with and without class-based weights. Five-fold cross-validation was implemented, and the best model was selected using the F1 score and area under the receiver operating characteristic curve (AUC) from cross-validation as the primary performance criteria. When applied to the test set, the ensemble of all five-folds was utilized for test set evaluation (average prediction of five-folds).

### Statistical analysis

The DICE coefficient,[39] a measure that describes spatial agreement between two image sets, was calculated to quantify the performance of the model compared with ground truth annotation from radiologists. To evaluate detection performance metrics at the lesion level, connected-components-3D (v3.12.1) was used to identify unique lesions in both the ground truth segmentations and model output. Next, each lesion was classified as a true positive (i.e., ground truth lesion correctly segmented by AI), false negative (i.e., ground truth lesion was not segmented by the model), or false positive (i.e., model segmented a lesion with no ground truth correlate) per scan. Each segmentation model's sensitivity, positive predictive value (PPV), and false positive trends were calculated and reported as summary statistics.

The relationship between tumor burden and NE scores was examined using two tests: (1) Spearman correlation analysis (SciPy v.1.11.1) for continuous NE scores, and (2) Wilcoxon rank sum tests (R v4.4.1) for binarized NE scores. For this analysis, one series per patient per scan date was selected. Tumor volume estimates were calculated using AI-predicted tumor regions from the final segmentation model.

Finally, the performance of each binary classifier was evaluated. Model accuracy, sensitivity, specificity, PPV, negative predictive value, F1 scores, and AUC were calculated (scikit-learn v1.2.2) and compared. Due to potential bias in multiple scans coming from the same CT study, bootstrap sampling was performed at the study level to select one scan per study per iteration randomly. The mean and 95% confidence intervals (CIs) of each performance metric are reported in the test set.

**Table 2.** Data splits per task

| Task | Level | Training/validation | Test¥ |
|---|---|---|---|
| **Segmentation†** | Pts | LiTS 121, mSCLC 62 | LiTS 10, mSCLC 18 |
| | Scans | LiTS 121, mSCLC 64 | LiTS 10, mSCLC 18 |
| | Series volumes | LiTS 121, mSCLC 64 | LiTS 10, mSCLC 18 |
| **Classification (mSCLC only)**** | Pts | 36 | 14 |
| | Scans | 92 | 26 |
| | Series volumes | 177 | 50 |

†, For the segmentation task, two training schemes were used: with and without mSCLC. For the training scheme without mSCLC, the entire cohort was reserved for testing. Labels are created at the series volume level. **, For the classification task, labels are assigned at the scan level. Patients can have multiple labels (i.e., positive or negative) depending on the time point at which the scan/sequencing was completed. ¥, The test set for mSCLC is preserved across both tasks on the patient-level (i.e., one unique set of patients was reserved for end-to-end testing of the segmentation and classification models). mSCLC, metastatic small cell lung cancer; LiTS, liver tumor segmentation.

**Table 3.** Comparison of lesion segmentation model performance presents the median DICE coefficient, sensitivity, PPV, and number of false positives per scan for all deep learning models implemented for hepatic lesion segmentation. Results are shown for the publicly available LiTS dataset and an in-house mSCLC dataset

| | U-Net | | SegResNet | | nnU-Net | | Finetuned nnU-Net | |
|---|---|---|---|---|---|---|---|---|
| **Metrics** | LiTS | mSCLC | LiTS | mSCLC | LiTS | mSCLC | LiTS | mSCLC |
| **Median DICE** | 0.439 | 0.297 | 0.417 | 0.418 | 0.750 | 0.607 | 0.771 | 0.640 |
| **Sensitivity** | 0.395 | 0.244 | 0.662 | 0.447 | 0.813 | 0.539 | 0.826 | 0.667 |
| **PPV** | 0.971 | 0.975 | 0.962 | 0.987 | 1.0 | 0.998 | 1.0 | 1.0 |
| **FP/scan** | 0.1 (0-1) | 0.1 (0-1) | 0.2 (0-1) | 0.1 (0-1) | 0 | 0.02 (0-1) | 0 | 0 |

PPV, positive predictive value; mSCLC, metastatic small cell lung cancer; LiTS, liver tumor segmentation.

## Results

The segmentation model utilized all LiTS data and incorporated a subset of scans from the mSCLC cohort. Of the 88 patients identified for possible study inclusion, a total of 86 patients were included in the final study cohort, with two exclusions due to insufficient data records (no sampling within the required timeframe from a scan date; no segmentations). Key characteristics of the mSCLC cohort are provided in Table 1. Image acquisition characteristics are summarized in Supplementary Table 1 for each cohort and task.

### Lesion detection

The cohort information for model training is shown in Table 2. In addition to the LiTS cohort, 82 annotated scans (82 unique studies) from the mSCLC dataset were utilized for the segmentation task, of which 50 scans were positive for containing hepatic lesions. First, models were trained only on the LiTS cohort and applied to mSCLC. Among the three deep learning models implemented for automated hepatic lesion detection and segmentation, the 3D *fullres* nnU-Net model provided the most accurate and robust results for the test set from both datasets. Its detection performance had a median DICE score of 0.75 and 0.607, lesion-level sensitivity of 0.813 and 0.539, and PPV of 1.0 and 0.99 for the LiTS and mSCLC test sets, respectively. The 3D *fullres* nnU-Net model also had a range of 0–1 false positive lesions per scan for both datasets. The U-Net and SegResNet models had median DICE scores of 0.439, 0.417 for the LiTS dataset, and 0.297, 0.418 for the mSCLC dataset. The U-Net model had lesion sensitivities of 0.395 and 0.244, whereas the SegResNet model achieved 0.662 and 0.447 for LiTS and mSCLC. A summary of each model's performance is provided in Table 3.

Next, an evaluation was performed to determine how finetuning of the nnU-Net for liver lesion segmentation on the mSCLC data may improve performance. Here, the 3D *fullres* nnU-Net model achieved a median DICE of 0.771 and 0.640, with sensitivities of 0.826 and 0.667 for the LiTS test set and mSCLC test set, respectively. For both datasets, the model had a PPV of 1.0 and 0 false positives per scan. Representative images of cases with high and low concordance between the AI-predicted and ground truth annotations of mSCLC liver lesions from the test set are provided in Figure 2.

### Correlation analysis of tumor burden and neuroendocrine status

A statistically significant correlation was found between tumor burden and NE scores for continuous data (Spearman: 0.252, *P* value: 0.0059) and binarized data (Wilcoxon rank sum *P* value: 0.028). Further analysis revealed that cfDNA samples had greater dependence on tumor volume, with a statistically significant correlation between tumor burden and cfDNA-derived NE scores (Spearman: 0.446, *P* value: 0.00057; Wilcoxon rank sum *P* value: 0.0013). Conversely, biopsy-derived NE scores did not have a statistically significant correlation with tumor volume (Spearman: –0.0738, *P* value: 0.5680; Wilcoxon rank sum *P* value: 0.253). Correlation plots are provided in Figure 3.

### Neuroendocrine phenotype classification

A total of 227 scans from 118 CT studies were used for the NE classification task after excluding scans with no hepatic lesions by radiologist read (n = 92 scans), scans with no corresponding RNA sequencing (RNAseq) or cfDNA data (n = 19), or false negatives by segmentation model (n = 6 volumes). Of the usable data, 172 scans (89 studies) were identified as NE positive and 55 scans (29 studies) were identified as NE negative (Table 1). The LASSO feature selection was performed within the training set, identifying 20/108 radiomic features correlated with NE-related tumor phenotypes for inclusion in the classification model. The distribution of selected radiomic feature categories was as follows: 30% shape, 25% glszm, 20% gldm, 15% first order statistics, and 10% glrlm. Among these, shape was the most dominant radiomic feature type found in the subset. The top five imaging feature characteristics determined during feature selection were minor axis length, maximum 2D diameter row, major axis length, gldm large dependence emphasis, and first order variance. A full list of the selected radiomic features and a distribution summary of the selected feature type are provided in Figure 4.

For all evaluated models, weighted training did not substantially boost the predictive performance of the algorithms (Supplementary Table 2), nor did they generalize better to the test set (Supplementary Table 3), regardless of the optimization strategy for determining the best weight. Based on the cross-validation performance, the random forest classifier had the highest F1 (mean 0.86; range 0.81, 0.91) and AUC (mean: 0.68; range: 0.56, 0.81) (Supplementary Table 2). All calculated metrics for the NE phenotype classification task are summarized in Table 4, where the logistic regression model generalized better than the random forest model by AUC (mean: 0.71; 95% CI: 0.59–0.82 vs. mean 0.58; CI: 0.48–0.69, respectively), though F1 and accuracy metrics were similar across all models, likely due to the imbalance in favor of the positive class (NE). The 50 scans in the test cohort represented 26 unique CT studies. Breaking accuracy down further by scan-level classification agreement across series volumes: 18/24 scans with multiple series volumes were correctly classified in

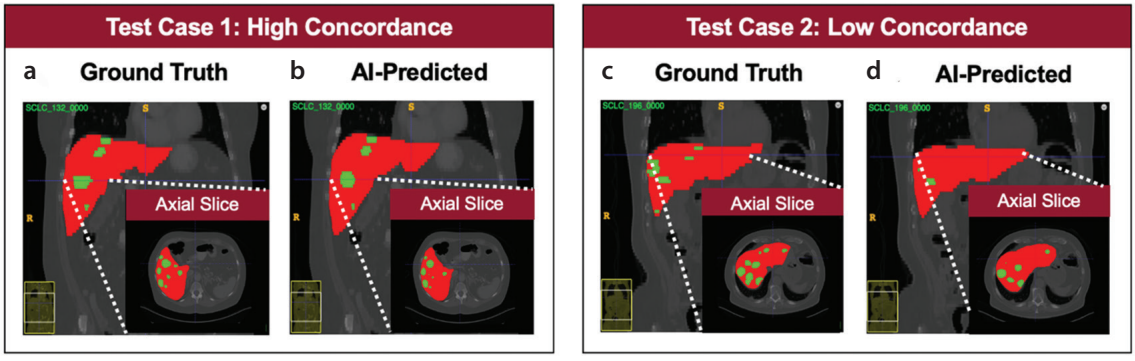



**Figure 2.** Representative results of the finetuned nnU-Net lesion detection model. Two metastatic small cell lung cancer (mSCLC) test cases are provided with liver (shown in red) and lesion (shown in green) segmentations from ground truth annotations and artificial intelligence (AI) predictions. A single axial slice from the three-dimensional segmentation is shown. **(a, b)** Example of an mSCLC test case with high concordance between the ground truth and AI-predicted lesions; DICE: 0.71. **(c, d)** Example of an mSCLC test case with low concordance between ground truth and AI-predicted lesions; DICE: 0.51.

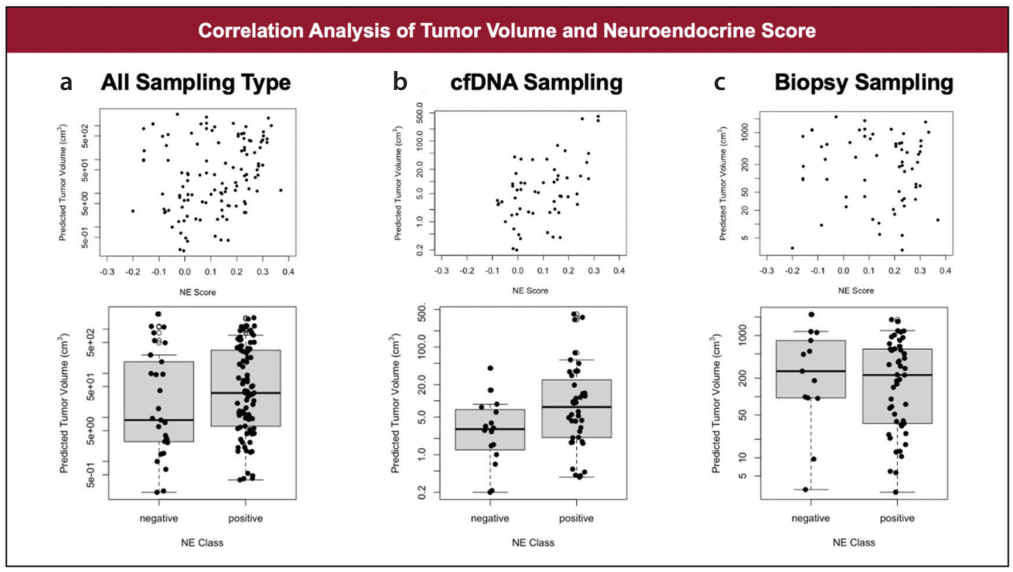
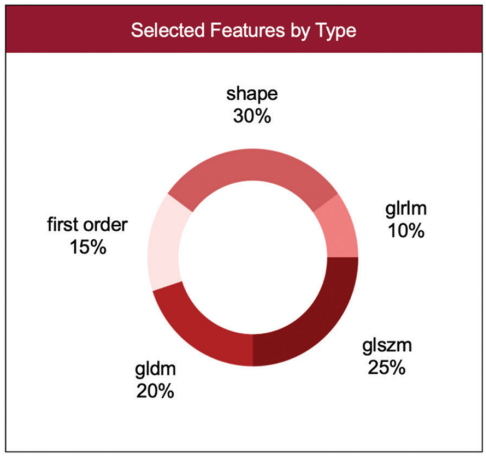



**Figure 3.** Correlation analysis of tumor burden to neuroendocrine scores. **(a)** There is a statistically significant correlation between tumor burden and neuroendocrine (NE) scores for both continuous data (Spearman: 0.252, *P* value: 0.0059) and binarized NE scores (Wilcoxon rank sum test statistic: 2.191, *P* value: 0.028). This analysis utilizes NE scores from both circulating free DNA (cfDNA) and biopsy samples. **(b)** Correlation plots of tumor burden and NE scores (top) and NE class (bottom) derived from cfDNA samples. There is a statistically significant (Spearman: 0.446, *P* value: 0.00057; Wilcoxon rank sum test statistic: 3.225, *P* value: 0.0013) tumor volume dependence in cfDNA-derived NE scores. **(c)** Correlation plots for biopsy-derived analysis of tumor burden and NE scores (top) or NE class (bottom) showing no statistically relevant relationship between biopsy-derived NE scores and tumor volume estimates.




| Group | Feature |
|---|---|
| Shape | Major Axis Length**<br>Maximum 2D Diameter Row**<br>Mesh Volume*<br>Minor Axis Length**<br>Surface Area<br>Voxel Volume |
| Gray Level Size Zone Matrix (glszm) | Gray Level Non-Uniformity*<br>Large Area Emphasis<br>Large Area High Gray Level Emphasis<br>Large Area Low Gray Level Emphasis<br>Zone Variance |
| Gray Level Dependence Matrix (gldm) | Dependence Non-Uniformity*<br>Gray Level Non-Uniformity<br>Large Dependence Emphasis**<br>Large Dependence High Gray Level Emphasis |
| First Order Statistics | Energy<br>Total Energy<br>Variance** |
| Gray Level Run Length Matrix (glrlm) | Gray Level Non-Uniformity*<br>Run Length Non-Uniformity* |
| ** included in top five imaging features<br>*included in top ten imaging features | |

**Figure 4.** Distribution of selected feature types and list of selected radiomics features. The distribution of each feature type is as follows: 30% shape, 25% gray level size zone matrix, 20% gray level dependent matrix, 15% first order statistics, 10% gray level run length matrix. Least absolute shrinkage and selection operator regression identified 20 features correlated to binarized neuroendocrine scores among 108 radiomic quantities. All selected radiomic features are listed by group.

**Table 4.** Scan-level mSCLC NE phenotype classification results in the test set, derived from ensemble of five-fold cross-validation. Reported are mean and 95% CIs from 1000 bootstrap samples, sampling one scan (series) per CT study date per iteration

| Model | Accuracy | PPV | NPV | Sensitivity | Specificity | F1 | AUC |
|---|---|---|---|---|---|---|---|
| **LR** | 0.79<br>(0.73, 0.85) | 0.81<br>(0.78, 0.83) | 0.64<br>(0.33, 1.00) | 0.95<br>(0.90, 1.00) | 0.25<br>(0.17, 0.33) | 0.87<br>(0.84, 0.91) | 0.71<br>(0.59, 0.82) |
| **RF** | 0.77<br>(0.69, 0.85) | 0.82<br>(0.77, 0.86) | 0.50<br>(0.25, 0.75) | 0.90<br>(0.85, 0.95) | 0.33<br>(0.17, 0.50) | 0.86<br>(0.81, 0.90) | 0.58<br>(0.48, 0.69) |
| **XGB** | 0.79<br>(0.73, 0.85) | 0.84<br>(0.81, 0.86) | 0.56<br>(0.40, 0.75) | 0.90<br>(0.85, 0.95) | 0.42<br>(0.33, 0.50) | 0.87<br>(0.83, 0.90) | 0.63<br>(0.48, 0.78) |

PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; mSCLC, metastatic small cell lung cancer; CT, computed tomography; CI, confidence interval

both series volumes, 4/24 scans were incorrectly classified in both series volumes, and 2/24 scans had different classification results across the different series. The two scans with only a single series volume available were both correctly classified.

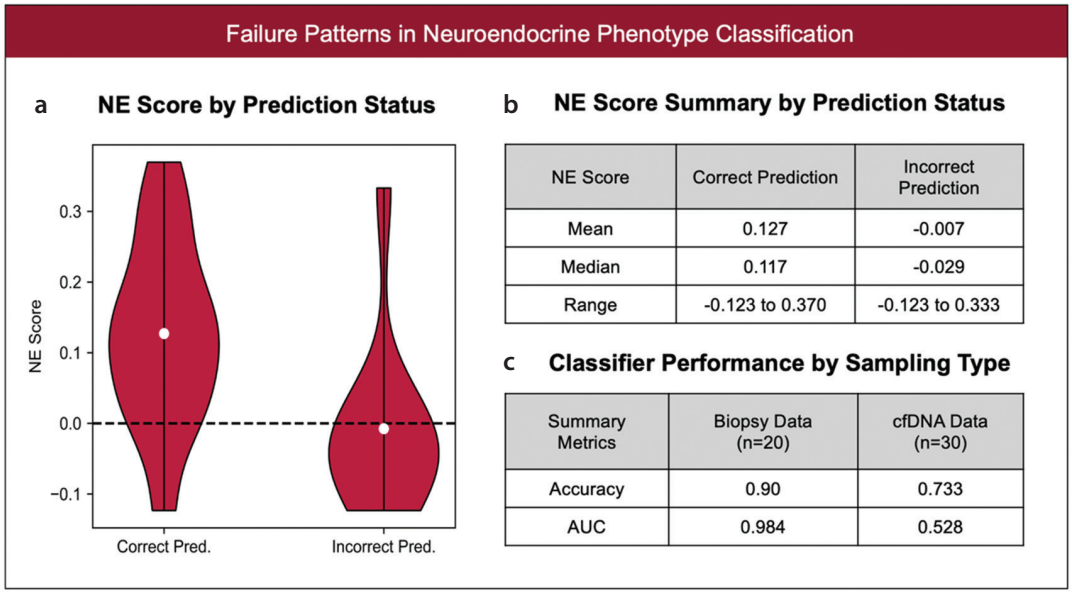
Failure analysis revealed that 8/10 of misclassifications were associated with NE scores near the boundary (NE score: 0) and with cfDNA-derived NE scores. Misclassified observations skewed to the NE-negative phenotype (Figure 5a), with a mean NE score of –0.007 as presented in Figure 5b. The NE classification results showed an AUC of 0.528 when evaluated on cfDNA-derived data alone, although they showed an AUC of 0.984 when strictly evaluated on biopsy-derived transcriptomic data, as shown in Figure 5c.

## Discussion

Patients with SCLC often present with hepatic metastases.[9,40] Characterizing NE profiles of mSCLC lesions offers a pathway to stratify patients based on distinct therapeutic vulnerabilities of their molecular subtypes.[13] We demonstrate the potential of deep learning on automated liver lesion detection and the feasibility of using radiomics to describe properties of mSCLC tumors. This framework can enable the determination of patients' NE status as positive or negative for bearing an NE-like phenotype without resorting to invasive biopsy procedures and relying on scans obtained as part of routine staging studies. Our 3D nnU-Net segmentation model demonstrated that hepatic lesions can be accurately detected for patient populations with highly variable disease characteristics (e.g., a wide range of tumor size, varying number of tumors in a single scan) using a fully automated platform. We also showed the possibility of stratifying patients with SCLC by their NE status using radiomic features extracted from routinely acquired abdominal CT scans of metastatic liver lesions.

The finetuned nnU-Net segmentation model was able to identify liver lesions with a median DICE score of 0.771 for the publicly available data (LiTS) and 0.640 for our internal mSCLC dataset. It also showed adeptness in locating regions with suspected lesions in CT images with high sensitivity (LiTS: 0.826 and mSCLC: 0.667) that outperformed previously reported models (ISBI 2017: 0.458, MICCAI 2017: 0.515, MICCAI 2018: 0.554)[25] while having low false positive rates (0–1 false positive per scan) and comparable tumor-level DICE scores (ISBI 2017: 0.674, MICCAI 2017: 0.702, MICCAI 2018: 0.739)[25] for the public dataset. Further characterization of the model's behavior on mSCLC cases without ground truth reports of a lesion (n = 94) revealed that false positive predictions typically were small in volume (0.638 ± 1.73 $cm^3$). Overall, our detection tool not only offers valuable improvements to previous liver lesion segmentation benchmarks but also enables automating the lesion annotation process for radiomics analysis.

Correlation analysis of AI-predicted lesions showed that tumor burden has a statistically significant association with NE status. Our results revealed that NE scores exhibit dependence on tumor volume, especially when expression profiles were derived from cfDNA samples. This finding is consistent with previous studies showing a correlation between tumor burden and cfDNA in NE-related[34,41] and lung[42] neoplasms, although their exact relationship remains elusive.[43] This is also supported by our radiomics analysis, shows that shape—which includes measurements of tumor size—is the most frequently selected radiomic feature type during model development. These findings indicate that tumor burden may be an indirect measure of NE expression in mSCLC tumors. Recent studies have demonstrated a strong correlation between cfDNA-derived and RNA-derived expression scores for NE phenotyping.[13,34-36] However, the same volume-based relationship was not observed in biopsy-based RNAseq determination of NE phenotype. This discrepancy may be explained by tumor heterogeneity, where cfDNA-derived expression metrics are an aggregate of multiple lesions throughout




| NE Score | Correct Prediction | Incorrect Prediction |
|---|---|---|
| Mean | 0.127 | -0.007 |
| Median | 0.117 | -0.029 |
| Range | -0.123 to 0.370 | -0.123 to 0.333 |

| Summary Metrics | Biopsy Data (n=20) | cfDNA Data (n=30) |
|---|---|---|
| Accuracy | 0.90 | 0.733 |
| AUC | 0.984 | 0.528 |

**Figure 5.** Failure analysis of the neuroendocrine (NE) phenotype classifier. **(a)** The NE score distribution based on prediction status (correct vs. incorrect). The mean NE score of observations associated with incorrect predictions is close to the NE score boundary (NE score: 0) that divides between NE-positive and NE-negative phenotypes. **(b)** Summary of the NE scores by prediction status. Correctly predicted observations had an NE score range of –0.123 to 0.370, a mean NE score of 0.127, and a median NE score of 0.117. Incorrectly predicted observations had a score range of –0.123 to 0.333, a mean NE score of –0.007, and a median NE score of –0.029. **(c)** Evaluation of classifier performance based on sampling type [biopsy vs. circulating free DNA (cfDNA)] revealed better predictive performance on biopsy data than cfDNA data. The NE phenotype classifier had an accuracy of 0.90 and an area under the receiver operating characteristic curve (AUC) of 0.984 for observations associated with biopsy sampling; the classifier achieved an accuracy of 0.733 and an AUC of 0.528 for cases associated with cfDNA tissue sampling.

the patient, whereas biopsy-derived RNAseq expression is sampled directly from a single lesion. This heterogeneity component can also explain why more errors in NE prediction were observed in cfDNA-derived samples, as metastatic lesions elsewhere in the body may contribute to this expression, despite the predominance of liver lesions in these patients. However, these hypotheses cannot be evaluated within the patients evaluated in this study due to the limited sample size and lack of multiple targeted biopsy-based samples for RNAseq expression.

To our knowledge, techniques for predicting patient-level SCLC NE status have yet to be explored. Our approach in integrating the molecular phenotypic landscape of SCLC with image-extracted tumor markers offers a path towards building translational computing workflows that may help tailor SCLC treatment. In this study, we show that a logistic regression classifier can distinguish NE phenotypes using radiomics data with 80% accuracy (0.79 ± 0.04) and an AUC of 0.73 (0.70 ± 0.08). We noted that the phenotype classifier demonstrated high sensitivity for NE-positive tumors but low specificity for NE-negative tumors. This classifier was trained and tested on heterogeneous image acquisition settings, leading to consistency in performance for multiple reconstructions (series volumes) within a single study; however, further validation is warranted.

Our approach has several limitations. This is a relatively small patient cohort of metastatic patients with SCLC; the classification algorithm was trained and validated in 227 scans from 50 patients, and further research is warranted. We were underpowered to perform classification based on the Riley et al.[44] criterion, increasing the likelihood of our algorithm overfitting to the training population. Our data utilize a mixture of biopsy-sampled and cfDNA-based NE scores. The cfDNA sampling is comparatively easier but captures signals from both cancerous and non-cancerous components. This impacts our ability to describe with confidence the NE expression profiles specific to each lesion. Next, our framework utilizes a cascaded algorithm that analyzes biomedical imaging data in a stepwise fashion. This workflow inherently propagates segmentation errors and failures at the first step to the radiomics analysis, which occurs downstream. Since radiomics is volume dependent, the effectiveness of the tumor phenotype classifier relies on the performance of the lesion detection tool. Our DICE performance indicates that the model did not perform as well in mSCLC cases (DICE <0.7) compared with LiTS (DICE >0.7). This may be influenced by high disease burden and complex anatomy from extensive prior treatments, or it can affect cases with very small lesions, which are difficult to identify. Further work on how changes in predicted volumes impact radiomics and downstream classification is needed. Similarly, further investigation into the effects of contrast use and acquisition parameters is warranted. Third, our NE classification task is conducted at the image level rather than the lesion level. This primarily describes the bulk NE profile rather than tumor-specific characteristics. Any heterogeneity reflected across the patients' disease burden cannot be evaluated through the current workflow. Future validation in lesion-based assessment for samples with targeted sequencing may provide more context to individual signatures. Finally, the data distribution of NE positive to NE negative is largely unbalanced, making it difficult to further optimize the machine learning classifier to better detect NE-negative phenotypes.

In conclusion, deep learning and radiomics-based analysis enable automated detection and characterization of SCLC liver metastasis. Using AI-based platforms, routinely acquired CT scans may be used to determine the NE status of patients with mSCLC liver lesions. This could enable clinicians to tailor SCLC treatments based on a patient's NE status and its associated molecular tumor profile.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

### Funding

This research was supported in part by the Center for Cancer Research, National Cancer Institute, National Institutes of Health Intramural Research Program project number ZIABC012163 and ZIABC011793. The research was supported in part by the NIH Undergraduate Scholarship Program (S.T.). The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

**Supplementary Tables:** https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/208802b3-a82d-4db5-b583-16e14886d8f3.pdf

## References

1. van Meerbeeck JP, Fennell DA, De Ruysscher DK. Small-cell lung cancer. *Lancet*. 2011;378(9804):1741-1755. [Crossref]
2. Thomas A, Pommier Y. Small cell lung cancer: time to revisit DNA-damaging chemotherapy. *Sci Transl Med*. 2016;8(346):346fs12. [Crossref]
3. Thomas A, Mohindroo C, Giaccone G. Advancing therapeutics in small-cell lung cancer. *Nat Cancer*. 2025;6(6):938-953. [Crossref]
4. George J, Lim JS, Jang SJ, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature*. 2015;524(7563):47-53. [Crossref]
5. Sher T, Dy GK, Adjei AA. Small cell lung cancer. *Mayo Clin Proc*. 2008;83(3):355-367. [Crossref]
6. Thomas A, Pattanayak P, Szabo E, Pinsky P. Characteristics and outcomes of small cell lung cancer detected by CT screening. *Chest*. 2018;154(6):1284-1290. [Crossref]
7. Herzog BH, Devarakonda S, Govindan R. Overcoming chemotherapy resistance in SCLC. *J Thorac Oncol.* 2021;16(12):2002-2015. [Crossref]
8. Jackman DM, Johnson BE. Small-cell lung cancer. *Lancet*. 2005;366(9494):1385-1396. [Crossref]
9. Cai H, Wang H, Li Z, Lin J, Yu J. The prognostic analysis of different metastatic patterns in extensive-stage small-cell lung cancer patients: a large population-based study. *Future Oncol*. 2018;14(14):1397-1407. [Crossref]
10. Kagohashi K, Satoh H, Ishikawa H, Ohtsuka M, Sekizawa K. Liver metastasis at the time of initial diagnosis of lung cancer. *Med Oncol*. 2003;20(1):25-28. [Crossref]
11. Rudin CM, Poirier JT, Byers LA, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer*. 2019;19(5):289-297. [Crossref]
12. Schroeder BA, Thomas A. SCLC subtypes and biomarkers of the transformative immunotherapy responses. *J Thorac Oncol.* 2023;18(9):1114-1117. [Crossref]
13. Lissa D, Takahashi N, Desai P, et al. Heterogeneity of neuroendocrine transcriptional states in metastatic small cell lung cancers and patient-derived models. *Nat Commun*. 2022;13(1):2023. [Crossref]
14. Thomas A, Takahashi N, Rajapakse VN, et al. Therapeutic targeting of ATR yields durable

regressions in small cell lung cancers with high replication stress. *Cancer Cell*. 2021;39(4):566-579.e7. [Crossref]

15. Takahashi N, Kim S, Schultz CW, et al. Replication stress defines distinct molecular subtypes across cancers. *Cancer Res Commun*. 2022;2(6):503-517. [Crossref]
16. Roper N, Velez MJ, Chiappori A, et al. Notch signaling and efficacy of PD-1/PD-L1 blockade in relapsed small cell lung cancer. *Nat Commun*. 2021;12(1):3880. [Crossref]
17. Gay CM, Stewart CA, Park EM, et al. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell*. 2021;39(3):346-360.e7. [Crossref]
18. Nabet BY, Hamidi H, Lee MC, et al. Immune heterogeneity in small-cell lung cancer and vulnerability to immune checkpoint blockade. *Cancer Cell*. 2024;42(3):429-443.e4. [Crossref]
19. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. A review of deep learning based methods for medical image multi-organ segmentation. *Phys Med*. 2021;85:107-122. [Crossref]
20. Jiang H, Zhou Y, Lin Y, Chan RCK, Liu J, Chen H. Deep learning for computational cytology: a survey. *Med Image Anal*. 2023;84:102691. [Crossref]
21. Alakwaa W, Nassef M, Badr A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Int J Adv Comput Sci Appl*. 2017;8(8). [Crossref]
22. Yu X, Jin F, Luo H, Lei Q, Wu Y. Gross tumor volume segmentation for stage III NSCLC radiotherapy using 3D ResSE-Unet. *Technol Cancer Res Treat*. 2022;21:15330338221090847. [Crossref]
23. Christ PF, Elshaer MEA, Ettlinger F, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Vol 9901. Lecture Notes in Computer Science. Springer International Publishing; 2016:415-423. [Crossref]
24. Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci Rep*. 2018;8(1):15497. [Crossref]
25. Bilic P, Christ P, Li HB, et al. The liver tumor segmentation benchmark (LiTS). *Med Image Anal*. 2023;84:102680. [Crossref]
26. Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. *Strahlenther Onkol*. 2020;196(10):879-887. [Crossref]
27. Gitto S, Interlenghi M, Cuocolo R, et al. MRI radiomics-based machine learning for classification of deep-seated lipoma and atypical lipomatous tumor of the extremities. *Radiol Med*. 2023;128(8):989-998. [Crossref]
28. Menon N, Guidozzi N, Chidambaram S, Markar SR. Performance of radiomics-based artificial intelligence systems in the diagnosis and prediction of treatment response and survival in esophageal cancer: a systematic review and meta-analysis of diagnostic accuracy. *Dis Esophagus*. 2023;36(6):doad034. [Crossref]
29. Abbaspour S, Barahman M, Abdollahi H, et al. Multimodality radiomics prediction of radiotherapy-induced the early proctitis and cystitis in rectal cancer patients: a machine learning study. *Biomed Phys Eng Express*. 2023;10(1). [Crossref]
30. Lubner MG, Stabo N, Lubner SJ, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging*. 2015;40(7):2331-2337. [Crossref]
31. Kiryu S, Akai H, Nojima M, et al. Impact of hepatocellular carcinoma heterogeneity on computed tomography as a prognostic indicator. *Sci Rep*. 2017;7(1):12689. [Crossref]
32. Akai H, Yasaka K, Kunimatsu A, et al. Predicting prognosis of resected hepatocellular carcinoma by radiomics analysis with random survival forest. *Diagn Interv Imaging*. 2018;99(10):643-651. [Crossref]
33. Chaudhary U, Desai PA, Takahashi N, et al. Automated detection and segmentation of small cell lung cancer liver metastases on CT. *Journal of Clinical Oncology*. 2022;40(16_suppl):e13555. [Crossref]
34. Fialkoff G, Takahashi N, Sharkia I, et al. Subtyping of small cell lung cancer using plasma cell-free nucleosomes. *bioRxiv*. 2022. [Crossref]
35. Zhang W, Girard L, Zhang YA, et al. Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. *Transl Lung Cancer Res*. 2018;7(1):32-49. [Crossref]
36. Hiatt JB, Doebley AL, Arnold HU, et al. Molecular phenotyping of small cell lung cancer using targeted cfDNA profiling of transcriptional regulatory regions. *Sci Adv*. 2024;10(15):eadk2082. [Crossref]
37. Medical Open Network for Artificial Intelligence (MONAI). Accessed March 11, 2024. [Crossref]
38. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. [Crossref]
39. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11(2):178-189. [Crossref]
40. Nakazawa K, Kurishima K, Tamura T, et al. Specific organ metastases and survival in small cell lung cancer. *Oncol Lett*. 2012;4(4):617-620. [Crossref]
41. Mettler E, Fottner C, Bakhshandeh N, Trenkler A, Kuchen R, Weber MM. Quantitative Analysis of Plasma Cell-Free DNA and Its DNA Integrity and hypomethylation status as biomarkers for tumor burden and disease progression in patients with metastatic neuroendocrine neoplasias. *Cancers (Basel)*. 2022;14(4):1025. [Crossref]
42. Hyun MH, Lee ES, Eo JS, et al. Clinical implications of circulating cell-free DNA quantification and metabolic tumor burden in advanced non-small cell lung cancer. *Lung Cancer*. 2019;134:158-166. [Crossref]
43. Nygaard AD, Holdgaard PC, Spindler KL, Pallisgaard N, Jakobsen A. The correlation between cell-free DNA and tumour burden was estimated by PET/CT in patients with advanced NSCLC. *Br J Cancer*. 2014;110(2):363-368. [Crossref]
44. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276-1296. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253489

ARTIFICIAL INTELLIGENCE AND INFORMATICS

COMMENTARY



# Turkish Society of Radiology artificial intelligence applications guide: a roadmap to help navigate the artificial intelligence landscape

Oğuz Dicle[1]
Fırat Atak[2]
Abdullah Utku Şenol[3]
Mustafa Nasuh Özmen[4]
Nur Hürsoy[5]
Naciye Sinem Gezer[1]
Üstün Aydıngöz[4,6]

[1]Dokuz Eylül University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[2]Mamak State Hospital, Clinic of Radiology, Ankara, Türkiye

[3]Akdeniz University Faculty of Medicine, Department of Radiology, Antalya, Türkiye

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

[5]Recep Tayyip Erdoğan University Faculty of Medicine, Department of Radiology, Rize, Türkiye

[6]Saint Louis University School of Medicine, Department of Radiology, St. Louis, USA

**Handling editor:** Burak Koçak

**Corresponding author:** Oğuz Dicle

**E-mail:** Odicle59@gmail.com

Received 10 June 2025; revision requested 01 July 2025; accepted 28 July 2025.

Epub: 16.10.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253489

This article introduces the artificial intelligence (AI) applications guide developed by the Turkish Society of Radiology (TSR) and provides a commentary on its framework, strengths, and areas for future development. The proliferation of accessible data, accelerated information processing capabilities, and the development of advanced algorithmic tools such as convolutional neural networks have led to a significant leap in AI studies. The field of radiology, with its growing imaging workload and reliance on archivable data, has naturally assumed a pioneering role in AI applications.

Amid ongoing concerns that AI may pose a threat to the profession, the TSR recognized its responsibility to support its members by providing a guide that promotes the responsible use of AI to enhance diagnostic speed, accuracy, and efficiency–ultimately benefiting patient care. The preparation of the guide involved a dedicated team effort, beginning with a comprehensive literature review, followed by expert-led identification of the main sections, and collaborative content development through group discussions.

This article introduces the guide, assesses its strengths and areas for improvement, and encourages feedback from the scientific community to inform future updates. The guide was published in both Turkish and English on the society's website on April 10, 2025.[1]

## Overview of the guide and key contents

The TSR AI applications guide is organized into six sections, structured around three core thematic pillars that provide a clear framework for understanding its content. First, the guide establishes foundational elements by outlining the rationale for its development and defining key AI terminology. These definitions are supported by the TSR Imaging Informatics Dictionary,[2] which provides a more comprehensive glossary. This section also presents a set of general principles in bullet-point format for clarity. These principles, developed in line with international recommendations, address the fundamental elements of health ethics–such as respect for autonomy, beneficence, informed consent, and confidentiality–in the context of AI.[3-6]

Second, the guide outlines the AI application development process, offering clear guidance and explanations for each phase. It covers the identification of clinical problems, data management (collection, curation, and privacy protection), the critical process of image labeling and annotation, and a comparison of traditional machine learning versus deep learning methodologies. It also outlines the selection of appropriate performance metrics for different tasks, emphasizes the need for model calibration, and highlights the importance of evaluating true clinical benefit in addition to discriminatory performance. Challenges related to AI interpretability and explainability are also addressed, along with strategies to improve transparency.

Third, the guide focuses on the clinical deployment and use of AI applications, as well as the responsibility issues that arise through their use. It lists principles to be followed when selecting and using AI applications in clinical settings and for informing patients about their

You may cite this article as: Dicle O, Atak F, Şenol AU, et al. Turkish Society of Radiology artificial intelligence applications guide: a roadmap to help navigate the artificial intelligence landscape. *Diagn Interv Radiol*. 2026;32(4):391-392.

use. This section also specifies the required reliability levels, provides considerations for quality control and continuous improvement, and addresses the responsibilities of the parties involved.

### Strengths and limitations

Combining a comprehensive framework with evidence-based content, the guide serves as an accessible and practical resource for radiologists, data scientists, researchers, and other stakeholders.

As with other rapidly developing technologies, one of the main difficulties in developing an AI guide lies in defining its theoretical scope while maintaining up-to-date information. To maintain clarity and focus, the guide intentionally omits extensive quantitative data and detailed application examples, which, though potentially helpful, would have compromised its conciseness. Furthermore, it should be noted that the guide does not fully encompass the complex and ongoing ethical and social discussions surrounding the clinical adoption of AI technologies.

### Future directions

A fundamental goal of the guide is to ensure periodic updates, informed by community feedback, to keep pace with rapid technological developments. Key areas for future consideration should include emerging topics such as best practices for clinical deployment strategies, the establishment of standardized benchmarking protocols for model comparison, and the impact of large-scale foundation models on radiology.

Future content should address the practical challenges of deployment, including information technology infrastructure requirements, integration with picture archiving and communication systems and electronic health records.

To ensure fair and transparent comparisons between AI models, the field needs standardized benchmarking protocols. Future guidance could explore the creation and use of curated, public datasets and consensus-based evaluation frameworks, allowing for reproducible and objective performance assessments of new technologies.

The paradigm is shifting from narrow, single-task algorithms to large-scale, pre-trained foundation models that can be adapted to multiple tasks. Subsequent versions of the guide should include recommendations on the unique challenges and opportunities presented by these models, including fine-tuning strategies, human–computer interactions (e.g., overreliance and automation bias), managing their inherent complexity (e.g., hallucinations), and ethical and societal implications, such as high costs and environmental impact.

In conclusion, this guide is more than a collection of recommendations–it provides a structured framework for standardizing the development and clinical integration of AI in radiology. By systematically addressing each stage of the AI implementation process, it offers a practical and actionable roadmap. Through its introduction in this journal and commitment to ongoing updates, we aim to foster responsible innovation among developers and clinicians alike.

### Footnotes

### Conflict of interest disclosure

Oğuz Dicle was not serving in any editorial role at the time of the manuscript's submission, evaluation, acceptance, or ahead of print publication. At the time of inclusion of the article in the regular issue, he is serving as Section Editor of the journal. Oğuz Dicle was not involved in the editorial evaluation, peer-review process, or publication decision of this manuscript. The other authors declare no conflict of interest.


### Funding

This work was supported by the TSR (Turkish Society of Radiology).


## References


1. Turkish Society of Radiology. Yapay Zeka Uygulamaları Rehberi. Turkish Society of Radiology website. Accessed September 30, 2025. [Crossref]
2. Turkish Society of Radiology. TRD-GBS Imaging Informatics Dictionary. Safari website. Published 2025. Accessed September 30, 2025. [Crossref]
3. OECD. OECD AI Principles. OECD.AI website. Published May 2019. Accessed September 30, 2025. [Crossref]
4. UNESCO. Recommendation on the Ethics of Artificial Intelligence. UNESCO website. Published November 2021. Accessed September 30, 2025. [Crossref]
5. Council of Europe. Guidelines on Artificial Intelligence and Data Protection. Council of Europe website. Published January 2019. Accessed September 30, 2025. [Crossref]
6. EU AI Act: [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263888

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR



# Reply: Comment on the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans

Olga Bayar Kapıcı[1]
Erman Altunışık[2]
Feyza Musabeyoğlu[2]
Şeyda Dev[2]
Ömer Kaya[3]

[1]Seyhan State Hospital, Clinic of Radiology, Adana, Türkiye

[2]University of Health Sciences Türkiye, Gaziantep City Hospital, Clinic of Neurology, Gaziantep, Türkiye

[3]Çukurova University Faculty of Medicine, Department of Radiology, Adana, Türkiye

**Dear Editor,**

We would like to thank the correspondent for their thoughtful and constructive comments on our article evaluating the diagnostic sensitivity of a multimodal large language model (MLLM) (ChatGPT-4V) for detecting intracranial hemorrhage in non-contrast cranial computed tomography.[1] We appreciate the opportunity to clarify the rationale behind our experimental design and to outline directions for future work.

First, we agree that subtle and borderline hemorrhagic findings represent a well-known diagnostic gray zone, even for human readers, and that discordance may partly reflect intrinsic interpretive uncertainty rather than a purely model-specific limitation.[1] In our dataset, ChatGPT-4V's performance was clearly influenced by lesion conspicuity; larger hemorrhage diameters were associated with higher correct classification rates, particularly for epidural and subdural hematomas.[2] This finding is consistent with the broader literature showing that MLLM performance using direct image inputs remains variable across tasks and settings and may be limited in the context of real-world radiologic interpretation.[3,4]

Second, we fully concur that clinical contexts can materially shape diagnostic reasoning.[1] Our study intentionally adopted an image-only framework to quantify baseline behavior under controlled conditions and to isolate the effect of the prompt structure. Specifically, after an initial open-ended prompt (Q2), we introduced a minimal, targeted clue ("There is bleeding…") (Q3) to test whether structured guidance influences performance.[2] The substantial improvement observed with this guided prompt supports the correspondent's emphasis on input conditions and prompt engineering.[2] It also aligns with published radiology-focused research indicating that prompt optimization (including structured prompting and few-shot approaches) can meaningfully influence LLM outputs and utility.[5]

Third, regarding the reliance on one or two preselected slices and the absence of dynamic window/level adjustments, we agree this differs from the routine radiologic workflow, in which multi-slice review and interactive windowing are integral, especially for subtle hemorrhage and artifact discrimination.[1] In our Methods section, we provided representative two-dimensional slices to approximate a best-case static-input scenario.[2] We acknowledge that a workflow-faithful evaluation would ideally allow multi-slice correlation (or a full-series review) and window/level control. These priorities are also reflected in broader multimodal GPT-4V radiology evaluations that highlight sensitivity to input presentation and context handling.[6]

Finally, we strongly support the safety considerations highlighted by the correspondent.[1] In our conclusion, we emphasized that the model is not suitable for autonomous radiologic interpretation and should be considered, at most, as a supervised adjunct within human-in-the-loop paradigms.[2] This caution is consistent with the emerging radiology-related literature emphasizing that MLLMs that use direct image input have not yet reached a level appropriate for unsupervised clinical deployment.[3,4,6]

**Handling editor:** Burak Koçak

**Corresponding author:** Olga Bayar Kapıcı

**E-mail:** olgasahbayar@gmail.com

Received 20 January 2026; accepted 31 January 2026.

Epub: 16.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263888

You may cite this article as: Kapıcı OB, Altunışık E, Musabeyoğlu F, Dev Ş, Kaya Ö. Reply: comment on the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol*. 2026;32(4):393-394.

We thank the correspondent again for their insightful remarks, which closely align with the key implications of our findings and help frame a clear agenda for clinically meaningful and safe evaluation of multimodal language–vision models in radiology.

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Kartal YC. Letter to the Editor: comment on the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol.* 2026;32(4):519-520. [Crossref]
2. Bayar-Kapıcı O, Altunışık E, Musabeyoğlu F, Dev Ş, Kaya Ö. Artificial intelligence in radiology: diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol*. 2026;32(1):27-32. [Crossref]
3. Zhou Y, Ong H, Kennedy P, et al. Evaluating GPT-V4 (GPT-4 with Vision) on detection of radiologic findings on chest radiographs. *Radiology*. 2024;311(2):e233270. Erratum in: *Radiology*. 2024;311(2):e249016. [Crossref]
4. Suh PS, Shim WH, Suh CH, et al. Comparing diagnostic accuracy of radiologists versus GPT-4V and Gemini Pro Vision using image inputs from diagnosis please cases. Radiology. 2024;312(1):e240273. [Crossref]
5. Russe MF, Reisert M, Bamberg F, Rau A. Improving the use of LLMs in radiology through prompt engineering: from precision prompts to zero-shot learning. Rofo. 2024;196(11):1166-1170. [Crossref]
6. Busch F, Han T, Makowski MR, Truhn D, Bressem KK, Adams L. Integrating text and image analysis: exploring GPT-4V's capabilities in advanced radiological applications across subspecialties. J Med Internet Res. 2024;26:e54948. Erratum in: J Med Internet Res. 2024;26:e64411. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.264022

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR



# Letter to the Editor: Rethinking reference accuracy in large language models for radiology

Hamza Eren Güzel

Ministry of Health, İzmir City Hospital, Department of Radiology, İzmir, Türkiye

**Dear Editor,**

I read with great interest the recent article by Güneş et al.[1] evaluating the reference accuracy of large language models (LLMs) in radiology. The authors should be congratulated for addressing a critical and timely issue—namely, the high rates of fabricated and inaccurate citations generated by contemporary LLMs. Their findings clearly demonstrate that, despite rapid advancements, substantial limitations remain in the reliability of LLM-generated academic references.

Although reference accuracy represents an important component of LLM evaluation, it may not fully capture the complexity of real-world model performance. One critical yet underexplored aspect is the temporal variability of LLM outputs. Due to stochastic decoding processes, model updates, and backend modifications, identical prompts may yield different responses across sessions or time points. This variability has direct implications for study design and interpretation. In the study by Güneş et al.,[1] each model was evaluated using a single response per query, which, although methodologically practical, may not fully reflect the range of possible outputs in real-world usage. Consequently, a model demonstrating acceptable reference accuracy in a single instance may still exhibit inconsistent or unreliable behavior across repeated interactions.

Another important consideration is that reference accuracy represents only one dimension of a broader construct encompassing clinical reasoning, contextual understanding, and decision relevance. In a recent study evaluating LLM performance in an examination-style radiology setting modeled after the European Diploma in Radiology, discrepancies were observed between diagnostic reasoning performance and the quality or validity of the supporting information.[2] Specifically, models were able to generate clinically plausible answers despite inconsistencies in explanations or evidentiary support, suggesting a disconnect between linguistic coherence and factual grounding.

This limitation is closely related to the phenomenon of hallucination, in which LLMs generate syntactically plausible but factually incorrect information. Multiple studies across different domains and model architectures have consistently demonstrated high rates of fabricated or inaccurate references in LLM-generated outputs.[3-5] Importantly, this issue is not confined to a single model family but appears to be a widely observed limitation of current generative artificial intelligence systems. As highlighted by Güneş et al.,[1] such inaccuracies may introduce misinformation into both clinical and academic contexts. Furthermore, erroneous citations may propagate through secondary referencing, ultimately distorting the scientific record and undermining evidence-based practice.[6]

These concerns extend beyond technical limitations and raise broader issues related to scientific integrity, reproducibility, and knowledge verification. As LLMs become increasingly integrated into academic writing and clinical decision support, the need for critical appraisal and human oversight becomes paramount. Without rigorous validation, reliance on generated content may inadvertently compromise both the quality of scientific output and patient safety.

**Corresponding author:** Hamza Eren Güzel

**E-mail:** hamzaerenguzel@gmail.com

Received 26 March 2026; accepted 30 March 2026.

Epub: 08.05.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.264022

You may cite this article as: Güzel HE. Letter to the editor: rethinking reference accuracy in large language models for radiology. *Diagn Interv Radiol*. 2026;32(4):395-396.

In this context, future evaluations of LLMs in radiology would benefit from approaches that account for response variability and integrate multiple performance dimensions, including clinical correctness, reasoning transparency, and hallucination risk. Such comprehensive frameworks may provide a more accurate representation of model capabilities and limitations, ultimately supporting safer and more effective implementation in both research and clinical practice.

In conclusion, the study by Güneş et al.[1] provides valuable insights into the current limitations of LLM-generated references. However, reference accuracy should be interpreted within a broader and temporally aware evaluation framework. Addressing these challenges will be essential for the responsible integration of LLMs into radiology.

### Footnotes

### Conflict of interest disclosure

The author declared no conflicts of interest.

## References


1. Güneş YC, Cesur T, Çamur E. Evaluating the reference accuracy of large language models in radiology: a comparative study across subspecialties. *Diagn Interv Radiol.* 2026;32(2):173-181. [Crossref]
2. Güzel HE, Oleaga L, Koç AM, Junquero V, Merino C. Large language models solving the European Diploma in Radiology: a comparative evaluation. *Acad Radiol*. 2026;33(5):1871-1878. [Crossref]
3. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus*. 2023;15(5):e39238. [Crossref]
4. Chelli M, Descamps J, Lavoué V, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. *J Med Internet Res.* 2024;26:e53164. [Crossref]
5. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep*. 2023;13:14045. [Crossref]
6. Dumas-Mallet E, Boraud T, Gonon F. Le mésusage des citations et ses conséquences en médecine [Citation misuse and its effects on public health]. *Med Sci (Paris).* 2021;37(11):1035-1041. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.264103

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR



# Reply: Evaluating the reference accuracy of large language models in radiology: a comparative study across subspecialties

Yasin Celal Güneş[1]
Turay Cesur[2]
Eren Çamur[3]

[1]Kırıkkale Yüksek İhtisas Hospital, Clinic of Radiology, Kırıkkale, Türkiye

[2]Mamak State Hospital, Clinic of Radiology, Ankara, Türkiye

[3]Ankara 29 Mayıs State Hospital, Clinic of Radiology, Ankara, Türkiye

**Dear Editor,**

We thank the authors for their thoughtful and well-articulated comments on our study.[1]

We agree that large language model (LLM) performance is inherently multidimensional and that temporal variability and response stochasticity are important considerations. As outlined in our Discussion, these aspects—including the use of a single response per query and the absence of repeated sampling—were explicitly acknowledged as limitations of our study.[2]

Our research was intentionally designed to provide a standardized baseline comparison across models. The single-response-per-query approach was adopted to ensure methodological consistency and comparability while recognizing that it does not capture the full variability of LLM outputs. In this context, the points raised by the authors are valid and consistent with the methodological considerations outlined in our manuscript.

We also concur that reference accuracy represents only one component of overall LLM performance. However, we believe it remains a particularly critical domain in radiology, where clinical and academic practice depends on accurate and verifiable sources.[3,4] From this perspective, our focused evaluation addresses a fundamental aspect of LLM reliability.

The authors' emphasis on hallucination is particularly relevant. Our findings are consistent with prior studies demonstrating that fabricated or inaccurate references remain a persistent limitation across current LLMs, reinforcing the need for careful validation and human oversight.[5-7]

We agree that future research incorporating repeated sampling and broader performance metrics will further enhance the understanding of LLM behavior. Within this context, we believe our study provides a necessary and timely benchmark in this domain.

We thank the authors again for their valuable contribution to this discussion.

## Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References

1. Güzel HE. Letter to the editor: rethinking reference accuracy in large language models for radiology. Diagn Interv Radiol. 2026;32(4):395-396. [Crossref]
2. Güneş YC, Cesur T, Çamur E. Evaluating the reference accuracy of large language models in radiology: a comparative study across subspecialties. *Diagn Interv Radiol*. 2026;32(2):173-181. [Crossref]
3. 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]

**Corresponding author:** Yasin Celal Güneş

**E-mail:** gunesyasincelal@gmail.com

Received 01 May 2026; accepted 05 May 2026.

Epub: 08.05.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.264103

You may cite this article as: Güneş YC, Cesur T, Çamur E. Reply: evaluating the reference accuracy of large language models in radiology: a comparative study across subspecialties. *Diagn Interv Radiol*. 2026;32(4):397-398.

4. Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. *Can Assoc Radiol J*. 2024;75(1):69-73. [Crossref]

5. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus*. 2023;15(5):e39238. [Crossref]

6. McGowan A, Gui Y, Dobbs M, et al. ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. *Psychiatry Res*. 2023;326:115334. [Crossref]

7. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep*. 2023;13(1):14045. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263831

BREAST IMAGING

ORIGINAL ARTICLE



# Comparative analysis of the diagnostic value of dynamic contrast-enhanced magnetic resonance imaging kinetic heterogeneity and apparent diffusion coefficient for grading invasive breast cancer

Xinyu Feng[1]*
Peiwei Ye[2,3]*
Hui Chen[1]
Changyu Liu[1]
Qingqiang Zhu[1]

[1]Northern Jiangsu People's Hospital Affiliated to Yangzhou University, Department of Medical Imaging, Yangzhou, China

[2]Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Department of Radiation Oncology, Shanghai, China

[3]Shanghai Key Laboratory of Proton-Therapy, Shanghai, China

**PURPOSE**

To quantitatively compare the diagnostic value of dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) kinetic heterogeneity and conventional diffusion-weighted imaging (DWI) for the extent of breast cancer infiltration.

**METHODS**

This study employed a retrospective analysis of DCE-MRI data of patients with invasive breast cancer (IBC) diagnosed by pathology in our hospital between January 2023 and February 2025. The aim was to obtain quantitative measures of kinetic heterogeneity and apparent diffusion coefficients (ADCs) from DWI, and to extract the six main parameters for lesion heterogeneity analysis from preoperative MRI data using MATLAB, SPM12, and R 4.4.1. The parameters included peak, enhancement volume, persistent fraction, plateau, washout, and heterogeneity. The diagnostic efficacy of DCE-MRI, conventional DWI, and their combination on the extent of IBC infiltration was compared by analyzing the receiver operating characteristic curves, sensitivities, specificities, and correlations among the parameters.

**RESULTS**

The high-grade group exhibited significantly higher peak, plateau, washout, and heterogeneity values, along with lower persistent and ADC values, compared with the low-grade group (all $P < 0.001$); tumor volume did not differ between groups ($P = 0.314$). ADC and persistent fractions were negatively correlated with pathological grade, whereas peak, plateau, washout, and heterogeneity were positively correlated. Receiver operating characteristic analysis showed that heterogeneity achieved a significantly higher area under the curve (AUC) than ADC [0.910, 95% confidence interval (CI): 0.857–0.948 vs. 0.808, 95% CI: 0.741–0.863; DeLong Z = 2.626, $P = 0.009$]. The AUC for the combined model of heterogeneity, peak value, and ADC was 0.969 (95% CI: 0.946–0.992), with a sensitivity of 95.5% and a specificity of 89.2%.

**CONCLUSION**

DCE-MRI combined with DWI has significant diagnostic value in identifying the extent of IBC infiltration.

**CLINICAL SIGNIFICANCE**

DCE-MRI kinetic heterogeneity combined with DWI enables noninvasive discrimination of the extent of IBC infiltration before surgery, facilitating personalized systemic therapy and nodal evaluation while avoiding overtreatment in patients at low risk. Computer-aided, whole-tumor heterogeneity analysis replaces limited region-of-interest sampling, significantly improving both efficiency and accuracy of IBC grading. Integration of kinetic heterogeneity plus diffusion parameters provides a panoramic view of tumor size, location, and perilesional relationships, empowering multidisciplinary teams to rapidly individualize surgical and adjuvant treatment strategies.

**KEYWORDS**

Invasive breast cancer, dynamic contrast-enhanced magnetic resonance imaging, kinetic heterogeneity, diffusion-weighted imaging, apparent diffusion coefficient

***Contributed equally to this work.**

**Handling editor:** Serap Gültekin

**Corresponding author:** Qingqiang Zhu

**E-mail:** zhuqingqiang1983@163.com

Received 25 December 2025; revision requested 11 January 2026; accepted 05 February 2026.

Epub: 26.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263831

You may cite this article as: Feng X, Ye P, Chen H, Liu C, Zhu Q. Comparative analysis of the diagnostic value of dynamic contrast-enhanced magnetic resonance imaging kinetic heterogeneity and apparent diffusion coefficient for grading invasive breast cancer. *Diagn Interv Radiol*. 2026;32(4):399-408.

Invasive breast cancer (IBC) accounts for 70%–80%[1] of breast cancers and poses an elevated recurrence risk due to stromal/vascular infiltration.[2] Accurate preoperative grading is essential for personalized therapy.[3] However, core biopsy—the reference standard—is invasive and limited by sampling error and inter-observer variability.[4,5] Multiparametric magnetic resonance imaging (MRI) offers a non-invasive alternative. Diffusion-weighted imaging (DWI) quantifies cellularity via apparent diffusion coefficient (ADC) values,[6] whereas dynamic contrast-enhanced MRI (DCE-MRI) extracts quantitative parameters based on the pharmacokinetic model of the contrast agent to reveal the perfusion and permeability characteristics of tumor microvessels.[7] Although these techniques reliably differentiate benign from malignant lesions,[8,9] their further application to breast cancer grading remains limited. This is because most studies rely on manual, single-layer region-of-interest (ROI) localization, thereby overlooking spatial heterogeneity.[10] Therefore, it is vital to systematically validate the value of DWI and DCE-MRI quantitative parameters in evaluating the degree of IBC infiltration across an entire tumor scale.

**Main points**

- High-grade invasive breast cancer (IBC) exhibited significantly higher peak, plateau, washout, and heterogeneity values than low-grade IBC. Conversely, high-grade IBC demonstrated lower persistent fraction and apparent diffusion coefficient values, with no discernible difference in volume values between the two groups.
- Apparent diffusion coefficient, persistent fraction, and IBC grading exhibited a negative correlation. Peak, plateau, washout, and heterogeneity demonstrated a positive correlation with IBC grading. However, no significant correlation was observed between enhancement volume and IBC grading.
- Dynamic contrast-enhanced magnetic resonance imaging kinetic heterogeneity combined with diffusion-weighted imaging has significant diagnostic value in identifying the extent of IBC infiltration.

Computer-aided diagnosis (CAD) overcomes this limitation by enabling voxel-wise quantification of entire tumors, capturing intratumoural heterogeneity from hemorrhage, necrosis, and microvascular variation.[11-13] Previous studies have shown that a stronger washout effect and higher kinetic heterogeneity on DCE-MRI can effectively distinguish benign from malignant breast lesions, reflecting increased angiogenesis and the degree of vascular leakage within the tumor.[14] Additionally, the CAD system automatically detects lesions, thereby reducing interpretation time and substantially lowering the false-positive rate below specific thresholds.[15] Building on these observations, the present study compares whole-tumor ADC and DCE-MRI quantitative parameters in terms of low- and high-grade IBC, evaluating their diagnostic efficacy.

## Methods

### Patients

This retrospective study analyzed clinical and imaging data from patients with breast cancer admitted to Northern Jiangsu People's Hospital between January 2023 and February 2025. A total of 203 patients were included (Figure 1). The inclusion criteria were as follows: 1) complete MRI data, including DWI and DCE-MRI; 2) postoperative pathological confirmation of IBC with defined invasive grade; and 3) no prior history of breast cancer surgery, neoadjuvant chemotherapy, or related interventions documented before the imaging examination. The exclusion criteria included 1) poor MRI image quality (e.g., significant artifacts), 2) non-lump cases and other types of breast malignancies, 3) factors impeding accurate image interpretation (including location in special anatomical sites, indistinct lesion boundaries, and lesion size < 5 mm), and 4) pregnancy. The grade of IBC infiltration was classified according to the Nottingham Combined Histological Grading system,[16] an amended version of the Schaff-Blum-Richardson system. IBC was classified into three grades based on ductal formation, nuclear

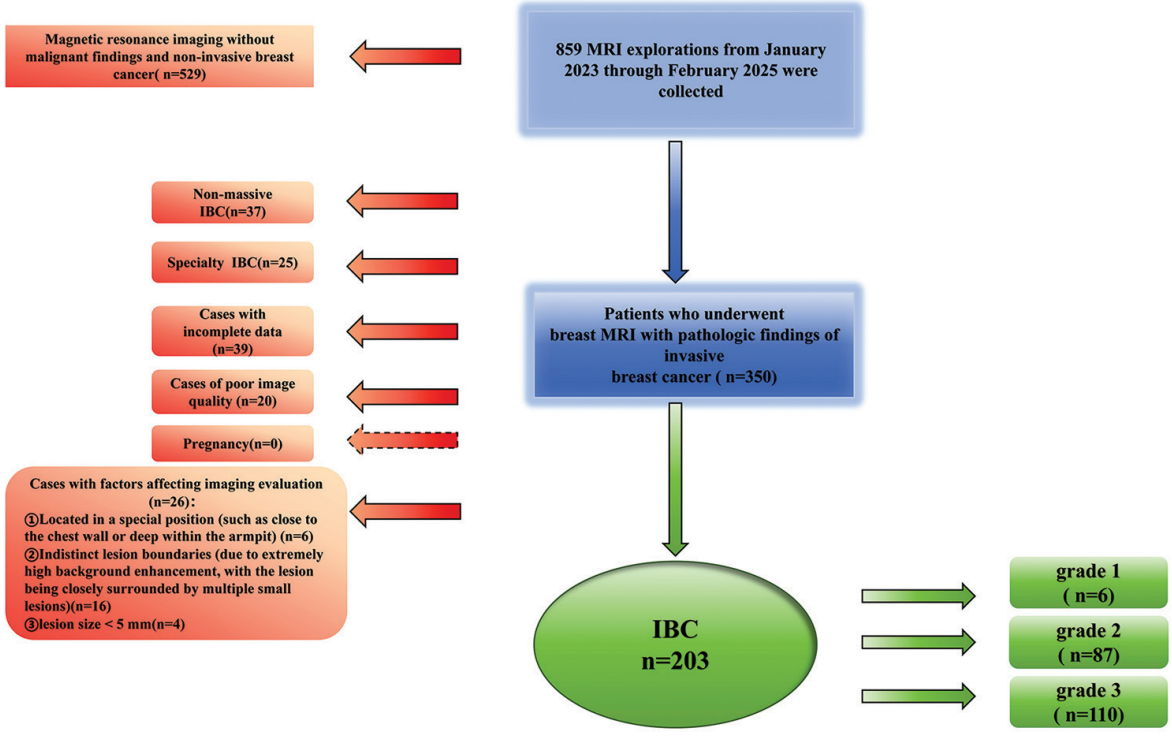



**Figure 1.** Flowchart showing the patient recruitment process. MRI, magnetic resonance imaging; IBC, invasive breast cancer.

pleomorphism, and mitotic count. According to prior studies, Grades I and II are grouped as low-grade IBC, and Grade III as high-grade IBC.[17] Immunohistochemistry (IHC) was performed using the streptavidin–biotin method to detect estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67 proliferation. HER2 expression status was further confirmed via fluorescence *in situ* hybridization (FISH). If the HER2 IHC result is 3+, it is classified as HER2-positive; if the result is 1+ or 0, it is classified as HER2-negative; if the result is 2+, FISH testing is required to determine the HER2 status.[18] ER and PR positivity is defined as ≥ 1% of tumor cell nuclei showing positive staining. The Ki-67 index is categorized as low proliferation (< 14%) or high proliferation (≥ 14%) based on a cut-off of 14%.[19] This study was approved by the Review Committee of Northern Jiangsu People's Hospital (approval number: 2024ky379, date: December 24, 2024). Informed consent was waived due to the study's retrospective design.

## Imaging acquisition

This study utilized the GE Healthcare Discovery MR750W 3.0 T (GE Healthcare, Chicago, IL, USA) MRI system for scanning. Patients were placed in the prone position, with both breasts naturally suspended within a dedicated eight-channel breast phased-array coil. The scan coverage included both breasts, the anterior chest wall, and the axillary regions. The imaging protocol commenced with a localization sequence, followed by axial fat-suppressed T2-weighted turbo spin-echo (TSE) imaging. DCE-MRI images were acquired using a three-dimensional T1-weighted fast gradient echo (LAVA/FLEX) sequence, comprising one pre-contrast and five post-contrast images. Following the pre-contrast acquisition, gadolinium diethyltripropionate (Omniscan, GE Healthcare) was bolus-injected at a flow rate of 2.5 mL/s via the antecubital vein using a high-pressure syringe, at a dose of 0.1 mmol/kg body weight. This was immediately followed by a 20-mL saline flush at the same flow rate. Following scan completion, images were uploaded to the GE Advantage Workstation for routine time-intensity curve (TIC) plotting of lesions. Prior to DCE-MRI, axial fat-suppressed single-shot echo-planar (SS-EPI) DWI was performed. Diffusion gradients were sequentially applied in three orthogonal directions, with an acquisition time of 222 s and b values of 0 and 1,000 s/mm$^2$. All sequences were acquired axially. Conventional T1-weighted imaging (WI) employed a TSE readout, with repetition time (TR): 4.4 ms, echo time (TE): 2.1 ms, flip angle: 15°, field of view (FOV): 320 × 320 mm$^2$, phase matrix: 320 × 320, and slice thickness: 1.3 mm. DCE T1-WI used an identical TSE readout, with TR: 5.9 ms, TE: 1.1 ms, flip angle: 15°, FOV: 320 × 320 mm$^2$, phase matrix: 240 × 240, and slice thickness: 1.3 mm. Fat-suppressed T2-WI was acquired using TR: 4,619 ms, TE: 85 ms, flip angle: 15°, FOV: 320 × 320 mm$^2$, phase matrix: 224 × 320, and slice thickness: 4 mm. DWI used an SS-EPI sequence in an oblique plane, with TR: 2,859 ms, TE: 88.5 ms, FOV: 320 × 320 mm$^2$, phase matrix: 128 × 128, and slice thickness: 10 mm.

## Image analysis

The DCE-MRI processing workflow was as follows. First, motion correction was performed on all phases of the DCE-MRI images using SPM 12 software, with the criteria that translational and rotational motion were < 1 mm and < 5°, respectively. The corrected images were then imported into ITK-SNAP software (version 3.8.0). Two radiologists (with 5 and 10 years of breast MRI experience, respectively) manually delineated the volume of interest (VOI) on second-phase DCE-MRI images acquired within 120 s of contrast agent injection to maximize lesion-background contrast and enhance delineation consistency.[20,21] Finally, MATLAB (version 2021a) was used to extract DCE parameters within the VOI. The parameters included peak, which is the maximum enhancement ratio (MER) within the tumor between the early peak enhancement phase (120 s post-contrast injection) and the pre-enhanced image. The MER calculation formula is as follows:[22]

$$MER=\frac{S_{peak}-S_{baseline}}{S_{baseline}}\times 100\%,$$

where $S_{baseline}$ = mean signal intensity of the first two pre-contrast phases, and Speak = maximum post-contrast signal intensity per voxel at 120 s.

Further parameters included the following. Enhancement volume, defined as the volume fraction of voxels within the tumor showing over 50% enhancement in the early peak enhancement phase compared with the pre-enhanced image. Persistent fraction, defined in terms of a > 10%-increase in voxel volume in the final phase compared with the early peak enhancement phase in terms of signal intensity. Plateau proportion, where the increase or decrease in the signal intensity ratio between the final phase and the early peak enhancement phase does not exceed 10% in the voxel volume. Washout proportion, defined as a decrease in voxel volume by more than 10% in the final phase compared with the early peak enhancement phase in terms of signal intensity. Heterogeneity, which is a measure of tumor heterogeneity based on the proportion of voxels exhibiting washout, plateau, or sustained enhancement. The formula for the calculation of heterogeneity is as follows:[23]

$$\text{heterogeneity}=-\sum_{i=1}^{k} Pi\log_k Pi$$

where Pi is the frequency of observations of each variable and k is the number of kinetic-class categories; here k: 3 (washout, plateau, persistent). The index ranges from 0 to 1, with higher values indicating greater heterogeneity.[24]

The DWI processing workflow was as follows: DWI data were converted to NIfTI format using MRIcron (https://www.nitrc.org/projects/mricron), then processed using the NeuDiLab software developed based on DIPY (https://dipy.org/). For conventional DWI, the quantitative metric ADC is generated via single-exponent fitting using the following formula:[25]

$$S(b) / S(0) = \exp(-b \cdot ADC),$$

where S(0) and S(b) represent the signal intensity of water molecule motion in the absence and presence of diffusion, respectively; b is the diffusion-sensitization factor determining the degree of diffusion weighting in signal intensity. To measure diffusion metrics, the VOI defined on DCE-MRI was used as a reference. This VOI was placed on the quantitative map of conventional DWI (i.e., ADC), positioned along the tumor margin with the largest dimension, and excluded adjacent normal breast and fatty tissue.

## Statistical analysis

All statistical analyses were performed using SPSS 25.0 (IBM Corp., Armonk, NY, USA), MedCalc 19.6 (MedCalc Software, Ostend, Belgium) and R 4.4.1 software (R Foundation for Statistical Computing, Vienna, Austria). Inter-reader reproducibility for continuous parameters was quantified using the intraclass correlation coefficient (ICC), with values > 0.75 indicating excellent agreement. Normality was first examined using the Shapiro–Wilk test. Continuous variables were expressed as mean ± standard deviation if normally distributed, and compared using the independent-sample t-test; otherwise, they were expressed as median (interquartile range) and compared using the Mann–Whitney U test.

To compare the distributions of categorical variables between two groups, the chi-square ($\chi^2$) test was used. When data in the contingency table were sparse—specifically, when more than 20% of expected cell counts were < 5 or any expected count was < 1—the Fisher–Freeman–Halton test (i.e., Fisher's exact test for R × C contingency tables) was used to obtain more reliable and precise *P* values. Univariate logistic regression was used to select initial parameters. Subsequently, a forward stepwise multivariable logistic regression analysis was performed on parameters with $P < 0.05$ to identify the simplest, clinically significant model for distinguishing low-grade from high-grade IBC. The optimal probability threshold for the combined diagnostic model was determined by maximizing Youden's index (J = Sensitivity + Specificity − 1) on the receiver operating characteristic (ROC) curve. To assess the risk of overfitting, stratified bootstrap resampling was used to assess area under the curve (AUC) sampling fluctuations, and the validated average AUC and its 95% confidence interval (CI) were calculated. The calibration curve was used to evaluate the consistency between the model's predicted and observed probabilities. Correlation analyses were performed using Pearson's r for normally distributed continuous variables and Spearman's ρ for non-normally distributed data. Two-tailed *P* values < 0.05 were considered statistically significant.

## Results

A total of 203 patients with pathologically confirmed IBC were included (Table 1). Among them, 93 patients (45.8%) were classified as low grade (Grade I, n = 6; Grade II, n = 87), and 110 (54.2%) as high grade. No significant difference was observed in the maximum tumor diameter (29.7 ± 23.0 vs. 24.4 ± 13.8 mm; $P = 0.585$) or PR expression between the low-grade and high-grade groups, whereas statistically significant differences were noted in ER and HER2 expression, and Ki-67 proliferation index ($P < 0.05$). TICs differed significantly between groups ($P = 0.010$): type III curves were more frequent in high-grade IBC (73.6% vs. 53.8%), whereas type I curves predominated in low-grade lesions (7.5% vs. 2.7%). The mean age was similar between the low-grade and high-grade groups (53 ± 10 vs. 55 ± 11 years, $P = 0.053$).

Inter-reader agreement for all quantitative metrics was excellent (ICC > 0.75, $P < 0.05$) (Table 2). High-grade IBC exhibited significantly higher peak enhancement (17.486 ± 23.881 vs. 7.025 ± 5.924, $P < 0.001$), plateau fraction (0.318 ± 0.182 vs. 0.137 ± 0.118, $P < 0.001$), washout fraction (0.210 ± 0.170 vs. 0.105 ± 0.146, $P = 0.007$), and kinetic heterogeneity (0.908 ± 0.191 vs. 0.493 ± 0.239, $P < 0.001$) than low-grade tumors (Table 1, Figure 2). Conversely, the persistent fraction (0.493 ± 0.239 vs. 0.729 ± 0.251, $P < 0.001$) and ADC value (0.770 ± 0.161 vs. 1.010 ± 0.221 × $10^{-3}$ $mm^2/s$, $P < 0.001$) were significantly lower in high-grade lesions. No between-group difference in lesion volume was observed ($P = 0.314$). Univariable logistic regression identified peak, persistent, plateau, washout, and heterogeneity as significant predictors of grading IBC ($P < 0.05$) (Table 3). These variables were subsequently entered into a forward stepwise multivariable logistic regression analysis. The final model identified kinetic heterogeneity [odds ratio

**Table 1.** General data characteristics, DCE-MRI kinetic analysis of each parameter and ADC values of patients with low-grade IBC vs. high-grade IBC

| | Low-grade IBC (n = 93) | High-grade IBC (n = 110) | *P* |
|---|---|---|---|
| Age | 53 ± 10 | 55 ± 11 | 0.053 |
| Tumor length diameter (mm) | 29.68 ± 22.99 | 24.41 ± 13.78 | 0.585 |
| ER | | | |
| + | 67 | 48 | < 0.001 |
| – | 26 | 62 | |
| PR | | | |
| + | 55 | 53 | 0.116 |
| – | 38 | 57 | |
| HER2 | | | |
| + | 76 | 39 | < 0.001 |
| – | 17 | 71 | |
| Ki-67 | | | |
| >14% | 46 | 72 | 0.018 |
| ≤14% | 47 | 38 | |
| TIC | | | |
| I | 7 (7.53%) | 3 (2.73%) | |
| II | 36 (38.71%) | 26 (23.64%) | 0.010 |
| III | 50 (53.76%) | 81 (73.63%) | |
| Peak | 7.025 ± 5.924 | 17.486 ± 23.881 | < 0.001 |
| Volume | 2,455.591 ± 5,343.670 | 4,852.045 ± 24,164.594 | 0.314 |
| Persistent | 0.729 ± 0.251 | 0.493 ± 0.239 | < 0.001 |
| Plateau | 0.137 ± 0.118 | 0.318 ± 0.182 | < 0.001 |
| Washout | 0.105 ± 0.146 | 0.210 ± 0.170 | 0.007 |
| Heterogeneity | 0.493 ± 0.239 | 0.908 ± 0.191 | < 0.001 |
| ADC | 1.010 ± 0.221 | 0.770 ± 0.161 | < 0.001 |

IBC, invasive breast cancer; ADC, apparent diffusion coefficient; TIC, time-intensity curve; HER2, human epidermal growth factor receptor 2; ER, estrogen receptor; PR, progesterone receptor; DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging.

**Table 2.** ICC values for each parameter of DCE-MRI kinetic analysis and ADC values between two observers

| Parameters | Two observer ICC values | *P* |
|---|---|---|
| Heterogeneity | 0.957 (0.935–0.971) | < 0.001 |
| ADC | 0.901 (0.883–0.926) | < 0.001 |
| Peak | 0.978 (0.928–0.968) | < 0.001 |
| Volume | 0.902 (0.837–0.928) | < 0.001 |
| Persistent | 0.912 (0.849–0.935) | < 0.001 |
| Plateau | 0.900 (0.835–0.925) | < 0.001 |
| Washout | 0.961 (0.930–0.970) | < 0.001 |

ADC, apparent diffusion coefficient; ICC, intraclass correlation coefficient; DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging.

(OR): 2.500; 95% CI: 1.848–3.373; $P < 0.001$] and peak (OR: 1.130; 95% CI: 1.026–1.244; $P = 0.012$) as independent positive predictors of high-grade IBC, whereas ADC was an independent negative predictor (OR: 0.318; 95% CI: 0.198–0.514; $P < 0.001$) (Figure 3).

The ROC analysis demonstrated good discriminatory power for all quantitative metrics (Figure 4). The AUC values were as follows: peak: 0.680 (95% CI: 0.611–0.744), persistent: 0.765 (95% CI: 0.700–0.821), plateau: 0.806 (95% CI: 0.744–0.858), washout: 0.717 (95% CI: 0.650–0.778), heterogeneity 0.914 (95% CI: 0.866–0.949), and ADC 0.819 (95% CI: 0.758–0.869) (Table 4). Corresponding sensitivities were 55.5%, 69.0%, 65.3%, 64.5%, 86.4%, and 90.3%, with specificities of 78.5%, 79.0%, 87.3%, 74.2%, 84.9%, and 53.6%, respectively. The combined model incorporating heterogeneity, peak, and ADC achieved a sensitivity of 95.5% and a specificity of 89.2% at the optimal threshold of 0.346. DeLong testing revealed that heterogeneity achieved a significantly higher AUC than ADC (0.914 vs. 0.819; $P = 0.009$). In evaluating the generalization ability and overfitting risk of the combined heterogeneity, peak, and ADC comprehensive diagnostic model, the average AUC obtained through stratified bootstrap resampling was 0.969 (95% CI: 0.933–0.985), which was significantly better than that of any single parameter. The calibration curve was plotted based on the final model (Figure 5). The predicted probabilities for each risk group were generally close to the ideal 45° reference line in the overall distribution, indicating acceptable calibration performance.

Correlation analyses were summarized and visualized using a heat map of Pearson coefficients (Figure 6). ADC and persistent fraction were negatively correlated with histological grade (r: −0.533, $P < 0.001$ and r: −0.414, $P < 0.001$, respectively), whereas peak, plateau, washout, and heterogeneity showed positive correlations (r: 0.279, $P < 0.001$; r: 0.505, $P < 0.001$; r: 0.314, $P < 0.001$; and r: 0.695, $P < 0.001$). No significant association was observed between enhancement volume and grade (r: 0.066, $P = 0.350$). Heterogeneity possessed the strongest correlation with invasive status, and the combination of heterogeneity plus ADC yielded the highest diagnostic performance.

## Discussion

This study demonstrated that high-grade IBC is characterized by significantly higher peak, plateau, and washout ratios on DCE-MRI, as well as greater heterogeneity, alongside lower persistent fraction ratios and ADC values. The combined diagnostic model incorporating heterogeneity, peak, and ADC achieved an AUC of 0.969, which was significantly superior to that of any single parameter. These findings indicate that combining DCE-MRI with DWI enables quantitative assessment of IBC grading. This approach can provide objective evidence to support

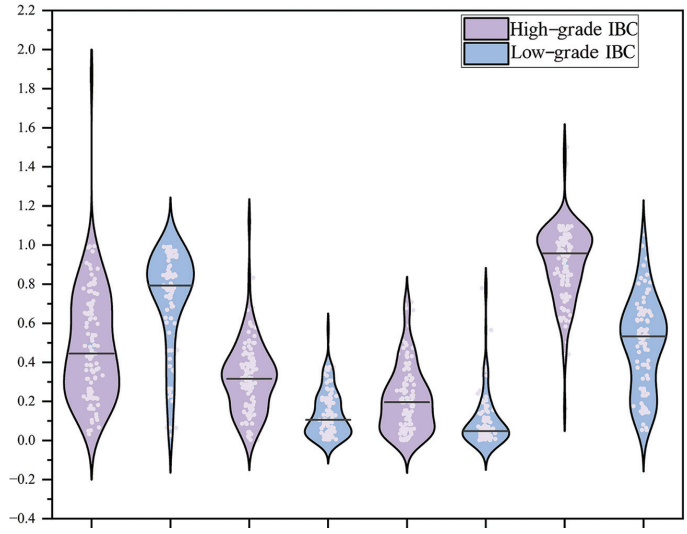



**Figure 2.** Violin graphs showing the differences in dynamic contrast-enhanced magnetic resonance imaging kinetic parameters and apparent diffusion coefficient values between low-grade and high-grade invasive breast cancer. IBC, invasive breast cancer.

**Table 3.** Univariable and multivariable logistic regression analyses of the relationship between low-grade and high-grade IBCs

| Parameters | Unit change | Univariable | | | Multivariable | | |
|---|---|---|---|---|---|---|---|
| | | OR | 95%CI | *P* | OR | 95% CI | *P* |
| ADC | 0.1 | 0.427 | 0.328–1.000 | < 0.001 | 0.318 | 0.198–0.514 | < 0.001 |
| Heterogeneity | 0.1 | 2.365 | 1.883–2.971 | < 0.001 | 2.500 | 1.848–3.373 | < 0.001 |
| Peak | 1 | 1.087 | 1.041–1.135 | < 0.001 | 1.130 | 1.026–1.244 | 0.012 |
| Persistent | 1 | 0.035 | 0.011–0.116 | < 0.001 | | | |
| Plateau | 1 | 4,615.207 | 369.461–57,651.9451 | < 0.001 | | | |
| Washout | 1 | 113.895 | 12.249–1,058.998 | < 0.001 | | | |

ADC, apparent diffusion coefficient; OR, odds ratio; CI, confidence interval; IBC, invasive breast cancer.

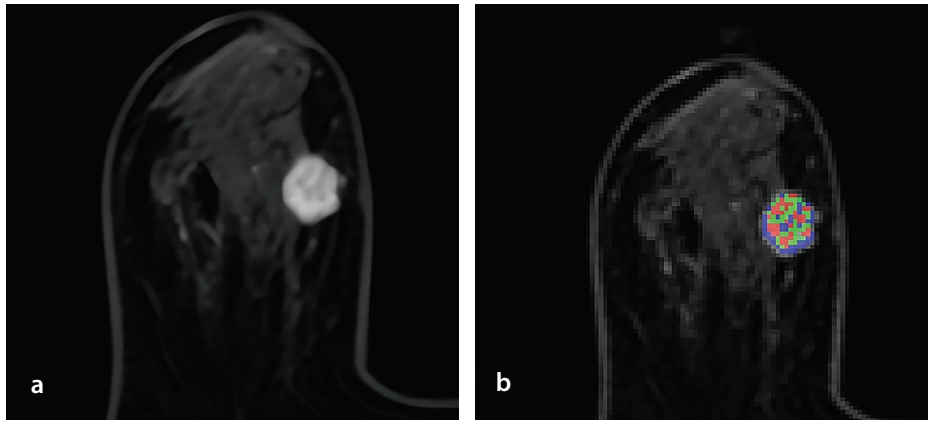



**Figure 3.** A 46-year-old female with a malignant tumor (high-grade IBC). BI-RADS:5. Dynamic contrast-enhanced kinetic analysis plot (**a** and **b**), peak: 5.983, persistent: 0.430, plateau: 0.312, washout: 0.257, and heterogeneity: 1.075. The blue, green, and red denote persistent, plateau, and washout, respectively. IBC, invasive breast cancer.

precise histological grading and guide individualized treatment planning and axillary management, thereby helping to avoid overtreatment in patients at low risk.

In agreement with the report by Li et al.,[26] which found that high-grade lesions exhibit elevated quantitative parameters, we observed significantly greater kinetic heterogeneity in high-grade IBC than in low-grade tumors (0.908 ± 0.191 vs. 0.493 ± 0.239, $P < 0.001$). Breast cancer heterogeneity is usually accompanied by regional hypoxia and microvascular abnormalities, including disorganized vessel architecture and increased permeability. We therefore postulated that greater kinetic heterogeneity reflects a more complex vascular network and more pronounced histopathological heterogeneity within the tumor microenvironment. Previous studies have demonstrated that tumor infiltration extent is associated with tumor cell growth rate, invasiveness to surrounding tissues, drug sensitivity, and treatment outcomes. Therefore, preoperative grading is of considerable importance in guiding treatment planning.[27] Conversely, ADC values were significantly lower in high-grade IBC than in the low-grade group (0.770 ± 0.161 vs. 1.010 ± 0.221 × $10^{-3}$ $mm^2/s$, $P < 0.001$) and showed a strong inverse correlation with invasive grade (r: −0.533, $P < 0.001$), which is in agreement with the findings of Mori et al.[28] This difference is likely attributable to the characteristic pathological features of high-grade tumors, such as marked nuclear pleomorphism, a high mitotic index, and densely packed tumor cells. Spearman correlation analysis further revealed that both ADC and persistent fraction were negatively correlated with histological grade (r: −0.533 and r: −0.414, respectively; $P < 0.001$). This indicates that increased cellular density and reduced sustained perfusion are associated with greater aggressiveness. In contrast, peak, plateau, washout, and heterogeneity were positively correlated with grade (r: 0.279, 0.505, 0.314, and 0.695, respectively; all $P < 0.001$), suggesting that greater perfusion heterogeneity and washout are indicative of higher malignant potential. In this study, the persistent, plateau, and washout parameters were represented by blue, green, and red, respectively. Compared with histogram analysis, which captures overall heterogeneity but

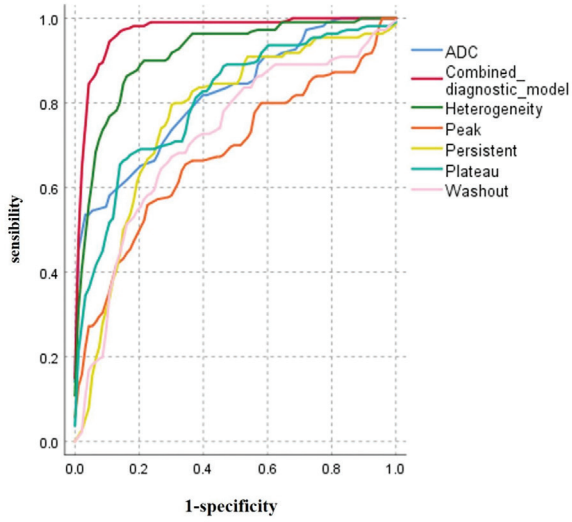



**Figure 4.** Receiver operating characteristic curve analysis for invasive breast cancer grade identification yielded an area under the curve of 0.910 (95% CI: 0.857–0.948) for heterogeneity, 0.808 (95% CI: 0.741–0.863) for apparent diffusion coefficient, and 0.969 (95% CI: 0.927–0.984) for the combined diagnostic model, demonstrating the superior discriminative performance of the combined model in assessing breast cancer invasiveness. ADC, apparent diffusion coefficient; CI, confidence interval.

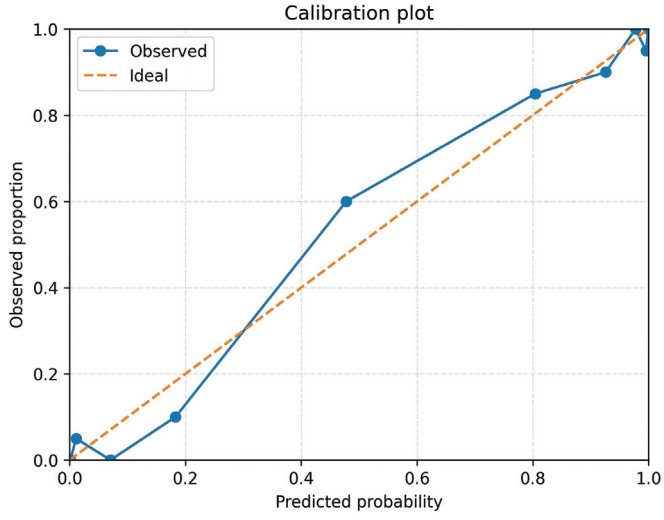



**Figure 5.** Calibration plot of the final multivariable model. The plot shows that the predicted probabilities across risk groups were generally close to the ideal 45° line, indicating acceptable calibration performance.

**Table 4.** Comparison of the diagnostic efficacy of DCE-MRI kinetic analysis of each parameter and ADC values between low-grade and high-grade IBCs

| Parameters | AUC | 95% CI | Sensitivity | Specificity | Cut-off value | Youden index |
|---|---|---|---|---|---|---|
| ADC | 0.819 | 0.758–0.869 | 90.30% | 53.60% | 0.735 | 0.525 |
| Heterogeneity | 0.914 | 0.866–0.949 | 86.40% | 84.90% | 0.709 | 0.713 |
| Peak | 0.68 | 0.611–0.744 | 55.50% | 78.50% | 10.322 | 0.340 |
| Persistent | 0.765 | 0.700–0.821 | 69.00% | 79.00% | 0.721 | 0.499 |
| Plateau | 0.806 | 0.744–0.858 | 65.30% | 87.30% | 0.263 | 0.524 |
| Washout | 0.717 | 0.650–0.778 | 64.50% | 74.20% | 0.122 | 0.387 |
| Combined diagnostic model | 0.969 | 0.946–0.992 | 95.50% | 89.20% | 0.346 | 0.847 |

ADC, apparent diffusion coefficient; AUC, area under the curve; CI, confidence interval; DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging; IBC, invasive breast cancer.

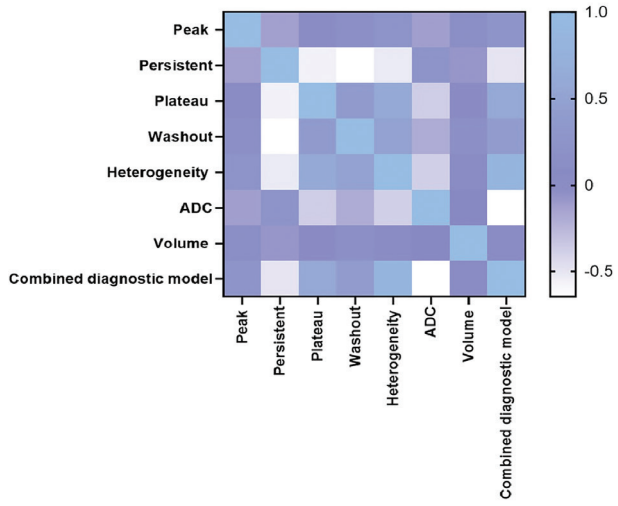



**Figure 6.** Heatmap of correlations among various parameters of dynamic contrast-enhanced magnetic resonance imaging kinetic analysis, apparent diffusion coefficient (ADC) values, and invasive breast cancer staging; ADC was negatively correlated with histological grade (r: –0.533, $P < 0.001$), whereas heterogeneity showed positive correlations (r: 0.695, $P < 0.001$).

loses spatial information, the core advantage of our model lies in preserving the spatial distribution map of heterogeneity. This enables us to identify and localize specific subregions within tumors that may represent the highest invasive potential or potential therapeutic resistance, which holds potential significance for guiding targeted biopsy or localized dose-fractionation radiotherapy.

This study is consistent with previous reports that combined DWI and DCE-MRI to improve the estimation of pathological grade in breast cancer.[29,30] The present study focused exclusively on invasive carcinoma of no special type (NST), the most common histological subtype, to eliminate biological heterogeneity among subtypes.[31] DCE-MRI combined with DWI technology has been applied in the preoperative evaluation of breast cancer to assess tumor size and internal structure. DCE imaging can demonstrate the extent of tumor lesions, providing reliable references for surgical planning. However, it is important to note that approximately 10%–15% of all IBCs, specifically invasive lobular carcinomas (ILCs),[32-36] exhibit delayed peak enhancement and infrequent contrast washout on DCE-MRI,[37] markedly differing from the rapid enhancement pattern seen in NST. The biological basis of ILC may relate to its diffuse growth pattern and reduced neovascularization requirements, manifested by lower vascular endothelial growth factor levels and more mature, low-permeability tumor vasculature,[38-40] resulting in overall weaker enhancement. Given the potential limitations of traditional assessment parameters for lesions with atypical enhancement patterns, the heterogeneity index used in this study quantifies voxel-level perfusion gradients to reflect spatial heterogeneity in microvascular function. Theoretically less dependent on lesion enhancement intensity or sharp margins, this index is hypothesized to retain discriminatory potential for ILC, although its value requires further validation through subgroup analysis. Fan et al.[41] constructed radiomic signatures from single-phase DCE-MRI and ADC maps, achieving AUCs of 0.811 and 0.816 for tumor grading, respectively; both values are significantly lower than the AUC of 0.969 obtained with the combined model of heterogeneity, peak, and ADC. A probable explanation is that static radiomics extracts thousands of features from a limited cohort, yielding features with questionable biological interpretability, whereas our kinetic heterogeneity metric quantifies time-resolved hypoxia and microvascular abnormalities across all perfusion phases, thereby providing complementary information to radiomics approaches. It should be noted that although this study demonstrates superior technical-level discriminative efficacy compared with conventional methods and attempts to provide biologically interpretable features, it remains within the scope of retrospective technical validation. Whether it can ultimately influence surgical decisions, assist in treatment selection, or improve patient outcomes requires validation in prospective clinical settings.

In the present study, high-grade IBC exhibited a significantly higher peak than low-grade tumors (17.486 ± 23.881 vs. 7.025 ± 5.924, $P < 0.001$), a finding consistent with previous work,[42] which demonstrated a strong association between greater peak enhancement and higher histological grade and advanced clinical stage. Peak enhancement reflects the early increase in signal intensity after contrast injection and thus mirrors the concentration of contrast medium in the intravascular and extravascular compartments.[43] We therefore postulate that the elevated peak in high-grade lesions is closely linked to altered tumor microvasculature. In this study of low-grade IBC, we found that 53.8% of cases exhibited a type III TIC, whereas previous similar studies reported approximately 34% of breast cancers presenting a type II TIC.[44] Furthermore, existing research indicates that distinguishing benign from malignant breast lesions often requires reliance on the most suspicious TIC classification,[45] suggesting that TIC generation is highly dependent on the ROI delineated by the operator. If the ROI is not accurately placed within the rapidly enhancing washout region, it may compromise the curve's representativeness and thus interfere with diagnostic judgment.[46] Therefore, precise ROI localization is critical for the reliability of diagnostic accuracy. However, this study overcomes the errors associated with manual placement through whole-tumor quantitative analysis, with results indicating that the proportion of persistent IBC was significantly higher in low-grade IBC than in high-grade IBC (0.729 ± 0.251 vs. 0.493 ± 0.239, $P < 0.001$), whereas the proportion of plateau-stage IBC was significantly lower in low-grade IBC than in high-grade IBC (0.137 ± 0.118 vs. 0.318 ± 0.182, $P < 0.001$). These findings indicate that high-grade IBC exhibits a marked reduction in persistent fraction and a corresponding increase in plateau. Consequently, the persistent and plateau parameters demonstrate potential utility in the graded diagnosis of IBC, offering crucial evidence for distinguishing low-grade from high-grade IBC. Research[47] has suggested that plateau may reflect the distribution and proportion of invasive tumor cells within breast cancer, presenting novel insights for noninvasive diagnosis; however, further clinical validation is required.

Immunohistochemical analysis in this study further revealed statistically significant differences in HER2 expression status and Ki-67 proliferation index across different histological grades, consistent with previous findings.[48,49] Prior studies have demonstrated that high Ki-67 proliferation correlates with aggressive growth and higher tumor grades in breast cancer, with tumor cell proliferation

often exhibiting spatially heterogeneous distribution.[50] Recent studies have utilized multiparametric MRI for noninvasive assessment of the biological characteristics of breast cancer.[51,52] This suggests that integrating complementary imaging and pathological information is crucial for developing predictive models that comprehensively interpret tumor heterogeneity, providing important insights for further refinement of this research. Additionally, although some early studies have reported a correlation between tumor volume and histological grade,[53] our analysis found no significant association between manually segmented tumor volume and grade ($P = 0.314$). This finding aligns with data from large-scale population studies. For example, in an analysis of 161,708 breast cancers from the SEER Program, Schwartz et al.[54] found that tumor size accounted for < 10% of the variance in histological grade. Notably, 15%–20% of tumors smaller than 1 cm were classified as Grade III, indicating that small lesions can be biologically aggressive. Similarly, Liu et al.[16] reported no correlation between invasive tumor size and any DCE-MRI perfusion parameter (all $P > 0.05$), and Schmitz et al.[55] found no association between macroscopic diameter and micro-vessel density on 3.0 T breast MRI. Collectively, these data suggest that invasive potential is driven more by intrinsic biological traits, such as angiogenic activity, cellular density, and genomic instability, than by physical tumor bulk. Furthermore, our volume segmentation deliberately excluded peritumoural edema, a methodological choice that may have reduced any incidental correlation between edema-rich large lesions and high-grade disease. Based on this evidence, we hypothesize that tumor size is not a determinant of angiogenic factors or blood perfusion in IBC.

Our study has several limitations. First, this was a single-center, exploratory study with a relatively small sample size. The model coefficients and calibration have not been verified in external cohorts, and the potential risk of overfitting cannot be completely ruled out. Second, this study was restricted to patients with NST lesions who had not received any breast cancer-related treatment, which may limit the generalizability of kinetic parameters to other IBC subtypes and introduce selection bias. Future research will further evaluate the diagnostic efficacy of DCE-MRI and DWI in assessing the response to neoadjuvant therapy. Third, tumor VOIs were manually delineated, introducing inter-observer variability. Artificial intelligence-based auto-segmentation should be explored to improve reproducibility and efficiency. Finally, heterogeneity analysis was performed using MATLAB and SPM12; inter-institutional differences in acquisition protocols and software packages hinder feature harmonization. Future work should establish standardized imaging protocols and consensus analysis pipelines to enhance the clinical portability of kinetic biomarkers.

In conclusion, this study employed volumetric quantitative DCE-MRI enhancement classification technology to visualize and quantitatively analyze the kinetic heterogeneity within tumors. Combined with ADC values, the model's diagnostic efficacy in evaluating IBC grades surpassed that of a single parameter. This provides a new and promising imaging biomarker for noninvasive assessment of tumor heterogeneity, although its exact clinical application value needs to be verified through subsequent prospective studies.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Yang WT, Bu H. [updates in the 5(th) edition of WHO classification of tumours of the breast]. *Zhonghua Bing Li Xue Za Zhi*. 2020;49(5):400-405. [Crossref]
2. Facina G, Oliveira VMD. Breast cancer care during the coronavirus pandemic. *Mastology*. 2020;30. [Crossref]
3. Wang J, Wu SG. Breast cancer: An overview of current therapeutic strategies, challenge, and perspectives. *Breast Cancer (Dove Med Press)*. 2023;15:721-730. [Crossref]
4. Wang W, Dou B, Wang Q, et al. Comparison of MUSE-DWI and conventional DWI in the application of invasive breast cancer and malignancy grade prediction: a comparative study. *Heliyon*. 2024;10(2):e24379. [Crossref]
5. Rakha EA, El-Sayed ME, Menon S, Green AR, Lee AH, Ellis IO. Histologic grading is an independent prognostic factor in invasive lobular carcinoma of the breast. *Breast Cancer Res Treat*. 2008;111(1):121-127. [Crossref]
6. Zhu G, Dong Y, Zhu R, et al. Dynamic contrast-enhanced magnetic resonance imaging parameters combined with diffusion-weighted imaging for discriminating malignant lesions, molecular subtypes, and pathological grades in invasive ductal carcinoma patients. *PLoS One*. 2025;20(4):e0320240. [Crossref]
7. Zhang D, Shen M, Zhang L, He X, Huang X. Establishment of an interpretable MRI radiomics-based machine learning model capable of predicting axillary lymph node metastasis in invasive breast cancer. *Sci Rep*. 2025;15(1):26030. [Crossref]
8. Gullo RL, Partridge SC, Shin HJ, Thakur SB, Pinker K. Update on DWI for breast cancer diagnosis and treatment monitoring. *AJR Am J Roentgenol*. 2024;222(1):e2329933. [Crossref]
9. Morris EA, Liberman L, Ballon DJ, et al. MRI of occult breast carcinoma in a high-risk population. *AJR Am J Roentgenol*. 2003;181(3):619-626. [Crossref]
10. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. *Radiology*. 2019;292(3):520-536. [Crossref]
11. Junttila MR, De Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501(7467):346-354. [Crossref]
12. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumour heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883-892. [Crossref] Erratum in: *N Engl J Med*. 2012;367(10):976.
13. Kim C, Suh JY, Heo C, et al. Spatiotemporal heterogeneity of tumour vasculature during tumour growth and antiangiogenic treatment: MRI assessment using permeability and blood volume parameters. *Cancer Med*. 2018;7(8):3921-3934. [Crossref]
14. Yao Y, Mou F, Kong J, Liu X. Kinetic heterogeneity improves the specificity of dynamic enhanced MRI in differentiating benign and malignant breast tumours. *Acad Radiol*. 2024;31(3):812-821. [Crossref]
15. Lehman CD, Peacock S, DeMartini WB, Chen X. A new automated software system to evaluate breast MR examinations: Improved specificity without decreased sensitivity. *AJR Am J Roentgenol*. 2006;187(1):51-56. [Crossref]
16. Liu L, Mei N, Yin B, Peng W. Correlation of DCE-MRI perfusion parameters and molecular biology of breast infiltrating ductal carcinoma. *Front Oncol*. 2021;11:561735. [Crossref]
17. Takahashi H, Oshi M, Asaoka M, Yan L, Endo I, Takabe K. Molecular biological features of nottingham histological grade 3 breast cancers. *Ann Surg Oncol*. 2020;27(11):4475-4485. [Crossref]
18. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Arch Pathol Lab Med*. 2014;138(2):241-256. [Crossref]
19. Nielsen TO, Leung SCY, Rimm DL, et al. Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. *J Natl Cancer Inst*. 2021;113(7):808-819. [Crossref]

20. Hu N, Zhao J, Li Y, et al. Breast cancer and background parenchymal enhancement at breast magnetic resonance imaging: A meta-analysis. *BMC Med Imaging*. 2021;21(1):32. [Crossref]
21. Giannotti E, James JJ, Chen Y, et al. Effectiveness of percutaneous vacuum-assisted excision (VAE) of breast lesions of uncertain malignant potential (B3 lesions) as an alternative to open surgical biopsy. *Eur Radiol*. 2021;31(12):9540-9547. [Crossref] Erratum in: *Eur Radiol*. 2022;32(1):742. [Crossref]
22. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*. 1999;211(1):101-110. [Crossref]
23. Kim JY, Kim JJ, Hwangbo L, et al. Kinetic heterogeneity of breast cancer determined using computer-aided diagnosis of preoperative MRI scans: relationship to distant metastasis-free survival. *Radiology*. 2020;295(3):517-526. [Crossref]
24. Zhao R, Ma WJ, Tang J, et al. Heterogeneity of enhancement kinetics in dynamic contrast-enhanced MRI and implication of distant metastasis in invasive breast cancer. *Clin Radiol*. 2020;75(12):961. [Crossref]
25. Mao C, Hu L, Jiang W, et al. Discrimination between human epidermal growth factor receptor 2 (HER2)-low-expressing and HER2-overexpressing breast cancers: a comparative study of four MRI diffusion models. *Eur Radiol*. 2024;34(4):2546-2559. [Crossref]
26. Li K, Machireddy A, Tudorica A, et al. Discrimination of malignant and benign breast lesions using quantitative multiparametric MRI: a preliminary study. *Tomography*. 2020;6(2):148-159. [Crossref]
27. Allarakha A, Gao Y, Jiang H, Wang PJ. Prediction and prognosis of biologically aggressive breast cancers by the combination of DWI/DCE-MRI and immunohistochemical tumor markers. *Discov Med*. 2019;27(146):7-15. [Crossref]
28. Mori N, Inoue C, Tamura H, et al. Apparent diffusion coefficient and intravoxel incoherent motion-diffusion kurtosis model parameters in invasive breast cancer: correlation with the histological parameters of whole-slide imaging. *Magn Reson Imaging*. 2022;90:53-60. [Crossref]
29. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann Oncol*. 2013;24(9):2206-2223. [Crossref]
30. Fan M, Yuan C, Huang G, et al. A framework for deep multitask learning with multiparametric magnetic resonance imaging for the joint prediction of histological characteristics in breast cancer. *IEEE J Biomed Health Inform*. 2022;26(8):3884-3895. [Crossref]
31. Jeong J, Park CS, Lee JW, et al. Computer-aided diagnosis parameters of invasive carcinoma of no special type on 3T MRI: correlation with pathologic immunohistochemical markers. *J Korean Soc Radiol*. 2022;83(1):149-162. [Crossref]
32. Parvaiz MA, Yang P, Razia E, et al. Breast MRI in invasive lobular carcinoma: a useful investigation in surgical planning? *Breast J*. 2016;22(2):143-150. [Crossref]
33. Hussien M, Lioe TF, Finnegan J, Spence RA. Surgical treatment for invasive lobular carcinoma of the breast. *Breast*. 2003;12(1):23-35. [Crossref]
34. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. *JAMA*. 2003;289(11):1421-1424. [Crossref]
35. Peiro G, Bornstein BA, Connolly JL, et al. The influence of infiltrating lobular carcinoma on the outcome of patients treated with breast-conserving surgery and radiation therapy. *Breast Cancer Res Treat*. 2000;59(1):49-54. [Crossref]
36. Li CI, Daling JR. Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004. *Cancer Epidemiol Biomarkers Prev*. 2007;16(12):2773-2780. [Crossref]
37. Mann RM, Hoogeveen YL, Blickman JG, Boetes C. MRI compared to conventional diagnostic work-up in the detection and evaluation of invasive lobular carcinoma of the breast: a review of existing literature. *Breast Cancer Res Treat*. 2008;107(1):1-14. [Crossref]
38. Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res*. 2004;6(3):R149-156. [Crossref]
39. Korhonen T, Huhtala H, Holli K. A comparison of the biological and clinical features of invasive lobular and ductal carcinomas of the breast. *Breast Cancer Res Treat*. 2004;85(1):23-29. [Crossref]
40. Molland JG, Donnellan M, Janu NC, Carmalt HL, Kennedy CW, Gillett DJ. Infiltrating lobular carcinoma--a comparison of diagnosis, management and outcome with infiltrating duct carcinoma. *Breast*. 2004;13(5):389-396. [Crossref]
41. Fan M, Yuan W, Zhao W, et al. Joint prediction of breast cancer histological grade and ki-67 expression level based on DCE-MRI and DWI radiomics. *IEEE J Biomed Health Inform*. 2020;24(6):1632-1642. [Crossref]
42. Nam SY, Ko ES, Lim Y, et al. Preoperative dynamic breast magnetic resonance imaging kinetic features using computer-aided diagnosis: Association with survival outcome and tumor aggressiveness in patients with invasive breast cancer. *PLoS One*. 2018;13(4):e0195756. [Crossref]
43. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging*. 1999;10(3):223-232. [Crossref]
44. Dietzel M, Zoubi R, Vag T, et al. Association between survival in patients with primary invasive breast cancer and computer aided MRI. *J Magn Reson Imaging*. 2013;37(1):146-155. [Crossref]
45. Cho N, Kim SM, Park JS, et al. Contralateral lesions detected by preoperative MRI in patients with recently diagnosed breast cancer: application of MR CAD in differentiation of benign and malignant lesions. *Eur J Radiol*. 2012;81(7):1520-1526. [Crossref]
46. Piccoli CW. Contrast-enhanced breast MRI: factors affecting sensitivity and specificity. *Eur Radiol*. 1997;7(Suppl 5):281-288. [Crossref]
47. Baltzer PAT, Zoubi R, Burmeister HP, et al. Computer assisted analysis of MR-mammography reveals association between contrast enhancement and occurrence of distant metastasis. *Technol Cancer Res Treat*. 2012;11(6):553-560. [Crossref]
48. El-Hawary AK, Abbas AS, Elsayed AA, Zalata KR. Molecular subtypes of breast carcinoma in Egyptian women: clinicopathological features. *Pathol Res Pract*. 2012;208(7):382-386. [Crossref]
49. Li J, Chen Z, Su K, Zeng J. Clinicopathological classification and traditional prognostic indicators of breast cancer. *Int J Clin Exp Pathol*. 2015;8(7):8500-8505. [Crossref]
50. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn HJ; Panel Members. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol*. 2015;26(8):1533-1546. [Crossref]
51. Yuan C, Jin F, Guo X, Zhao S, Li W, Guo H. Correlation Analysis of breast cancer DWI combined with DCE-MRI imaging features with molecular subtypes and prognostic factors. *J Med Syst*. 2019;43(4):83. [Crossref]
52. Cao M, Liu X, Yang A, Xu Y, Zhang Q, Cao Y. Prediction of HER-2 expression status in breast cancer based on multi-parameter MRI intratumoral and peritumoral radiomics. *Magn Reson Imaging*. 2025;122:110434. [Crossref]
53. Chen ST, Lai HW, Tseng HS, Chen LS, Kuo SJ, Chen DR. Correlation of histologic grade with other clinicopathological parameters, intrinsic subtype, and patients' clinical outcome

in Taiwanese women. *Jpn J Clin Oncol*. 2011;41(12):1327-1335. [Crossref]

54. Schwartz AM, Henson DE, Chen D, Rajamarthandan S. Histologic grade remains a prognostic factor for breast cancer regardless of the number of positive lymph nodes and tumor size: a study of 161 708 cases of breast cancer from the SEER Program. *Arch Pathol Lab Med*. 2014;138(8):1048-1052. [Crossref]

55. Schmitz AC, Peters NHGM, Veldhuis WB, et al. Contrast-enhanced 3.0-T breast MRI for characterization of breast lesions: Increased specificity by using vascular maps. *Eur Radio*. 2008;18(2):355-364. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.252878

CARDIOVASCULAR IMAGING

INVITED REVIEW



# Computed tomography and magnetic resonance imaging findings in congenital cardiovascular anomalies

Mustafa Koplay
Nusret Seher

Selçuk University Faculty of Medicine, Department of Radiology, Konya, Türkiye

**ABSTRACT**

The incidence of congenital heart disease (CHD) is approximately 6-7 per 1,000 newborns. With advanced diagnostic, medical, and surgical methods, survival of CHD is increasing, as is the number of people living with CHD. Echocardiography is a useful modality in non-invasive imaging, whereas magnetic resonance imaging (MRI), cardiac MR (CMR), cardiac computed tomography (CT), and CT angiography (CTA) are increasingly gaining ground in congenital cardiac imaging with developing technology. Considering the limited postoperative use of echocardiography, these techniques have assumed vital roles with the increasing population of CHD in children and adults. CMR and cardiac CT can complement the information obtained with echocardiography and invasive cardiac catheterization and can sometimes provide more detail. In postoperative imaging of CHD, CMR allows an evaluation of anatomy, especially with spin echo MRI techniques, whereas cine MRI, created in gradient echo sequences, allows functional data to be obtained. Phase contrast CMR data provides information on flow direction and flow rate, allowing accurate measurement of regurgitation and shunt volume. In addition, in post-gadolinium imaging, data such as on MR angiography, myocardial perfusion, and fibrosis can be obtained with CMR. Cardiac CT and CTA provide great advantages, especially in newborns, by almost completely reducing movement and respiratory artifacts through capabilities such as high spatial and temporal resolution, fast acquisition, and short acquisition time. Three-dimensional reformatted images with contrast-enhanced CMR or cardiac CT/CTA provide excellent visualization of vascular structures in complex CHDs. Coronary imaging can be viewed more easily with CT imaging. This article reviews the literature to provide an overview of the diagnostic value, relative advantages, and overall evaluation of CMR and cardiac CT examinations in the diagnosis and postoperative follow-up of CHD.

**KEYWORDS**

Cardiac imaging, congenital heart disease, cardiac computed tomography, cardiac magnetic resonance imaging, computed tomography angiography, magnetic resonance angiography

Congenital cardiac malformations occur in 6-7 out of every 1,000 live births.[1] Major advances in the diagnosis and treatment of congenital heart disease (CHD) in recent years have led to increased survival rates, even in patients with the most complex malformations. Studies conducted in light of these developments indicate that the number of adults living with CHD has surpassed the number of children with CHD.[2] Echocardiography and cardiac angiography are the primary traditional imaging techniques used to evaluate patients with CHD.[3] However, these techniques have certain limitations. Echocardiography is limited by its narrow field of view, the need for an adequate acoustic window for imaging, and operator experience. Catheter angiography is limited by the superimposition of structures being examined, the risk of complications associated with the catheter, and its suboptimal evaluation of extracardiac structures. Studies have demonstrated that cardiac catheterization, the gold standard for cardiac imaging in terms of hemodynamic assessment, causes death in 1% of newborns with complex CHD.[4] In recent years, the use of cardiac computed tomography (CT) and magnetic resonance imaging (MRI) in the evaluation of this condition has increased substantially. Considering the long acquisition times of MRI and the age group of patients, MRI may require sedation and can lead to artifacts and potential anesthesia-related

**Handling editor:** Furkan Ufuk

**Corresponding author:** Mustafa Koplay

**E-mail:** koplaymustafa@hotmail.com

Received 09 July 2025; revision requested 28 July 2025; last revision received 08 September 2025; accepted 05 October 2025.

Epub: 17.11.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.252878

You may cite this article as: Koplay M, Seher N. Computed tomography and magnetic resonance imaging findings in congenital cardiovascular anomalies. *Diagn Interv Radiol*. 2026;32(4):409-425.

complications. However, recent studies have reported that the data obtained from compressed sensing four-dimensional (4D) flow MRI shows no significant difference compared with the reference standard two-dimensional (2D) phase contrast.[5] CT scans can be performed in seconds, thereby limiting the need for sedation in patients who are clinically incompatible. With isovolumetric imaging in CT, high-quality multiplanar reconstructions (MPR) and three-dimensional (3D) images can now be created. Finally, CT provides more detailed information about extracardiac structures than MRI. Imaging modalities for early and accurate diagnosis of CHD, along with advancements in surgical and medical treatments, have led to improved long-term survival.[6] To ensure that this multidisciplinary approach is properly implemented, it is essential that radiology, cardiology, and cardiovascular surgery units use a common language in classifying and defining these diseases so that they can play a role in the diagnosis and follow-up of CHD. In this review, CT and MRI findings in CHD are discussed.

## Computed tomography in congenital heart disease

Multi-detector CT (MDCT) scanners quickly collect volumetric data with high spatial resolution. The increased scanning speed reduces motion and breathing artifacts during imaging, enabling images to be obtained at high heart rates.[7] By completing the scan in a very short time, it also eliminates the need for sedation.[8] Although recent advancements in scanners and protocols have reduced radiation exposure, radiation remains a key concern for children. In addition to cardiac anatomy, CT angiography (CTA) with CT imaging helps evaluate the lung parenchyma, chest wall structures, tracheobronchial structures, and related pathologies of the aorta–pulmonary vascular structure.[9] Data from CT is analyzed in three orthogonal planes (axial, coronal, and sagittal) using MPR. The volume rendering technique allows for 3D visualization of cardiac and extracardiac morphology. Cardiac CTA (cCTA) is a highly successful imaging modality in pediatric cardiac imaging; however, it has poorer temporal resolution than other cardiac imaging techniques, such as echocardiography, cardiac catheterization, and cardiac MR (CMR).[10] In the literature, a 64-slice MDCT system is considered the minimum adequate modality for the evaluation of coronary heart disease and coronary artery disease, as well as for optimizing image quality through radiation dose-reduction and noise-reduction algorithms.[11] In pediatric populations, however, given the higher likelihood of heart rate and respiratory artifacts, faster scanners equipped with wider detector coverage are recommended.[12] Advances in MDCT technology, including faster and broader collimation as well as higher gantry rotation speeds, have led to reduced scan times, enhanced temporal resolution, and, as a result, diminished motion and respiratory artifacts.[13] Functional assessment is generally feasible with MDCT scanners with 256 slices or more. Next-generation MDCT scanners, including 256–320/640-slice systems and dual-source MDCT, achieve a 50%–70% reduction in radiation exposure through shorter scan times, reaching as low as 0.12 seconds.[13,14] Functional cardiac assessment is generally feasible with scanners of 256 slices or more.[15] Moreover, recent studies in pediatric cardiac imaging indicate that photon-counting CT offers superior resolution and a reduced radiation dose when compared with dual-source CT.[16] Contrast medium (CM) is required for all scans, and the CM dose is determined based on body weight. For imaging without coronary arteries, the routine dose is 1.0 mL/kg of CM containing 300 mg of iodine per mL, and for imaging with coronary arteries, the dose is 1.0–1.2 mL/kg of CM containing 350–370 mg of iodine per mL.[17]

**Main points**

- The use of cardiac computed tomography (CT) and magnetic resonance imaging (MRI) in the diagnosis and postoperative follow-up of congenital heart disease (CHD) is steadily increasing.
- Cardiac CT, with its high spatial and temporal resolution, rapid acquisition, and reduced motion artifacts, plays a crucial role, particularly in newborns and in the evaluation of complex vascular anomalies.
- Cardiac MRI, as a radiation-free modality, is advantageous for long-term follow-up and provides valuable functional information, including ventricular performance, flow quantification, shunt assessment, and myocardial fibrosis detection.
- Both modalities complement the limitations of echocardiography and invasive angiography, offering critical data for surgical planning and postoperative management.
- Ongoing technological advances have enabled lower radiation doses, shorter imaging times, and improved image quality, thereby contributing to the prognosis and long-term outcomes of pediatric and adult patients with CHD.

## Cardiac computed tomography angiography and computed tomography angiography techniques

There are primarily two imaging techniques: non-electrocardiogram (ECG)-triggered and ECG-triggered imaging. Non-ECG-triggered CTA is primarily used for evaluating non-cardiac structures, whereas ECG-triggered imaging is primarily used for coronary artery imaging.[15] ECG-triggered imaging has two modes: prospective and retrospective ECG-triggered imaging.[14] Retrospective ECG-triggered imaging acquires data for the R–R interval of a complete cardiac cycle, thereby enabling multisegmental reconstruction, improved temporal resolution, and assessment of left ventricular function. However, despite these advantages, it results in 3–5 times the radiation exposure of prospective imaging.[13] Additionally, it allows imaging at heart rates of up to 170 bpm. In prospective ECG-triggered imaging, high pitch values and dual-source devices enable imaging. This technique can also be referred to as the "step and shoot" technique. It is necessary for the heart rate to remain low, and more importantly, stable.[18,19] In this technique, the CT scanner is positioned during the R–R cycle of the heart, applying maximum tube current only at end diastole in adults and early systole in children and completely stopping radiation during the remainder of the cardiac cycle, thereby reducing the dose.[20] The use of a low tube current and the inability to visualize all phases of the heart limit functional assessment.[18] Table 1 presents the technical specifications and protocols of next-generation CT systems.[16]

## Cardiac magnetic resonance imaging in congenital heart disease

In recent years, CMR has gained importance in the evaluation of complex malformations and the morphological and functional assessment of CHDs.[21] Due to the absence of ionizing radiation, it is ideal for long-term CHD follow-up. Moreover, it provides blood flow data, and detailed functional assessment can be performed using cine images.[18] CMR provides images that can serve as standard references for the accurate and reproducible quantification of right and left ventricular volumes, mass, and function.[22] The steady-state free precession (SSFP) sequence is preferred in CMR examinations because it provides good contrast between blood and the myocardium. The 2D "white blood" SSFP sequence can provide dynamic information about the heart and large vessels.[23] The "black blood" spin echo sequences

**Table 1.** Technical characteristics of new-generation computed tomography scanners and recommended parameters for contrast media injection

| | DSCT (2 × 192) | PCCT (2 × 144) |
|---|---|---|
| Extracardiac structure | Non-ECG gated | Non-ECG gated |
| Intracardiac structure/coronary arteries | ECG gated | ECG gated |
| Spatial resolution (mm) | 0.24 | 0.11 |
| Temporal resolution (msec) | 66 | 66 |
| Gantry rotation time (sec) | 0.25 | 0.25 |
| Maximum scan speed (mm/sec) | 737 | 737 |
| Tube voltage (kVp) | 70 | 90 |
| Tube current (mA) | Automatic modulation | Automatic modulation |
| Pitch | 3.0 | 3.0 |
| Iterative reconstruction | Available | Available |
| Mean dose (mSv) | 0.50 | 0.52 |
| Contrast medium | Iodine | Iodine |
| Concentration | 320–370 mg/mL | 320–370 mg/mL |
| Dose | 1–1.2 mL/kg | 1–1.2 mL/kg |
| Injection rate | 1.5–4 mL/second | 1.5–4 mL/second |

DSCT, dual-source computed tomography; PCCT, photon-counting computed tomography; kVp, kilovoltage peak; mA: milliampere; mSv, milisievert; ECG, electrocardiogram.

obtained with the double inversion recovery technique provide static high-resolution images. Another advantage is the minimal presence of metallic artifacts in these sequences. Images obtained with the 3D isotropic SSFP sequence allow evaluation in different planes without loss of resolution.[24] CMR is the preferred technique for assessing valvular insufficiency, with shunt quantification based on pulmonary-to-systemic flow ratios (Qp/Qs) achievable through flow measurements in the main pulmonary artery and aorta.[25] Additional advantages of CMR include comprehensive evaluation and grading of valvular dysfunction, accurate shunt quantification, detailed flow assessment, and non-invasive tissue characterization.[26] Recent studies have demonstrated that advancements in 4D phase contrast velocity mapping yield novel insights into cardiovascular hemodynamics, which are of particular relevance for surgical reconstruction. This is especially crucial in patients with single-ventricle physiology undergoing cavopulmonary connection. Furthermore, 4D flow MRI has been established as the non-invasive gold standard for postoperative follow-up of procedures such as the Fontan operation.[27] For contrast-enhanced MR angiography (MRA), a breath-hold 3D spoiled gradient echo sequence without ECG triggering is preferred.[28] It is particularly useful in evaluating abnormal pulmonary or systemic venous return and aortic coarctation (CoA), as well as in postoperative follow-up, and it correlates well with conventional angiography and CTA in identifying stenotic vascular structures.[29] In cases where contrast is contraindicated, the 3D SSFP sequence can be used. In delayed contrast studies, it enables the assessment of myocardial scarring and fibrosis, providing prognostic information for patients with ventricular dysfunction or myocardial evaluation following surgical repair.[30] Table 2 outlines the clinical scenarios in which specific CMR sequences should be applied.[31] Recent studies have highlighted the value of T1, T2, and extracellular volume mapping and dark- or gray-blood late gadolinium enhancement techniques for tissue characterization and the detection of subendocardial scarring in CHD.[32] Although CMR has become safer, it remains challenging in patients with metallic implants and pacemakers. Therefore, all cardiac imaging modalities are complementary, and it is essential to understand their strengths and weaknesses to optimize the use of various imaging modalities (Table 3).[2]

## Acyanotic heart defects: conditions and imaging considerations

### Atrial septal defect

Atrial septal defects (ASDs) are the most common congenital heart defects detected in adults (Figure 1).[33] There are three main types: secundum ASDs account for 75%, primum ASDs for 20%, and sinus venosus ASDs for 5%.[34] Coronary sinus-type ASDs are the rarest. cCTA and CMR are useful for detecting associated abnormal pulmonary veins located above the level of the azygos vein in the superior vena cava. The presence of an ASD causes left-to-right shunting, leading to atrial and ventricular enlargement on the right side. ASDs are also an independent risk factor for thromboembolic stroke.

### Ventricular septal defect

Ventricular septal defects (VSDs) are a shunt that occurs between the right ventricles (RVs) and left ventricles (LVs) through a defect in the ventricular septum (Figure 2). VSDs are the most common type of CHD detected in newborns,[35] and they can occur in isolation or with other anomalies. When associated with multiple anomalies, if the most notable hemodynamic abnormality is the shunt associated with the VSD, the VSD is considered the primary anomaly.[36] VSD classifications are based on the location of the defect, and the most common type is the perimembranous type. Other types include the muscular, outlet, and inlet types. Both MRI and CT are useful in determining the location and type of a VSD.[37]

### Atrioventricular septal defect

Atrioventricular septal defects (AVSDs) are also known as endocardial cushion defects or atrioventricular canal defects (Figure 3). These types of defects involve the atrial and ventricular septum, with one or both of the mitral and tricuspid valves.[38] Clinical findings may present as complete defects (ASDs and VSDs) or partial defects (only ASDs).[39] AVSDs result from the underdevelopment of the apical portion of the atrial septum, the basal portion of the ventricular septum, and the AV valves. In this defect, the four chambers of the heart are generally interconnected, allowing shunts to form from left to right or right to left. There is a strong association between AV canal defects and trisomy 21, with a risk of trisomy 21 being as much as 50% in fetuses with AV canal defects.[40] Classification is based on complete and partial types. In the complete type, a posterior or inlet VSD is accompanied by a primum ASD and a common AV valve. In the partial type, a VSD is not present, and there is a single AV valve annulus with two separate valve orifices.[41] The most common partial form involves primum ASDs and a cleft in the anterior leaflet of the mitral valve.[38]

### Patent ductus arteriosus

The ductus arteriosus (DA), a normal vascular structure connecting the pulmonary artery and the aorta during fetal life, allows blood from the RV to pass into the aorta. The DA accounts for 41% of total cardiac output.[42]

**Table 2.** Indications for the use of specific sequences in cardiac magnetic resonance imaging

| Sequence | Indication |
|---|---|
| Cine bSSFP | RV/LV size, shape, and volume; atrial size and shape; ventricular function; mass and EF; LVOT/RVOT obstruction; assessment of pulmonary artery anatomy and pulsatility; aortic valve morphology; aortic anatomy; measurement of aortic root; ascending aorta and aortic arch |
| Tagging | Screen for fusion of pericardium and myocardium in constriction |
| Phase contrast flow | Pulmonary regurgitation, differential pulmonary blood flow, aortopulmonary collateral flow measurement, Qp/Qs measurement in the context of a suspected shunt, valve dysfunction |
| Black blood imaging (especially with metallic artifacts) | Arterial or venous evaluation |
| Contrast-enhanced magnetic resonance angiography | Indicated as part of first study: pulmonary artery and vein anatomy, extracardiac vascular anatomy |
| T1- or T2-weighted with post-gadolinium spin echo or cine SSFP | Pericardial layer or fluid characterization |
| 3D bSSFP imaging | Intracardiac anatomy |
| 4D flow imaging | Assessment of pulmonary arterial pressures, wall shear stress, RV diastolic function |
| T2 imaging and/or mapping | Myocardial edema |
| T1 mapping (native) | Myocardial characterization |
| T1 mapping (post contrast) | Myocardial characterization, extracellular volume |
| Early gadolinium enhancement | Hyperemia |
| Late gadolinium enhancement | Myocardial inflammation, fibrosis, fat |

3D, three-dimensional; 4D, four-dimensional; EF, ejection fraction; LV, left ventricular; RV, right ventricular; bSSFP, balanced steady-state free precession; LVOT, left ventricular outflow tract; RVOT, right ventricular outflow tract; Qp/Qs, pulmonary–systemic flow ratio.

**Table 3.** Relative advantages of CMR and CT/CTA

| | CMR | CT/CTA |
|---|---|---|
| Morphology | ++++ | ++ |
| Function | ++++ | ++ (with ECG gated) |
| Extracardiac structure | ++ | ++++ |
| Ventricular volumes | ++++ | ++ |
| Flow quantification | ++++ | – |
| Myocardial perfusion | ++++ | ++ |
| Myocardial viability | ++++ | ++ |
| Intracardiac anatomy | ++++ | ++ |
| Coronary artery evaluation | + | ++++ |
| Radiation | – | ++++ |
| Examination time | ++++ | + |
| Contrast allergy | + | ++ |
| Emergency case | ++ | ++++ |
| Real-time viewing | ++++ | – |
| Calcium imaging | + | ++++ |
| Long-axis resolution | + | ++++ |
| Trained radiologists | + | ++ |

CMR, cardiac magnetic resonance; CT, computed tomography; CTA, computed tomography angiography; ECG, electrocardiogram.

Unlike other vascular structures, the media layer of the DA contains smooth muscle.[43] After birth, this smooth muscle contracts and thickens, thereby reducing the lumen size and causing the functional closure of the DA. Functional closure occurs in 44% of infants within 24 hours and in 88% within 48 hours. Failure of the DA to close postnatally results in patent DA (PDA) (Figure 4). The incidence of PDA in term infants ranges from 3 to 8 per 10,000 live births and is more common in female infants.[44] Prematurity is the most important risk factor for PDA, and the risk of PDA in premature infants is generally more than 10 times higher.[45] Although both right and left PDA are possible, the most common type is left-sided PDA between the main pulmonary artery and the proximal descending aorta.[46] PDAs are classified into five types based on their angiographic appearance, as described by Krichenko et al.[47] The hemodynamic outcomes of PDA vary depending on the extent of the left-to-right shunt. In a small PDA, the pulmonary–systemic flow ratio (Qp/Qs) is <1.5, in a medium-sized PDA, it is <2.2, and in a large PDA, it is >2.2.[48]

## Partial anomalous pulmonary venous drainage

A defect in the normal embryonic development of the pulmonary veins can lead to congenital anomalies ranging from partial anomalous pulmonary venous connection (PAPVC) to total anomalous pulmonary venous connection. Abnormal pulmonary venous variations can be seen in up to 38% of the population.[49] The most common type is drainage of the right pulmonary vein into the superior vena cava or right atrium (Figure 5).[50] The anomaly may be associated with the presence of an associated ASD or may occur as part of a complex CHD. Associated congenital anomalies serve as the main determinants of morbidity and mortality in PAPVC.[51] Currently, MDCT and MRA have replaced traditional imaging techniques and provide optimal evaluation capabilities.

## Aortic stenosis

Left ventricular outflow tract obstructions account for 6% of CHD cases (Figure 6),[26]

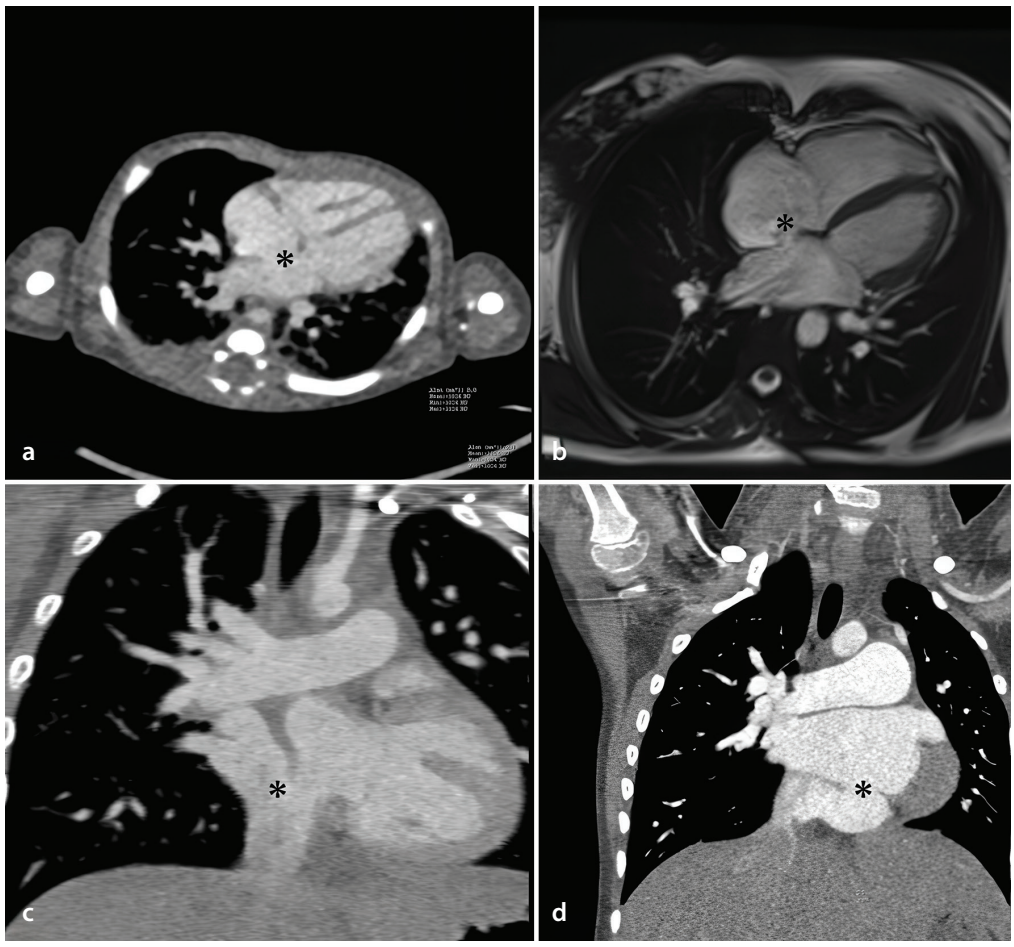



**Figure 1.** Primum-type atrial septal defect (ASD; black star) in an axial computed tomography angiography (CTA) image **(a)**, secundum-type ASD (black star) in a four-chamber cine cardiac magnetic resonance image **(b)**, sinus venosus-type ASD (black star) in a coronal CTA image **(c)**, and coronary sinus-type ASD (black star) in a coronal CTA image **(d)**.

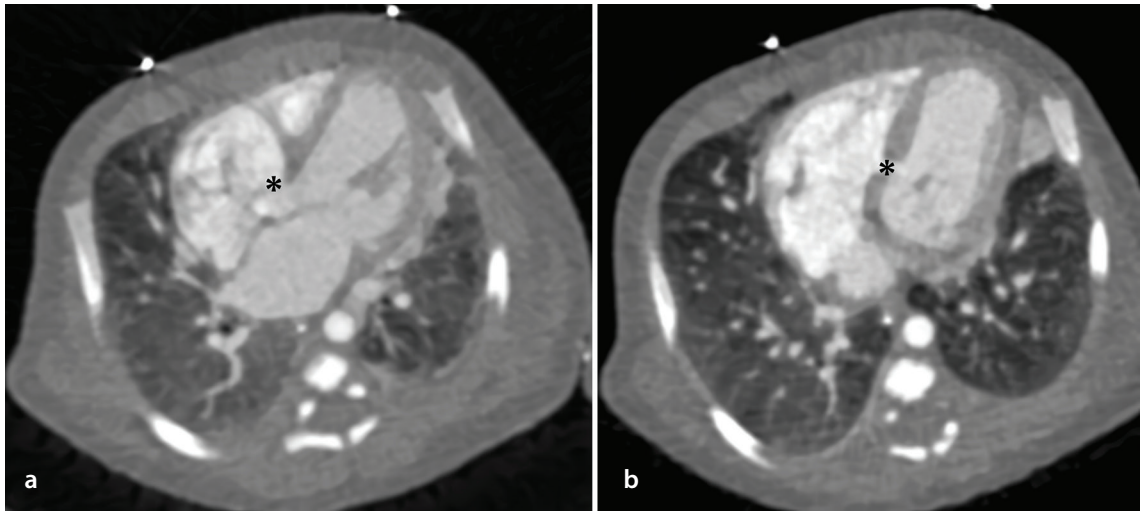



**Figure 2.** Axial computed tomography angiography images: **(a)** perimembranous-type (black star) and **(b)** muscular-type ventricular septal defects.

with valvular aortic stenosis (AS) accounting for 70%.[52] Valvular AS most commonly occurs secondary to a bicuspid aortic valve, and it may be accompanied by Turner syndrome. Critical neonatal AS is characterized by a myxomatous valve and an aortic valve opening as small as a needle tip. The aortic valve is typically hypoplastic, and mitral valve hypoplasia, left ventricular hypoplasia, and VSDs frequently accompany it.[40] Supravalvular AS is rarely seen, and Williams syndrome is present in 30%–50% of cases. The subvalvular type accounts for 10%–20% of cases. It causes narrowing due to a simple membrane or a fibromuscular structure resembling a long tunnel in the left ventricular outflow tract. It is often associated with other left ventricular anomalies, including Shone complex.[40]

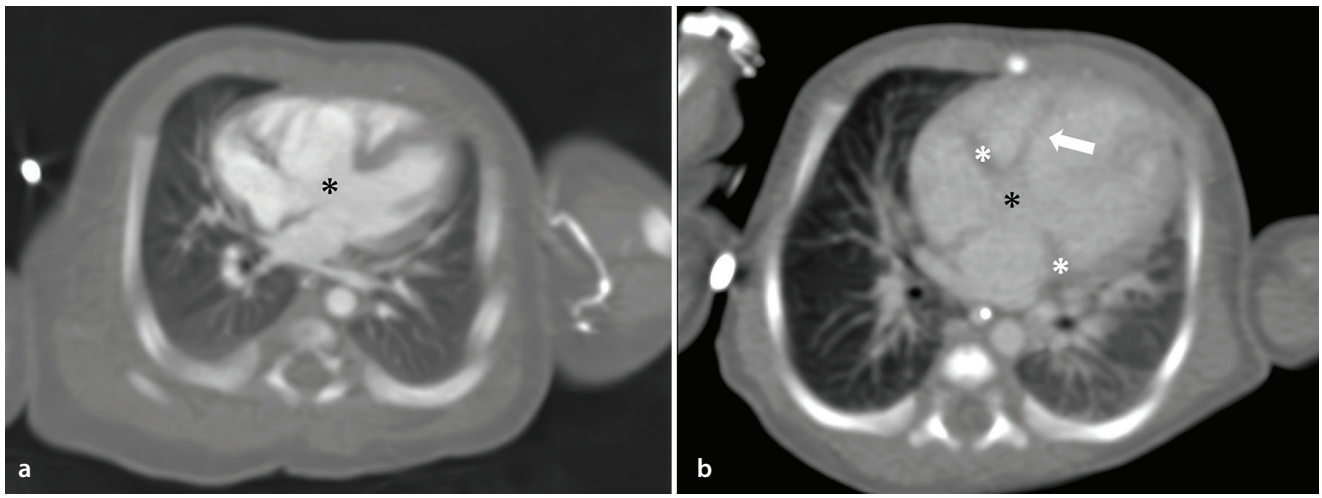



**Figure 3.** Axial computed tomography angiography images: complete-type **(a)** atrioventricular septal defect (AVSD; black star) [primum-type atrial septal defect (ASD)], membranous-type ventricular septal defect, single AV valve) and incomplete-type **(b)** AVSD (black star) [primum-type ASD and two separate valves (white stars)]. The interventricular septum (white arrow) is intact.


**Figure 4.** Type A patent ductus arteriosus (PDA; black arrow) in a three-dimensional (3D) volume rendered image **(a)**, Type B PDA (black arrow) in a sagittal reformatted computed tomography angiography image **(b)**, and Type E PDA (black arrow) in a 3D volume rendered image **(c)**.

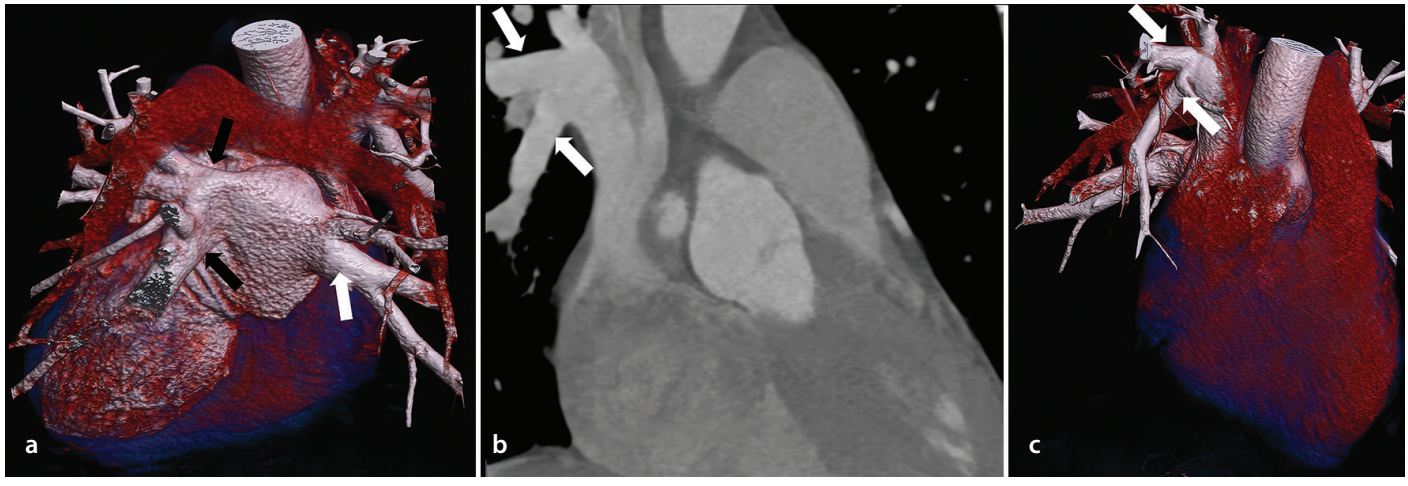



**Figure 5.** In the three-dimensional (3D) volume rendered cardiac computed tomography (CT) **(a)** image of a 25-year-old male patient, the left pulmonary veins (black arrows) and the right inferior pulmonary vein (white arrow) open into the left atrium, whereas in the coronal cardiac CT **(b)** and 3D volume rendered **(c)** images, the right upper pulmonary vein, formed by the union of two veins, opens into the superior vena cava (white arrows).

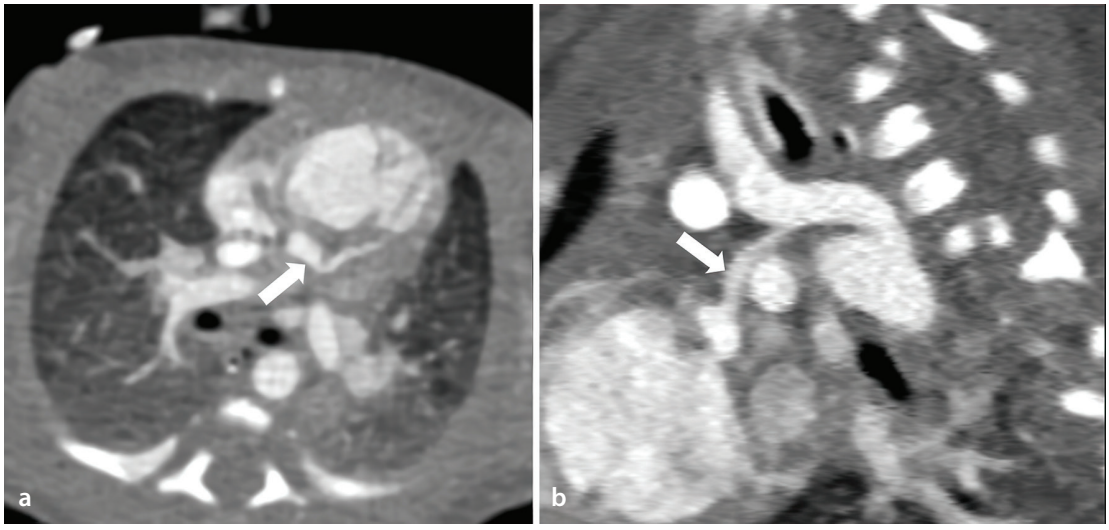



**Figure 6.** Axial **(a)** and sagittal reformatted **(b)** computed tomography angiography images of a 6-day-old female patient showing supravalvular diffuse aortic stenosis (white arrows).

### Pulmonary stenosis

Pulmonary stenosis (PS) accounts for 8%–12% of CHDs,[40] with valvular PS the most common type (Figure 7). The pulmonary valve has thickened leaflets, and the commissures may be absent or fused. In most patients, the pulmonary valve appears dome shaped. Severe PS is associated with right ventricular and infundibular muscle hypertrophy.[39] Critical PS is the most serious type of PS, resulting in inadequate pulmonary blood flow, and is seen in 10%–20% of patients. Dysplastic pulmonary valve is associated with Noonan syndrome.[53] Tricuspid atresia, tetralogy of Fallot (ToF), and transposition of the great arteries (TGA) may accompany PS. Alagille, Williams, and congenital rubella syndromes are associated with peripheral-type PS.[54]

### Aortic coarctation

CoA accounts for 5%–8% of CHDs (Figure 8),[55] and may occur in isolation or in association with other cardiovascular anomalies, such as bicuspid aortic valve (50%–75%), aortic arch hypoplasia, subaortic stenosis, mitral valve abnormalities, AVSDs, and PDAs.[56] CoA is generally sporadic and may be accompanied by extracardiac anomalies. Cerebral aneurysms are observed in 2.5%–10% of patients with CoA.[57] Preductal CoA occurs proximal to the DA and is typically seen in newborns.[58] Moreover, preductal CoA is the most common type of coarctation in Turner syndrome.[59] Postductal coarctation, however, typically occurs after the neonatal period and is anatomically located distal to the insertion of the DA. Patients with postductal coarctation are generally asymptomatic and are diagnosed based on decreased femoral artery pressure and systemic arterial hypertension.[56]

### Interrupted aortic arch

Interrupted aortic arch is a rare congenital anomaly, with an incidence of approximately 2 in 100,000.[60] It is characterized by the complete absence of connection between the ascending and descending aorta and may be connected by a remnant fibrous band. Blood flow to the descending aorta is usually via a

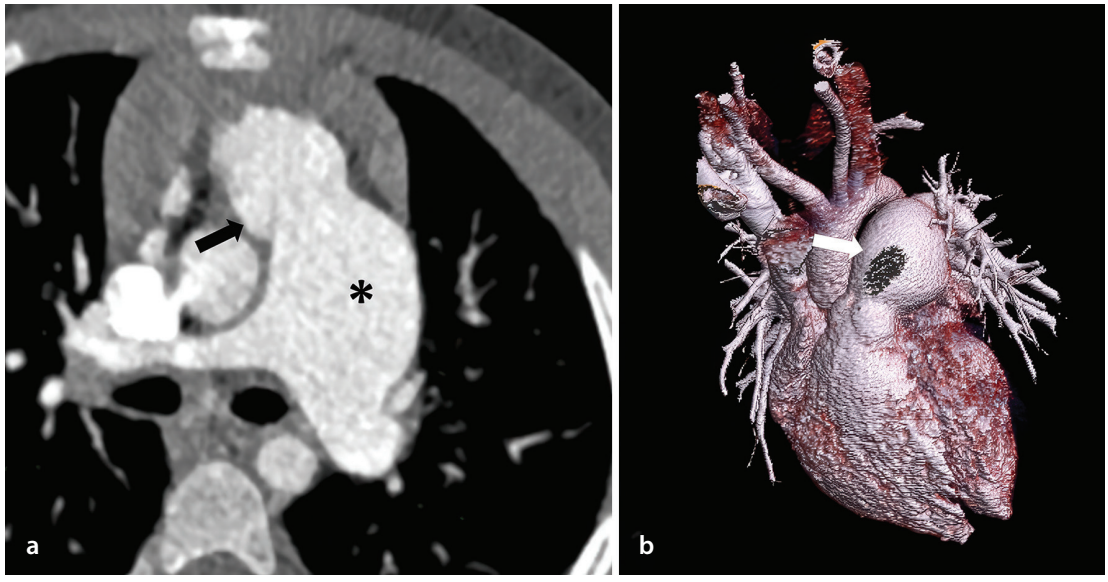



**Figure 7.** In the axial cardiac computed tomography angiography image **(a)** of a 6-year-old girl, stenosis (black arrow) and poststenotic dilatation (star) at the pulmonary valve level are observed. In the three-dimensional volume rendered image **(b)**, poststenotic dilatation is seen more clearly (white arrow).

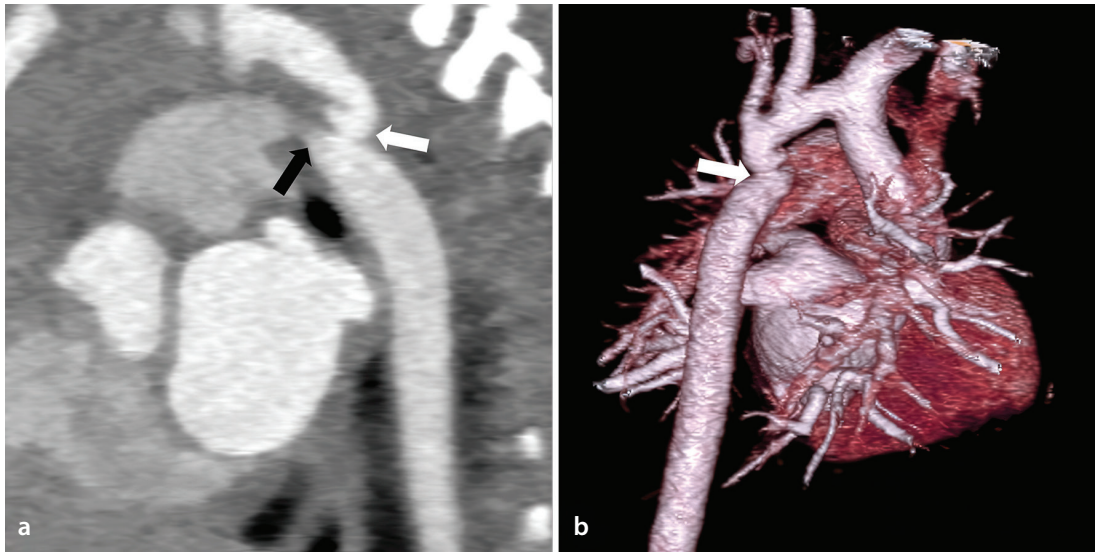



**Figure 8.** A 10-day-old male patient: sagittal computed tomography angiography image **(a)** showing a coarcted segment (white arrow) and patent ductus arteriosus at the level of the aortic isthmus (black arrow). In the three-dimensional volume rendered image **(b)**, the coarcted segment can be clearly seen (white arrow).

PDA. Interrupted aortic arch clinically resembles CoA, but the underlying mechanisms are thought to be different.[61] Interrupted aortic arch is classified into three types (A, B, and C) based on the location of the aortic interruption relative to the arch vessels. Type A occurs when the interruption is distal to the left subclavian artery (Figure 9), type B when it is between the carotid and subclavian arteries, and type C when it is between the carotid arteries. Subtypes are classified based on the normal or abnormal origin of the right subclavian artery. In the normal origin of the right subclavian artery (subtype 1), the aberrant right subclavian artery originates distally from the left subclavian artery (subtype 2), and the isolated right subclavian artery originates from the right DA (subtype 3).[62] Interrupted aortic arch is typically associated with a large VSD or PDA or, less commonly, with an aortopulmonary window or truncus arteriosus (TA).[61]

## Cyanotic heart defects: conditions and imaging considerations

ToF is the most common cyanotic CHD, occurring in approximately 1 in 3,600 live births.[63] There are four anatomical features: a VSD, an overriding aorta, a pulmonary outflow tract obstruction, and right ventricular hypertrophy (Figure 10).[64] Anterior deviation of the conal septum and infundibular hypoplasia cause pulmonary outlet obstruction. The primary problem in ToF is the inadequate development of the pulmonary infundibulum,[64] and the pulmonary valve is often abnormal. In patients with atresia, pulmonary blood flow is maintained via a PDA. The VSD is typically located in the membranous septum. In the variant of ToF with absent pulmonary valve, RV and pulmonary arterial dilatation are observed. When ToF is accompanied by an ASD, it is named Fallot's pentalogy. The degree of pulmonary regurgitation plays a key role in ToF follow-up (Figure 11).[65] In particular, quantification of RV volumes and function using CMR is indispensable for treatment protocols.[66]

## Transposition of the great arteries

TGAs refers to a condition in which the aorta originates from the RV and the main

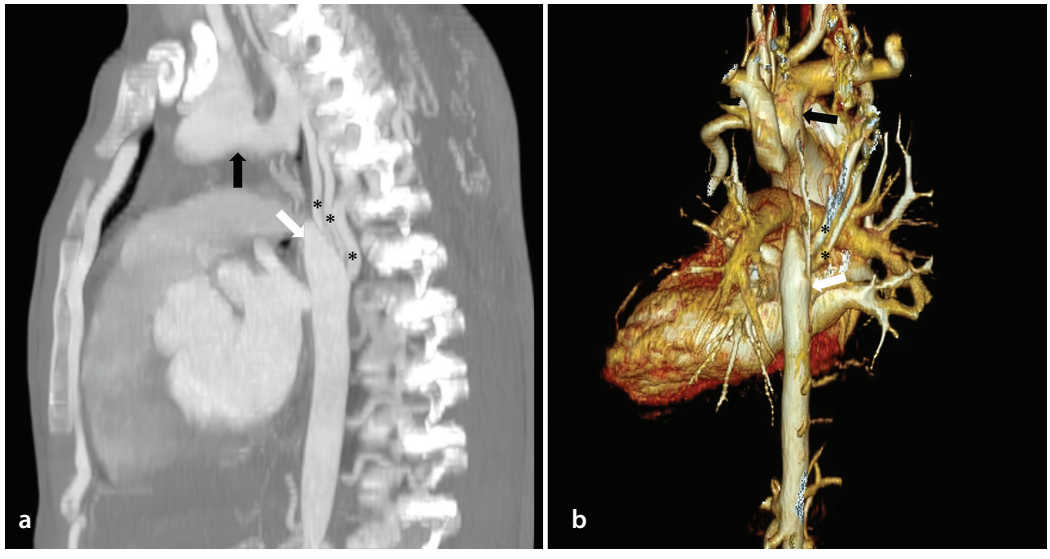



**Figure 9.** A 19-year-old male patient diagnosed with interrupted aortic arch (Type A): the aortic arch is normal (black arrows) in the sagittal reformatted computed tomography angiography image **(a)** and three-dimensional volume rendered image **(b)**, but an interruption is observed between the descending aorta (white arrows) and the left subclavian artery distal to it. Numerous collateral vascular structures draining into the descending aorta are observed (black stars).

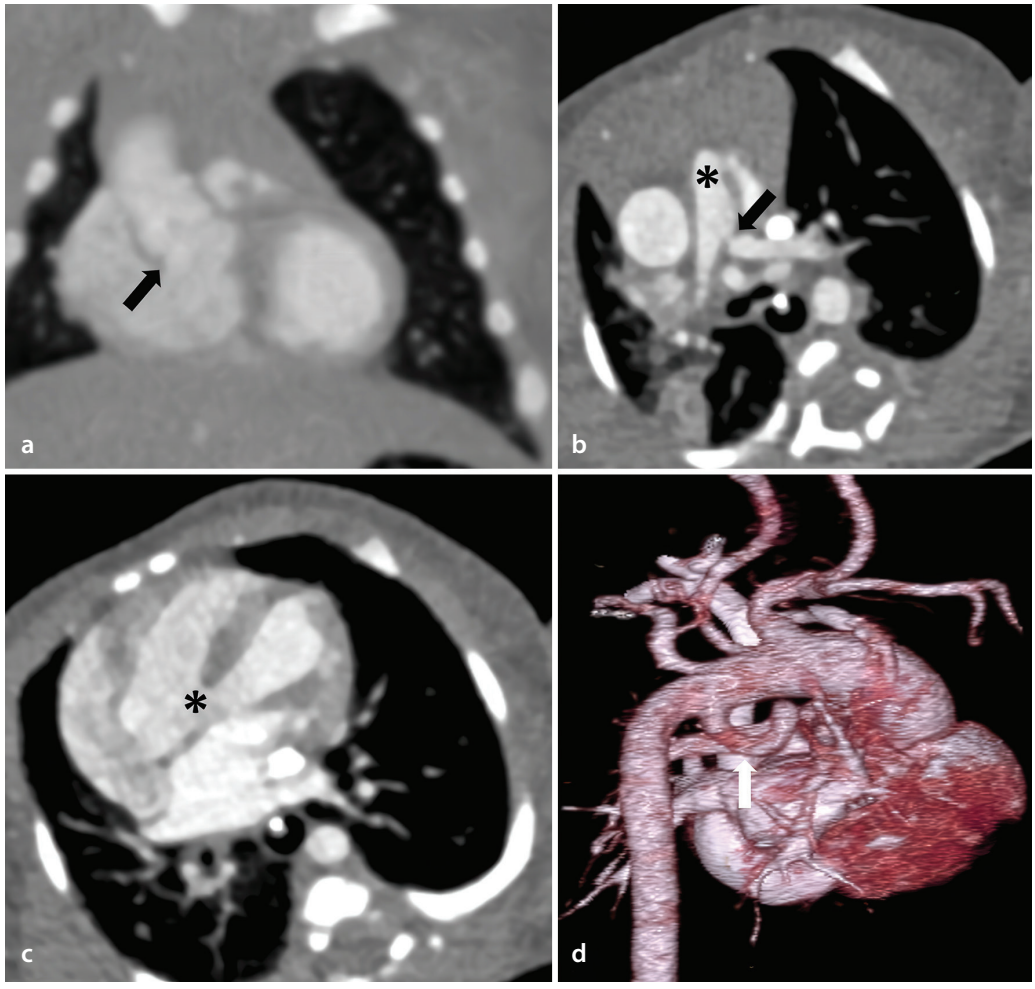



**Figure 10.** Male patient diagnosed with 2-day-old tetralogy of Fallot: dextroposed aorta is observed in a coronal computed tomography angiography (CTA) image **(a)** (black arrow), pulmonary trunk hypoplasia (star) and notable stenosis in the left pulmonary artery root (black arrow) in an axial CTA image **(b)**, perimembranous ventricular septal defect (black star) in an axial CTA image **(c)**, and patent ductus arteriosus in a three-dimensional volume rendered image **(d)** (white arrow).

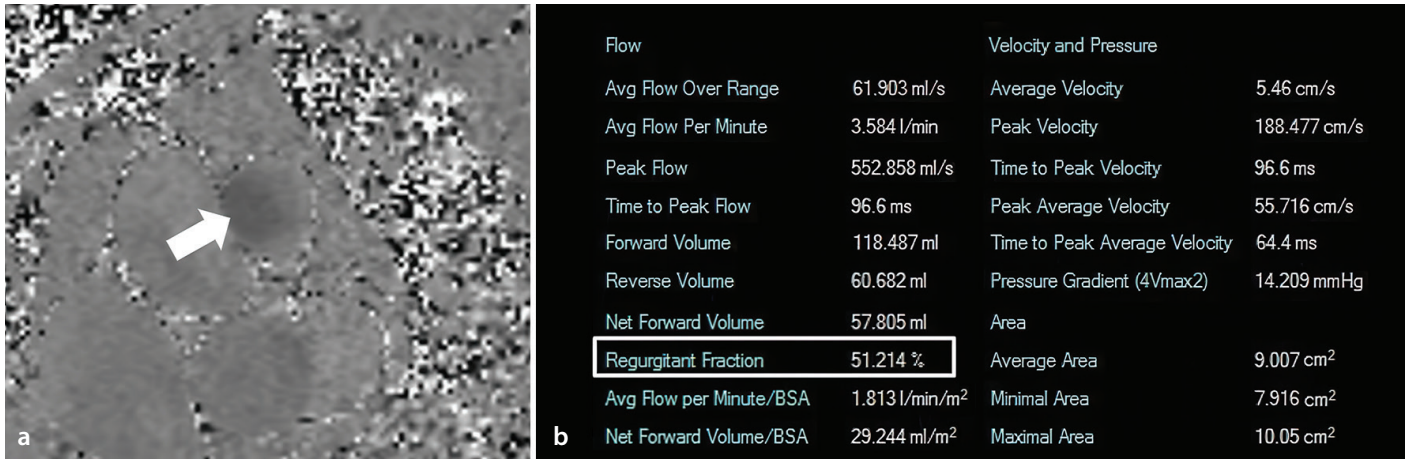



**Figure 11.** Control cardiac magnetic resonance image of a 62-year-old male patient who underwent surgery for tetralogy of Fallot 40 years ago: phase contrast flow examination **(a)** showing notable insufficiency findings in the pulmonary valve. Calculations made from the pulmonary valve **(b)** show a 51% regurgitation fraction in the pulmonary valve.

pulmonary artery originates from the LV (ventriculoarterial discordance).[67] The most common form is known as complete transposition and is often referred to as dextro-TGA or d-loop TGA (Figure 12). Complete transposition is characterized by AV concordance (the LV is connected to the left atrium and the pulmonary artery, whereas the RV is connected to the right atrium and the aorta). Congenital corrected TGA, also known as l-loop or L-TGA, is characterized by AV discordance (the left atrium is morphologically connected to the RV, and the right atrium is morphologically connected to the LV).[68] The incidence of TGA is 1 in 2,000 to 5,000 live births.[63,69] TGAs

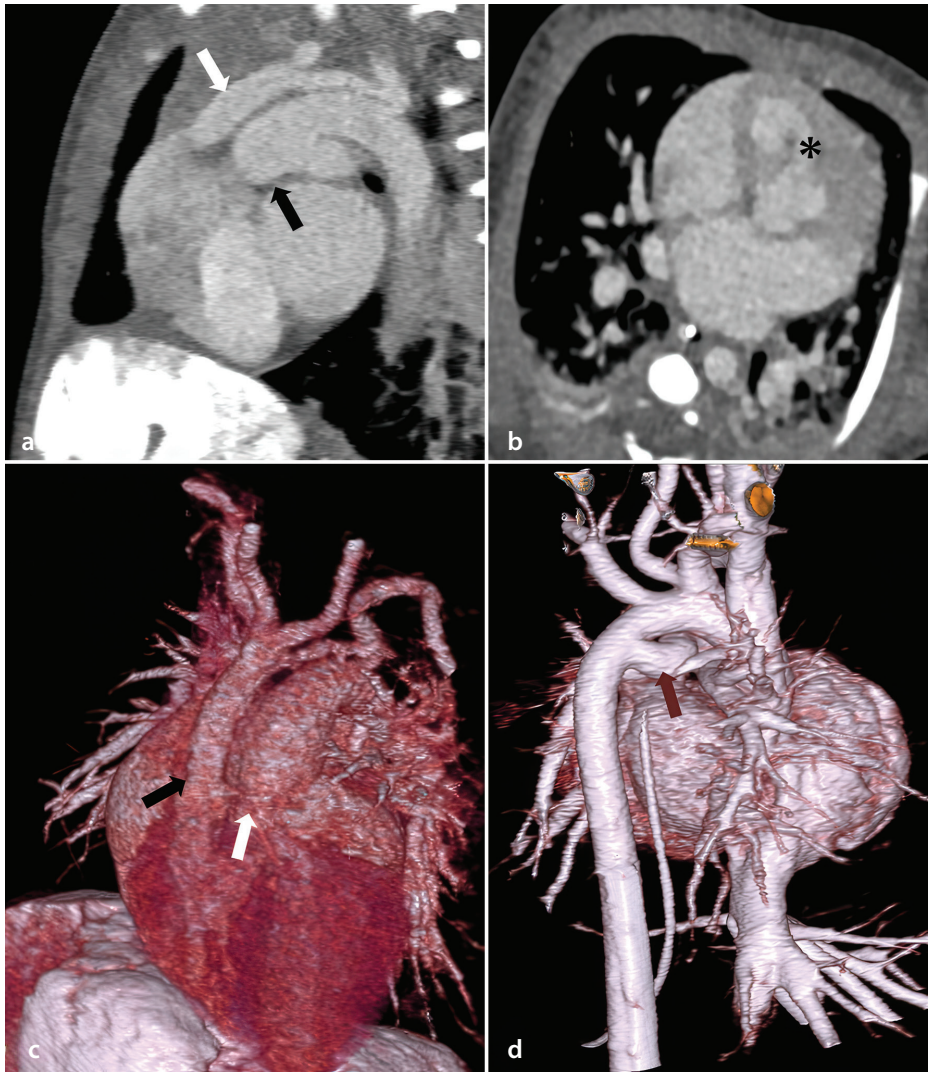



**Figure 12.** Patient diagnosed with 1-day transposition of the great arteries: sagittal computed tomography angiography (CTA) image **(a)** showing the aorta originating from the right ventricle and anteriorly (white arrow) and the pulmonary artery originating from the left ventricle and posteriorly (black arrow); axial CTA image **(b)** showing a membranous-type ventricular septal defect (black star). In three-dimensional volume rendered images **(c, d)**, the aorta originates from the right ventricle and anteriorly (black arrow), the pulmonary trunk originates from the left ventricle and posteriorly (white arrow), and there is a distinct patent ductus arteriosus (red arrow) between the aorta and pulmonary artery.

accounts for 5% to 7% of CHDs and 10% of all neonatal cyanotic CHDs.[70] A VSD is the most common associated anomaly (50%), and an ASD and PDA may also be present. It is the cyanotic CHD that presents with the earliest findings in newborns. Thoracic CTA is a highly useful imaging modality for diagnosis, as it can be performed rapidly and provides evaluation of vascular structures and surrounding tissues. Moreover, CMR can be used preoperatively to measure LV mass, function, and volume and to assess associated anomalies.[67]

## Truncus arteriosus

TA is a rare cyanotic CHD in which a single common trunk emerges from the heart instead of the aorta and pulmonary artery.[71] A VSD is always present, and the atria and ventricles are usually normal.[72] TA accounts for 1% to 2% of CHDs,[62] and is associated with DiGeorge syndrome. It is classified into subtypes according to the Collet and Edwards classification. In Type I (50%–70%), the pulmonary trunk arises from the common trunk and divides into the right and left pulmonary arteries (Figure 13). In Type II (30%–50%), the right and left pulmonary arteries arise separately from the posterior portion of the common trunk. In Type III (6%–10%), the pulmonary arteries arise from the lateral aspect of the common trunk.[69] In Type IV, pulmonary atresia (PA) is present and may be accompanied by VSDs, PDA, and major aortopulmonary collateral arteries (MAPCAs). In the revised classification by Van Praagh, Type IV is not considered a subtype of TA but rather a variant of PA.[64] Patients are initially asymptomatic due to high pulmonary vascular resistance but begin to develop symptoms as pulmonary vascular resistance decreases over several weeks.[72] Cyanosis is typically minimal but increases as pulmonary artery stenosis becomes more pronounced.[73] Preoperative CT or MRI is generally helpful in distinguishing the branching of the pulmonary arteries and the presence of MAPCAs, pulmonary venous drainage, aortic arch anomalies, and complex anatomy in the newborn.[74]

## Double outlet right ventricle

Double outlet RV (DORV) is a CHD in which both the aorta and the pulmonary trunk originate from the RV (Figure 14),[75] and a VSD is always present. Classification is based on the relationship of the VSD to the great vessels. The type associated with subaortic VSDs (50%) is the most common. The type associated with subpulmonary VSDs is also known as the Taussing–Bing anomaly. The type associated with bilateral VSDs and the type as-

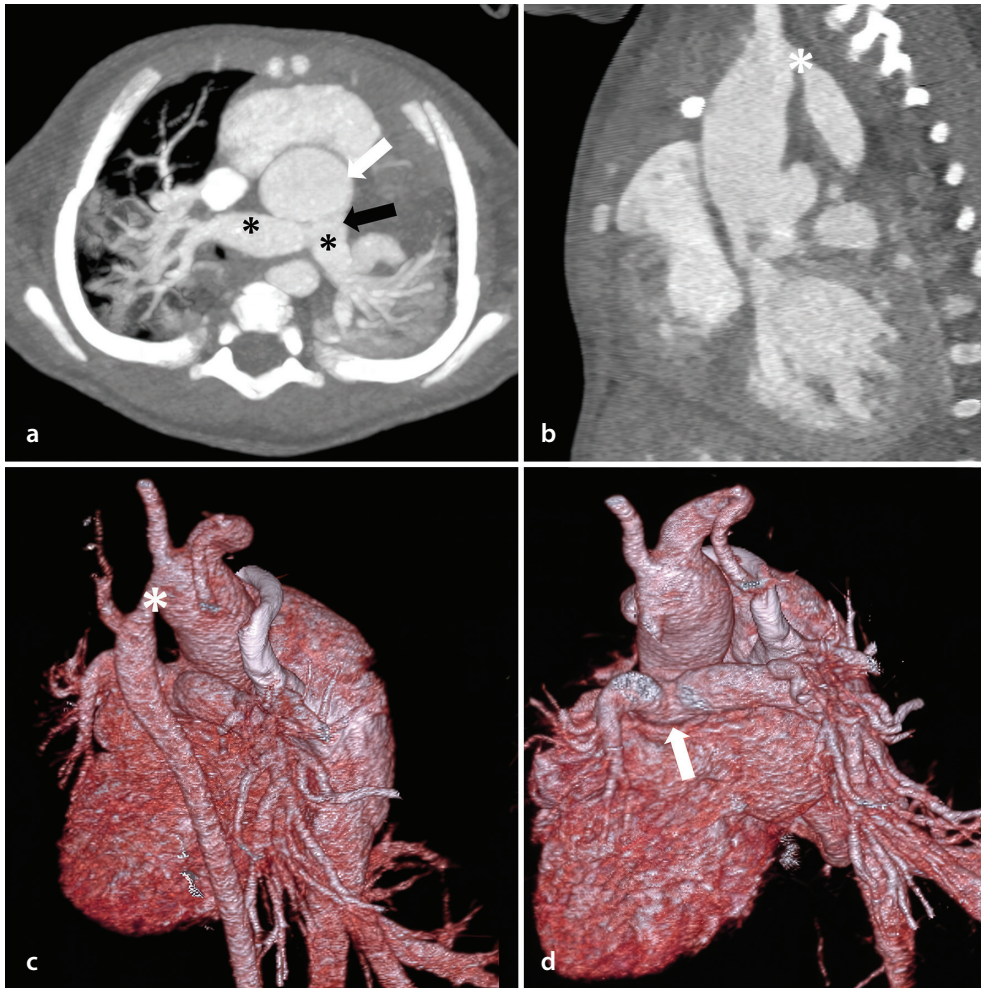



**Figure 13.** A 3-day-old male patient diagnosed with truncus arteriosus Type I: the common trunk (white arrow), the pulmonary trunk branching from the trunk (black arrow), and both main pulmonary arteries are seen in the axial computed tomography angiography (CTA) image **(a)** (black stars). In the sagittal CTA image **(b)** and three-dimensional (3D) volume rendering image **(c)**, the accompanying aortic coarctation is seen (white stars). In the 3D volume rendering image **(d)**, the pulmonary trunk arising from the common trunk is clearer (white arrow).

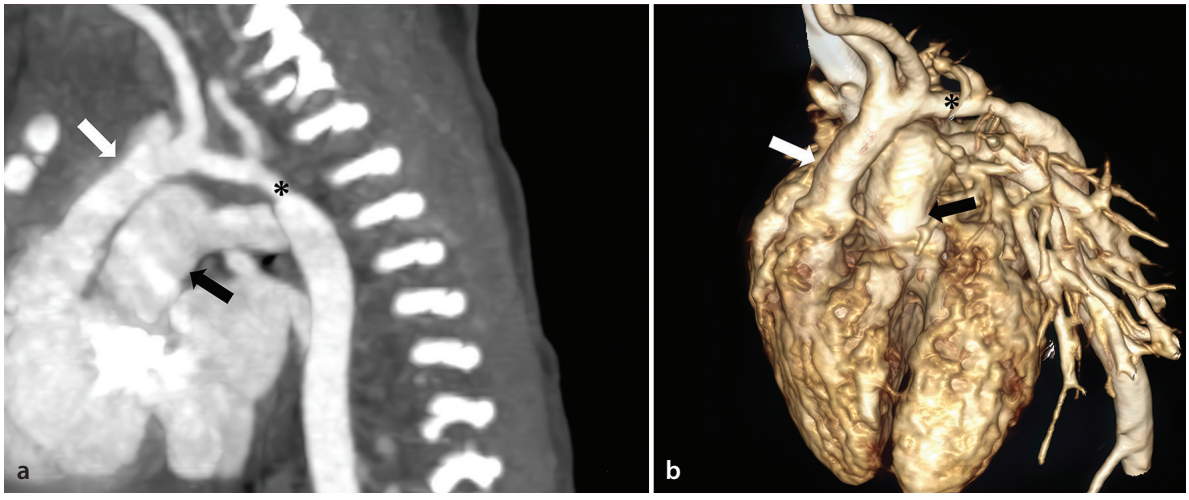



**Figure 14.** Female patient diagnosed with 3-day double outlet right ventricle: sagittal computed tomography angiography (CTA) **(a)** and three-dimensional volume rendered **(b)** thoracic CTA images show the aorta (white arrows) and pulmonary trunk (black arrows) originating from the right ventricle and accompanying hypoplasia (black star) in the aorta.

sociated with distally located VSDs are rare.[76] Ventricular function is important to monitor in patients with DORV, so CMR is particularly important.[77]

### Tricuspid atresia

Tricuspid atresia is a CHD characterized by agenesis or congenital absence of the tricuspid valve.[73] It accounts for 1.4% of patients with CHD, is the third most common type of cyanotic CHD (Figure 15), and is the most common cause of cyanosis and left ventricular hypertrophy.[78] It is often associated with right ventricular hypoplasia and VSDs, and pulmonary blood flow is shunted via the DA. Tricuspid atresia classification is subdivided into subtypes based on PS and the position of the great arteries. In Type 1, the great arteries are normal (the most common type); in Type 2, ToF is present; and in Type 3, complex anomalies such as TA or AVSD accompany ToF or malposition.[72]

### Ebstein anomaly

Ebstein anomaly (EA) is a malformation of the tricuspid valve (Figure 16). The septal and posterior leaflets of the dysplastic tricuspid valve extend toward the RV and are usually associated with tricuspid regurgitation.[79] EA is rare, with a prevalence of 5.2 per 100,000 births. It accounts for 1% of CHD but is consequential because it constitutes approximately 40% of congenital tricuspid valve malformations.[66,80] The severity of the condition depends on the extent to which the RV is filled by the enlarged right atrium, a condition known as RV atrialization.[73] There are four types, classified from A to D, based on the degree of RV atrialization. A patent foramen ovale (PFO) or ostium secundum ASD is always present.[79] As a result, pressure in the right atrium increases, leading to a right-to-left shunt.

### Total anomalous pulmonary venous connection

Total anomalous pulmonary venous connection is characterized by abnormal drainage of the pulmonary veins into the systemic circulation.[81] It is a rare cardiac malformation occurring in approximately 7 per 100,000 live births.[82] Typically, the pulmonary veins drain into the left brachiocephalic vein (most common) (Figure 17), the superior or inferior vena cava, the coronary sinus, the portal vein, or, rarely, into other systemic veins. Less commonly, the pulmonary veins drain directly into the right atrium,[83] resulting in a left-to-right shunt. This condition is classified based on the location where the abnormal venous connection occurs relative to the heart. The supracardiac type accounts for 50%, the cardiac type for 20%, and the infracardiac type for 20%.[81] The mixed type is observed in 10% of cases. Physiologically, it can be characterized as obstructive or non-obstructive based on the condition of the pulmonary venous flow. Obstruction is commonly seen in the infracardiac type.[84]

### Hypoplastic left heart syndrome

Hypoplastic left heart syndrome (HLHS) is a broad spectrum of anomalies characterized by hypoplasia of the left-sided structures of the heart, including the aorta, mitral valve, LV, aortic root, and ascending aorta.[85] The left atrium is typically small but may be normal or enlarged in size (Figure 18). By contrast, the right heart structures are markedly enlarged. The PFO and PDA are critical to the systemic circulation. Despite a prevalence of 1 in 5,000 live births, HLHS accounts for 7% to 9% of all CHDs diagnosed before the age of 1 year.[73] Additionally, it is the fourth most common CHD presenting in the first year of life. It may be associated with syndromes such as Noonan, Holt–Oram, Turner, or Smith–Lemli–Oplitz syndromes.[86] CT is preferred preoperatively to detect anomalies such as CoA and VSD, which may accompany HLHS. CMR is an extremely

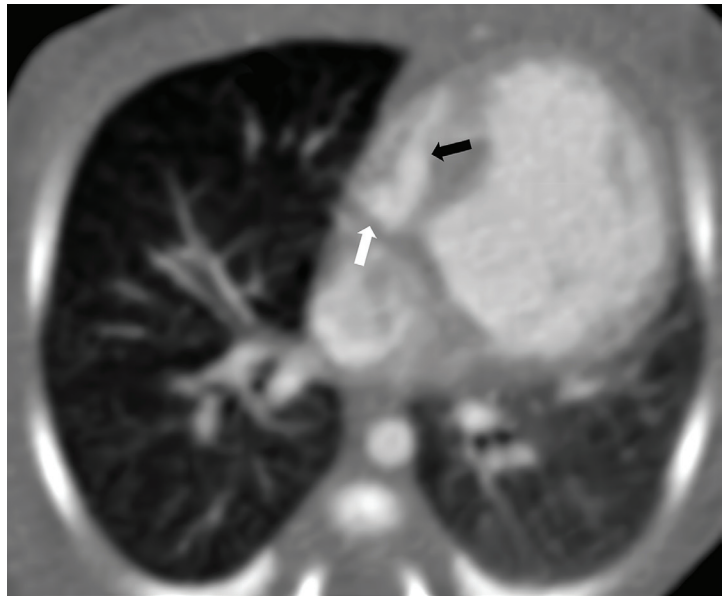


**Figure 15.** Tricuspid valve atresia (white arrow) and accompanying right ventricular hypoplasia (black arrow) in a 5-day-old girl with a diagnosis of tricuspid atresia.

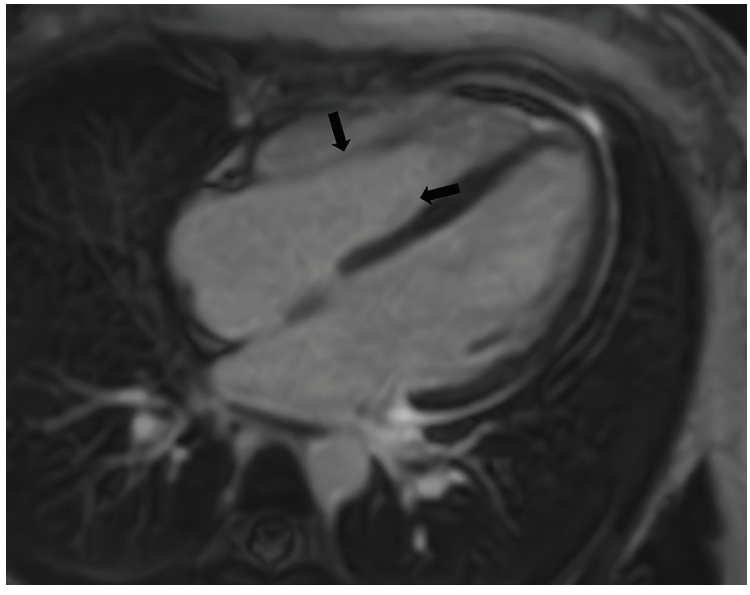


**Figure 16.** In the cardiac magnetic resonance examination performed on a 26-year-old female patient, a distinct extension of the tricuspid valve cusps toward the right ventricle was observed and a diagnosis of Ebstein anomaly was made (black arrows).

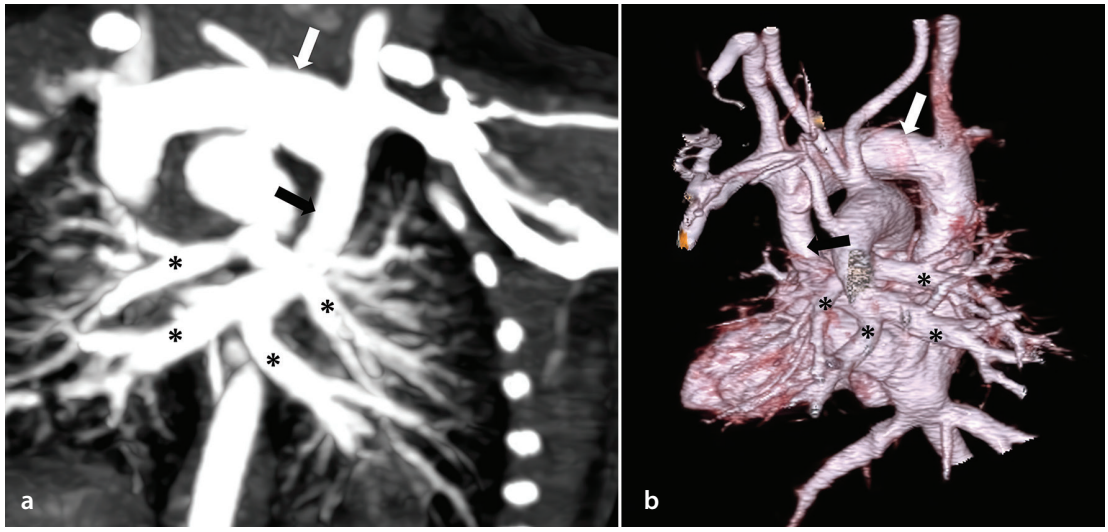


**Figure 17.** Coronal reformatted computed tomography angiography **(a)** and three-dimensional volume rendered image **(b)** of a 3-day-old male patient showing the pulmonary veins (black stars) opening into the left brachiocephalic vein (white arrows) via a vertical venous connection (black arrows).

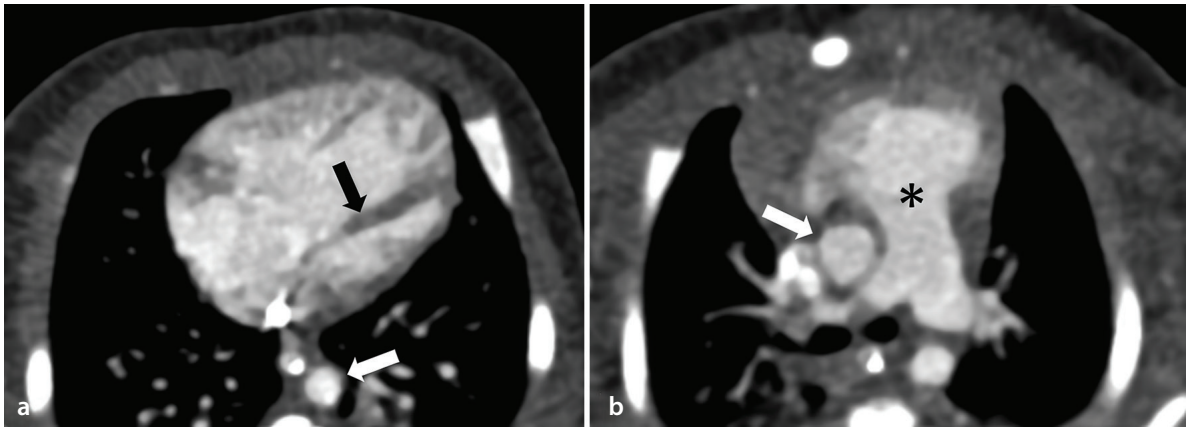


**Figure 18.** Axial computed tomography angiography **(a,b)** images of a 2-day-old female patient showing left ventricular hypoplasia (black arrow), aortic hypoplasia (white arrows), and accompanying pulmonary artery enlargement (black star).

important imaging modality for detecting pathologies such as endocardial fibroelastosis that may accompany HLHS, for quantitative measurement of ventricular volumes pre- and postoperatively, for functional assessment of the aortic and mitral valves, and, particularly, for detecting fibrosis, which may develop in the ventricles during postoperative follow-up.

### Pulmonary atresia with intact ventricular septum

PA with intact ventricular septum (PA–IVS) is characterized by the absence of communication between the RV and LV and the presence of PA (Figure 19).[87] Moreover, the RV is hypoplastic to varying degrees. This condition is rare, occurring in 5 per 100,000 live births and accounting for 1% to 3% of CHDs.[88] Another important feature of PA–IVS is that it may be associated with other cardiac anomalies; PFO, coronary arteriovenous fistulas (10%–50%), and tricuspid insufficiency secondary to RV dilatation, as well as MAPCAs, which provide pulmonary vascularization, may be present.[88] CT is an extremely successful technique for identifying any additional pathologies, fistulas, or collateral vascular structures that may be present.[89] CMR is vital in the postoperative period for evaluating pulmonary valve and RV function and monitoring the presence of fibrosis.[90]

### Single ventricle

The primary definition of single ventricle is the presence of a single ventricle into which both atria drain and the systemic and pulmonary circulation provided by this ventricle; it is a CHD (Figure 20).[91] Depending on the ventricular morphology, it is referred to as double-inlet LV, double-inlet RV, or, when differentiation is not possible, primitive ventricle.[92] The incidence of the LV type is 5–10 per 100,000 live births, making it the most common functional single ventricle defect.[93] Associated conditions include TGA (70%, typically L-type), subvalvular PS or atresia (66%), subaortic obstruction, and CoA.[72] CT and MRI are superior to echocardiography in evaluating extracardiac vascular structures, identifying associated anomalies, and assessing complex anatomical relationships. CMR has a distinct role in evaluating postoperative valve and ventricular function.[94,95]

In conclusion, CHDs represent a wide spectrum, and accurate diagnosis is vital. This is because treatment plans are specific to the patient's diagnosis. Advances in cardiac imaging in recent years have been instrumental in reducing morbidity rates and increasing life expectancy in individuals with CHD. In the evaluation of imaging methods, radiography is non-specific but can assist in determining heart size and pulmonary vas-

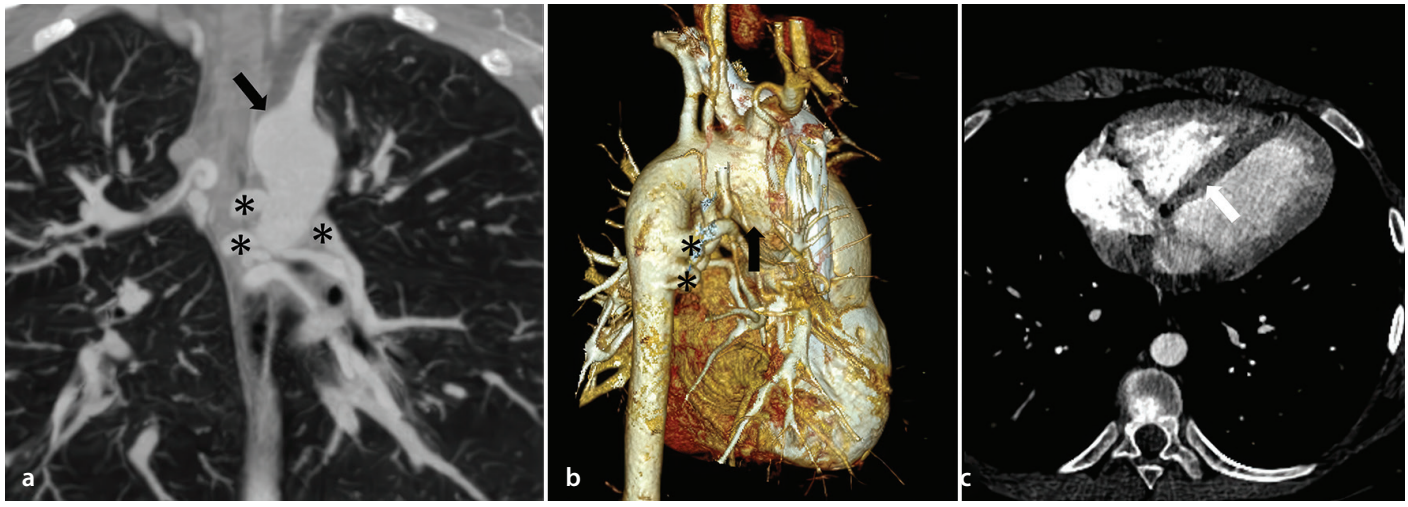



**Figure 19.** A 19-year-old male patient with isolated pulmonary atresia: the aorta (black arrows) is observed, but the pulmonary artery is not observed in coronal reformatted computed tomography angiography (CTA) **(a)** and three-dimensional volume rendered **(b)** images. Multiple accompanying major aortopulmonary collateral arteries are observed (black stars). In the axial CTA image **(c)**, the interventricular septum is seen intact (white arrow).

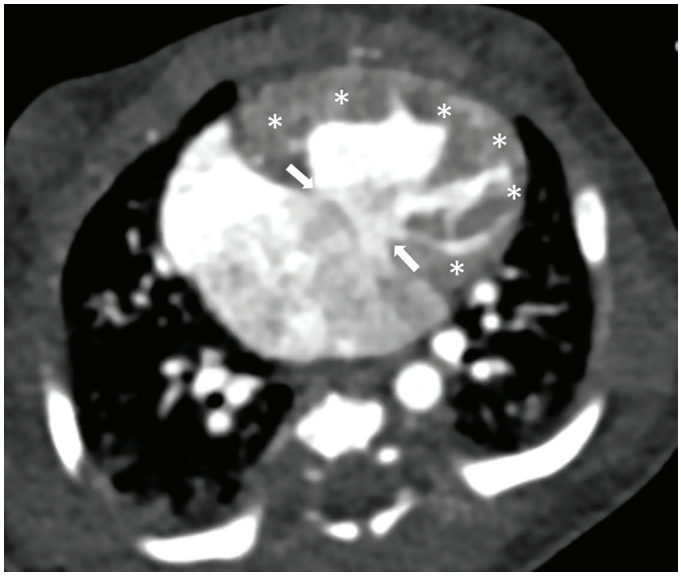


**Figure 20.** Axial computed tomography angiography image of a 6-day-old girl showing atrioventricular septal defect, single atrioventricular valve (white arrows), and single ventricle morphology (white stars).

cularity, whereas echocardiography is user dependent and has a limited role in evaluating extracardiac structures. Advances in imaging have led to an increasing role for CT and MRI in addition to cardiac evaluation, including the visualization of extracardiac vascular structures and postoperative follow-up. This has substantially reduced the role of invasive cardiac catheterization in the diagnosis of CHD. Ongoing technological innovations have led to non-invasive modalities with improved image quality, lower ionizing radiation and contrast doses, and reduced imaging time, thereby decreasing the need for anesthesia. Diverse modalities for detecting cyanotic CHD, along with ongoing advances in surgical procedures, will result in more patients with CHD surviving into adulthood in the future. Therefore, the role of CT and MRI will continue to expand. The appropriate selection of imaging methods is guided by the clinical context and the relative strengths of various imaging modalities. CT is preferred for the initial assessment of large vessel anatomy in newborns, especially when functional information is not important. CMR is critical for the preoperative and postoperative assessment of ventricular and valve function, measurement of flow parameters in existing shunts, and detection of myocardial fibrosis. With this approach, the proper use of imaging modalities provides the detailed anatomy and post-surgical follow-up of the patient that are important for surgical planning.

## Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Stout KK, Daniels CJ, Aboulhosn JA, et al. 2018 AHA/ACC Guideline for the Management of Adults with Congenital Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;73(12):e81-e192. [Crossref]
2. Koplay M. Congenital cardiovascular anomalies: CT and MRI imaging. 1st ed. Ankara: Dünya Tıp Kitabevi; 2020.
3. Chung JH, Gunn ML, Godwin JD, Takasugi J, Kanne JP. Congenital thoracic cardiovascular anomalies presenting in adulthood: a pictorial review. *J Cardiovasc Comput Tomogr*. 2009;3(1 Suppl):S35-S46. [Crossref]
4. Tsai IC, Chen MC, Jan SL, et al. Neonatal cardiac multidetector row CT: why and how we do it. *Pediatr Radiol*. 2008;38(4):438-451. [Crossref]
5. Panayiotou HR, Mills LK, Broadbent DA, et al. Comprehensive neonatal cardiac, feed and wrap, non-contrast, non-sedated, free-breathing compressed sensing 4D flow MRI assessment. *J Magn Reson Imaging*. 2023;57(3):789-799. [Crossref]
6. Lapierre C, Garel L, El-Jalbout R, Damphousse A, Déry J. Cardiac CT and MRI of cardiac malformations: how to interpret them? *Diagn Interv Imaging*. 2016;97(5):519-530. [Crossref]
7. Frommelt P, Lopez L, Dimas VV, et al. Recommendations for Multimodality Assessment of Congenital Coronary Anomalies: A Guide from the American Society of Echocardiography: Developed in Collaboration with the Society for Cardiovascular Angiography and Interventions, Japanese Society of

Echocardiography, and Society for Cardiovascular Magnetic Resonance. *J Am Soc Echocardiogr*. 2020;33(3):259-294. [Crossref]

8. Tricarico F, Hlavacek AM, Schoepf UJ, et al. Cardiovascular CT angiography in neonates and children: image quality and potential for radiation dose reduction with iterative image reconstruction techniques. *Eur Radiol*. 2013;23(5):1306-1315. [Crossref]
9. Siripornpitak S, Pornkul R, Khowsathit P, Layangool T, Promphan W, Pongpanich B. Cardiac CT angiography in children with congenital heart disease. *Eur J Radiol*. 2013;82(7):1067-1082. [Crossref]
10. Bailliard F, Hughes ML, Taylor AM. Introduction to cardiac imaging in infants and children: techniques, potential, and role in the imaging work-up of various cardiac malformations and other pediatric heart conditions. *Eur J Radiol*. 2008;68(2):191-198. [Crossref]
11. Hellinger JC, Pena A, Poon M, Chan FP, Epelman M. Pediatric computed tomographic angiography: imaging the cardiovascular system gently. *Radiol Clin North Am*. 2010;48(2):439-467. [Crossref]
12. Schicchi N, Fogante M, Esposto Pirani P, et al. Third-generation dual-source dual-energy CT in pediatric congenital heart disease patients: state-of-the-art. *Radiol Med*. 2019;124(12):1238-1252. [Crossref]
13. Krishnamurthy R. Neonatal cardiac imaging. *Pediatr Radiol*. 2010;40(4):518-527. [Crossref]
14. Kravchenko D, Hart C, Garbe S, et al. Image quality and radiation dose of dual source high pitch computed tomography in pediatric congenital heart disease. *Sci Rep*. 2022;12(1):9934. [Crossref]
15. Sun Z. Multislice CT angiography in cardiac imaging: prospective ECG-gating or retrospective ECG-gating? *Biomed Imaging Interv J*. 2010;6(1):e4. [Crossref]
16. Dirrichs T, Tietz E, Rüffer A, et al. Photon-counting versus dual-source CT of congenital heart defects in neonates and infants: initial experience. *Radiology*. 2023;307(5):e223088. [Crossref]
17. Dillman JR, Hernandez RJ. Role of CT in the evaluation of congenital cardiovascular disease in children. *AJR Am J Roentgenol*. 2009;192(5):1219-1231. [Crossref]
18. Stolzmann P, Leschka S, Scheffel H, et al. Dual-source CT in step-and-shoot mode: non-invasive coronary angiography with low radiation dose. *Radiology*. 2008;249(1):71-80. [Crossref]
19. Paul JF, Abada HT. Strategies for reduction of radiation dose in cardiac multislice CT. *Eur Radiol*. 2007;17(8):2028-2037. [Crossref]
20. Han BK, Rigsby CK, Hlavacek A, et al. Computed tomography imaging in patients with congenital heart disease part I: rationale and utility. an expert consensus document of the society of cardiovascular computed tomography (SCCT): endorsed by the Society of Pediatric Radiology (SPR) and the North American Society of Cardiac Imaging (NASCI). *J Cardiovasc Comput Tomogr*. 2015;9(6):475-492. [Crossref]
21. Zamorano JL, Bax J, Lancellotti P, Knuuti J, Badano L. The ESC Textbook of Cardiovascular Imaging: Oxford University Press, USA; 2015. [Crossref]
22. Kilner PJ, Geva T, Kaemmerer H, Trindade PT, Schwitter J, Webb GD. Recommendations for cardiovascular magnetic resonance in adults with congenital heart disease from the Respective Working Groups of the European Society of Cardiology. *Eur Heart J*. 2010;31(7):794-805. [Crossref]
23. Biglands JD, Radjenovic A, Ridgway JP. Cardiovascular magnetic resonance physics for clinicians: part II. *J Cardiovasc Magn Reson*. 2012;14(1):66. [Crossref]
24. Ridgway JP. Cardiovascular magnetic resonance physics for clinicians: part I. *J Cardiovasc Magn Reson*. 2010;12(1):71. [Crossref]
25. Vahanian A, Alfieri O, Andreotti F, et al. Guidelines on the management of valvular heart disease (version 2012): the Joint Task Force on the Management of Valvular Heart Disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *Eur J Cardiothorac Surg*. 2012;42(4):1-44. [Crossref]
26. Di Salvo G, Miller O, Babu Narayan S, et al. Imaging the adult with congenital heart disease: a multimodality imaging approach-position paper from the EACVI. *Eur Heart J Cardiovasc Imaging*. 2018;19(10):1077-1098. [Crossref]
27. Lee GH, Koo HJ, Park KJ, Yang DH, Ha H. Characterization of baseline hemodynamics after the Fontan procedure: a retrospective cohort study on the comparison of 4D flow MRI and computational fluid dynamics. *Front Physiol*. 2023;14:1199771. [Crossref]
28. Pednekar AS, Wang H, Flamm S, Cheong BY, Muthupillai R. Two-center clinical validation and quantitative assessment of respiratory triggered retrospectively cardiac gated balanced-ssfp cine cardiovascular magnetic resonance imaging in adults. *J Cardiovasc Magn Reson*. 2018;20(1):44. [Crossref]
29. Taylor AM, Dymarkowski S, Hamaekers P, et al. MR coronary angiography and late-enhancement myocardial MR in children who underwent arterial switch surgery for transposition of great arteries. *Radiology*. 2005;234(2):542-547. [Crossref]
30. Bonnichsen C, Ammash N. Choosing between MRI and CT imaging in the adult with congenital heart disease. *Curr Cardiol Rep*. 2016;18(5):45. [Crossref]
31. Dorfman AL, Geva T, Samyn MM, et al. SCMR expert consensus statement for cardiovascular magnetic resonance of acquired and non-structural pediatric heart disease. *J Cardiovasc Magn Reson*. 2022;24(1):44. [Crossref]
32. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update : Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. *J Cardiovasc Magn Reson*. 2020;22(1):19. [Crossref]
33. Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol*. 2002;39(12):1890-1900. [Crossref]
34. Puranik R, Muthurangu V, Celermajer DS, Taylor AM. Congenital heart disease and multi-modality imaging. *Heart Lung Circ*. 2010;19(3):133-144. [Crossref]
35. Wald RM, Powell AJ. Simple congenital heart lesions. *J Cardiovasc Magn Reson*. 2006;8(4):619-631. [Crossref]
36. Khanna A, Litt HI. ECG-gated MDCT of left-to-right shunt flow through a ventricular septal defect. *AJR Am J Roentgenol*. 2004;183(6):1838. [Crossref]
37. Ropers D, Regenfus M, Wasmeier G, et al. [Ventricular septum defect: imaging using computer tomography, magnetic resonance tomography and color doppler echocardiography]. *Dtsch Med Wochenschr*. 2007;132(14):769-770. [Crossref]
38. Backer CL, Stewart RD, Mavroudis C. Overview: history, anatomy, timing, and results of complete atrioventricular canal. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu*. 2007:3-10. [Crossref]
39. Allen HD, Driscoll DJ, Shaddy RE, Feltes TF, editors. Moss & Adams' heart disease in infants, children, and adolescents: including the fetus and young adult. 8th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. [Crossref]
40. Park MK. Pediatric cardiology for practitioners. 5th ed. Part 2: Special tools in evaluation of cardiac patients. Chapter 6: Noninvasive techniques. Philadelphia: Mosby Elsevier; 2008. [Crossref]
41. Saremi F, Hassani C, Sánchez-Quintana D. Septal atrioventricular junction region: comprehensive imaging in adults. *Radiographics*. 2016;36(7):1966-1986. [Crossref]
42. Prsa M, Sun L, van Amerom J, et al. Reference ranges of blood flow in the major vessels of the normal human fetal circulation at term by phase-contrast magnetic resonance imaging. *Circ Cardiovasc Imaging*. 2014;7(4):663-670. [Crossref]
43. Körperich H, Gieseke J, Barth P, et al. Flow volume and shunt quantification in pediatric congenital heart disease by real-time magnetic

resonance velocity mapping: a validation study. *Circulation*. 2004;109(16):1987-1993. [Crossref]

44. Moore P, Brook M. Patent Ductus Arteriosus and Aorticopulmonary Window. In: Allen Hd, Driscoll Dj, Shaddy Re, Feltes Tf, eds. Moss and Adams' Heart Disease. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2013. [Crossref]
45. Takahashi Y, Roman C, Chemtob S, et al. Cyclooxygenase-2 inhibitors constrict the fetal lamb ductus arteriosus both *in vitro* and *in vivo*. *Am J Physiol Regul Integr Comp Physiol*. 2000;278(6):R1496-R1505. [Crossref]
46. Burkett DA. Common left-to-right shunts. *Pediatr Clin North Am*. 2020;67(5):821-842. [Crossref]
47. Krichenko A, Benson LN, Burrows P, Möes CA, McLaughlin P, Freedom RM. Angiographic classification of the isolated, persistently patent ductus arteriosus and implications for percutaneous catheter occlusion. *Am J Cardiol*. 1989;63(12):877-880. [Crossref]
48. Schena F, Francescato G, Cappelleri A, et al. Association between hemodynamically significant patent ductus arteriosus and bronchopulmonary dysplasia. *J Pediatr*. 2015;166(6):1488-1492. [Crossref]
49. Kato R, Lickfett L, Meininger G, et al. Pulmonary vein anatomy in patients undergoing catheter ablation of atrial fibrillation: lessons learned by use of magnetic resonance imaging. *Circulation*. 2003;107(15):2004-2010. [Crossref]
50. Alsoufi B, Cai S, Van Arsdell GS, Williams WG, Caldarone CA, Coles JG. Outcomes after surgical treatment of children with partial anomalous pulmonary venous connection. *Ann Thorac Surg*. 2007;84(6):2020-2026. [Crossref]
51. Furlanetto G, Furlanetto BH, Henriques SR, et al. Mixed type total anomalous pulmonary venous connection: early results and surgical techniques. *World J Pediatr Congenit Heart Surg*. 2015;6(1):26-32. [Crossref]
52. Liu CW, Hwang B, Lee BC, Lu JH, Meng LC. Aortic stenosis in children: 19-year experience. *Zhonghua Yi Xue Za Zhi (Taipei)*. 1997;59(2):107-113. [Crossref]
53. Lancellotti P, Zamorano JL, Habib G, Badano L. The EACVI Textbook of Echocardiography: Oxford University Press; 2016. [Crossref]
54. Marchini F, Meossi S, Passarini G, Campo G, Pavasini R. Pulmonary valve stenosis: from diagnosis to current management techniques and future prospects. *Vasc Health Risk Manag*. 2023;19:379-390. [Crossref]
55. Singh S, Hakim FA, Sharma A, et al. Hypoplasia, pseudocoarctation and coarctation of the aorta - a systematic review. *Heart Lung Circ*. 2015;24(2):110-118. [Crossref]
56. Karaosmanoglu AD, Khawaja RD, Onur MR, Kalra MK. CT and MRI of aortic coarctation: pre- and postsurgical findings. *AJR Am J Roentgenol*. 2015;204(3):W224-W233. [Crossref]
57. Nie P, Wang X, Cheng Z, et al. The value of low-dose prospective ECG-gated dual-source CT angiography in the diagnosis of coarctation of the aorta in infants and children. *Clin Radiol*. 2012;67(8):738-745. [Crossref]
58. Hughes D Jr, Siegel MJ. Computed tomography of adult congenital heart disease. *Radiol Clin North Am*. 2010;48(4):817-835. [Crossref]
59. Kim HK, Gottliebson W, Hor K, et al. Cardiovascular anomalies in Turner syndrome: spectrum, prevalence, and cardiac MRI findings in a pediatric and young adult population. *AJR Am J Roentgenol*. 2011;196(2):454-460. [Crossref]
60. Hanneman K, Newman B, Chan F. Congenital variants and anomalies of the aortic arch. *Radiographics*. 2017;37(1):32-51. [Crossref]
61. Gruber PJ, Epstein JA. Development gone awry: congenital heart disease. *Circ Res*. 2004;94(3):273-283. [Crossref]
62. Mulay AV, Watterson KG. Isolated Right Subclavian Artery, Interrupted Aortic Arch, and Ventricular Septal Defect. *Ann Thorac Surg*. 1997;63(4):1163-1165. [Crossref]
63. Apitz C, Webb GD, Redington AN. Tetralogy of fallot. *Lancet*. 2009;374(9699):1462-1471. [Crossref]
64. Syed MA, Mohiaddin RH. Magnetic resonance imaging of congenital heart disease: Springer Nature; 2023. [Crossref]
65. Moscatelli S, Pergola V, Motta R, et al. Multimodality imaging assessment of tetralogy of fallot: from diagnosis to long-term follow-up. *Children (Basel)*. 2023;10(11):1747. [Crossref]
66. Warmerdam EG, Neijzen RL, Voskuil M, Leiner T, Grotenhuis HB. Four-dimensional flow CMR in tetralogy of fallot: current perspectives. *Br J Radiol*. 2022;95(1133):20210298. [Crossref]
67. Cohen MS, Eidem BW, Cetta F, et al. Multimodality imaging guidelines of patients with transposition of the great arteries: a report from the American Society of Echocardiography Developed in Collaboration with the Society for Cardiovascular Magnetic Resonance and the Society of Cardiovascular Computed Tomography. *J Am Soc Echocardiogr*. 2016;29(7):571-621. [Crossref]
68. Canan A, Ashwath R, Agarwal PP, François C, Rajiah P. Multimodality imaging of transposition of the great arteries. *Radiographics*. 2021;41(2):338-360. [Crossref]
69. Chan FP, Hanneman K. Computed tomography and magnetic resonance imaging in neonates with congenital cardiovascular disease. *Semin Ultrasound CT MR*. 2015;36(2):146-160. [Crossref]
70. Syamasundar Rao P. Diagnosis and management of cyanotic congenital heart disease: part II. *Indian J Pediatr*. 2009;76(3):297-308. [Crossref]
71. Goo HW. Cardiac MDCT in children: CT technology overview and interpretation. *Radiol Clin North Am*. 2011;49(5):997-1010. [Crossref]
72. Rao PS. Diagnosis and management of cyanotic congenital heart disease: part I. *Indian J Pediatr*. 2009;76(1):57-70. [Crossref]
73. Rao PS. Consensus on timing of intervention for common congenital heart diseases: part II - cyanotic heart defects. *Indian J Pediatr*. 2013;80(8):663-674. [Crossref]
74. Koplay M, Cimen D, Sivri M, et al. Truncus arteriosus: diagnosis with dual-source computed tomography angiography and low radiation dose. *World J Radiol*. 2014;6(11):886-889. [Crossref]
75. Singh AAV, Yoo SJ, Seed M, Lam CZ, Valverde I. Recent advances in multimodal imaging in tetralogy of fallot and double outlet right ventricle. *Curr Opin Cardiol*. 2024;39(4):323-330. [Crossref]
76. Frank L, Dillman JR, Parish V, et al. Cardiovascular MR imaging of conotruncal anomalies. *Radiographics*. 2010;30(4):1069-1094. [Crossref]
77. Dewi DK, Perdhana DP. Double outlet right ventricle presenting in an adult woman: a case report. *Radiol Case Rep*. 2022;17(5):1413-1415. [Crossref]
78. Zucker EJ, Koning JL, Lee EY. Cyanotic congenital heart disease: essential primer for the practicing radiologist. *Radiol Clin North Am*. 2017;55(4):693-716. [Crossref]
79. Malik SB, Kwan D, Shah AB, Hsu JY. The right atrium: gateway to the heart--anatomic and pathologic imaging findings. *Radiographics*. 2015;35(1):14-31. [Crossref]
80. Negoi RI, Ispas AT, Ghiorghiu I, et al. Complex Ebstein's malformation: defining preoperative cardiac anatomy and function. *J Card Surg*. 2013;28(1):70-81. [Crossref]
81. Files MD, Morray B. Total anomalous pulmonary venous connection: preoperative anatomy, physiology, imaging, and ınterventional management of postoperative pulmonary venous obstruction. *Semin Cardiothorac Vasc Anesth*. 2017;21(2):123-131. [Crossref]
82. Seale AN, Uemura H, Webber SA, et al. Total anomalous pulmonary venous connection: morphology and outcome from an international population-based study. *Circulation*. 2010;122(25):2718-2726. [Crossref]
83. Gaca AM, Jaggers JJ, Dudley LT, Bisset GS, 3rd. Repair of congenital heart disease: a primer--part 2. *Radiology*. 2008;248(1):44-60. [Crossref]
84. Karamlou T, Gurofsky R, Al Sukhni E, et al. Factors associated with mortality and reoperation in 377 children with total anomalous pulmonary venous connection. *Circulation*. 2007;115(12):1591-1598. [Crossref]
85. Pedra SRFF. Imaging for hypoplastic left heart syndrome. *World J Pediatr Congenit Heart Surg*. 2022;13(5):571-575. [Crossref]

86. Connor JA, Thiagarajan R. Hypoplastic left heart syndrome. *Orphanet J Rare Dis*. 2007;2:23. [Crossref]

87. Chikkabyrappa SM, Loomba RS, Tretter JT. Pulmonary Atresia with an intact ventricular septum: preoperative physiology, imaging, and management. *Semin Cardiothorac Vasc Anesth*. 2018;22(3):245-255. [Crossref]

88. Uribe S, Bächler P, Valverde I, et al. Hemodynamic assessment in patients with one-and-a-half ventricle repair revealed by four-dimensional flow magnetic resonance imaging. *Pediatr Cardiol*. 2013;34(2):447-451. [Crossref]

89. Hascoet S, Combelles S, Acar P. Cardiac computed tomography of multiple coronary arteries to right ventricle fistulas in a newborn with pulmonary atresia and intact ventricular septum. *Can J Cardiol*. 2014;30(2):247. [Crossref]

90. Liang XC, Lam WW, Cheung EW, Wu AK, Wong SJ, Cheung YF. Restrictive right ventricular physiology and right ventricular fibrosis as assessed by cardiac magnetic resonance and exercise capacity after biventricular repair of pulmonary atresia and intact ventricular septum. *Clin Cardiol*. 2010;33(2):104-110. [Crossref]

91. Wilkinson JL, Anderson RH. Anatomy of functionally single ventricle. *World J Pediatr Congenit Heart Surg*. 2012;3(2):159-164. [Crossref]

92. Festa P, Ait Ali L, Bernabei M, De Marchi D. The role of magnetic resonance imaging in the evaluation of the functionally single ventricle before and after conversion to the Fontan circulation. *Cardiol Young*. 2005;15(Suppl 3):51-56. [Crossref]

93. Güvenç O, Saygi M, Şengül FS, Ayyıldız P, Güzeltaş A. Double inlet left ventricle-ventriculoarterial discordance without surgical treatment. *Pediatr Int*. 2016;58(6):509-511. [Crossref]

94. Ito D, Shiraishi J, Noritake K, Kohno Y. Multidetector computed tomography demonstrates double-inlet, double-outlet right ventricle. *Intern Med*. 2011;50(18):2053-2054. [Crossref]

95. Rathod RH, Prakash A, Kim YY, et al. Cardiac magnetic resonance parameters predict transplantation-free survival in patients with fontan circulation. *Circ Cardiovasc Imaging*. 2014;7(3):502-509. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263576

CARDIOVASCULAR IMAGING

PICTORIAL ESSAY



# The imaging spectrum of myocardial infarction and its associated complications: a contemporary pictorial review of computed tomography and magnetic resonance imaging

Yu Zhang[1]
Sung Min Ko[2]

[1]Yuhuangding Hospital, Clinic of Radiology, Yantai, China

[2]Wonju Severance Christian Hospital, Yonsei University Wonju Faculty of Medicine, Department of Radiology, Wonju, South Korea

**ABSTRACT**

Myocardial infarction (MI) is a leading cause of morbidity and mortality worldwide. Although rapid diagnosis and reperfusion in the acute setting rely primarily on clinical assessment, electrocardiography, echocardiography, and invasive coronary angiography, advanced cardiac imaging plays an essential role beyond the hyperacute phase. Cardiac computed tomography (CCT) and cardiac magnetic resonance (CMR) imaging provide complementary information that extends from coronary anatomy to myocardial tissue characterization and functional assessment. CMR imaging enables the comprehensive, multiparametric evaluation of MI, including left ventricular function, myocardial edema and area at risk, infarct size and transmural extent, microvascular obstruction, intramyocardial hemorrhage, myocardial viability, and ischemia using cine imaging, T1/T2 and T2* mapping, perfusion imaging, and late gadolinium enhancement. These features support an accurate differentiation between acute and chronic infarction, an assessment of myocardial salvage, and prognostic stratification. CCT offers a rapid, non-invasive assessment of coronary artery stenosis and plaque characteristics and has expanded to include an evaluation of ventricular function, myocardial perfusion, and delayed-enhancement patterns. When combined with CT-derived fractional flow reserve or myocardial perfusion imaging, CT allows for an integrated anatomic and functional assessment of the myocardium, particularly for non-culprit lesions, following MI. Myocardial delayed-enhancement CT can visualize the infarcted myocardium and microvascular injury in select patients, though it remains complementary to magnetic resonance imaging. This pictorial essay illustrates the imaging spectrum of MI and its major mechanical and thromboembolic complications, including ventricular rupture, septal defects, papillary muscle rupture, aneurysm formation, left ventricular thrombi, and pericardial disease. By highlighting the strengths and limitations of CCT and CMR and providing practical guidance for modality selection, this article aims to support informed clinical decision-making in the contemporary management of patients with MI.

**KEYWORDS**

Myocardial infarction, magnetic resonance imaging, computed tomography, delayed enhancement, viability, mapping, microvascular obstruction

**Handling editor:** Furkan Ufuk

**Corresponding author:** Sung Min Ko

**E-mail:** ksm9723@yahoo.co.kr

Received 10 November 2025; revision requested 08 December 2025; accepted 01 February 2026.

Epub: 26.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263576

Myocardial infarction (MI) is a leading cause of morbidity and mortality worldwide.[1] Although rapid reperfusion strategies rely primarily on clinical assessment, electrocardiography (ECG), echocardiography, and invasive coronary angiography, advanced cardiac imaging plays a pivotal role beyond the hyperacute phase. In particular, cardiac magnetic resonance (CMR) imaging and cardiac computed tomography (CCT) provide complementary anatomical, functional, and tissue-level information that informs prognosis, risk stratification, and management after MI. CMR imaging is the reference standard for myocardial tissue characterization, allowing for a multiparametric evaluation of ventricular function, infarct size, myocardial edema, microvascular injury, intramyocardial hemorrhage, and myocardial viability. CCT, which has traditionally focused on coronary anatomy, has evolved to enable functional assessment using CT-derived fractional flow reserve CT myocardial perfusion imaging (CT-MPI), and myocardial delayed-enhancement CT (MDE-CT).[2,3] Recent ad-

You may cite this article as: Zhang Y, Ko SM. The imaging spectrum of myocardial infarction and its associated complications: a contemporary pictorial review of computed tomography and magnetic resonance imaging. *Diagn Interv Radiol*. 2026;32(4):426-436.

vances incorporating artificial intelligence (AI)-based noise suppression and automated infarct segmentation have further expanded the capabilities of CT while reducing radiation exposure.[4] This pictorial essay reviews the imaging spectrum of MI and its complications, emphasizing quantitative biomarkers, modality-specific strengths and limitations, and practical guidance for selecting CCT, CMR imaging, or echocardiography in common clinical scenarios.

**Main points**

- Myocardial infarction (MI) remains a leading cause of morbidity and mortality worldwide.
- Cardiac magnetic resonance (CMR) imaging enables a detailed evaluation of ventricular function, myocardial viability, area at risk, microvascular obstruction, and intramyocardial hemorrhage through multiparametric imaging.
- Cardiac computed tomography (CCT) allows a rapid, non-invasive assessment of coronary stenosis, myocardial damage, and post-MI complications, aiding prompt clinical decisions.
- Together, CCT and CMR imaging play complementary roles in comprehensive MI evaluation, guiding risk stratification and personalized management.

## Overview of myocardial infarction

MI is commonly classified according to the extent of myocardial involvement as either transmural or subendocardial infarction and by its electrocardiographic presentation as either ST-segment elevation MI (STEMI) or non-STEMI.[5] In the setting of prolonged but incomplete ischemia, the myocardium may exhibit reversible contractile dysfunction despite the preservation of myocardial viability, retaining the potential for functional recovery after reperfusion. Accordingly, timely revascularization is critical to limiting irreversible myocardial injury, preserving viable myocardial tissue, and improving clinical outcomes.[6]

## Cardiac magnetic resonance imaging

### Multiparametric cardiac magnetic resonance imaging protocols

CMR imaging enables a comprehensive evaluation of MI through a multiparametric approach. Standard protocols integrate cine imaging to assess global and regional left ventricular (LV) function; T2-weighted imaging (T2WI) and T2 mapping for myocardial edema and area at risk (AAR); native T1 mapping and extracellular volume (ECV) quantification for diffuse myocardial injury; first-pass perfusion imaging for microvascular integrity; late gadolinium enhancement (LGE) for infarct size, transmural extent, and microvascular obstruction (MVO); and T2* mapping for intramyocardial hemorrhage (IMH). This integrated examination allows for the simultaneous assessment of myocardial injury and salvage and prognostic markers in a single session (Figure 1).[7]

### Quantitative cardiac magnetic resonance imaging biomarkers and prognostic implications

Quantitative CMR imaging biomarkers provide clinically relevant prognostic information beyond a visual assessment. Elevated native T1 values (e.g., > 1,250 ms at 1.5 T) and an increased ECV (> 30%) have been associated with adverse LV remodeling. MVO, visualized as a non-enhancing region within the infarcted myocardium, is a strong predictor of poor outcome, with an MVO burden exceeding approximately 1.4% of the LV mass linked to an increased risk of heart failure and major adverse cardiovascular events (MACE) (Figure 2). Moreover, T2* values below 20 ms indicate an IMH and are associated with unfavorable remodeling (Figure 3).[8-10] Texture analysis and radiomics applied to LGE and mapping images have shown promise for predicting LV remodeling and MACE, though these techniques remain investigational.[11]

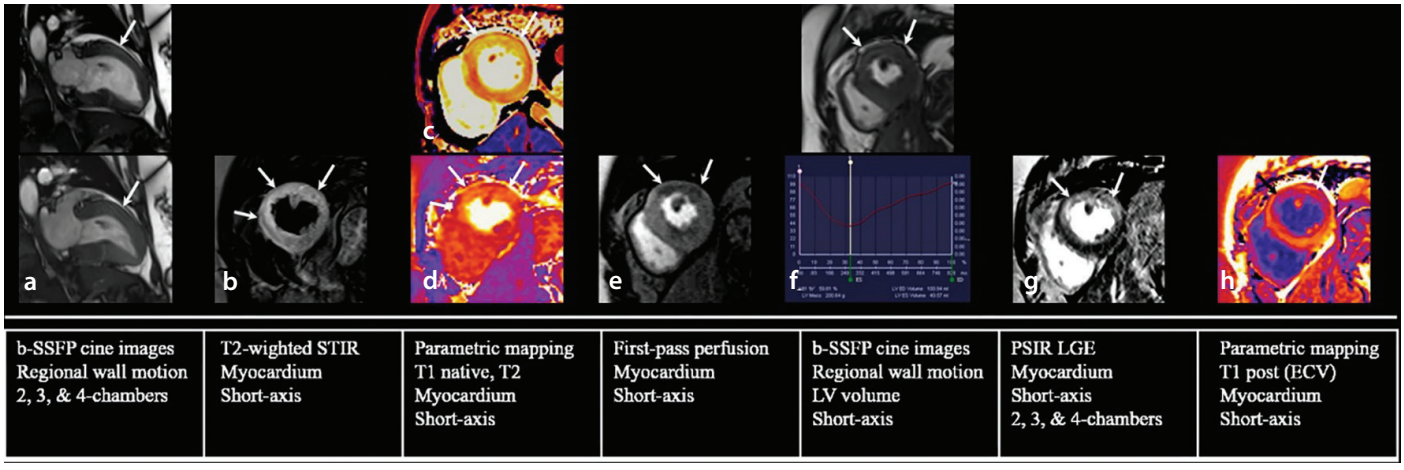



**Figure 1.** Schematic of the cardiac magnetic resonance protocol workflow in a 65-year-old man with reperfused STEMI. (**a**): Two-chamber b-SSFP cine images at end diastole (upper) and end systole (lower) demonstrate hypokinesia of the mid-anterior left ventricular (LV) segment (arrow). (**b**): T2-weighted STIR image shows transmural high signal intensity consistent with infarct-related myocardial edema in mid-anterior, anteroseptal, and anterolateral LV segments (arrows). (**c, d**): Native T1 (c, upper) and T2 (d, lower) mapping images demonstrate elevated T1 (1,590–1,620 ms) and T2 (62–71 ms) values in the same LV segments (arrows). (**e**): First-pass perfusion image using TurboFLASH shows transmural perfusion defects in mid-anterior and anterolateral LV segments (arrows). (**f**): Quantitative assessment of global LV function on short-axis b-SSFP cine images shows normal LV end-diastolic and end-systolic volumes (101 and 40.6 mL) and a preserved LV ejection fraction of 59.8%. Regional hypokinesia is noted in the mid-anterior and anterolateral segments on the systolic short-axis image (arrows). (**g**): Late gadolinium enhancement image with PSIR demonstrates transmural delayed enhancement in LAD and LCX territories (arrows). (**h**): Post-contrast T1 mapping image shows reduced T1 values (325–360 ms) in the same LV segments as on native T1 mapping (arrows). STEMI, ST-segment elevation myocardial infarction; b-SSFP, balanced steady state-free precession; ECV, extracellular volume; STIR, short tau inversion recovery; FLASH, fast low-angle shot; PSIR, phase-sensitive inversion recovery; LAD, left anterior descending coronary artery; LCX, left circumflex coronary artery.

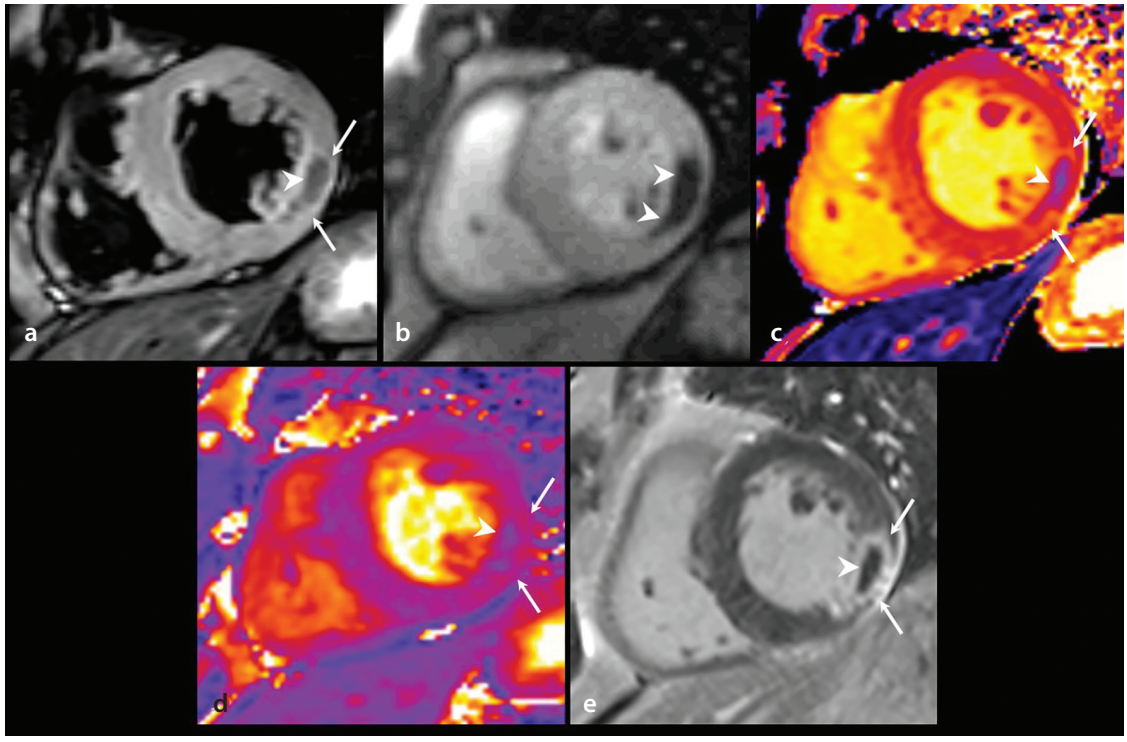



**Figure 2.** Evaluation of MVO in a 49-year-old man with reperfused STEMI using various CMR imaging techniques (**a**): Short-axis T2-weighted STIR image shows transmural high signal intensity (arrows) and transmural dark signal intensity (arrowhead) in the mid-inferolateral LV segment. (**b**): Short-axis first-pass perfusion image shows a transmural perfusion defect (arrowheads) in the mid-inferolateral LV segment. (**c, d**): Short-axis T1 (**c**) and T2 mapping (**d**) images show elevated T1 (1,379–1,435 ms) and T2 (46–54 ms) values along the infarct border (arrows) and shortened T1 (1,093–1,127 ms, c) and T2 values (27–33 ms, d) in the infarct core (arrowheads). (**e**): Short-axis LGE image shows transmural inferolateral infarction (arrows) with MVO in the infarct core (arrowhead). CMR, cardiac magnetic resonance; MVO, microvascular occlusion; STEMI, ST-segment elevation myocardial infarction; LV, left ventricular; LGE, late gadolinium enhancement; STIR, short tau inversion recovery.

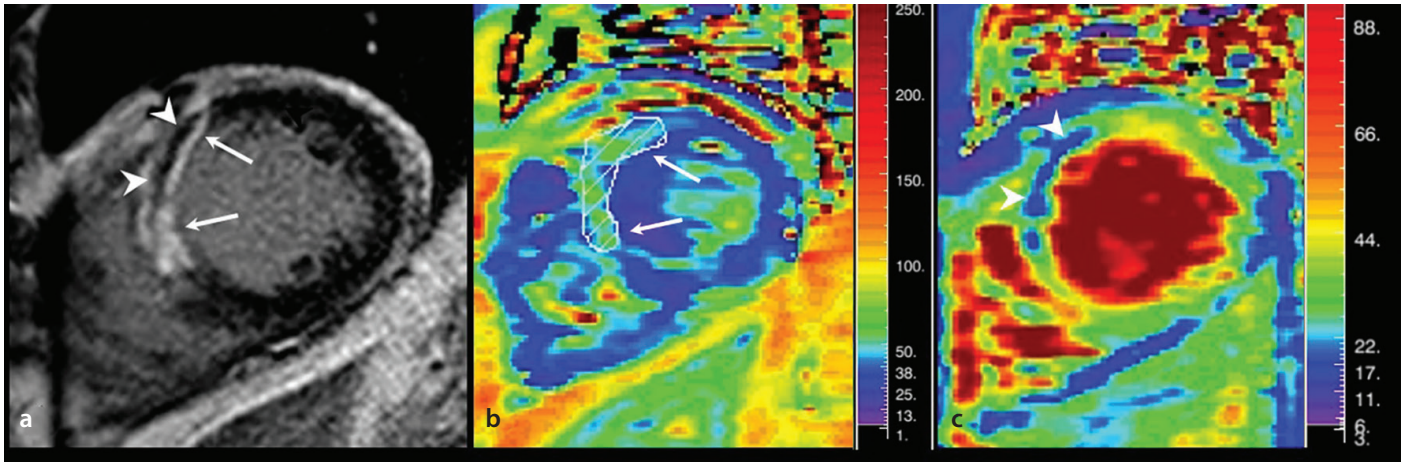



**Figure 3.** Intramyocardial hemorrhage in a 56-year-old man with acute MI. (**a**): Short-axis LGE image shows mid-anterior and anteroseptal infarction (arrows) with MVO in the infarct core (arrowheads). (**b**): Short-axis T2 mapping image shows lower T2 values (38–42 ms) in the mid-anteroseptal and anterior LV wall (arrows) than in the remote myocardium (50–58 ms). (**c**): Short-axis T2* mapping image shows a shorter T2* time (< 20 ms) in the infarct core (arrowheads), indicating IMH. Photographs courtesy of Jung Im Jung, MD, Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea. MI, myocardial infarction; IMH, intramyocardial hemorrhage; LGE, late gadolinium enhancement; MVO, microvascular occlusion; LV, left ventricular.

## Myocardial viability, ischemia, and stress cardiac magnetic resonance imaging

LGE-CMR imaging remains the reference standard for the assessment of myocardial viability. Segments with < 25% of LGE involvement demonstrate a high likelihood of functional recovery, whereas those with > 75% involvement rarely recover.[12] Stress CMR imaging, and particularly adenosine stress perfusion imaging, enables a robust functional assessment of myocardial ischemia and may be useful for evaluating residual ischemia and non-culprit lesions after MI,[13] and T2 mapping and T2WI help differentiate acute from chronic MI by depicting infarct-related myocardial edema (Figure 4).

## Area at risk and myocardial salvage

A primary goal in MI management is to limit infarct size and preserve the AAR. In acute MI, T2WI displays myocardial edema that typically extends beyond the infarct core identified on LGE, with the difference representing salvageable myocardium (Figure 5). Furthermore, T1 and T2 mapping en-

able quantitative assessment of the AAR and myocardial salvage, thus reducing observer dependency and improving reproducibility. A greater myocardial salvage index has been associated with favorable LV remodeling and functional recovery, highlighting the clinical relevance of mapping-based assessment in post-infarction risk stratification and therapeutic decision making.[7,12]

### Differentiation between microvascular obstruction, intramyocardial hemorrhage, and myocardial dissecting hematoma

CMR imaging plays a critical role in differentiating between MVO, IMH, and myocardial dissecting hematomas, and accurate distinction between them directly influences their management. MVO appears as a non-enhancing area within the hyperenhanced, infarcted myocardium on LGE and is best delineated on first-pass perfusion imaging. IMH typically co-localizes with MVO and is characterized by shortened T2* values (< 20 ms) resulting from hemosiderin deposition (Figures 2 and 3).[7] By contrast, a myocardial dissecting hematoma presents as an intramyocardial cavity with variable signal intensity, a preserved epicardial contour, and a lack of typical infarct-related enhancement, allowing for its differentiation from reperfusion-related injury and thus guiding appropriate clinical management.[14]

### Cardiac computed tomography

Coronary CT angiography (CCTA) allows for a rapid, non-invasive evaluation of coronary stenosis and plaque morphology, including high-risk features associated with acute coronary syndromes (Figure 6). When integrated with CCTA, CT-MPI enables combined anatomic and functional assessment and may be useful for evaluating residual ischemia and non-culprit lesions after MI treated with percutaneous coronary intervention.[15] Dynamic stress and rest CT-MPI examinations enable a quantitative assessment of myocardial blood flow and coronary flow reserve, with reduced coronary flow reserve values (typically < 2.0) commonly used as a marker of myocardial ischemia (Figure 7).[16]

MDE-CT visualizes the infarcted myocardium as delayed iodine enhancement and may identify non-enhancing regions consistent with MVO. Compared with LGE-CMR, MDE-CT has inherent limitations, including a lower contrast-to-noise ratio, greater radiation exposure, a lack of standardized acquisition protocols, and limited availability. Accordingly, MDE-CT should be regarded as a complementary and largely investigational technique that is primarily considered when CMR imaging is contraindicated or when a CT scan is already being performed for coronary assessment (Figure 8). An MDE-CT image is typically acquired 5–10 minutes after contrast administration using low-kV, ECG-triggered delayed scanning with or without a minimal, additional contrast injection.[17] Recent technical advances, including dual-energy iodine mapping and iterative reconstruction, have improved contrast and image quality. In addition, AI-based deep learning denoising enables automated infarct segmentation, enhances the visualization capabilities of MDE, and may facilitate an acceptable image quality at lower radiation doses (Figure 9).[18]

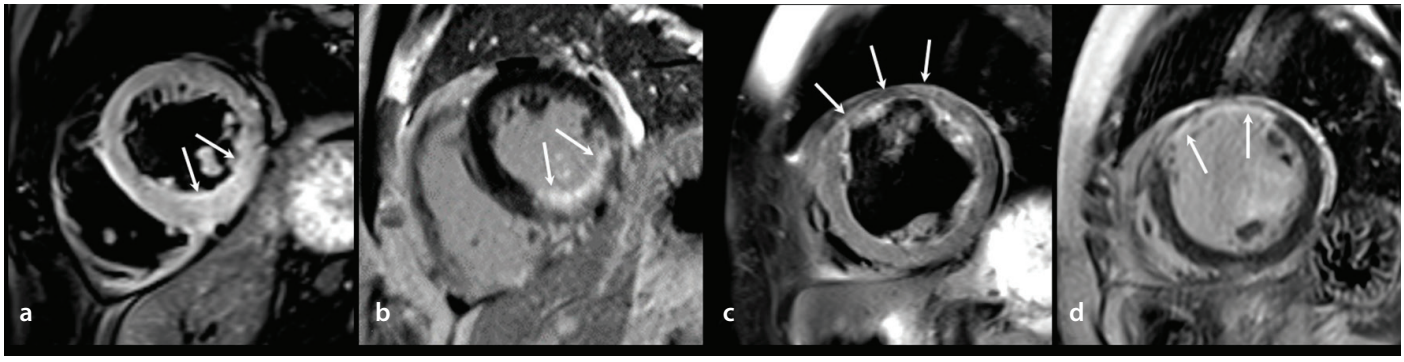



**Figure 4.** Difference in CMR imaging findings between acute and chronic MI. (**a, b**): Short-axis T2-weighted STIR (**a**) and LGE (**b**) images show transmural high signal intensity (**a**) and delayed enhancement (**b**) in the mid-inferior and inferoseptal LV segments without myocardial thinning (arrows), consistent with acute MI. (**c, d**): Short-axis T2-weighted STIR (**c**) and LGE (**d**) images show a normal T2 signal with myocardial thinning (**c**) and transmural delayed enhancement (**d**) in the mid-anterior LV segment (arrows), consistent with chronic MI. Although T2-weighted STIR alone cannot reliably distinguish acute from chronic MI, the combination of LGE and T2-weighted STIR enables the identification of infarct-related myocardial edema indicative of acute MI. CMR, cardiac magnetic resonance; MI, myocardial infarction; LV, left ventricular; LGE, late gadolinium enhancement; STIR, short tau inversion recovery.

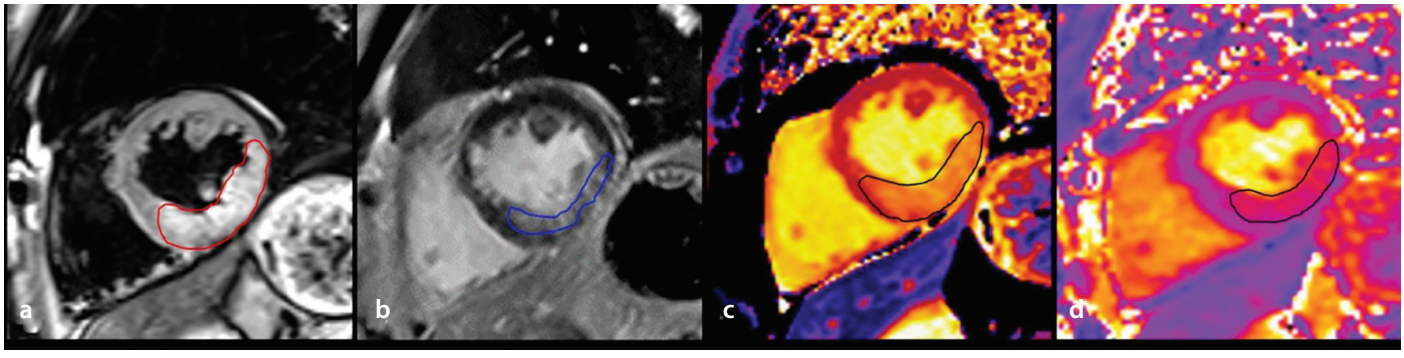



**Figure 5.** Evaluation of myocardial salvage in a 62-year-old man with reperfused STEMI using T2-weighted STIR and LGE. (**a, b**): Short-axis T2-weighted STIR (**a**) and LGE (**b**) images show transmural high signal intensity (red circle, **a**) and delayed enhancement (blue circle, **b**) in the mid-inferior and inferoseptal LV segments. Myocardial salvage is defined as the hyperintense area on T2-weighted STIR minus infarct size on LGE. (**c, d**): The region of elevated T1 values (1,479–1,535 ms; black circle, (**c**) corresponds to the T2-weighted STIR hyperintense area (**a**) and is larger than the regions of elevated T2 values (50–56 ms; black circle, **d**) and LGE (**b**). STEMI, ST-segment elevation myocardial infarction; LV, left ventricular; LGE, late gadolinium enhancement; STIR, short tau inversion recovery.

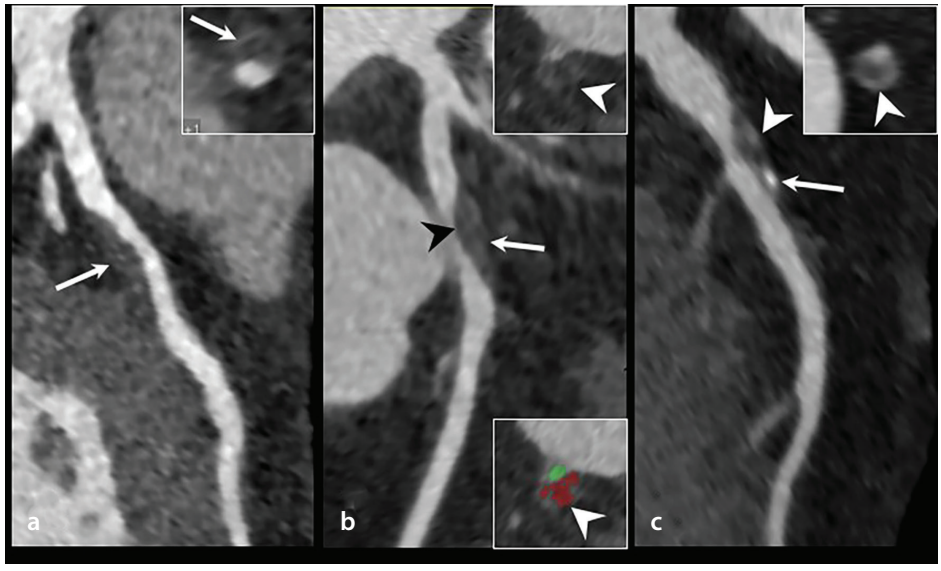



**Figure 6.** Analysis of vulnerable atherosclerotic plaques using coronary computed tomography angiography. High-risk (vulnerable) plaque features are classified as (**a**) positive remodeling (arrow); (**b**) positive remodeling with low-attenuation plaque (arrowhead); and (**c**) positive remodeling with spotty calcification (arrow) and low-attenuation plaque with a napkin-ring sign (non-calcified plaque with an enhancing ring, arrowhead).

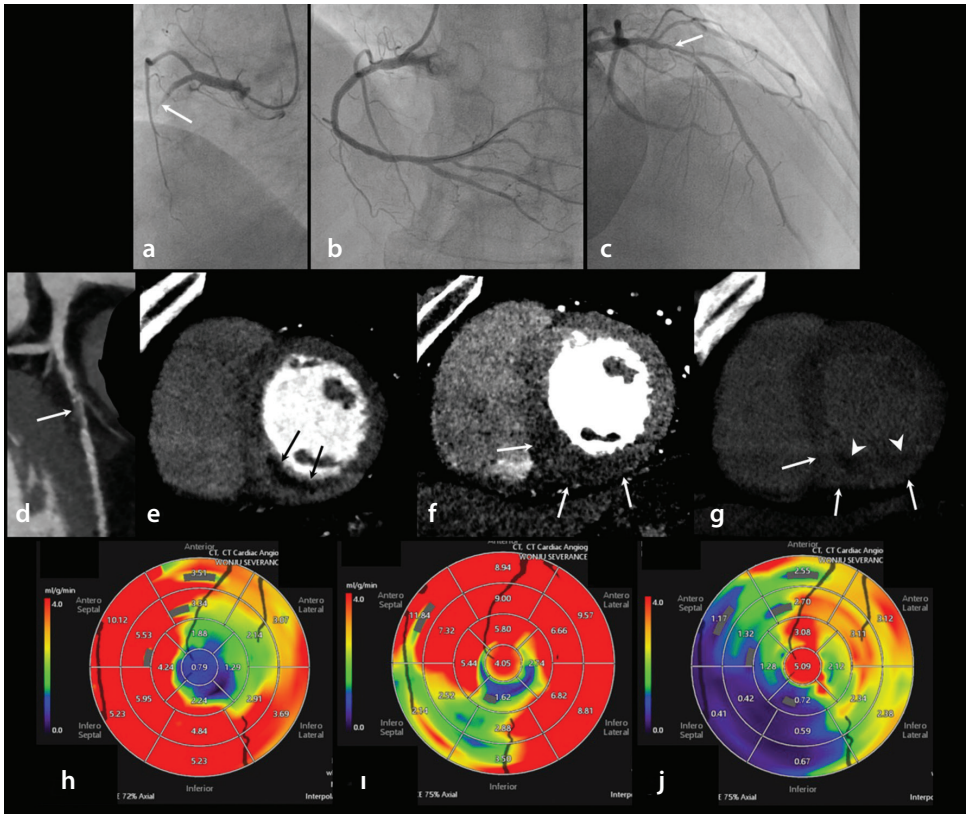



**Figure 7.** Dynamic CT-MPI to assess the hemodynamic significance of non-infarct-related arteries in reperfused STEMI. (**a–c**): Invasive coronary angiography shows complete thrombotic occlusion of the proximal RCA (**a**, arrow), successfully treated with PCI (**b**). A non-culprit lesion with approximately 70% luminal narrowing is noted in the proximal LAD (**c**, arrow). (**d**): Curved MPR CCTA image shows significant (> 70%) stenosis with non-calcified plaque in the proximal LAD (arrow). (**e, f**): Dynamic stress (**e**) and rest (**f**) CT-MPI shows more pronounced perfusion defects during stress than at rest in the mid-inferior and inferoseptal LV segments (arrows), with no perfusion defects in the LAD territory. (**g**): Myocardial delayed-enhancement short-axis MPR CT image demonstrates transmural delayed enhancement (arrows) with subendocardial non-enhancing areas consistent with MVO (arrowheads) in the RCA territory, indicating acute transmural MI. (**h–j**): Polar maps of dynamic stress (**h**) and rest (**i**) CT-MPI examinations and CFR (**j**). Per-segment $CFR_{CT}$ is calculated as the ratio of hyperemic $MBF_{CT}$ to resting $MBF_{CT}$. A $CFR_{CT}$ image of the anteroseptal LV wall is reduced (< 1.4) compared with the lateral LV wall (> 2.0). Using the optimal CFRCT cut-off value of 2.0, ischemia is seen in the LAD territory. Invasive FFR was 0.75 in the proximal LAD. Staged physiology-guided PCI was performed on the proximal LAD, and the FFR was increased to 0.91 (not shown). CT, computed tomography; MPI, myocardial perfusion imaging; STEMI, ST-segment elevation myocardial infarction; MPR, multiplanar reformation; LV, left ventricular; CCTA, coronary CT angiography; RCA, right coronary artery; LAD, left anterior descending coronary artery; PCI, percutaneous coronary intervention; MVO, microvascular obstruction; CFR, coronary flow reserve; MBF, myocardial blood flow; FFR, fractional flow reserve.

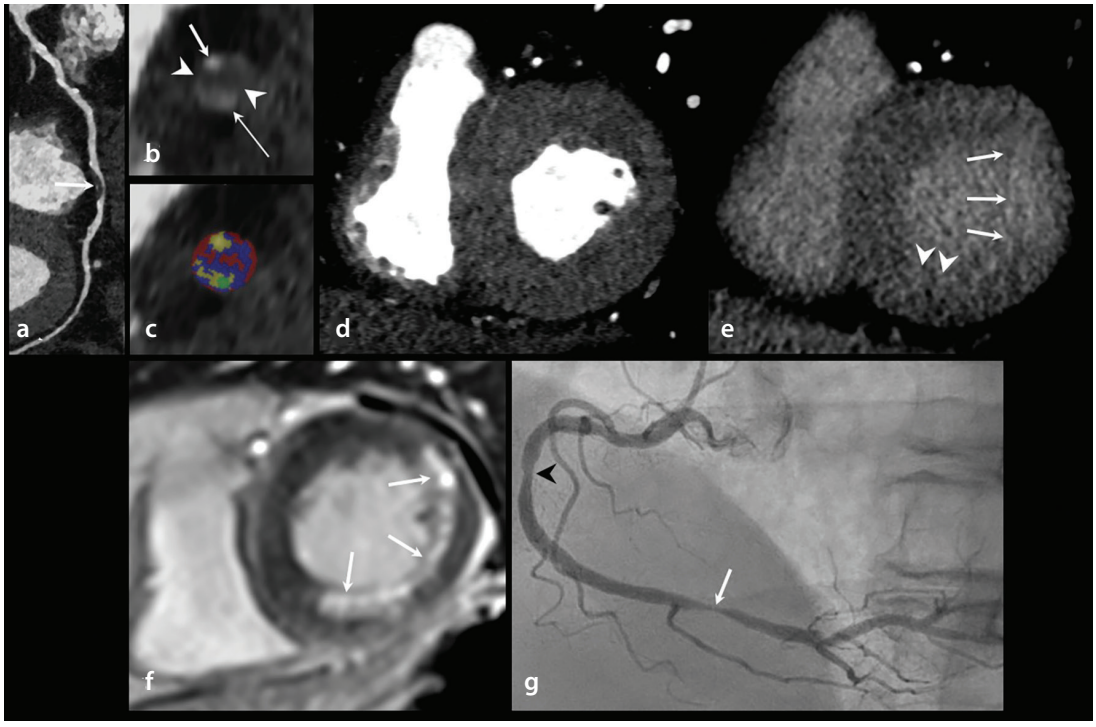



**Figure 8.** Comprehensive coronary CT angiography assessment of acute MI in a 62-year-old man with NSTEMI. (**a**): Curved multiplanar reformation image of the RCA shows mixed calcified and non-calcified plaques with positive remodeling in the distal RCA (arrow). (**b, c**): Cross-sectional images demonstrate very low-attenuation non-calcified plaque (mean attenuation: 2 ± 34 HU, arrowheads) adjacent to calcified plaque (arrow) and contrast-enhanced lumen (long, thin arrow) (**b**). In acute MI, this finding suggests high-risk plaque with a possible superimposed thrombus, though CT-based differentiation is limited. Quantitative plaque analysis software (**c**) depicts soft (red), fibrous (blue), and calcified (yellow) plaque components. (**d, e**): Early- and delayed-phase short-axis CT images show no early hypoenhancement (**d**), but delayed transmural hyperenhancement (arrows) in the basal lateral LV and subendocardial hyperenhancement (arrowheads) in the inferior LV segments (**e**) are visible, consistent with acute MI. (**f**): Delayed-enhancement CMR image confirms corresponding myocardial enhancement patterns (arrows). (**g**): Invasive coronary angiography shows 80% stenosis in the distal RCA (arrow) and 50% stenosis in the mid-RCA (arrowhead), with a hyperdominant RCA supplying the lateral LV wall. NSTEMI, non-ST-segment elevation myocardial infarction; CT, computed tomography; RCA, right coronary artery; LV, left ventricular; CMR, cardiac magnetic resonance; HU, Hounsfield unit.

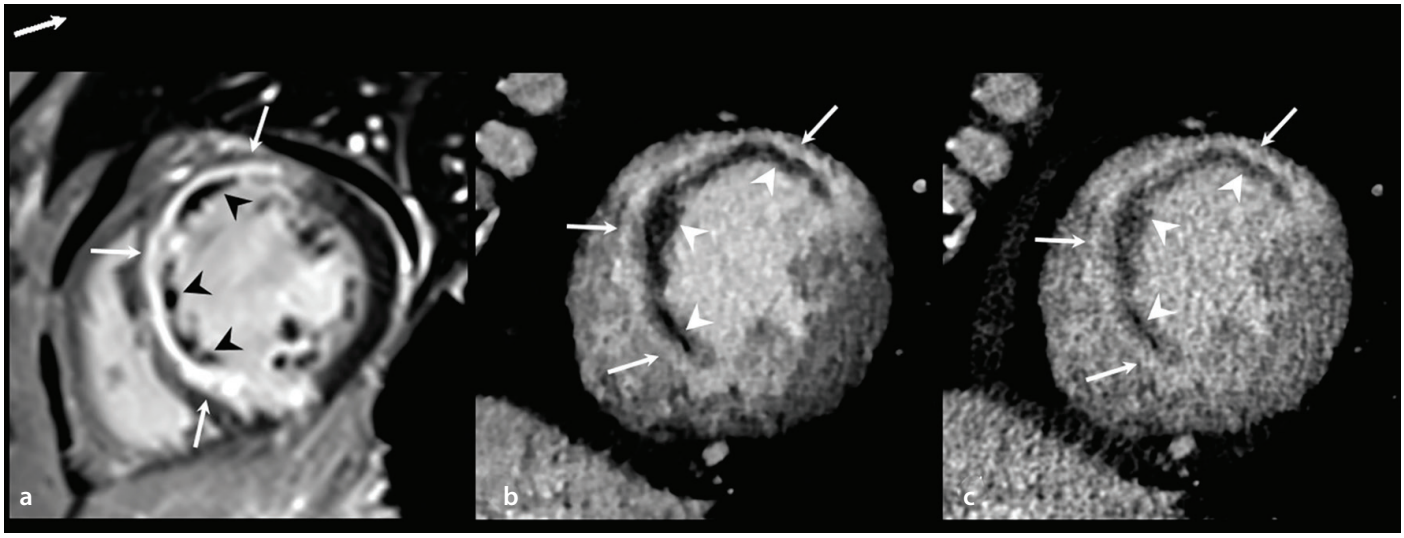



**Figure 9.** Short-axis MDE-CT images of a 64-year-old man with reperfused STEMI. (**a**): Short-axis LGE-CMR image demonstrates transmural delayed enhancement (arrows) with central non-enhancing areas consistent with MVO (arrowheads) involving the mid-anterior, anterolateral, anteroseptal, and inferoseptal LV segments. (**b, c**): Artificial intelligence-based, contrast-enhanced, and denoised CT image (**b**) shows reduced image noise and improved visualization of transmural delayed enhancement (arrows) and non-enhancing regions consistent with MVO (arrowheads) compared with the standard CT image (**c**). MDE, myocardial delayed enhancement; CT, computed tomography; STEMI, ST-segment elevation myocardial infarction; CMR, cardiac magnetic resonance; LV, left ventricular; LGE, late gadolinium enhancement; MVO, microvascular obstruction; AI, artificial intelligence.

Despite these advances, further external validation and protocol standardization are required before broader clinical adoption can proceed.

## Imaging of post-myocardial infarction complications

A variety of mechanical and thromboembolic complications may occur after MI, particularly with delayed or inadequate reperfusion, and are associated with hemodynamic instability and adverse outcomes.[19,20]

Free wall rupture is a catastrophic complication of STEMI. CCT enables rapid assessment, displaying hemopericardium, pericardial effusion, and rupture tracts, whereas LGE-CMR confirms myocardial discontinuity, irregular margins, and associated thrombi, facilitating prompt diagnosis and management (Figure 10).

Ventricular septal rupture is an uncommon but life-threatening complication of MI. CCT provides the rapid, non-invasive detection of septal defects, whereas CMR imaging offers a detailed evaluation of defect size and location and surrounding myocardial integrity, which is essential for surgical planning and prognostic assessment (Figure 11).

Papillary muscle rupture results in acute, severe mitral regurgitation. Echocardiography remains the first-line modality for hemodynamic assessment, whereas CCT and CMR imaging allow for the visualization of papillary muscle integrity and global LV function. In chronic mitral regurgitation, CMR imaging enables an accurate quantification of regurgitant volume and an assessment of myocardial fibrosis using LGE and T1 mapping (Figure 12).

Differentiation between a true aneurysm and a pseudoaneurysm relies heavily on imaging. CCT and CMR imaging assess wall continuity, neck morphology, thrombus formation, and pericardial involvement, which are critical for guiding management decisions (Figures 13 and 14).

LV thrombi are a common post-MI complication. On LGE-CMR, a thrombus appears as a non-enhanced filling defect, with long inversion time imaging improving thrombus conspicuity by suppressing myocardial and blood pool signals. Cine imaging allows for the assessment of thrombus mobility, which aids risk stratification and anticoagulation planning (Figure 13).

Pericardial complications, including pericarditis and effusion, are readily evaluated with CCT for anatomic assessment. CMR imaging enables characterization of pericardial inflammation and enhancement. Fat-suppressed LGE imaging improves visualization of pericardial inflammation by suppressing the adjacent epicardial fat signal, thereby facilitating differentiation between simple effusion and inflammatory pericarditis (Figure 15).

Heart failure represents a major long-term consequence of MI. CMR imaging enables a comprehensive evaluation of ventricular function, valvular mechanics, and myocardial tissue characteristics to support longitudinal assessment and management.

## Clinical integration and modality selection

Rather than competing modalities, CCT, CMR imaging, and echocardiography provide synergistic information.[14,16,19,20] Modality selection should be guided by the clinical question at hand, patient stability, contraindications, and local expertise. A practical clinical algorithm summarizing modality choice is provided in Table 1.

## Ethics and consent

This pictorial essay was conducted in accordance with institutional ethics standards. Formal institutional review board approval and informed consent were waived owing to the essay's retrospective use of anonymized data.

This pictorial essay does not provide pooled diagnostic accuracy metrics, and quantitative thresholds may vary by scanner, field strength, and post-processing procedures. Advanced CT techniques, such as MDE-CT and CT-MPI, and emerging AI-based and radiomics approaches are not yet widely standardized—a circumstance that could potentially limit the generalizability of this essay's conclusions. Therefore, this article focuses on established and clinically applicable imaging findings.

CCT and CMR imaging provide complementary information in the evaluation of MI beyond the acute phase. CMR imaging enables detailed tissue characterization, viability assessment, and evaluation of post-infarction complications. CCT allows for rapid coronary assessment and functional evaluation and plays an important role in assessing select post-infarction complications. Integrated with echocardiography and clinical context, multimodality imaging allows for the characterization of infarct-related injury and residual ischemia. A scenario-based imaging approach supports informed clinical decision-making in patients with MI.

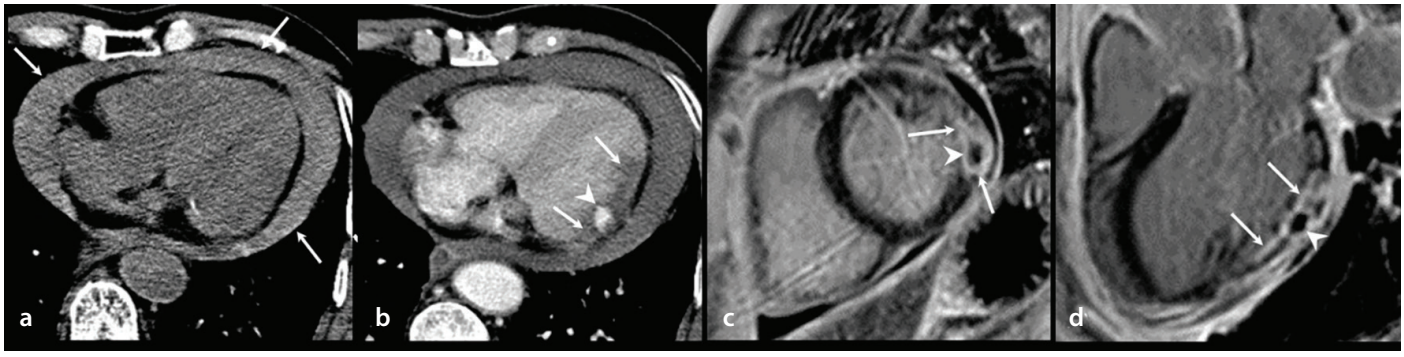



**Figure 10**. Post-MI LV free wall rupture in a 79-year-old man with NSTEMI. (**a**): Non-contrast axial CT image shows massive acute hemopericardium (arrows). (**b**): Contrast-enhanced axial CT image shows a patchy, contrast-filled outpouching lesion (arrowhead) in the lateral LV wall with poor enhancement (arrows). (**c, d**): Short-axis (**c**) and three-chamber (**d**) LGE-CMR images show transmural delayed enhancement (arrows) with central low signal intensity (arrowheads) in the basal to mid-lateral LV wall. Oval-shaped low signal intensity lesion was considered a thrombus within the ruptured LV wall. The patient underwent open-heart surgery, and an LV free wall rupture of the oozing type was noted in the inferolateral LV wall. NSTEMI, non-ST-segment elevation myocardial infarction; CT, computed tomography; LV, left ventricular; CMR, cardiac magnetic resonance; LGE, late gadolinium enhancement.

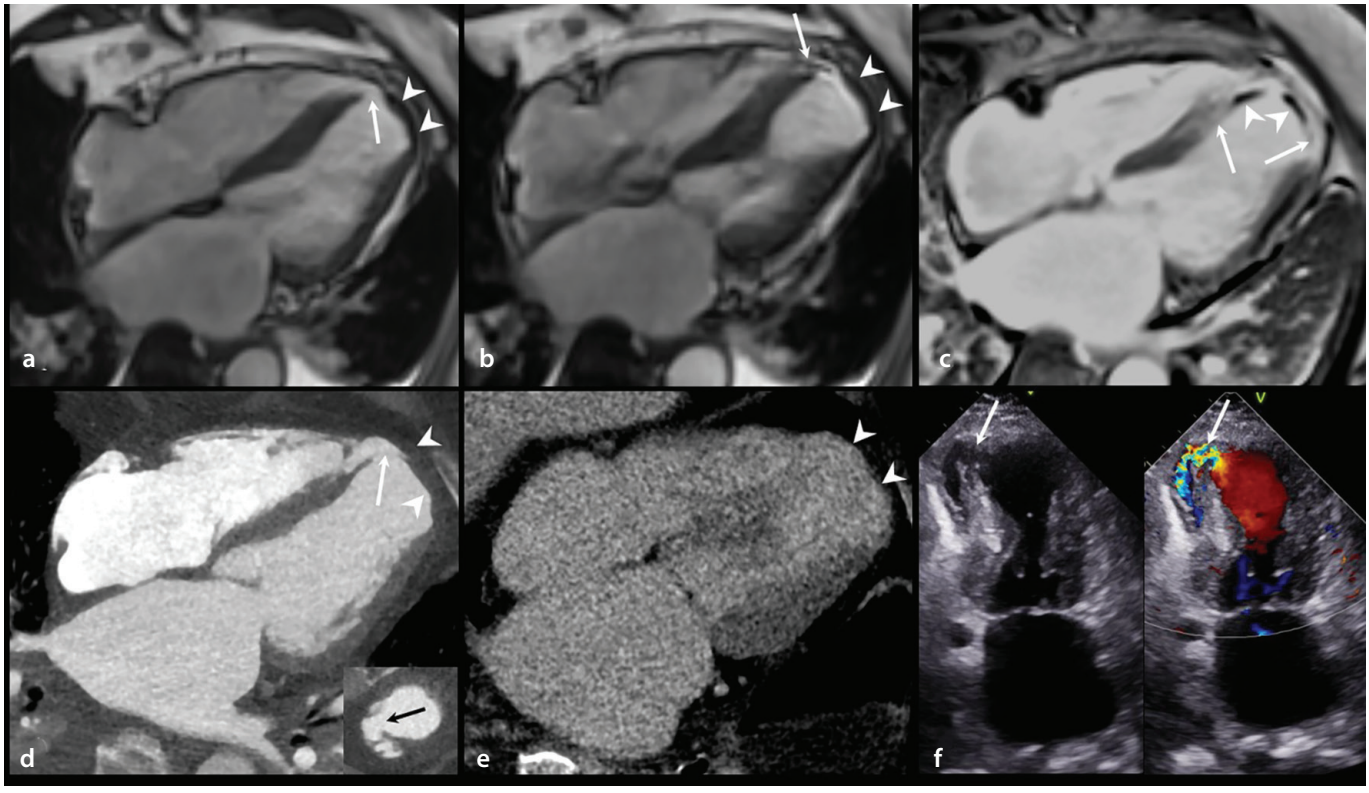



**Figure 11**. Post-MI LV septal rupture in an 80-year-old woman with STEMI. (**a, b**): Four-chamber cine CMR images in the end-diastolic (a) and end-systolic (**b**) phases demonstrate LV apical akinesia with increased wall thickness (arrowheads) and a pencil-like signal void (**b**, arrow) through a focal defect in the apical septum (**a**, arrow). (**c**): Four-chamber LGE-CMR image shows transmural delayed enhancement with patchy MVO (arrowheads) in the apex as well as in the apical septal and lateral and mid-septal LV walls (arrows). (**d, e**): Four-chamber MPR arterial phase (**d**) and delayed-phase (**e**) CT images show focal thinning with an apical septal defect (**d**, arrow) and increased wall thickness (**d**, arrowheads) of the apex and apical lateral LV walls with delayed hyperenhancement (e, arrowheads). (**f**): Apical four-chamber view of color Doppler echo images shows shunt flow through a focal defect in the apical septum (arrow). The patient underwent open-heart surgery, and a 3-cm defect was found in the lower part of the apical septum and sutured with a bovine patch. Apical aneurysmal change was removed and reinforced using felt material. STEMI, ST-segment elevation myocardial infarction; b-SSFP, balanced steady state-free precession; CMR, cardiac magnetic resonance; CT, computed tomography; LGE, late gadolinium enhancement; MPR, multiplanar reformation

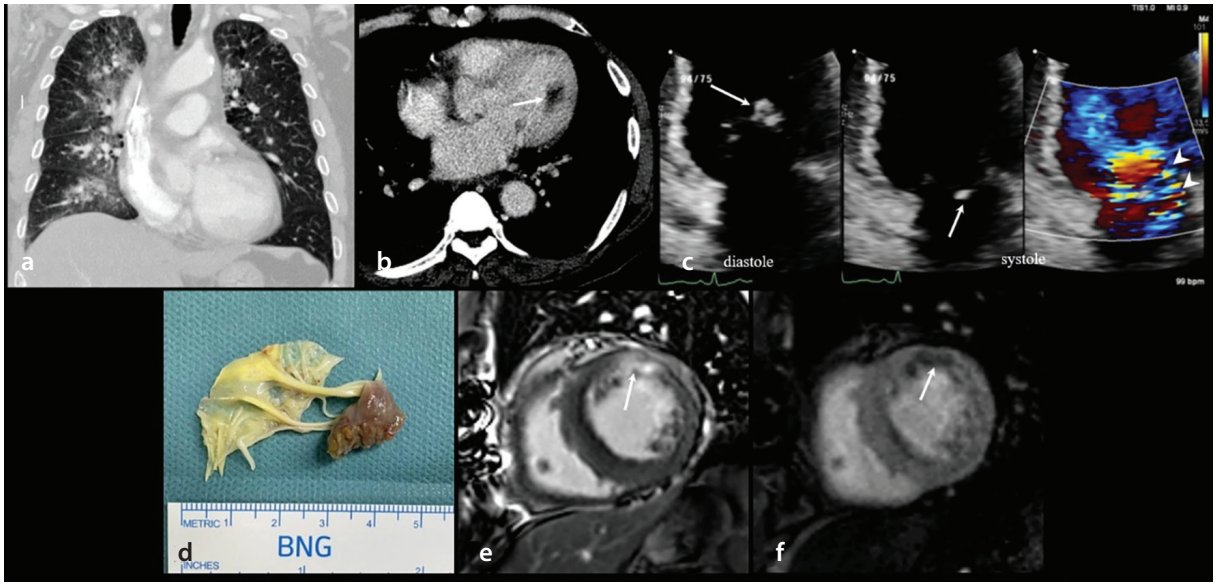



**Figure 12**. Acute severe MR resulting from papillary muscle rupture in a 68-year-old man who presented with sudden shortness of breath. (**a**): Coronal MPR image of a chest CT scan shows pulmonary edema in both lungs (primarily in the right lung). (**b**): Contrast-enhanced axial chest CT image shows hypoenhancement in the anterolateral papillary muscle (arrow). (**c**): Apical four-chamber view of color Doppler echo images shows severe eccentric MR (arrowheads) resulting from an anterolateral papillary muscle rupture (arrows) and associated mitral valve prolapse (mainly A2). The patient immediately underwent emergency surgery for mechanical mitral valve replacement, and an anterolateral papillary muscle rupture was discovered (**d**). Two weeks later, the patient underwent stress perfusion CMR imaging to evaluate the underlying cause of the anterolateral papillary muscle rupture. (**e, f**): LGE (**e**) and stress perfusion (**f**) CMR images show transmural delayed enhancement (**e**, arrow) and stress-induced, patchy transmural perfusion defects (**f**, arrow) in the mid-anterior and anterolateral LV segments. MR, mitral regurgitation; MPR, multiplanar reformation; CMR, cardiac magnetic resonance; CT, computed tomography; LGE, late gadolinium enhancement.

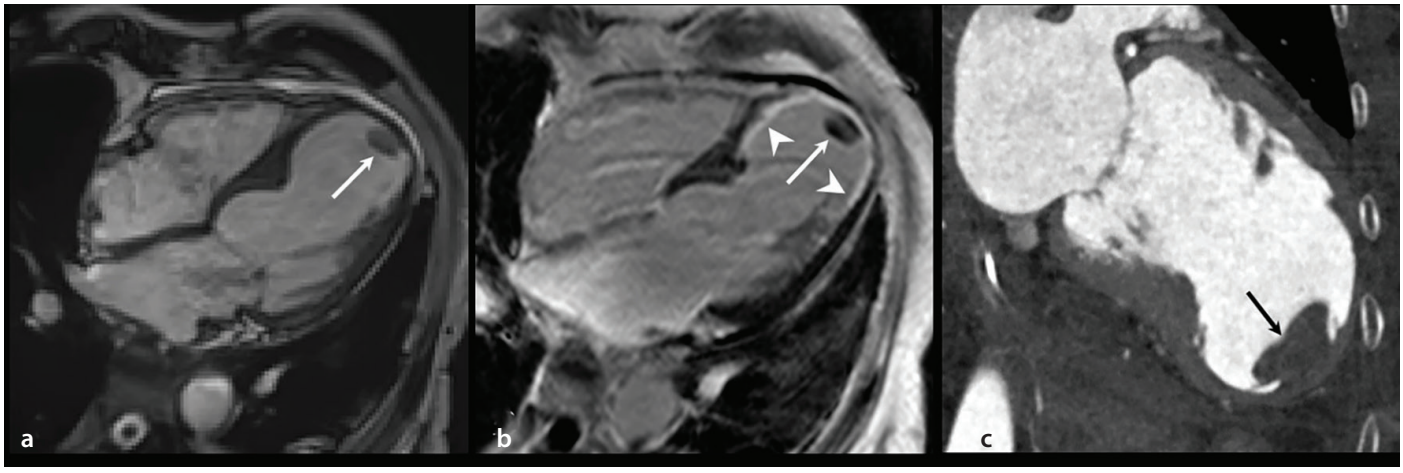



**Figure 13.** True LV aneurysm in a 75-year-old woman with a history of reperfused STEMI. (**a–c**): Four-chamber cine (**a**) and LGE-CMR images (**b**) and a two-chamber MPR CT image (**c**) show an apical aneurysm of the left ventricle with transmural delayed enhancement (**b**, arrowheads) and an apical thrombus (arrows). STEMI, ST-segment elevation myocardial infarction; LV, left ventricular; MPR, multiplanar reformation; CMR, cardiac magnetic resonance; CT, computed tomography; LGE, late gadolinium enhancement.

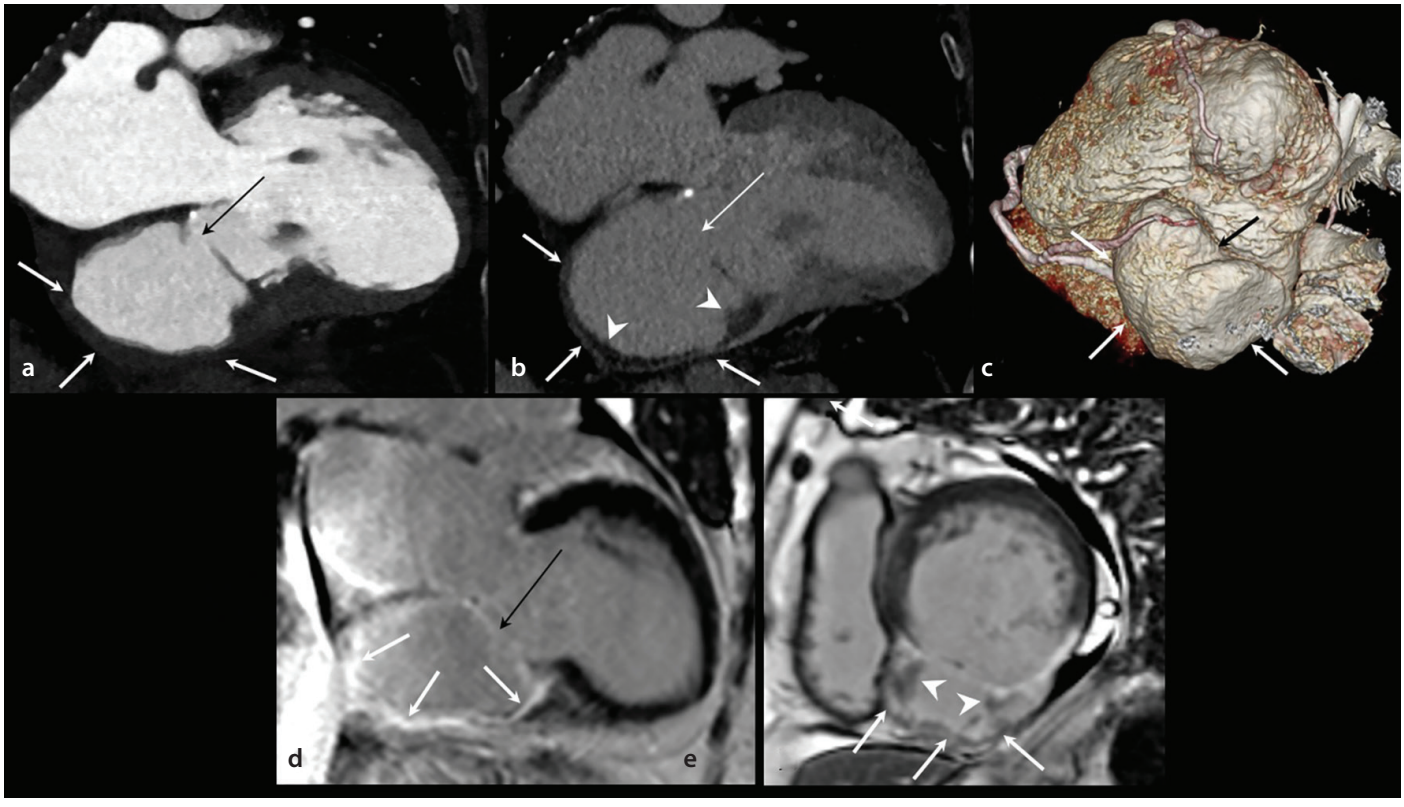



**Figure 14.** Left ventricular pseudoaneurysm in an 80-year-old woman with a history of reperfused STEMI. (**a, b**): Two-chamber MPR early contrast-enhanced (**a**) and MDE (**b**) CT images demonstrate a pseudoaneurysm of the basal to mid-inferior wall of the left ventricle with a narrow neck (long arrow), very thin wall (arrows), and intracavitary thrombus (arrowheads). (**c**): Three-dimensional, volume-rendering CT image of the heart shows a large aneurysm (arrows) arising from the inferior wall of the left ventricle. (**d, e**): Two-chamber (**d**) and short-axis (**e**) LGE-CMR images demonstrate a large inferior pseudoaneurysm with a narrow neck (**d**, long arrow), diffuse LGE of the pseudoaneurysm wall (arrows), and a thrombus (**e**, arrowheads). The surgical finding was consistent with a pseudoaneurysm caused by a rupture of the LV wall secondary to MI. STEMI, ST-segment elevation myocardial infarction; LV, left ventricular; MPR, multiplanar reformation; CMR, cardiac magnetic resonance; CT, computed tomography; LGE, late gadolinium enhancement; MDE, myocardial delayed enhancement.

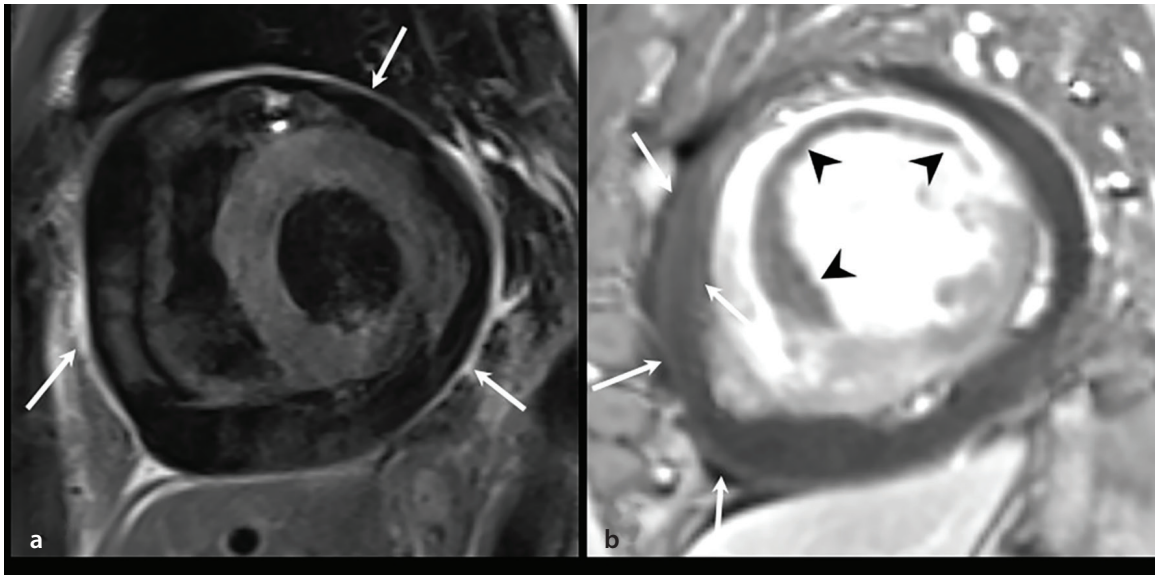



**Figure 15**. Post-MI pericarditis in a 62-year-old man with reperfused STEMI. (**a**): Short-axis T2-weighted STIR CMR image shows increased pericardial signal intensity (arrows) reflecting pericardial edema. A moderate amount of pericardial effusion is also seen. (**b**): Short-axis LGE-CMR image shows increased pericardial enhancement (arrows) suggestive of pericardial inflammation and extensive MVO (arrowheads) in the mid-anterior, anterolateral, and anteroseptal LV walls. STEMI, ST-segment elevation myocardial infarction; LGE, late gadolinium enhancement; STIR, short tau inversion recovery; CMR, cardiac magnetic resonance; LV, left ventricular; MVO, microvascular obstruction.

**Table 1.** Algorithm for clinical imaging modality selection following myocardial infarction

| Clinical scenario | Preferred modality | Rationale |
|---|---|---|
| Hemodynamically unstable acute MI | Echocardiography | Bedside, real-time assessment of ventricular function and mechanical complications. |
| Post-reperfusion viability and area at risk | CMR imaging | Gold standard for assessing infarct size, transmurality, area at risk, and MVO. |
| Non-culprit lesion significance/residual ischemia | Stress CMR or cardiac CT (CCTA + CT-derived MPI/ CT-FFR) | Functional assessment of ischemia with an integrated anatomic–functional assessment. |
| Suspected LV thrombus | CMR (long-TI LGE) imaging | Highest diagnostic accuracy for thrombus detection and characterization. |
| MRI contraindicated | Cardiac CT (MDE-CT) | Complementary alternative for myocardial viability assessment. |
| Mechanical complications | CT + echocardiography + CMR | Comprehensive evaluation of anatomy, hemodynamics, and myocardial integrity. |
| Remodeling/prognosis | CMR (T1 mapping, ECV, MVO) | Quantitative biomarkers for risk stratification and outcome prediction. |

MI, myocardial infarction; CMR, cardiac magnetic resonance; MVO, microvascular obstruction; CT, computed tomography; CCTA, coronary CT angiography; MPI, myocardial perfusion imaging; FFR, fractional flow reserve; LV, left ventricular; TI, inversion time; LGE, late gadolinium enhancement; MRI, magnetic resonance imaging; MDE, myocardial delayed enhancement; ECV, extracellular volume.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

### Funding

This work was supported by the Natural Science Foundation of Shandong Province.

## References

1. Salari N, Morddarvanjoghi F, Abdolmaleki A, et al. The global prevalence of myocardial infarction: a systematic review and meta-analysis. *BMC Cardiovasc Disord*. 2023;23(1):206. [Crossref]
2. Rajiah P, Desai MY, Kwon D, Flamm SD. MR imaging of myocardial infarction. *RadioGraphics*. 2013;33(5):1383-1412. [Crossref]
3. Peper J, Suchá D, Swaans M, Leiner T. Functional cardiac CT-going beyond anatomical evaluation of coronary artery disease with cine CT, CT-FFR, CT perfusion and machine learning. *Br J Radiol*. 2020;93(1113):20200349. [Crossref]
4. Williams MC, Weir-McCall JR, Baldassarre LA, et al. Artificial intelligence and machine learning for cardiovascular computed tomography (CCT): A White Paper of the Society of Cardiovascular Computed Tomography (SCCT). *J Cardiovasc Comput Tomogr*. 2024;18(6):519-532. [Crossref]
5. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *Circulation*. 2018;138(20):e618-e651. [Crossref]
6. Kloner RA. Stunned and hibernating myocardium: where are we nearly 4 decades later? *J Am Heart Assoc*. 2020;9(3):e015502. [Crossref]
7. Bulluck H, Dharmakumar R, Arai AE, Berry C, Hausenloy DJ. Cardiovascular magnetic resonance in acute ST-segment-elevation myocardial infarction: recent advances, controversies, and future directions. *Circulation*. 2018;137(18):1949-1964. [Crossref]
8. Liu D, Borlotti A, Viliani D, et al. CMR native T1 mapping allows differentiation of reversible

versus irreversible myocardial damage in ST-segment-elevation myocardial infarction: an OxAMI Study (Oxford Acute Myocardial Infarction). *Circ Cardiovasc Imaging*. 2017;10(8):e005986. [Crossref]

9. Chen Y, Ren D, Guan X, et al. Quantification of myocardial hemorrhage using T2* cardiovascular magnetic resonance at 1.5T with ex-vivo validation. *J Cardiovasc Magn Reson*. 2021;23(1):104. [Crossref]
10. Bulluck H, Carberry J, Carrick D, et al. A noncontrast CMR risk score for long-term risk stratification in reperfused ST-segment elevation myocardial infarction. *JACC Cardiovasc Imaging*. 2022;15(3):431-443. [Crossref]
11. Ma Q, Ma Y, Wang X, et al. A radiomic nomogram for prediction of major adverse cardiac events in ST-segment elevation myocardial infarction. *Eur Radiol*. 2021;31(2):1140-1150. [Crossref]
12. Garcia MJ, Kwong RY, Scherrer-Crosbie M, et al. State of the art: imaging for myocardial viability: a scientific statement from the American Heart Association. *Circ Cardiovasc Imaging*. 2020;13(7):e000053. [Crossref]
13. Nagel E, Greenwood JP, McCann GP, et al. Magnetic resonance perfusion or fractional flow reserve in coronary disease. *N Engl J Med*. 2019;380(25):2418-2428. [Crossref]
14. Alyousef T, Malhotra S, Iskander F, et al. Left ventricular intramyocardial dissecting hematoma: a multimodality imaging diagnostic approach. *Circ Cardiovasc Imaging*. 2021;14(7):e012410. [Crossref]
15. Pontone G, Baggiano A, Andreini D, et al. Stress computed tomography perfusion versus fractional flow reserve CT derived in suspected coronary artery disease: the PERFECTION Study. *JACC Cardiovasc Imaging*. 2019;12(8 Pt 1):1487-1497. [Crossref]
16. Dewey M, Siebes M, Kachelrieß M, et al. Clinical quantitative cardiac imaging for the assessment of myocardial ischaemia. *Nat Rev Cardiol*. 2020;17(7):427-450. [Crossref]
17. Rodriguez-Granillo GA. Delayed enhancement cardiac computed tomography for the assessment of myocardial infarction: from bench to bedside. *Cardiovasc Diagn Ther*. 2017;7(2):159-170. [Crossref]
18. Nishii T, Kobayashi T, Tanaka H, et al. Deep Learning-based post hoc CT denoising for myocardial delayed enhancement. *Radiology*. 2022;305(1):82-91. [Crossref]
19. Stephens NR, Restrepo CS, Saboo SS, Baxi AJ. Overview of complications of acute and chronic myocardial infarctions: revisiting pathogenesis and cross-sectional imaging. *Postgrad Med J*. 2019;95(1126):439-450. [Crossref]
20. Damluji AA, van Diepen S, Katz JN, et al. Mechanical complications of acute myocardial infarction: a scientific statement from the American Heart Association. *Circulation*. 2021;144(2):e16-e35. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253550

CHEST IMAGING

INVITED REVIEW



# Photon-counting detector computed tomography in thoracic oncology: revolutionizing tumor imaging through precision and detail

Masahiro Yanagawa[1]
Midori Ueno[1]
Rintaro Ito[2]
Daiju Ueda[3]
Tsukasa Saida[4]
Ryo Kurokawa[5]
Koji Takumi[6]
Kentaro Nishioka[7]
Shunsuke Sugawara[8]
Satoru Ide[9]
Maya Honda[10]
Mami Iima[11]
Mariko Kawamura[12]
Akihiko Sakata[13]
Keitaro Sofue[14]
Seitaro Oda[15]
Tadashi Watabe[1]
Kenji Hirata[16]
Shinji Naganawa[12]

[1]The University of Osaka Graduate School of Medicine, Department of Diagnostic and Interventional Radiology, Osaka, Japan

[2]Nagoya University Graduate School of Medicine, Department of Innovative BioMedical Visualization, Nagoya, Japan

[3]Osaka Metropolitan University Graduate School of Medicine, Department of Artificial Intelligence, Osaka, Japan

[4]University of Tsukuba Institute of Medicine, Department of Radiology, Tsukuba, Japan

[5]The University of Tokyo Graduate School of Medicine, Department of Radiology, Tokyo, Japan

[6]Kagoshima University Graduate School of Medical and Dental Sciences, Department of Radiology, Kagoshima, Japan

[7]Hokkaido University Faculty of Medicine, Global Center for Biomedical Science and Engineering, Division of Radiation Oncology, Sapporo, Japan

[8]National Cancer Center Hospital, Department of Diagnostic Radiology, Tokyo, Japan

[9]University of Occupational and Environmental Health, Department of Radiology, Kitakyushu, Japan

[10]Kyoto University Hospital, Preemptive Medicine and Lifestyle-Related Disease Research Center, Kyoto, Japan

[11]Nagoya University Graduate School of Medicine, Department of Fundamental Development for Advanced Low Invasive Diagnostic Imaging, Nagoya, Japan

[12]Nagoya University Graduate School of Medicine, Department of Radiology, Nagoya, Japan

[13]Kyoto University Graduate School of Medicine, Department of Diagnostic Imaging and Nuclear Medicine, Kyoto, Japan

[14]Kobe University Graduate School of Medicine, Department of Radiology, Kobe, Japan

[15]Kumamoto University Faculty of Life Sciences, Department of Diagnostic Radiology, Kumamoto, Japan

[16]Hokkaido University Graduate School of Medicine, Department of Diagnostic Imaging, Sapporo, Japan

**ABSTRACT**

Photon-counting detector computed tomography (PCD-CT) is an emerging imaging technology that promises to overcome the limitations of conventional energy-integrating detector (EID)-CT, particularly in thoracic oncology. This narrative review summarizes technical advances and clinical applications of PCD-CT in the thorax with emphasis on spatial resolution, dose–image-quality balance, and intrinsic spectral imaging, and it outlines practical implications relevant to thoracic oncology. A literature review of PubMed through May 31, 2025, was conducted using combinations of "photon counting," "computed tomography," "thoracic oncology," and "artificial intelligence." We screened the retrieved records and included studies with direct relevance to lung and mediastinal tumors, image quality, radiation dose, spectral/iodine imaging, or artificial intelligence-based reconstruction; case reports, editorials, and animal-only or purely methodological reports were excluded. PCD-CT demonstrated superior spatial resolution compared with EID-CT, enabling clearer visualization of fine pulmonary structures, such as bronchioles and subsolid nodules; slice thicknesses of approximately 0.4 mm and *ex vivo* resolvable structures approaching 0.11 mm have been reported. Across intraindividual clinical comparisons, radiation-dose reductions of 16%–43% have been achieved while maintaining or improving diagnostic image quality. Intrinsic spectral imaging enables accurate iodine mapping and low-keV virtual monoenergetic images and has shown quantitative advantages versus dual-energy CT in phantoms and early clinical work. Artificial intelligence-based deep-learning reconstruction and super-resolution can complement detector capabilities to reduce noise and stabilize fine-structure depiction without increasing dose. Potential reductions in contrast volume are biologically plausible given improved low-keV contrast-to-noise ratio, although clinical dose-finding data remain limited, and routine K-edge imaging has not yet translated to clinical thoracic practice. In conclusion, PCD-CT provides higher spatial and spectral fidelity at lower or comparable doses, supporting earlier and more precise tumor detection and characterization; future work should prioritize outcome-oriented trials, protocol harmonization, and implementation studies aligned with "Green Radiology".

**KEYWORDS**

Artificial intelligence, cancer, computed tomography, diagnosis, iodine, radiation dose reduction, spatial resolution, spectral imaging

**Handling editor:** Furkan Ufuk

**Corresponding author:** Masahiro Yanagawa

**E-mail:** m-yanagawa@radiol.med.osaka-u.ac.jp

Received 17 July 2025; revision requested 11 August 2025; accepted 11 September 2025.

Epub: 24.09.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253550

You may cite this article as: Yanagawa M, Ueno M, Ito R, et al. Photon-counting detector computed tomography in thoracic oncology: revolutionizing tumor imaging through precision and detail. *Diagn Interv Radiol*. 2026;32(4):437-447.

Computed tomography (CT) plays a pivotal role in thoracic imaging, offering invaluable insights for the diagnosis, staging, and monitoring of various pulmonary conditions, particularly in oncology. However, conventional energy-integrating detector (EID)-CT systems have limitations, including restricted spatial resolution, higher image noise, and an inability to differentiate materials based on their spectral properties. These limitations can hinder the precise characterization of subtle pathological abnormalities, such as early-stage lung nodules and complex interstitial lung diseases. Photon-counting detector CT (PCD-CT) has emerged as a revolutionary imaging modality that addresses many of these challenges, promising to alter clinical practice in thoracic oncology fundamentally. Conventional EID-CT converts X-rays into scintillation light, requiring separators that limit spatial resolution and efficiency; furthermore, light-to-current conversion introduces electronic noise, reducing image quality and accuracy. In contrast, PCD-CT converts photons directly into electrical signals without noise, offering theoretically superior spatial resolution (0.11 mm in-plane, 0.20 mm along the body axis) and multi-energy information for precise X-ray energy analysis (Figure 1). As a result of these structural differences, PCD-CT offers a range of advantages, including significantly improved spatial resolution, reduced image noise, and multi-energy spectral imaging capabilities.[1-10] These advancements result in clearer, more detailed images, even at reduced radiation exposure levels. The clinical utility of PCD-CT in chest imaging is rapidly expanding, with reports demonstrating its efficacy in improving visualization of bronchi and vessels,[7,9,11] enhancing the detection of emphysema and post-coronavirus disease-2019 lung abnormalities,[6,10,12,13] increasing diagnostic confidence in interstitial pneumonia-related findings, such as reticulation,[6,10,14,15] and improving image quality for pulmonary embolism assessment.[10,16] Beyond these immediate clinical benefits, PCD-CT's inherent capabilities for a sizeable radiation dose reduction and potential for lower contrast agent usage align with the growing concept of Green Radiology. This paradigm emphasizes environmentally sustainable and patient-safe radiological practices, making PCD-CT a key technology for minimizing the ecological footprint and patient burden of medical imaging. Lung cancer, with its diverse histological types and heterogeneous nature, presents a major diagnostic challenge. Accurate preoperative assessment of lung tumors, including the grading of lung adenocarcinoma and the estimation of invasive components, is crucial for guiding treatment strategies and predicting patient prognosis.[17-19] Although conventional CT has offered valuable insights, its ability to evaluate these subtle tumor characteristics precisely has been limited. Furthermore, the integration of cutting-edge artificial intelligence (AI) with PCD-CT is poised to revolutionize tumor imaging further. Deep learning reconstruction (DLR) algorithms, for instance, can markedly reduce image noise while preserving spatial resolution, surpassing the capabilities of traditional iterative reconstruction algorithms.[11,20] This synergistic collaboration between advanced detector technology and AI holds immense potential to enhance image quality and diagnostic confidence in lung tumor evaluation, paving the way for more precise and personalized oncology care.

**Main points**

- Photon-counting detector computed tomography (PCD-CT) offers markedly improved spatial and spectral resolution compared with conventional energy-integrating detector CT, allowing for more accurate detection and characterization of thoracic tumors, including subsolid nodules and early-stage lung cancer.
- PCD-CT substantially reduces radiation doses by up to 40% without compromising image quality and shows promise in reducing contrast agent volume, contributing to safer and more sustainable imaging.
- The integration of artificial intelligence and spectral imaging techniques with PCD-CT enhances diagnostic precision and aligns with the principles of Green Radiology, supporting both clinical excellence and environmental responsibility.

Although PCD-CT holds promise across a wide range of anatomical regions, this review particularly emphasizes its application in thoracic oncology imaging due to its clinical importance. The aim is to outline comprehensively the transformative impact of PCD-CT in this field, highlighting advancements in spatial resolution, the critical balance between image quality and radiation dose reduction, and the novel insights enabled by its spectral imaging capabilities. By exploring these key areas, we highlight how PCD-CT is revolutionizing tumor imaging through unprecedented precision and detail, ultimately leading to improved diagnostic performance and patient outcomes.

## Methods of literature review

This work is a narrative review focusing on PCD-CT in thoracic oncology, synthesizing technical advances in spatial resolution, the balance between dose and image quality, and spectral/iodine imaging, and considering their clinical implications for lung and mediastinal tumors. The PubMed database was searched through May 31, 2025, using combinations of "photon counting," "photon-counting detector," "computed tomography," "thoracic," "lung," "mediastinum," "oncology," and "artificial intelligence." After deduplication, titles and abstracts were screened, and full texts were assessed. The priorities were as follows: comparisons of intra-individual PCD-CT versus EID-CT; thoracic oncology endpoints (nodule characterization, staging adjuncts, treatment response, surveillance/complications); trade-offs between dose and image quality; spectral/iodine quantification; and AI/DLR analyses. In total, 323 records were screened, and 68 studies were included in the qualitative synthesis; to avoid implying a systematic review, a flow diagram is not provided, and a meta-analysis was not performed. Animal-only studies, purely methodological papers without thoracic relevance, editorials/case reports/letters, and duplicate analyses without additional information were excluded.

## Comparison of spatial resolution between energy-integrating detector and photon-counting detector computed tomography

Building upon the foundational advancements outlined in the introduction, one of the most compelling advantages of PCD-CT in thoracic imaging is its inherently superior spatial resolution compared with traditional EID-CT (Figure 2). This enhanced capability stems directly from the ability of PCD-CT to count individual photons and precisely categorize them by energy. This fundamental difference in detection methodology minimizes signal loss and electronic noise, thereby allowing for the acquisition of data with finer detail and leading to a notable improvement in image sharpness and the ability to discern minute anatomical structures.[1-12] Clinical studies have provided robust evidence of superior spatial resolution on PCD-CT and its direct impact on diagnostic image quality. Inoue et al.[12] conducted a prospective clinical study in adult patients undergoing low-dose lung cancer screening, in which three readers directly compared PCD-CT and EID-CT using a 5-point Likert scale (−2 to +2).

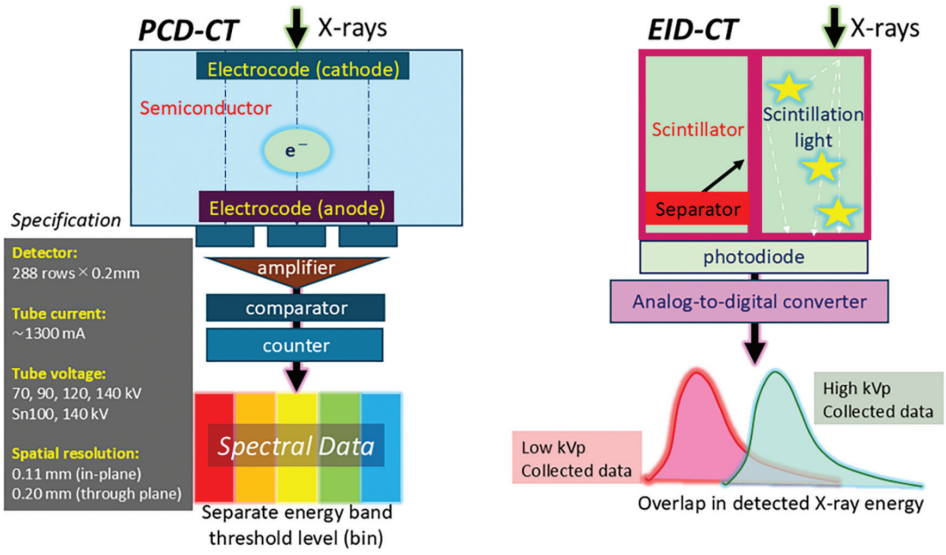



**Figure 1.** Comparison of signal conversion processes in energy-integrating detector computed tomography (EID-CT) and photon-counting detector (PCD)-CT. Conventional scintillator-type detector (EID)-CT devices convert incident X-rays into scintillation light containing all spectral information for processing, and separators are required to prevent light from diffusing to adjacent detection elements. The thickness of the separator limits spatial resolution and reduces X-ray usage efficiency, and the electrical noise generated when converting light into electric current degrades image quality and quantitative performance. In comparison, PCD-CT collects X-rays by energy and converts electrons directly into an electric current; therefore, there is no electrical noise. Theoretically, it provides higher spatial resolution than conventional CT (0.11 mm in plane, 0.20 mm in body axis direction) and multi-energy information for each X-ray energy, enabling highly accurate X-ray energy analysis.

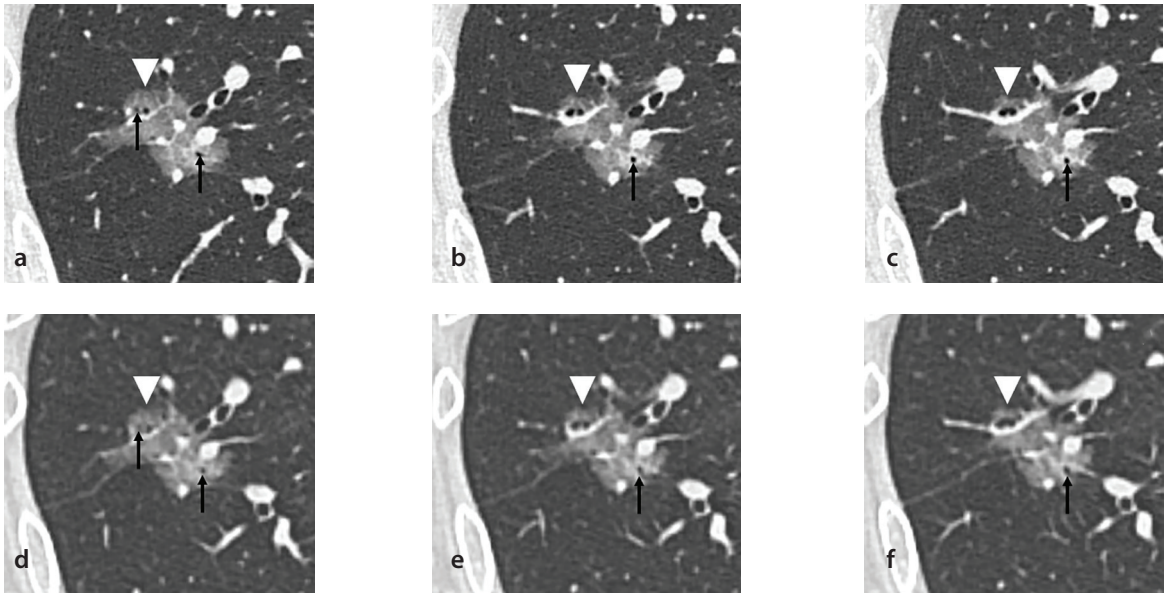



**Figure 2.** Comparison of photon-counting computed tomography (CT) images (**a-c**) and energy-integrating detector (EID)-CT images (**d-f**) in a 72-year-old woman with lung adenocarcinoma. The photon-counting CT images [CT dose index volume (CTDIvol), 9.2 mGy; matrix size, 512; and slice thickness, 0.4 mm] demonstrate clearer depiction of air bronchogram (black arrows) within the nodule and the nodule margins (white arrow heads) than the EID-CT images (CTDIvol, 12.8 mGy; matrix size, 512; and slice thickness, 0.625 mm).

Their findings revealed that PCD-CT provided significantly superior delineation of lung nodule boundaries (+0.8 ± 0.9, $P < 0.001$) and improved visualization of emphysema (+0.3 ± 0.6, $P < 0.001$). PCD-CT was consistently rated superior for image sharpness (mean score +0.8, $P < 0.001$), despite the use of a similar reconstruction kernel. Two of the three readers also reported significantly higher overall diagnostic quality (+0.3 to +0.7, $P < 0.001$). Similarly, Wang et al.[21] demonstrated PCD-CT's advanced capability in characterizing subsolid nodules, noting its superiority over EID-CT in depicting subtle nodule features, with the exception of lobulation. This study also highlighted a crucial technical aspect, concluding that 0.4 mm slice thickness PCD-CT images achieved an optimal balance between ultra-high resolution and subjective diagnostic image quality, which is crucial for guiding clinical imaging protocols. Furthermore, Bartlett et al.[9] investigated the impact of higher matrix reconstruction (e.g., 1,024 matrix) with PCD-CT, reporting improved visualization of higher-order bronchi and bronchial wall clarity, suggesting the clinical benefit of the enhanced technical resolution capabilities inherent to PCD-CT systems.

Further compelling evidence for the high spatial resolution of PCD-CT comes from *ex vivo* studies using human tissues, which bridge the gap between technical capabilities and direct clinical relevance. Hata et al.[11] performed a detailed comparative analysis of PCD-CT and EID-CT using inflated cadaveric human lungs. This unique study design allowed for the precise evaluation of the technologies on human tissue without the confounding effects of physiological motion. The study confirmed that PCD-CT depicted lung nodules and airway microstructure with significantly greater clarity than EID-CT. A pivotal finding was the quantification of PCD-CT's detection limit, demonstrating its capability to detect nodules and airways with a median diameter of approximately 600 µm. The maximum spatial resolution achieved in this study was 0.11 mm, emphasizing its extraordinary ability to resolve microscopic anatomical details.[11] These findings underscore the current ability of PCD-CT to provide unprecedented information on subtle pulmonary structures. In a clinical context,[6,20] this means a higher potential for earlier detection and more precise characterization of a wide range of lung diseases, including incipient lung cancers, thereby facilitating earlier intervention and improved patient outcomes. In the future, this enhanced spatial resolution is expected to become a cornerstone for refined disease staging, personalized treatment planning, and potentially even the identification of novel imaging biomarkers for prognosis and treatment response in thoracic oncology. See "Thoracic Oncology Tasks and PCD-CT Advantages: Evidence Map" in Table 1.

### Radiation dose reduction without degradation of spatial resolution in photon-counting detector computed tomography

Researchers have conducted many studies on conventional EID-CT and have investigated and implemented radiation dose reduction strategies, implying the importance of dose management.[22-33] Radiation dose is a critical consideration in all CT examinations, and minimizing patient exposure is a fundamental principle of radiological practice.

Recent clinical studies have provided compelling evidence that PCD-CT can achieve substantial radiation dose reductions compared with EID-CT.[12,21,34,35] These reductions are clinically meaningful, especially in the context of lung cancer screening programs and long-term oncologic surveillance, where cumulative dose becomes a significant concern. As noted in the previous section, Inoue et al.[12] revealed the utility of PCD-CT in evaluating nodule boundaries and visualization of emphysema. Crucially, these improvements were achieved concurrently with a significantly lower average CT dose index volume, demonstrating an approximate 16.4% reduction (0.61 mGy vs. 0.73 mGy, $P < 0.001$) compared with EID-CT. Similarly, Wang et al.[21] further corroborated PCD-CT's benefit in characterizing subsolid nodules. Their study demonstrated that PCD-CT not only improved lesion depiction but also delivered a significantly reduced effective dose, quantifying an approximate 17.5% reduction (1.79 ± 0.39 mSv vs. 2.17 ± 0.57 mSv, $P < 0.001$) compared with EID-CT. Overall, various clinical studies have demonstrated significant dose reductions with PCD-CT compared with EID-CT, generally ranging from approximately 16% to over 40% depending on the specific comparison and protocol. The dose-reduction benefit of PCD-CT extends to specific vulnerable patient populations and advanced imaging protocols where dose optimization is critical. Early studies with PCD-CT in pediatric chest imaging of cystic fibrosis have reported average effective doses as low as 0.12 mSv,[36] a remarkable reduction compared with even ultra-low-dose EID-CT protocols, which can be approximately 0.15 mSv for combined inspiratory and expiratory chest CT in pediatric populations.[37] For adults with cystic fibrosis, PCD-CT has demonstrated effective doses as low as 0.55 mSv, representing a 42% reduction compared with EID-CT protocols for similar diagnostic tasks.[34] PCD-CT has notably demonstrated a 31% lower effective dose to red bone marrow than EID-CT[34] in younger cohorts, where the risk of radiation-induced malignancy is a significant concern.[38,39] Although conventional low-dose EID-CT protocols for chest imaging typically report effective radiation doses ranging from 1.5 mSv to 2.0 mSv,[40,41] PCD-CT has shown substantial advancements; its substantial dose reduction capabilities offer promising benefits for enhancing patient safety in routine clinical practice.[42] Furthermore, the advantages of PCD-CT also extend to comparisons with other advanced CT modalities, such as dual-energy CT (DECT). Hagen et al.[43] performed a clinical comparison of contrast-enhanced chest imaging, a cornerstone for evaluating tumor vascularity and treatment response, using PCD-CT versus a second-generation dual-source DECT in oncology patients. They reported that PCD-CT provided significantly higher tumor-to-lung parenchyma contrast ratios, indicating enhanced tumor conspicuity and delineation. Importantly, PCD-CT maintained equivalent or superior image quality while achieving a substantial 43% reduction in radiation dose compared with a second-generation dual-source DECT.[43] This robust clinical evidence directly confirms PCD-CT's impressive dose efficiency and image quality benefits in a demanding patient cohort undergoing oncologic imaging.

PCD-CT has the potential to reduce contrast agent usage in thoracic oncology diagnosis. Sawall et al.[44] compared PCD-CT with EID-CT using phantoms containing various iodine concentrations (ICs) and demonstrated improvements in the contrast-to-noise ratio (CNR) of up to 30% with a single energy bin and up to 37% with optimal two-bin weighting, corresponding to a potential reduction in contrast medium dose of up to 37%.[44] In addition, several studies have reported that PCD-CT enables a 25%–50% reduction in contrast medium dose in head and neck CT angiography (CTA) or coronary CTA while maintaining diagnostic image quality or CNR.[45,46] Such a contrast medium dose reduction may be particularly beneficial for patients with impaired renal function and elderly populations.

These clinical investigations collectively demonstrate that PCD-CT offers a compelling combination of maintained or improved image quality and significant dose reduction across various clinical scenarios pertinent to thoracic oncology. This improved benefit–risk ratio represents a major advancement in CT, positioning PCD-CT as an invaluable tool, particularly for lung cancer screening programs and long-term oncologic follow-up, where repeated examinations are common. Crucially, this unique synergy of imaging quality, dose efficiency, and potential for contrast medium dose reduction also strongly aligns with the burgeoning principles of Green Radiology, emphasizing the minimization of patient exposure and the promotion of environmentally sustainable imaging practices. This comprehensive advantage solidifies PCD-CT's role as a cornerstone for optimizing patient care and advancing diagnostic capabilities in thoracic imaging.

### Artificial intelligence technology for super-resolution

This AI-based super-resolution subsection may be useful to clarify achievable EID-CT resolution, complement PCD-CT in practice, and support detector-aware harmonization across EID-CT and PCD-CT data.

The field of medical imaging continues to advance rapidly, driven by innovations in detector hardware and post-processing algorithms.[20,28,47-51]

**Table 1.** Thoracic oncology tasks and photon-counting detector computed tomography (PCD-CT) advantages: evidence map

| Clinical task | Specific question | Typical comparator | Study type (examples) | Primary metrics (examples) | Reported signal | Clinical implication | References |
|---|---|---|---|---|---|---|---|
| **Lung cancer screening** (low-dose nodule detection) | Does PCD-CT improve detection and clarity of small pulmonary nodules at low radiation doses, compared with standard low-dose computed tomography (CT)? | Conventional low-dose energy-integrating detector CT (EID-CT) | Prospective intra-individual comparisons; feasibility trials in screening populations | Nodule detection rate; small nodule boundary visibility; image noise; CT dose index volume (radiation dose) | ↑ (Improved) | Earlier detection of lung cancer with sharper nodule depiction at reduced dose, enhancing screening efficacy while lowering radiation risk | [8,12] |
| **Subsolid nodule characterization** (invasive vs. indolent) | Can PCD-CT better characterize subsolid (ground-glass or part-solid) lung nodules and assess invasive components versus conventional CT? | Thin-section EID-CT (standard high-res) | Paired same-day patient studies (PCD-CT vs. EID-CT in each patient) | Lesion conspicuity and demarcation; CT attenuation of solid component; radiologist invasiveness classification; effective dose | ↑ (Improved) | More accurate non-invasive assessment of tumor invasiveness, aiding surgical decision-making (e.g., suitability for sublobar resection) while also reducing patient dose | [21,58] |
| **Tumor morphology and margins** (solid tumor evaluation) | Does PCD-CT provide clearer visualization of tumor margins and fine morphological features (e.g., spiculation, lobulation) compared with conventional CT? | High-resolution EID-CT (ultra-thin slices with iterative recon) | *Ex vivo* cadaver lung comparisons; prospective clinical studies with artificial intelligence super-resolution reconstruction | Spatial resolution (achieved voxel size); minimum detectable lesion size; clarity of edge features (spiculation, air bronchograms); reader confidence scores | ↑ (Improved) | Enhanced delineation of tumor extent and characteristics, improving lesion characterization and aiding precise surgical or radiotherapy planning | [9,20] |
| **Lymph node staging** (mediastinal metastasis detection) | Can PCD-CT spectral imaging improve identification of metastatic lymph nodes in thoracic malignancies compared with standard CT or dual-energy CT? | Contrast-enhanced EID-CT or dual-energy CT (with iodine mapping) | Retrospective clinical spectral CT studies (with surgical pathology as reference); phantom iodine-detection experiments | Iodine concentration in nodes (correlation with malignancy); virtual monoenergetic imaging clarity; minimum detectable iodine level (phantom) | ↑ (Improved) | Potential for more accurate non-invasive nodal staging – improved detection of nodal metastases due to superior contrast and spectral sensitivity, possibly reducing need for invasive sampling | [63,64] |
| **Treatment response assessment** (therapy monitoring) | Does PCD-CT improve the evaluation of tumor response to therapy (e.g., changes in size or enhancement) compared with conventional CT? | Standard contrast-enhanced CT (EID-CT or latest dual-source CT) | Preliminary clinical comparisons in oncology patients (PCD-CT vs. dual-source CT) | Tumor-to-lung contrast ratio; lesion conspicuity on follow-up; image noise and SNR; dose per scan | ↑ (Improved) | Clearer visualization of residual tumor and subtle changes in enhancement or size, facilitating more confident assessment of treatment efficacy while minimizing radiation and contrast burden | [43] |

SNR, signal-to-noise ratio.

Among the most promising developments is the application of AI, particularly super-resolution techniques based on deep learning, to improve the spatial resolution of CT imaging beyond the physical limits of existing systems.[52,53] These methods can reconstruct high-resolution images from lower-resolution data, effectively enhancing fine structural details essential for accurate diagnosis and treatment planning (Figure 3).

Although advances in CT spatial resolution have often focused on detector innovations, particularly photon-counting systems, conventional EID-CT has also demonstrated substantial diagnostic capabilities by employing high-matrix imaging. For instance, Yanagawa et al.[54] used a 2,048 × 2,048 matrix with a 0.25 mm slice thickness on an EID-CT to evaluate invasive lung adenocarcinomas. Their findings showed excellent diagnostic performance for identifying features such as disrupted air bronchograms and solid components >0.8 cm, with a sensitivity of 97% and specificity of 86% [area under the curve (AUC), 0.94]. These results suggest that enhancing spatial resolution could lead to more accurate lung cancer evaluations. In addition, high-resolution EID-CT imaging may influence computational approaches, such as radiomics. Ninomiya et al.[55] demonstrated that CT images reconstructed with a 1,024 matrix improved the radiomic prediction of solid and micropapillary components in invasive lung adenocarcinoma, highlighting the potential benefits of high-spatial-resolution input data in image-based feature analysis.

Parallel to these CT hardware-based improvements, AI-based super-resolution techniques have emerged as a promising approach to further enhance CT image quality. Kim et al.[56] demonstrated that applying a three-dimensional (3D) deep learning super-resolution algorithm to thick-slice CT data significantly reduced volumetric measurement error (from 52.2% to 15.7%) and improved Lung Imaging Reporting and Data System categorization accuracy (from 72.7% to 94.5%). Similarly, super-resolution radiomics has been shown to increase the predictive accuracy of histologic subtypes in lung cancer. In a cohort of 245 patients, Xing et al.[57] reported an AUC improvement from 0.761 to 0.819 for the detection of micropapillary and solid components. A multicenter study also found that a super-resolution CT pipeline combined with an SE-ResNet50 model achieved superior prediction of spread through air spaces, with an AUC of 0.806 compared with 0.695 for standard models.[58]

Among these AI-based approaches, Precise IQ Engine (PIQE; Canon Medical Systems, Otawara, Japan) represents a clinically implemented, super-resolution DLR technique tailored for EID-CT systems. The PIQE system uses a 3D convolutional neural network trained on paired low- and high-resolution images derived from ultra-high-resolution CT, enabling the reconstruction of low-noise, high-resolution images without increased radiation dose. Notably, it supports high-matrix reconstruction (up to 1,024 × 1024), providing sharper spatial detail and preserving image texture.[52,53] Although most validation studies of PIQE have focused on coronary and abdominal CT, its technical design suggests considerable potential in thoracic oncology. Enhanced spatial resolution and reduced blooming artifacts may improve the delineation of lesion margins, internal heterogeneity, and subtle features, such as broncho-

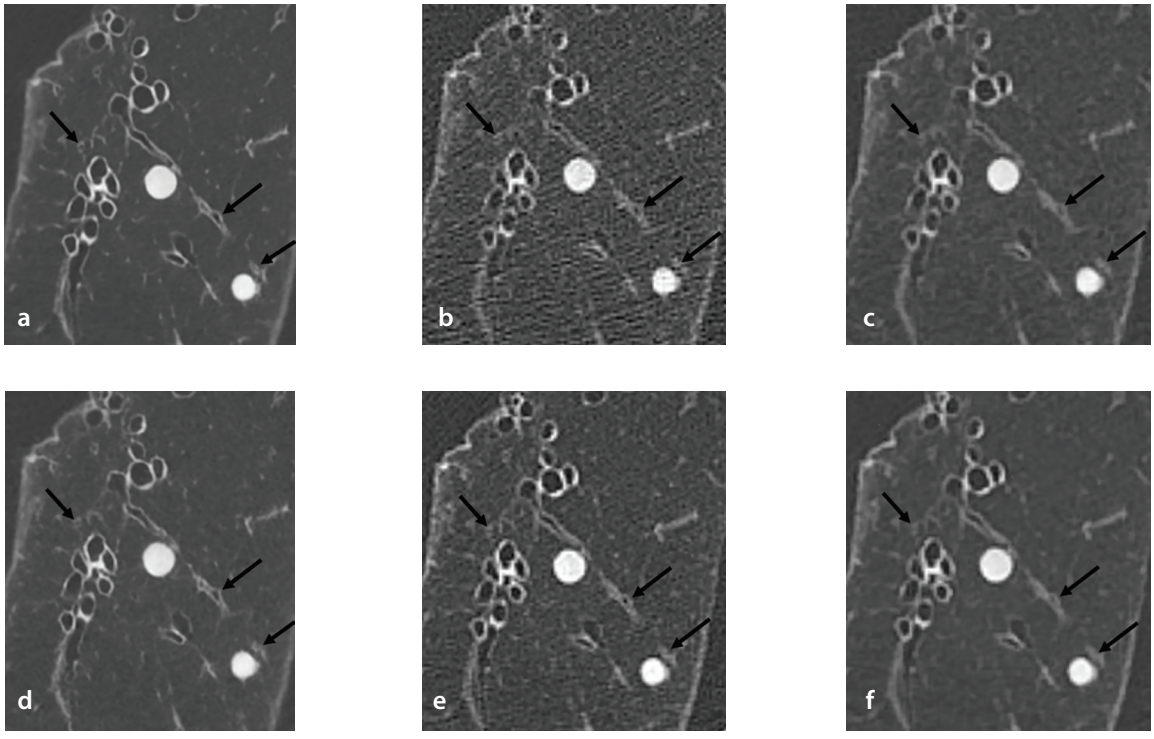



**Figure 3.** Comparison of images acquired at 8.5 mGy and 2.2 mGy, where the dose refers to the computed tomography dose index volume, using different reconstruction methods in a human cadaveric lung with lung metastases. Images include 8.5 mGy-FBP (matrix size, 512; slice thickness, 0.5 mm; (**a**), 8.5 mGy-AiCE (matrix size, 512; slice thickness, 0.5 mm; (**b**), 8.5 mGy-PIQE (matrix size, 1024; slice thickness, 0.5 mm; (**c**), 2.2 mGy-FBP (matrix size, 512; slice thickness, 0.5 mm; (**d**), 2.2 mGy-AiCE (matrix size, 512; slice thickness, 0.5 mm; (**e**), and 2.2 mGy-PIQE [matrix size, 1024; slice thickness, 0.5 mm; (**f**)]. 8.5 mGy-AiCE (**b**) reduces image noise compared with 8.5 mGy-FBP (**a**). 8.5 mGy-PIQE (**c**) further enhances the visualization of bronchiolar walls (black arrows) due to the higher matrix size and higher noise reduction. At a lower dose, 2.2 mGy-FBP (**d**) shows increased noise, whereas 2.2 mGy-AiCE (**e**) mitigates it. Notably, 2.2 mGy-PIQE (**f**) preserves superior delineation of bronchiolar structures despite the reduced dose. FBP, filtered back projection; AiCE, advanced intelligent Clear-IQ Engine, which is a deep learning reconstruction (DLR) technique; PIQE, precise IQ Engine, which is a super-resolution DLR technique.

vascular invasion. These improvements are particularly relevant for image-based staging and treatment planning in lung cancer. Moreover, the use of high-fidelity input data is increasingly recognized as critical in radiomics and AI-based prediction models. Thus, PIQE-enabled super-resolution CT may not only enhance visual interpretation but also improve the performance of computational tools in thoracic oncology.

Although PCD-CT inherently offers enhanced spatial resolution through its advanced detector design, recent studies have highlighted the added value of integrating AI-based reconstruction algorithms. Sasaki et al.[20] investigated a cadmium zinc telluride-based PCD-CT system combined with DLR and demonstrated that ultra-high-resolution images provided clearer delineation of key morphological features, such as spiculation and lobulation. These features are particularly important for the characterization of lung cancer. The observed improvements contributed to greater diagnostic confidence and suggest that AI-based super-resolution techniques can further enhance the clinical utility of PCD-CT in thoracic oncology. Together, these developments demonstrate that AI-enabled super-resolution contributes not only to improved image aesthetics but also to enhanced diagnostic precision and prognostic assessment. Whether applied to EID-CT or PCD-CT, and whether implemented through high-matrix acquisition or DLR, these techniques enable better visualization of subtle but clinically significant features, support more robust radiomic analyses, and allow for earlier and more accurate diagnosis of thoracic malignancies. As AI-driven reconstruction becomes increasingly integrated into clinical workflows, further validation across diverse patient populations and imaging settings will be essential. Nonetheless, the convergence of advanced detector technologies, high-resolution matrix design, and deep learning-based image enhancement represents a promising foundation for the next generation of precision imaging in thoracic oncology.

## Spectral imaging with photon-counting detector computed tomography: technical advances and clinical implications in comparison with dual-energy computed tomography

Spectral imaging has emerged as a pivotal advancement in CT, enabling compositional and quantitative assessments that surpass the capabilities of conventional, attenuation-based imaging. DECT, which achieves spectral differentiation through dual-source acquisition, rapid kilovoltage switching, or layered detectors, has long been used in clinical settings. In thoracic oncology, DECT has proven particularly useful in iodine quantification, which aids in tumor characterization and assessment of treatment response. However, DECT is limited by binary energy separation, spectral overlap, and reduced material specificity.[59-63] PCD-CT represents a transformative advancement in spectral imaging. Unlike EID-CT, PCD-CT systems utilize semiconductor materials, such as cadmium zinc telluride, to directly count individual X-ray photons and classify them by energy.[1-10] This novel detector architecture enables simultaneous multi-energy acquisition with superior energy resolution and reduced electronic noise. Furthermore, PCD-CT inherently provides higher spatial resolution due to smaller detector elements and reduced cross-talk. These features collectively contribute to more accurate material decomposition, enhanced signal-to-noise ratio (SNR), and the potential for lower radiation doses. These advantages are expected to enhance diagnostic performance across a broad range of clinical applications, including thoracic oncology. A comparison of the spectral imaging capabilities of DECT and PCD-CT is summarized in Table 2, illustrating the technical distinctions and clinical implications relevant to thoracic tumor evaluation. As the table indicates, PCD-CT surpasses DECT in energy resolution, spatial resolution, signal-to-noise characteristics, and spectral quantification accuracy, representing a next-generation platform for spectral thoracic oncology imaging (Figures 4, 5).

**Table 2.** Technical comparison between dual-energy computed tomography (DECT) and photon-counting detector computed tomography (PCD-CT) for spectral thoracic imaging

| Feature | DECT | PCD-CT |
|---|---|---|
| Detector type | Energy-integrating detector (EID) | Photon-counting semiconductor detector (e.g., cadmium zinc telluride) |
| Energy resolution | Binary (low) | Multi-bin (high) |
| Spectral acquisition | Two spectra via dual-source/kVp switching, dual-layer/split filter | Simultaneous multi-energy per photon |
| Virtual monochromatic imaging quality | Available (noisy at low keV) | Available (low-noise, high fidelity) |
| Material decomposition | Limited to two-material basis | Multi-material decomposition feasible |
| K-edge imaging | Not supported | Supported (e.g., iodine, gadolinium, gold) |
| Spatial resolution | Limited by detector and noise | Higher due to small pixel and low noise |
| Signal-to-noise ratio | Moderate | High |
| Quantitative accuracy | Limited by beam hardening, overlap | Superior accuracy, artifact-resistant |
| Contrast efficiency | Requires higher dose | Enhanced at lower dose |
| Dose efficiency | Variable, protocol-dependent | Low-dose with high image quality |
| Tumor evaluation | Effective for iodine concentration (IC), Zeff, and extracellular volume fraction in mediastinal tumors | Superior delineation, quantification, and functional imaging |
| Nodal evaluation | Correlation between IC and nodal metastasis shown | Potential for improved nodal characterization with higher resolution |
| Artificial intelligence integration | Available | Available and ongoing |

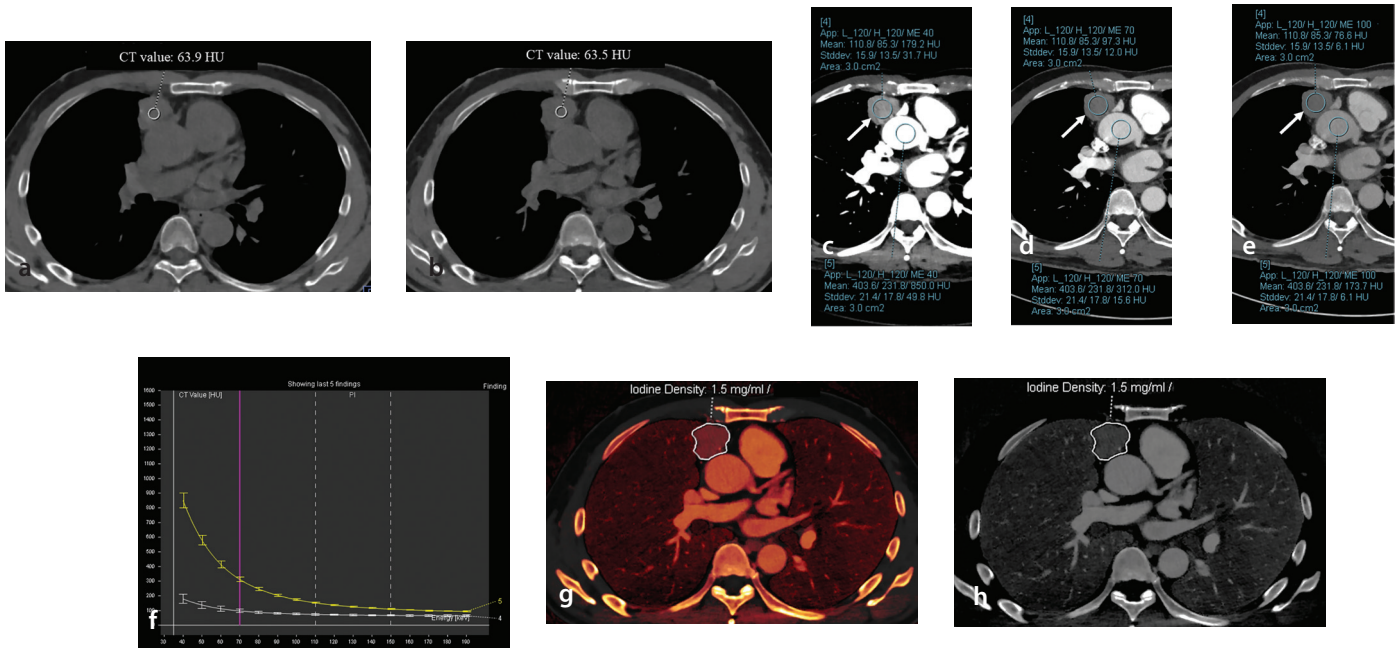



**Figure 4.** Spectral images from photon-counting computed tomography (CT) in a 51-year-old man with thymoma. On the true non-contrast image (**a**) in the mediastinal window, the CT value within the region of interest (circle) is 63.9 Hounsfield units (HU). On the virtual non-contrast image (**b**) reconstructed from the 120-second delayed scan, the CT value within the region of interest (circle) is 63.5 HU, which is nearly identical to that of the true non-contrast image. Monochromatic images reconstructed of 40 keV (**c**), 70 keV (**d**), 100 keV (**e**) from the 120-second delayed scan, and spectral HU curve (**f**) are shown. At 40 keV (**c**), the contrast of the anterior mediastinal mass (white arrows) and blood vessels is enhanced compared with 70 keV (**d**) and 100 keV (**e**). The spectral HU curve (**f**) indicates a change of CT values at each keV level in the tumor (white line) and the aorta (yellow line). The pink line indicates the 70 keV energy level. Iodine maps at a 120-second delay (**g**, color map; h, gray scale map) show an iodine concentration of 1.5 mg/mL in the region of interest.

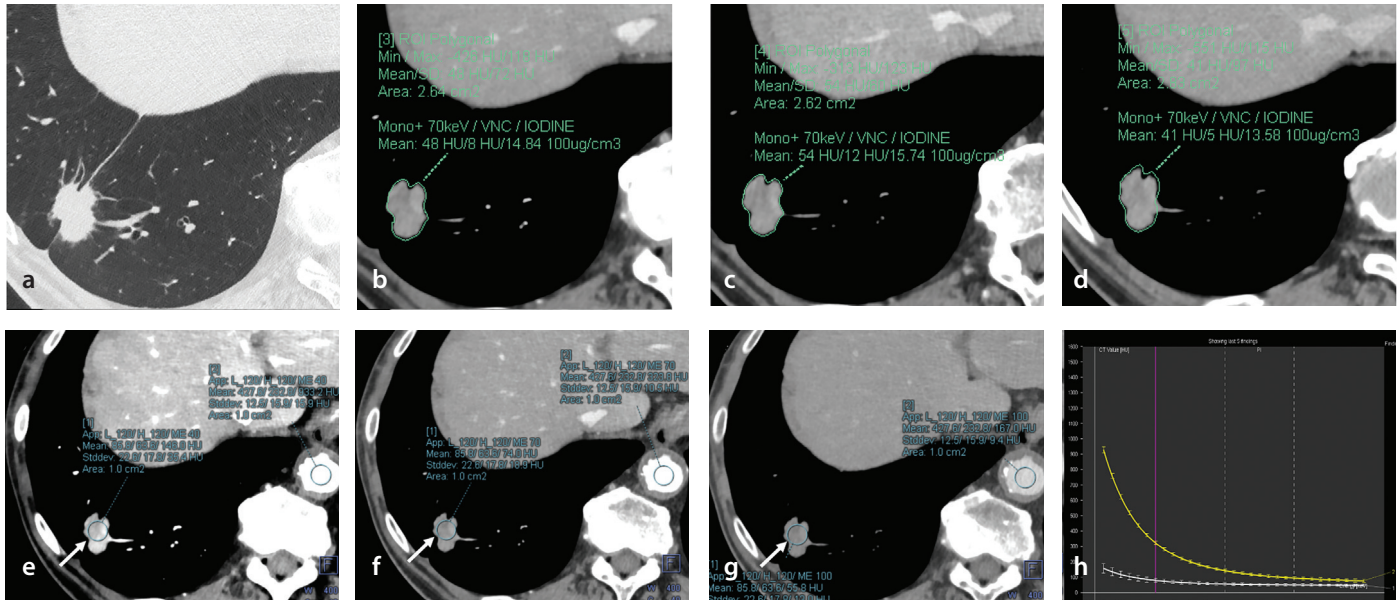



**Figure 5.** Photon-counting detector computed tomography (CT) images in an 80-year-old man with lung adenocarcinoma. (**a**) Shows the lung window CT images (matrix size, 512; and slice thickness, 0.4 mm). (**b-d**) Shows the mediastinal window CT images obtained at 60 seconds (**b**), 120 seconds (**c**), and 180 seconds (**d**) after contrast administration. At 60 seconds (**b**), the mean CT value of the nodule is 48 Hounsfield units (HU), and the iodine concentration (IC) is 1.484 mg/cm$^3$ (originally expressed in units of 100 μg/cm$^3$). At 120 seconds (**c**), the mean CT value is 54 HU, with an IC of 1.574 mg/cm$^3$. At 180 seconds (d), the mean CT value is 41 HU, and the IC is 1.358 mg/cm$^3$. Monochromatic images reconstructed of 40 keV (**e**), 70 keV (**f**), 100 keV (**g**) from the 60-second delayed scan, and spectral HU curve (**h**) are shown. At 40 keV (**e**), the contrast of the lung nodule (white arrows) and blood vessels is enhanced compared with 70 keV (**f**) and 100 keV (**g**). The spectral HU curve (**h**) indicates a change of CT values at each keV level in the nodule (white line) and the aorta (yellow line). The pink line indicates the 70 keV energy level.

DECT has emerged as a valuable tool in thoracic oncology, with iodine-based quantification playing a key role in its diagnostic utility. For example, Deng et al.[61] demonstrated that DECT-derived spectral parameters, including IC, effective atomic number, and spectral slope, enabled differentiation between invasive thymic epithelial tumors and mediastinal lung cancers with high diagnostic accuracy (AUC: 0.88). Similarly, Takumi et al.[62] reported that IC and extracellular volume fraction, measured during the equilibrium phase using dual-layer DECT, were significantly elevated in thymic carcinoma. These findings support their usefulness in tumor subtyping. Additionally, DECT-based iodine mapping has been utilized for nodal staging. Huang et al.[63] showed that IC measurements correlated well with histopathologic nodal status in patients with lung adenocarcinoma, suggesting a role for spectral imaging in non-invasive lymph node evaluation. Building upon these advances, PCD-CT offers further improvements through its inherently higher spectral and spatial resolution. In a phantom study of mediastinal lesions, Centen et al.[64] demonstrated that PCD-CT could detect ICs as low as 0.238 mg/mL in 5 mm lesions when using low-keV virtual monoenergetic imaging and a high-resolution matrix. This performance significantly surpassed that of conventional CT, even when the radiation dose was reduced by 66%. These dose-saving and contrast-optimization capabilities not only improve diagnostic safety but also align with the principles of Green Radiology, which emphasize environmentally sustainable and patient-centered imaging practices. Furthermore, Vrbaski et al.[65] showed that PCD-CT achieved more accurate iodine quantification than DECT under low-dose conditions. The bias observed with PCD-CT was substantially lower, indicating a higher level of quantitative precision. Collectively, these studies highlight the growing importance of spectral CT imaging in thoracic oncology and suggest that PCD-CT is a promising next-generation imaging modality. With its excellent energy discrimination, improved image fidelity, and dose efficiency, PCD-CT will be a powerful tool for tumor characterization, lymph node evaluation, and personalized treatment planning in future clinical settings.

In addition to quantitative improvements, PCD-CT offers high spectral resolution by capturing individual photon energies, enabling advanced imaging techniques such as K-edge imaging. This method takes advantage of the sudden increase in photoelectric absorption at the K-edge of specific high-Z elements (e.g., gold, gadolinium, iodine), allowing their selective detection and quantification. In a foundational phantom study, Si-Mohamed et al.[66] demonstrated the feasibility of multi-material decomposition using spectral PCD-CT. They achieved accurate quantification and discrimination of mixed contrast agents—including iodine, gadolinium, and gold nanoparticles—with high linearity (Pearson correlation coefficient ≥0.97) and low cross-contamination (root mean square error ≤0.47 mg/mL). This "multi-color" imaging was unattainable with conventional CT or DECT, highlighting the advantage of PCD-CT for detecting multiple agents within a single scan.[66] However, a practical limitation of K-edge imaging with PCD-CT is that clinically approved doses of contrast agents, such as gadolinium, may not produce a sufficient SNR for accurate material decomposition.[67] Phantom studies have generally required higher-than-clinical concentrations to achieve reliable K-edge detection. This raises concerns about the feasibility of directly applying these protocols in clinical practice. Although K-edge imaging with PCD-CT has shown clear feasibility and high quantitative accuracy in phantom and preclinical studies, clinical translation remains limited. This is mainly due to the lack of contrast agents specifically approved for K-edge imaging and the current need for elevated doses to ensure adequate image quality. At present, no clinical studies have reported the use of K-edge imaging in human patients.

### Summary and future directions

PCD-CT is redefining thoracic oncologic imaging by delivering unmatched spatial and spectral resolution, improved diagnostic confidence, and substantial reductions in radiation dose. Its ability to detect individual X-ray photons and measure their energy enables more accurate tumor characterization, earlier detection of small lesions, and advanced spectral imaging. These strengths are further enhanced by AI-based reconstruction techniques that improve image quality while maintaining diagnostic precision. PCD-CT is expected to play a central role in the future of precision imaging. Techniques such as K-edge imaging and multi-material decomposition may offer new insights into tumor biology, although further work is needed to adapt K-edge contrast agents for routine clinical use. Moreover, PCD-CT is well aligned with the concept of Green Radiology, which promotes safe, sustainable, and patient-centered imaging practices. As Rockall et al.[68] have noted, radiology must transition toward low-carbon and climate-resilient systems, with innovations that reduce emissions and environmental impact while preserving the quality of care. PCD-CT exemplifies this shift by delivering high-quality images at lower radiation doses and reduced contrast volume. It is particularly valuable in screening and long-term surveillance, where cumulative exposure is a concern. Future directions include broader clinical validation, integration with AI-based workflows, and incorporation into guidelines that reflect diagnostic excellence and environmental stewardship, while taking into account both the advantages of PCD-CT and its challenges, such as high cost and substantial power consumption.

In conclusion, PCD-CT represents not only a technological innovation but also a meaningful step toward more responsible and forward-looking radiologic practice. Its continued development will help ensure that thoracic oncology imaging remains both clinically effective and environmentally conscious.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Pourmorteza A, Symons R, Sandfort V, et al. Abdominal imaging with contrast-enhanced photon-counting CT: first human experience. *Radiology*. 2016;279(1):239-245. [Crossref]
2. Willemink MJ, Persson M, Pourmorteza A, Pelc NJ, Fleischmann D. Photon-counting CT: technical principles and clinical prospects. *Radiology*. 2018;289(2):293-312. [Crossref]
3. Nakamura Y, Higaki T, Kondo S, Kawashita I, Takahashi I, Awai K. An introduction to photon-counting detector CT (PCD CT) for radiologists. *Jpn J Radiol*. 2023;41(3):266-282. [Crossref]
4. Higaki F, Hiramatsu M, Yasuhara T, et al. Cranial and spinal computed tomography (CT) angiography with photon-counting detector CT: comparison with angiographic and operative findings. *Jpn J Radiol*. 2025;43(2):143-151. [Crossref]
5. Flohr T, Schmidt B. Technical basics and clinical benefits of photon-counting CT. *Invest Radiol*. 2023;58(7):441-450. [Crossref]
6. Fletcher JG, Inoue A, Bratt A, et al. Photon-counting CT in thoracic imaging: early clinical evidence and incorporation into clinical practice. *Radiology*. 2024;310(3):e231986. [Crossref]

7. Si-Mohamed S, Boccalini S, Rodesch PA, et al. Feasibility of lung imaging with a large field-of-view spectral photon-counting CT system. *Diagn Interv Imaging*. 2021;102(5):305-312. [Crossref]
8. Symons R, Cork TE, Sahbaee P, et al. Low-dose lung cancer screening with photon-counting CT: a feasibility study. *Phys Med Biol*. 2017;62(1):202-213. [Crossref]
9. Bartlett D, Koo CW, Bartholmai BJ, et al. High-resolution chest computed tomography imaging of the lungs: impact of 1024 matrix reconstruction and photon-counting detector computed tomography. *Invest Radiol*. 2019;54(3):129-137. [Crossref]
10. Schiebler ML, Jinzaki M, Yanagawa M, et al. Future applications of cardiothoracic CT. *Radiology*. 2025;315(3):e240085. [Crossref]
11. Hata A, Yanagawa M, Ninomiya K, et al. Photon-counting detector CT radiological-histological correlation in cadaveric human lung nodules and airways. *Invest Radiol*. 2025;60(2):151-160. [Crossref]
12. Inoue A, Johnson TF, Walkoff LA, et al. Lung cancer screening using clinical photon-counting detector computed tomography and energy-integrating-detector computed tomography: a prospective patient study. *J Comput Assist Tomogr*. 2023;47(2):229-235. [Crossref]
13. Prayer F, Kienast P, Strassl A, et al. Detection of post-COVID-19 lung abnormalities: photon-counting CT versus same-day energy-integrating detector CT. *Radiology*. 2023;307(1):e222087. [Crossref]
14. Inoue A, Johnson TF, White D, et al. Estimating the clinical impact of photon-counting-detector CT in diagnosing usual interstitial pneumonia. *Invest Radiol*. 2022;57(11):734-741. [Crossref]
15. Yanagawa M, Han J, Wada N, et al. Advances in concept and imaging of interstitial lung disease. *Radiology*. 2025;315(2):e241252. [Crossref]
16. Pannenbecker P, Huflage H, Granz JP, et al. Photon-counting CT for diagnosis of acute pulmonary embolism: potential for contrast medium and radiation dose reduction. *Eur Radiol*. 2023;33(11):7830-7839. [Crossref]
17. Hattori A, Suzuki K, Matsunaga T, et al. Is limited resection appropriate for radiologically "solid" tumors in small lung cancers? *Ann Thorac Surg*. 2012;94(1):212-215. [Crossref]
18. Hattori A, Matsunaga T, Takamochi K, Oh S, Suzuki K. Importance of ground glass opacity component in clinical stage IA radiologic invasive lung cancer. *Ann Thorac Surg*. 2017;104(1):313-320. [Crossref]
19. Koike H, Ashizawa K, Tsutsui S, et al. Surgically resected lung adenocarcinoma: do heterogeneous GGNs and part-solid nodules on thin-section CT show different prognosis? *Jpn J Radiol*. 2023;41(2):164-171. [Crossref]
20. Sasaki T, Kuno H, Nomura K, et al. CZT-based photon-counting-detector CT with deep-learning reconstruction: image quality and diagnostic confidence for lung tumor assessment. *Jpn J Radiol*. 2025;43(7):1132-1144. [Crossref]
21. Wang J, Huang Z, Zhu Z, et al. Photon-counting detector CT provides superior subsolid nodule characterization compared to same-day energy-integrating detector CT. *Eur Radiol*. 2025;35(6):2979-2989. [Crossref]
22. Prakash P, Kalra MK, Ackman JB, et al. Diffuse lung disease: CT of the chest with adaptive statistical iterative reconstruction technique. *Radiology*. 2010;256(1):261-269. [Crossref]
23. Funama Y, Taguchi K, Utsunomiya D, et al. Combination of a low-tube-voltage technique with hybrid iterative reconstruction (iDose) algorithm at coronary computed tomographic angiography. *J Comput Assist Tomogr*. 2011;35(4):480-485. [Crossref]
24. Gervaise A, Osemont B, Lecocq S, et al. CT image quality improvement using adaptive iterative dose reduction with wide-volume acquisition on 320-detector CT. *Eur Radiol*. 2012;22(2):295-301. [Crossref]
25. Katsura M, Matsuda I, Akahane M, et al. Model-based iterative reconstruction technique for ultralow-dose chest CT: comparison of pulmonary nodule detectability with the adaptive statistical iterative reconstruction technique. *Invest Radiol*. 2013;48(4):206-212. [Crossref]
26. Yanagawa M, Gyobu T, Leung AN, et al. Ultra-low-dose CT of the lung: effect of iterative reconstruction techniques on image quality. *Acad Radiol*. 2014;21(6):695-703. [Crossref]
27. Araki S, Kitagawa K, Kokawa T, et al. Radiation exposure in cardiac computed tomography imaging in Mie prefecture in 2021. *Jpn J Radiol*. 2023;41(6):596-604. [Crossref]
28. Hamabuchi N, Ohno Y, Kimata H, et al. Effectiveness of deep learning reconstruction on standard to ultra-low-dose high-definition chest CT images. *Jpn J Radiol*. 2023;41(12):1373-1388. [Crossref]
29. Hosokawa T, Kawakami H, Tanabe Y, et al. Left atrial strain assessment using cardiac computed tomography in patients with hypertrophic cardiomyopathy. *Jpn J Radiol*. 2023;41(8):843-853. [Crossref]
30. Murota M, Norikane T, Yamamoto Y, et al. An analysis of the left top pulmonary vein and comparison with the right top pulmonary vein for lung resection by three-dimensional CT angiography and thin-section images. *Jpn J Radiol*. 2023;41(9):965-972. [Crossref]
31. Fukamatsu F, Yamada A, Yamada K, et al. Serial assessment of computed tomography angiography for pulmonary and systemic arteries using a reduced contrast agent dose for the diagnosis of systemic artery-to-pulmonary artery shunts. *Jpn J Radiol*. 2024;42(5):460-467. [Crossref]
32. Kobayashi T, Kunihiro Y, Uehara T, Tanabe M, Ito K. Volume changes of diseased and normal areas in progressive fibrosing interstitial lung disease on inspiratory and expiratory computed tomography. *Jpn J Radiol*. 2024;42(8):832-840. [Crossref]
33. Ohno Y, Aoki T, Endo M, et al. Machine learning-based computer-aided simple triage (CAST) for COVID-19 pneumonia as compared with triage by board-certified chest radiologists. *Jpn J Radiol*. 2024;42(3):276-290. [Crossref]
34. Frings M, Welsner M, Mousa C, et al. Low-dose high-resolution chest CT in adults with cystic fibrosis: intraindividual comparison between photon-counting and energy-integrating detector CT. *Eur Radiol Exp*. 2024;8(1):105. [Crossref]
35. Graafen D, Emrich T, Halfmann MC, et al. Dose reduction and image quality in photon-counting detector high-resolution computed tomography of the chest: routine clinical data. *J Thorac Imaging*. 2022;37(5):315-322. [Crossref]
36. Horst KK, Hull NC, Thacker PG, et al. Pilot study to determine whether reduced-dose photon-counting detector chest computed tomography can reliably display Brody II score imaging findings for children with cystic fibrosis at radiation doses that approximate radiographs. *Pediatr Radiol*. 2023;53(6):1049-1056. [Crossref]
37. Bayfield KJ, Weinheimer O, Boyton C, et al. Implementation and evaluation of ultra-low dose CT in early cystic fibrosis lung disease. *Eur Respir J*. 2023;62(1):2300286. [Crossref]
38. Bosch de Basea M, Thierry-Chef I, Harbron R, et al. Risk of hematological malignancies from CT radiation exposure in children, adolescents and young adults. *Nat Med*. 2023;29(12):3111-3119. [Crossref]
39. Johnson JN, Hornik CP, Li JS, et al. Cumulative radiation exposure and cancer risk estimation in children with heart disease. *Circulation*. 2014;130(2):161-167. [Crossref]
40. National Lung Screening Trial Research Team; Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409. [Crossref]
41. Zensen S, Bos D, Opitz M, et al. Radiation exposure and establishment of diagnostic reference levels of whole-body low-dose CT for the assessment of multiple myeloma with second- and third-generation dual-source CT. *Acta Radiol*. 2022;63(4):527-535. [Crossref]
42. El Bakkali J, Doudouh A. Comparison between InterDosi and MCNP in the estimation of photon SAFs on a series of ICRP pediatric voxelized phantoms. *Jpn J Radiol*. 2023;41(12):1420-1430. [Crossref]
43. Hagen F, Walder L, Fritz J, et al. Image quality and radiation dose of contrast-enhanced chest-CT acquired on a clinical photon-counting detector CT vs. second-

generation dual-source CT in an oncologic cohort: preliminary results. *Tomography*. 2022;8(3):1466-1476. [Crossref]

44. Sawall S, Klein L, Amato C, et al. Iodine contrast-to-noise ratio improvement at unit dose and contrast media volume reduction in whole-body photon-counting CT. *Eur J Radiol*. 2020;126:108909. [Crossref]
45. Zhang M, Nie K, Yuan D, et al. Reduction of radiation dose and contrast medium volume in photon-counting detector computed tomography head and neck angiography: a comparison to energy-integrating detector computed tomography. *Acta Radiol*. 2025;S1076-6332(25)00660-9. [Crossref]
46. Emrich T, O'Doherty J, Schoepf UJ, et al. Reduced iodinated contrast media administration in coronary CT angiography on a clinical photon-counting detector CT system: a phantom study using a dynamic circulation model. *Invest Radiol*. 2023;58(2):148-155. [Crossref]
47. Nakaura T, Ito R, Ueda D, et al. The impact of large language models on radiology: a guide for radiologists on the latest innovations in AI. *Jpn J Radiol*. 2024;42(7):685-696. [Crossref]
48. Jiang B, Li N, Shi X, et al. Deep learning reconstruction shows better lung nodule detection for ultra-low-dose chest CT. *Radiology*. 2022;303(1):202-212. [Crossref]
49. Kawata N, Iwao Y, Matsuura Y, et al. Prediction of oxygen supplementation by a deep-learning model integrating clinical parameters and chest CT images in COVID-19. *Jpn J Radiol*. 2023;41(12):1359-1372. [Crossref]
50. Svalkvist A, Fagman E, Vikgren J, et al. Evaluation of deep-learning image reconstruction for chest CT examinations at two different dose levels. *J Appl Clin Med Phys*. 2023;24(3):e13871. [Crossref]
51. Tang R, Li J, Zhao P, et al. Utility of machine learning for identifying stapes fixation on ultra-high-resolution CT. *Jpn J Radiol*. 2024;42(1):69-77. [Crossref]
52. Nagayama Y, Emoto T, Kato Y, et al. Improving image quality with super-resolution deep-learning-based reconstruction in coronary CT angiography. *Eur Radiol*. 2023;33(12):8488-8500. [Crossref]
53. Nakamoto A, Onishi H, Ota T, et al. Contrast-enhanced thin-slice abdominal CT with super-resolution deep learning reconstruction technique: evaluation of image quality and visibility of anatomical structures. *Jpn J Radiol*. 2025;43(3):445-454. [Crossref]
54. Yanagawa M, Tsubamoto M, Satoh Y, et al. Lung adenocarcinoma at CT with 0.25-mm section thickness and a 2048 matrix: high-spatial-resolution imaging for predicting invasiveness. *Radiology*. 2020;297(2):462-471. [Crossref]
55. Ninomiya K, Yanagawa M, Tsubamoto M, et al. Prediction of solid and micropapillary components in lung invasive adenocarcinoma: radiomics analysis from high-spatial-resolution CT data with 1024 matrix. *Jpn J Radiol*. 2024;42(6):590-598. [Crossref]
56. Kim D, Ahn C, Kim JH. Impact of deep learning 3D CT super-resolution on AI-based pulmonary nodule characterization. *Tomography*. 2025;11(2):13. [Crossref]
57. Xing X, Li L, Sun M, et al. Deep-learning-based 3D super-resolution CT radiomics model: predict the possibility of the micropapillary/solid component of lung adenocarcinoma. *Heliyon*. 2024;10(13):e34163. [Crossref]
58. Wang S, Liu X, Jiang C, et al. CT-based super-resolution deep learning models with attention mechanisms for predicting spread through air spaces of solid or part-solid lung adenocarcinoma. *Acad Radiol*. 2024;31(6):2601-2609. [Crossref]
59. Flohr TG, McCollough CH, Bruder H, et al. First performance evaluation of a dual-source CT (DSCT) system. *Eur Radiol*. 2006;16(2):256-268. Erratum in: *Eur Radiol*. 2006;16(6):1405. [Crossref]
60. Karçaaltıncaba M, Aktaş A. Dual-energy CT revisited with multidetector CT: review of principles and clinical applications. *Diagn Interv Radiol*. 2011;17(3):181-194. [Crossref]
61. Deng L, Yang J, Jing M, et al. Differentiating invasive thymic epithelial tumors from mediastinal lung cancer using spectral CT parameters. *Jpn J Radiol*. 2023;41(9):973-982. [Crossref]
62. Takumi K, Nagano H, Myogasako T, et al. Feasibility of iodine concentration and extracellular volume fraction measurement derived from the equilibrium phase dual-energy CT for differentiating thymic epithelial tumors. *Jpn J Radiol*. 2023;41(1):45-53. [Crossref]
63. Huang HC, Huang YS, Chang YC, et al. Dual energy computed tomography for evaluating nodal staging in lung adenocarcinoma: correlation with surgical pathology. *Jpn J Radiol*. 2024;42(6):468-475. [Crossref]
64. Centen JR, Greuter MJW, Prokop M. Detectability of iodine in mediastinal lesions on photon counting CT: a phantom study. *Diagnostics (Basel)*. 2025;15(6):696. [Crossref]
65. Vrbaski S, Bache S, Rajagopal J, Samei E. Quantitative performance of photon-counting CT at low dose: virtual monochromatic imaging and iodine quantification. *Med Phys*. 2023;50(9):5421-5433. [Crossref]
66. Si-Mohamed S, Bar-Ness D, Sigovan M, et al. Multicolour imaging with spectral photon-counting CT: a phantom study. *Eur Radiol Exp*. 2018;2(1):34. [Crossref]
67. Kravchenko D, Gnasso C, Schoepf UJ, et al. Gadolinium-based coronary CT angiography on a clinical photon-counting-detector system: a dynamic circulating phantom study. *Eur Radiol Exp*. 2024;8(1):118. [Crossref]
68. Rockall AG, Allen B, Brown MJ, et al. Sustainability in radiology: position paper and call to action from ACR, AOSR, ASR, CAR, CIR, ESR, ESRNM, ISR, IS3R, RANZCR, and RSNA. *Radiology*. 2025;314(3):e250325. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263936

CHEST IMAGING

ORIGINAL ARTICLE



# Diagnostic value of portable handheld digital radiography in a beagle model of thoracic trauma

Gaofeng Liu[1]
Xueqi Fang[2]
Sujuan Cui[1]
Li Zhou[1]
Yong Zhang[1]
Qingyuan Li[1]
Xingjian Fang[1]
Yutao Yan[3]

[1]The 988th Hospital of PLA Joint Logistics Support Force, Department of Cardiothoracic Surgery, Zhengzhou, China

[2]Medical Service Training Center, 988th Hospital of PLA Joint Logistic Support Force, Zhengzhou, China

[3]The 988th Hospital of PLA Joint Logistics Support Force, Department of Imaging Medicine, Zhengzhou, China

**PURPOSE**

To assess the diagnostic value of portable handheld digital radiography (DR) in a beagle model of thoracic trauma.

**METHODS**

Twenty-seven beagles were randomly assigned to three experimental groups: the pneumothorax group (induced by intrapleural air injection at 50 mL/kg), the pleural effusion group (induced by intrapleural normal saline injection at 30 mL/kg), and the rib fracture group (created by surgical transection of the 6th rib). All animals underwent three imaging examinations in a randomized order: conventional chest X-ray, mobile DR, and portable handheld DR. Detection rates (using surgical outcomes as the gold standard), image quality, and total examination time (from positioning to image acquisition) were compared among the three modalities.

**RESULTS**

A total of 21 animals completed the full protocol (7 per group). There was no significant difference in detection rates among the three examination methods ($P > 0.050$). The image quality of both mobile DR and portable handheld DR was significantly superior to that of conventional chest X-ray ($P = 0.021$). Examination times for mobile DR (8.37 ± 0.80 minutes) and portable handheld DR (7.07 ± 0.67 minutes) were significantly shorter than for conventional chest X-ray (10.40 ± 0.96 minutes) ($P < 0.001$). Furthermore, portable handheld DR had a significantly shorter examination time than mobile DR ($P < 0.05$).

**CONCLUSION**

Portable handheld DR provides detection rates comparable with mobile DR and conventional chest X-ray for thoracic trauma, with the advantages of superior image quality over conventional chest X-ray and the shortest examination time. Its user-friendly operation and high portability make it a valuable tool for emergency imaging in austere environments.

**CLINICAL SIGNIFICANCE**

Based on diagnostic results in a beagle model of thoracic trauma, this study demonstrates that portable handheld DR can provide reliable methodological support for real-time imaging and rapid triage in harsh environments such as field and post-disaster settings.

**KEYWORDS**

Portable handheld digital radiography, thoracic trauma, emergency rescue, diagnostic value, canine model

**Handling editor:** Sonay Aydın

**Corresponding author:** Yutao Yan

**E-mail:** hnsyt66@163.com

Received 12 February 2026; revision requested 23 February 2026; accepted 21 March 2026.

Epub: 06.04.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263936

You may cite this article as: Liu G, Fang X, Cui S, et al. Diagnostic value of portable handheld digital radiography in a beagle model of thoracic trauma. *Diagn Interv Radiol*. 2026;32(4):448-452.

Thoracic trauma often leads to varying degrees of damage to multiple organs and systems, presenting as critical and complex conditions with a high mortality rate. Because it frequently involves compound injuries, symptoms can easily mask each other, posing major challenges for diagnosis and treatment.[1] Such injuries may include rib fractures, pulmonary contusions, and pneumothorax, characterized by rapid onset and progression and often involving multiple injuries.[2] If not diagnosed and treated promptly, they can lead to severe complications such as massive hemorrhage, infection, or shock. Therefore, rapid and early diagnosis, along with effective treatment, are critical for saving lives, and there is an urgent need for quicker and more efficient diagnostic methods to improve treatment success rates.

Digital radiography (DR) is a commonly used conventional imaging technique widely applied in the diagnosis of acute thoracic trauma. It features low cost, fast examination speed, clear images, good contrast, simple operation, and low radiation dose. It can assist in diagnosis through various image post-processing functions.[3] However, in special scenarios such as field rescue, emergency response at primary healthcare facilities, and disaster sites in remote areas, large-scale equipment such as chest DR and computed tomography scanners is bulky and immobile, relying on fixed power sources and professional operating environments, which makes them difficult to rapidly deploy and use for examinations.[4] Although conventional DR is relatively common in clinical applications, its limited portability restricts its widespread use in emergency medical rescue. In contrast, portable DR considerably enhances equipment mobility and utilization efficiency, opening up a new direction for the development of emergency medical rescue equipment. In this context, how to utilize accessible small-scale portable imaging devices on-site to achieve rapid preliminary diagnosis has become a critical issue in improving the efficiency of emergency medical treatment.

This study systematically investigates the diagnostic value of portable handheld DR in acute thoracic trauma using a beagle model, with the primary objective of delivering efficient and convenient imaging support for battlefield casualties with thoracic trauma, thereby enhancing diagnostic capabilities for injured patients.

**Main points**

- This study directly compares portable handheld digital radiography (DR), mobile DR, and conventional chest X-ray in beagle models of pneumothorax, pleural effusion, and rib fracture.
- Key findings demonstrate that although all three methods show comparable diagnostic accuracy, portable handheld DR offers significantly superior image quality compared with conventional chest X-ray and the shortest examination time among all modalities.
- These results confirm the high clinical value of portable handheld DR. Its combination of diagnostic efficacy, speed, and portability provides crucial support for emergency imaging in challenging environments, such as disaster zones and field medicine.

## Methods

### Experimental animals

A total of 27 beagles aged 1–2 years and weighing 12–15 kg were purchased from the Henan Provincial Experimental Animal Center (Zhengzhou, China). All experimental procedures were conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals. The animals were housed individually in a barrier facility under controlled environmental conditions: temperature 22 °C ± 2 °C, relative humidity 50%–60%, and a 12-hour light–dark cycle. They had free access to water and were fed a standard canine diet at regular times daily. A 7-day acclimatization period was provided before the experiment.

The sample size of 27 beagles was determined based on animal availability and adherence to the 3R principles. This sample size was consistent with those used in similar exploratory studies in the field of veterinary imaging.

### Experimental methods

The experimental protocol was approved by the Ethics Committee of the 988th Hospital of the Joint Logistic Support Force of the People's Liberation Army (decision number: 988YY20240033LLSP, date: March 6, 2024). All experimental procedures strictly complied with the Regulations for the Administration of Laboratory Animals and the Animal Research: Reporting of *In Vivo* Experiments guidelines to minimize animal suffering.

### Animal model establishment

Pneumothorax model (n = 9): Beagles were intravenously anesthetized with propofol (6–8 mg/kg) and underwent endotracheal intubation, with anesthesia maintained using 1.5%–2.5% isoflurane. During the procedure, heart rate, blood pressure, respiratory rate, oxygen saturation, electrocardiogram (ECG), and body temperature were continuously monitored. Butorphanol (0.2–0.4 mg/kg) was administered for analgesia. The beagles were placed in the left lateral decubitus position. A puncture drainage catheter was used to perform thoracentesis in the right sixth intercostal space. Filtered air was slowly injected into the pleural cavity at a volume of 3–5 mL/kg using a 50-mL syringe. After catheter removal, the puncture site was sealed with adhesive tape.

Pleural effusion model (n = 9): Beagles were intravenously anesthetized with propofol (6–8 mg/kg) and underwent endotracheal intubation, with anesthesia maintained using 1.5%–2.5% isoflurane. Intraoperative monitoring included heart rate, blood pressure, respiratory rate, oxygen saturation, ECG, and body temperature. Butorphanol (0.2–0.4 mg/kg) was administered for analgesia. The beagles were positioned in the left lateral decubitus position. Thoracentesis was performed in the right sixth intercostal space using a puncture drainage catheter. Normal saline was slowly injected into the pleural cavity at a volume of 3–5 mL/kg using a 50-mL syringe. The catheter was then removed, and the puncture site was sealed with adhesive tape.

Rib fracture model (n = 9): Beagles were intravenously anesthetized with propofol (6–8 mg/kg) and underwent endotracheal intubation, with anesthesia maintained using 1.5%–2.5% isoflurane. Vital signs, including heart rate, blood pressure, respiratory rate, oxygen saturation, ECG, and body temperature, were continuously monitored. Butorphanol (0.2–0.4 mg/kg) was administered for perioperative analgesia. The animals were placed in the left lateral decubitus position. The right chest wall was disinfected with povidone-iodine. At approximately 10 cm from the sternal end of the 5th intercostal space, the 4th, 5th, and 6th ribs were transected. Before closing the chest, residual air was evacuated from the right pleural cavity to prevent iatrogenic pneumothorax.

### Humane end points

The following humane end points were pre-established: (1) severe respiratory distress (respiratory rate > 60 breaths/minute or cyanosis); (2) uncontrolled persistent bleeding; (3) inability to stand or cessation of spon-

taneous eating and drinking within 24 hours; and (4) body weight loss exceeding 20%. If any of these end points occurred, the animal was immediately euthanized by an overdose of sodium pentobarbital (100 mg/kg, intravenous injection).

Initially, a total of 27 beagles were included in the experiment, and by the end of the study, 21 animals were successfully modeled. The exclusions were as follows: in the pneumothorax group, two beagles were excluded due to incomplete imaging data; in the pleural effusion group, one beagle died from an anesthetic accident, and one was excluded due to incomplete data; in the rib fracture group, one beagle was excluded due to postoperative wound infection, and one died from an anesthetic accident (Figure 1).

### X-ray examination

All beagles underwent conventional chest X-ray, mobile DR, and portable handheld DR examinations.

Conventional chest X-ray was performed using a F51-5C Medical diagnostic X-ray machine (Wandong, Beijing, China) at 66 kV, 200 mA, with a single exposure time of 0.03 seconds. Mobile DR (DRX-Rise) was performed at 85 kV, 320 mA, with an exposure time of 0.02 seconds. Portable handheld DR (70 kV, 2 mA) was performed with a single exposure time of 0.2 seconds; the device weighed ≤ 2.5 kg (Figure 2). All images were acquired in a standard posteroanterior projection during end-inspiration, with the focal distance set at 150 cm for conventional chest X-ray and mobile DR and at 100 cm for handheld DR. To ensure model stability, the interval between each examination was controlled within 5 minutes.

In strict accordance with standard X-ray examination protocols, conventional posteroanterior projections of the thoracoabdominal region were obtained for all beagles. Following image acquisition, all radiographs were independently reviewed by at least two senior radiologists who were blinded to the imaging modality to avoid observer bias. The reading sequence was randomized for each radiologist, and the two radiologists reviewed the images independently without any communication or mutual influence to ensure the independence of the reading process. To assess interobserver consistency, the kappa (κ) statistic was used: a κ value > 0.75 indicated excellent consistency, 0.40–0.75 indicated moderate to good consistency, and < 0.40 indicated poor consistency. A final diagnosis was established upon consensus between the two radiologists. In cases of discrepancy, a third senior radiologist was invited to review the images jointly to reach a final consensus.

### Observation indicators

### Detection rates

Imaging data from all animals were collected and independently reviewed by two senior radiologists using a double-blind method. Using surgical findings as the gold standard, the number of true-positive detections for each of the three examination methods was recorded, and the detection rate was calculated as follows: detection rate = (number of positive detections/ total number of cases) × 100%. In cases of disagreement between the two radiologists regarding interpretation, a consensus was reached through group discussion to ensure the accuracy and objectivity of the statistical results.

### Grading evaluation of image quality

The evaluation criteria were as follows: Grade A (fully met diagnostic requirements, correct positioning, clear image with optimal contrast, and free from blurring or artifacts); Grade B (met diagnostic requirements, correct positioning, relatively clear image, with minimal blurring or artifacts); Grade C (basically met diagnostic requirements, suboptimal positioning, moderate image clarity and discernibility, with noticeable blurring, artifacts, or noise); non-diagnostic (did not meet diagnostic requirements, poor positioning, severely blurred image, with significant artifacts or noise). The same two radiologists independently graded all images based on these criteria. The proportions of Grade A, Grade B, Grade C, and non-diagnostic images for each of the three examination methods were calculated to compare imaging quality differences across the different devices.

### Examination times

The examination times for the three methods were recorded and compared. Examination time was defined as the entire process from the start of animal positioning to the completion of image acquisition and confirmation that the images met diagnostic quality standards. By analyzing and comparing differences in operational efficiency among the three devices, their applicability in emergency or bedside settings was evaluated.

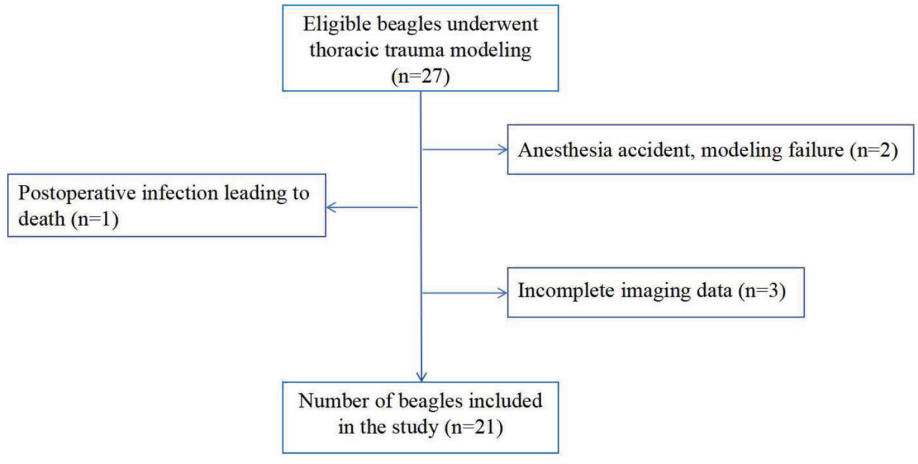



**Figure 1.** Animal model establishment.

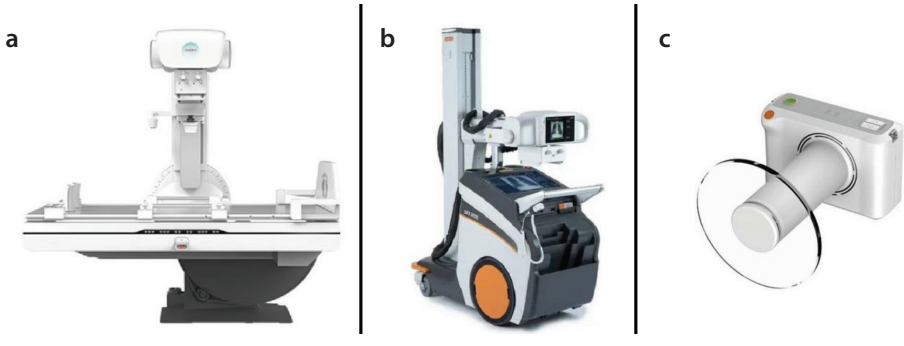



**Figure 2.** Imaging of devices: conventional chest X-ray device **(a)**, mobile DR **(b)**, and a portable handheld DR **(c)**. DR, digital radiography.

### Statistical analysis

Statistical analysis was performed using SPSS version 26.0. Count data were presented as frequencies and percentages [n (%)] and analyzed using Cochran's Q test. Ranked data were assessed using the nonparametric Friedman test. Measurement data were expressed as mean ± standard deviation and analyzed using repeated-measures analysis of variance, with post-hoc pairwise comparisons conducted using the least significant difference t-test. A value of $P < 0.05$ was considered statistically significant.

## Results

### Detection rates of thoracic trauma

The detection rates for thoracic trauma varied across the three examination methods (Table 1). Mobile DR demonstrated the highest detection rate at 85.71%, followed by portable handheld DR (80.95%) and conventional chest X-ray (66.67%). However, despite the apparent numerical advantage of the mobile and portable handheld DR methods, the differences among the three groups did not reach statistical significance ($P > 0.05$). This suggests that although portable handheld DR may offer comparable diagnostic performance, the current sample size may be insufficient to detect a true difference.

### Image quality assessment

Image quality grades for the three examination methods are presented in Table 2. Both mobile DR and portable handheld DR demonstrated significantly higher image quality than conventional chest X-ray. The differences were statistically significant ($P < 0.05$, Table 2). However, the difference in image quality between mobile DR and portable handheld DR was not statistically significant ($P > 0.05$), indicating that both mobile DR and portable handheld DR provided comparable image quality that was superior to conventional chest X-ray.

### Examination time

The total examination times for the three examination methods are presented in Table 3. Conventional chest X-ray had the longest examination time (10.40 ± 0.96 minutes), followed by mobile DR (8.37 ± 0.80 minutes), whereas portable handheld DR was the fastest (7.07 ± 0.67 minutes). The overall differences among the three methods were highly significant ($P < 0.001$). Pairwise comparisons revealed that both mobile DR and portable handheld DR significantly reduced examination time compared with conventional chest X-ray ($P < 0.05$). Furthermore, portable handheld DR was significantly faster than mobile DR ($P < 0.05$).

## Discussion

In this study, the fracture detection rate of portable handheld DR showed no statistically significant difference compared with conventional chest X-ray or mobile DR, and its image quality was comparable to that of conventional DR ($P > 0.05$). However, it demonstrated a significantly faster examination speed. Acute thoracic trauma is often more complex, progresses more rapidly, and is more difficult to diagnose than ordinary trauma, necessitating accurate and efficient diagnostic tools. The resulting critical conditions are frequently life-threatening, and delayed diagnosis and treatment may lead to death from hemorrhagic shock or respiratory failure within a short period.[5] DR technology enables rapid acquisition of high-resolution images without requiring special patient cooperation. Its simple workflow, low radiation dose, and ease of storage substantially streamline emergency care and save valuable time for patient management.

Previous studies have shown that portable handheld DR devices could be used for tuberculosis screening, with image quality comparable with that of digital X-ray machines commonly used in medical facilities. Moreover, portable handheld DR devices were easy to operate and could be transported to any location.[4] Chest X-rays obtained using camera-type portable X-ray devices could be used for clinical evaluation of indwelling instruments.[6] Due to their high mobility and operational simplicity, portable DR devices enable rapid diagnosis in challenging conditions. Their outstanding emergency response capability makes them the preferred imaging modalities for military and disaster rescue scenarios.[7] The present study found no significant difference between portable handheld DR and mobile DR in terms

**Table 1.** Comparison of detection rates among the three examination methods

| Method | Conventional X-ray | Mobile DR | Portable handheld DR | Q | *P* |
|---|---|---|---|---|---|
| Pneumothorax (n = 7) | 5 | 6 | 5 | 1.000 | 0.607 |
| Pleural effusion (n = 7) | 5 | 6 | 6 | 0.500 | 0.779 |
| Rib fracture (n = 7) | 4 | 6 | 6 | 2.000 | 0.368 |
| Missed diagnosis | 7 | 3 | 4 | | |
| Positive detection rate [% (n/m)] | 66.67% (14/21) | 85.71% (18/21) | 80.95% (17/21) | 2.600 | 0.273 |
| 95% CI | (0.447, 0.887) | (0.694, 1.020) | (0.626, 0.810) | | |

DR, digital radiography; CI, confidence interval.

**Table 2.** Comparison of the image quality of the three examination methods

| Method | Grade A | Grade B | Grade C | Non-diagnostic | 95% CI |
|---|---|---|---|---|---|
| Conventional X-ray | 14 | 5 | 2 | 0 | (1.12, 1.74) |
| Mobile DR | 20 | 1 | 0 | 0 | (0.95, 1.15) |
| Portable handheld DR | 19 | 2 | 0 | 0 | (0.96, 1.23) |
| $X^2$ | | | 7.750 | | |
| *P* | | | 0.021 | | |

DR, digital radiography; CI, confidence interval.

**Table 3.** Comparison of examination times among the three examination methods

| Method | Examination time (min) | 95% CI |
|---|---|---|
| Conventional X-ray | 10.40 ± 0.96 | (9.968, 10.845) |
| Mobile DR | 8.37 ± 0.80[a] | (8.008, 8.737) |
| Portable handheld DR | 7.07 ± 0.67[ab] | (6.771, 7.378) |
| F | 89.367 | |
| *P* | <0.001 | |

[a]$P < 0.05$ vs. conventional X-ray; [b]$P < 0.05$ vs. mobile DR. DR, digital radiography; CI, confidence interval.

of fracture detection rate and image quality. Research by Audin et al.[8] also confirmed that wireless portable DR can provide clinical information equivalent or superior to computed radiography in bedside chest radiography in the intensive care unit (ICU). Furthermore, the latest portable dynamic DR technology has expanded the functional applications of bedside chest imaging, offering new tools for the rapid assessment of trauma patients.[9] Collectively, these studies support the clinical value of portable DR in the diagnosis of thoracic trauma, particularly its advantages in specialized settings such as emergency departments, ICUs, and disaster sites. The device features an integrated design with low radiation dose (leakage ≤ 2.4 μGy/hour at 1 minute), high safety, lightweight construction, and palm-sized operability, providing reliable imaging support for emergency scenarios.

This study has several limitations. First, the diagnostic value of portable handheld DR was investigated only in a beagle animal model, and no human clinical studies have been conducted, which limits the generalizability of the results to clinical practice. In addition, the long-term stability and practicality of the device in extreme environments, complex trauma types, and mass-casualty scenarios have not been thoroughly evaluated. Further validation through large-sample, multicenter clinical studies is warranted.

## Conclusion

In summary, portable handheld DR offers simple operation and high portability, enabling rapid deployment in complex emergency settings. While ensuring safety and diagnostic efficacy, it can meet basic imaging needs in first-aid scenarios. Its significant application potential in environments lacking fixed equipment, such as disaster relief sites and field medical rescue operations, provides crucial support for optimizing imaging equipment allocation and emergency procedures in these contexts.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.


### Funding

This study was supported by the Key Discipline Independent Research Project of the 988th Hospital (grant number: YNZX2024006).


## References


1. Dogrul BN, Kiliccalan I, Asci ES, Peker SC. Blunt trauma related chest wall and pulmonary injuries: an overview. *Chin J Traumatol*. 2020;23(3):125-138. [Crossref]
2. Lodhia JV, Eyre L, Smith M, Toth L, Troxler M, Milton RS. Management of thoracic trauma. *Anaesthesia*. 2023;78(2):225-235. [Crossref]
3. Zhang ZL. Analysis and research on the role and value of DR in the diagnosis of rib fractures. *Chin Med Device Inform*. 2023;29:110-112. [Crossref]
4. Kamal R, Singh M, Roy S, et al. A comparison of the quality of images of chest X-ray between handheld portable digital X-ray & routinely used digital X-ray machine. *Indian J Med Res*. 2023;157(2&3):204-210. [Crossref]
5. Yahyoev SMO, Khodzhimatov GM. Diagnostic methods and surgical treatment of patients with thoracoabdominal traumas (review of the literature). *Kardiochir Torakochirurgia Pol*. 2022;19(4):226-231. [Crossref]
6. Kim SJ, Yong HS, Kang EY, Yang Z, Kim JY, Yoon YH. Is a camera-type portable X-Ray device clinically feasible in chest imaging?: Image quality comparison with chest radiographs taken with traditional mobile digital X-Ray devices. *J Korean Soc Radiol*. 2024;85(1):138-146. [Crossref]
7. Ma X, Zhang B, Qi Z, et al. Applicability of field portable digital radiology on the plateau. *Aviation Medicine of Air Force*. 2022;39(2):179-180. [Crossref]
8. Audin CR, Aran S, Muse VV, et al. Bedside chest radiographs in the intensive care setting: wireless direct radiography compared to computed radiography. *Curr Probl Diagn Radiol*. 2018;47(6):397-403. [Crossref]
9. Cè M, Oliva G, Rabaiotti FL, et al. Portable dynamic chest radiography: literature review and potential bedside applications. *Med Sci (Basel)*. 2024;12(1):10. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253814

GENERAL RADIOLOGY

LETTER TO THE EDITOR



# Letter to the editor: Beyond publication rates: improving the quality and impact of radiology residency research

Hakan Ayyıldız[1]
Şükrü Mehmet Ertürk[2]

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

[2]İstanbul University, İstanbul Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

**Dear Editor,**

I congratulate Salbas and Koc on their pioneering analysis of Turkish radiology residency theses.[1] It demonstrates improved thesis publication rates but also highlights several ongoing challenges in the academic training ecosystem.

Although 37.1% of theses have been published, only around 20.8% of all theses yielded a Science Citation Index Expanded (SCIE)-indexed article, with just 6.2% in a top-quartile (Q1/Q2) journal.[1] In short, much of the residency research fails to reach high-impact journals.

Another notable finding is the paucity of high-level evidence studies. In total, 86.7% of theses were retrospective, with only 13% prospective.[1] Prospective (especially multicenter) projects are harder to conduct during residency, yet such studies achieve higher publication rates and tend to appear in better journals. This dominance of retrospective work suggests systemic barriers (time, resources, mentorship) limiting more ambitious prospective research.

The study also found disparities by institution and authorship. University hospital theses received significantly more citations than those from training hospitals, indicating an academic resource gap.[1] Moreover, although residents were first authors on 76.4% of papers, these papers took longer to publish and garnered fewer citations, whereas papers with advisors as the first author more often appeared in SCIE journals and were published faster.[1] This finding underscores the critical role of active senior mentorship and guidance throughout the publication process, as demonstrated in prior research.[2]

Important subspecialties are underrepresented: only 1.6% of theses focused on imaging physics and radiation safety and just 1.6% on obstetric radiology.[1] Limited resident exposure and reliance on other departments for certain services (e.g., obstetric ultrasound) likely contribute to this gap. However, this is concerning, as physics and radiation safety are fundamental for imaging quality, and inadequate obstetric imaging training can affect patient care.[3-5] These niche but critical fields deserve greater attention in thesis research.

I propose several remedies: 1) implement national mentorship and writing workshops for residents to strengthen research design and manuscript skills—experienced supervision is associated with higher thesis publication success;[6] 2) incentivize prospective and multicenter thesis projects (via funding, protected time, or academic credit), as rigorous studies are more likely to yield high-impact publications; 3) establish a national thesis registry or standards to ensure methodological quality, avoid duplication, and foster multicenter collaborations. These steps would help elevate the scholarly output and impact of residency research.

In conclusion, this timely analysis by Salbas and Koc[1] illuminates the progress and challenges in radiology residency research in Türkiye. By strengthening mentorship, promoting prospective multi-institution studies, and instituting national quality standards, we can enhance the academic output of future radiologists and their contributions to the literature. Future studies could also explore additional individual-level factors, such as thesis length, residents' prior publication experience, time since medical school graduation, and residents' age, which may further influence thesis-to-publication conversion and scientific impact. Un-

**Handling editor:** Mehmet Ruhi Onur

**Corresponding author:** Hakan Ayyıldız

**E-mail:** hakanayyildiz77@gmail.com

Received 17 December 2025; accepted 23 December 2025.

Epub: 04.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253814

You may cite this article as: Ayyıldız H, Ertürk ŞM. Letter to the editor: beyond publication rates: improving the quality and impact of radiology residency research. *Diagn Interv Radiol*. 2026;32(4):453-454.

derstanding these factors may help tailor mentorship strategies and optimize research training during residency.

### Conflict of interest disclosure

Şükrü Mehmet Ertürk was not serving in any editorial role at the time of the manuscript's submission, evaluation, acceptance, or ahead of print publication. At the time of inclusion of the article in the regular issue, he is serving as Editor-in-Chief of the journal. Şükrü Mehmet Ertürk was not involved in the editorial evaluation, peer-review process, or publication decision of this manuscript. The other author declare no conflict of interest.

## References


1. Salbas A, Koc AM. Bibliometric analysis of radiology residency theses in Türkiye: publication metrics and trends. *Diagn Interv Radiol*. 2026;32(1):71-80. [Crossref]
2. Özgen Ü, Eğri M, Aktaş M, et al. Publication pattern of Turkish medical theses: analysis of 22.625 medical theses completed in years 1980-2005. *Turkiye Klinikleri J Med Sci*. 2011;31(5):1122-1131. [Crossref]
3. Hendee WR. Teaching physics to radiology residents. *AJR Am J Roentgenol*. 2009;192(4):855-858. [Crossref]
4. Özkan Ş, Aba G, Tekinsoy B. The importance of radiation safety in terms of hospital administration and research on the awareness stage of radiology technicians. *J Acad Res Med.* 2016;6(3):162-169. [Crossref]
5. Kasales CJ, Coulson CC, Mauger D, Chertoff JD, Matthews A. Training in obstetric sonography for radiology residents and fellows in the United States. *AJR Am J Roentgenol*. 2001;177(4):763-767. [Crossref]
6. Brunod I, Rességuier N, Fabre A. Medical thesis publication and academic productivity of pediatric residents at the Medical University of Marseille: associated factors and evolution over 20 years. *Arch Pediatr*. 2020;27(8):408-415. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263862

GENERAL RADIOLOGY

LETTER TO THE EDITOR



# Reply: Beyond publication rates: improving the quality and impact of radiology residency research

Ali Salbas
Ali Murat Koc

İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Radiology, İzmir, Türkiye

**Dear Editor,**

We thank the authors for their valuable and constructive comments on our article.[1,2] The points highlighted in their letter strongly support one of the primary objectives of our study: improving the academic training ecosystem in radiology.

The high proportion of retrospective studies identified in our research, together with publication trends showing residents as first authors, confirms the need for mentorship mentioned in the letter. In particular, the approximately 5.5-fold increase in radiology residency quotas since 2022 is expected to disproportionately increase the number of residents relative to faculty members.[3] Furthermore, according to data from the Organisation for Economic Co-operation and Development, Türkiye ranked first worldwide in 2024 for the number of magnetic resonance imaging (MRI) and computed tomography examinations per 1,000 population, as well as the number of examinations per MRI scanner.[4] Given this escalating clinical workload, maintaining a sustainable balance between residency education and clinical service obligations is becoming increasingly challenging. This imbalance may be associated with reduced academic productivity and could plausibly affect the quality of scientific output in radiology. Therefore, effective mentorship practices, together with strengthened institutional research infrastructure, are likely to become even more critical in the future.

We suggest that proposals to encourage prospective and multicenter projects, along with the establishment of a national thesis registry and standardized thesis guidelines, would substantially improve the quality of academic output. In addition, addressing the low representation rate (1.6%) in subspecialties such as imaging physics and obstetric radiology should be a priority in residency training policies. These fields are essential for ensuring radiation protection, quality assurance, and maternal–fetal safety. Notably, our study found that despite being the least chosen topic, imaging physics and radiation safety had the highest publication rate, at 51.0% (25/49).[1] This suggests considerable academic potential in this area, possibly related to the availability of standardized datasets and objective measurement parameters in physics-based research. Residents should be actively encouraged to pursue research in this area.

In conclusion, we thank the authors once again for their insightful contributions. We believe that such dialogue will meaningfully enhance the academic quality and discourse within the Turkish radiology community. Their suggested parameters for future research are highly valuable for expanding the existing literature. Finally, we reiterate that residency theses should not be viewed merely as graduation requirements, but as foundational opportunities to foster scientific thinking, research skills, and sustained academic engagement.

**Handling editor:** Mehmet Ruhi Onur

**Corresponding author:** Ali Salbas

**E-mail:** dralisalbas@gmail.com

Received 07 January 2026; revision requested 01 February 2026; accepted 02 February 2026.

Epub: 19.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263862

You may cite this article as: Salbas A, Koc AM. Reply: beyond publication rates: improving the quality and impact of radiology residency research. *Diagn Interv Radiol*. 2026;32(4):455-456.

## Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Salbas A, Koc AM. Bibliometric analysis of radiology residency theses in Türkiye: publication metrics and trends. Diagn Interv Radiol. 2026;32(1):71-80. [Crossref]
2. Ayyıldız H, Ertürk ŞM. Letter to the editor: beyond publication rates: improving the quality and impact of radiology residency research. Diagn Interv Radiol. 2026;32(4):453-454. [Crossref]
3. ÖSYM. Student Selection and Placement Center (ÖSYM) – Official Website [Internet]. Ankara: Student Selection and Placement Center of Turkey; c2022. [Crossref]
4. Republic of Türkiye Ministry of Health. Health Statistics Yearbook 2024 [Internet]. Ankara: Ministry of Health; 2025. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253536

HEAD AND NECK IMAGING

ORIGINAL ARTICLE



# T2 signal ratio enhances the diagnostic performance of apparent diffusion coefficient in differentiating orbital lymphoma from inflammatory mimickers

Elif Günay Bulut[1]
Ekim Gümeler[1]
Jale Karakaya[2]
Levent Kılıç[3]
Ömer Karadağ[3]
Kader Karlı Oğuz[4]

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

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

[3]Hacettepe University Faculty of Medicine, Department of Internal Medicine, Division of Rheumatology, Ankara, Türkiye

[4]University of California Davis Medical Center, Department of Radiology, Division of Neuroradiology, California, United States of America

**PURPOSE**

Orbital lymphomas share overlapping clinical and radiological features with immunoglobulin G4-related orbital disease (IgG4-ROD) and granulomatous orbital diseases, which may lead to diagnostic delays. This study aims to evaluate the added diagnostic value of the transverse relaxation time (T2) signal ratio when combined with apparent diffusion coefficient (ADC) measurements in distinguishing orbital lymphoma from inflammatory mimickers.

**METHODS**

In this retrospective study, two blinded radiologists independently measured T2 signal ratios (lesion to cerebral cortex), ADC values, and ADC ratios on pretreatment orbital magnetic resonance imaging scans of 58 patients (21 lymphomas, 21 IgG4-ROD, 16 granulomatous inflammation). Measurements were performed on axial images at the lesion's maximal diameter. Regions of interest were manually drawn to cover the entire lesion, avoiding necrosis and edges. Diagnostic performance was assessed using receiver operating characteristic curve analysis, with optimal cut-off values determined by Youden's index. The ADC ratio (≤1) and T2 signal ratio (>0.88) were combined using OR (either positive) and AND (both positive) rules. Interobserver agreement was evaluated using intraclass correlation coefficients (ICC).

**RESULTS**

All measurements showed statistically significant differences between the two cohorts. Reviewer 1's ADC ratio measurements demonstrated excellent diagnostic performance, with an area under the curve (AUC) of 0.920 (85.7% sensitivity and 86.5% specificity at the optimal cut-off of ≤1). In comparison, T2 signal ratios showed moderate diagnostic value (AUC: 0.726; 80.95% sensitivity and 64.86% specificity at a cut-off >0.88). The combination of both parameters significantly improved diagnostic accuracy: the OR rule (ADC ≤1 or T2 >0.88) increased sensitivity to 95.2%, whereas the AND rule (ADC ≤1 and T2 >0.88) increased specificity to 94.6%. Interobserver reliability was excellent, with ICC values ranging from 0.969 to 0.985.

**CONCLUSION**

Although diffusion imaging remains the primary discriminator for orbital lymphoma, the T2 signal ratio considerably enhances diagnostic confidence, particularly in borderline ADC cases.

**CLINICAL SIGNIFICANCE**

Incorporating T2 signal ratio measurements adds minimal workflow burden in routine clinical practice and provides a practical tool for differentiating lymphoma from IgG4-ROD and granulomatous inflammatory mimics.

**KEYWORDS**

Granulomatous inflammation, immunoglobulin G4-related disease, lymphoma, magnetic resonance imaging, orbit

**Handling editor:** Erkan Gökçe

**Corresponding author:** Elif Günay Bulut

**E-mail:** drelifbulut@yahoo.com

Received 28 June 2025; revision requested 23 July 2025; accepted 08 September 2025.

Epub: 06.10.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253536

You may cite this article as: Bulut EG, Gümeler E, Karakaya J, Kılıç L, Karadağ Ö, Karlı Oğuz K. T2 signal ratio enhances the diagnostic performance of apparent diffusion coefficient in differentiating orbital lymphoma from inflammatory mimickers. *Diagn Interv Radiol*. 2026;32(4):457-462.

Orbital lymphoma is the most common orbital neoplasm in adults, with extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT) being the most frequent histopathological subtype.[1] Differentiating orbital lymphoma from inflammatory mimickers is crucial because treatment implications and prognosis differ. Orbital lymphomas are primarily treated with low-dose radiotherapy (20–30 Gy), whereas combined-modality therapy with systemic agents (e.g., rituximab or chemotherapy) is required for aggressive subtypes. In contrast, inflammatory conditions [e.g., immunoglobulin G4-related orbital disease (IgG4-ROD)] often respond well to corticosteroid therapy.[2,3]

Clinical manifestations are largely non-discriminating, and intraorbital location may interfere with obtaining adequate tissue samples for histopathologic evaluation. Magnetic resonance imaging (MRI) has an established role in preoperative discrimination. Several prior studies have reported descriptive and quantitative MRI features for differentiating orbital lymphoma from inflammatory diseases, including margin characteristics, diffusion, and perfusion patterns.[3-6] Nevertheless, studies comparing lymphoma with specific entities of orbital inflammation remain sparse.[7]

Orbital inflammation is an umbrella term that includes idiopathic orbital inflammation (IOI) and specific entities with established diagnostic criteria, such as IgG4-ROD. IOI can be more readily differentiated from orbital lymphoma both clinically and radiologically, as it typically presents with pain and demonstrates an infiltrative pattern without restricted diffusion.[8] However, IgG4-ROD and granulomatous orbital inflammation share overlapping clinical and radiological features with lymphoma, such as mass-like involvement and relatively low apparent diffusion coefficient (ADC) values.[9-11] Therefore, the role of diffusion-weighted imaging (DWI) in distinguishing orbital lymphoma from these specific inflammatory entities may be limited, necessitating additional imaging discriminators.

IgG4-ROD and granulomatous inflammations usually appear hypointense on transverse relaxation time (T2)-weighted images due to fibrosis and granulomas, respectively.[12] In contrast, orbital lymphomas have mostly been reported to demonstrate iso- or hyperintense T2 signals in previous studies.[1,13] Thus, T2 signal intensity may serve as a useful discriminator between orbital lymphoma and these inflammatory entities.

In this study, we aim to retrospectively assess the diagnostic performance of ADC values, T2 signal intensity measurements, and their combination in differentiating orbital lymphoma from its most challenging mimickers: IgG4-ROD and granulomatous diseases.

**Main points**

- Apparent diffusion coefficient (ADC) values demonstrate superior diagnostic performance for lymphoma detection, but the overlap with immunoglobulin G4-related orbital disease (IgG4-ROD) and granulomatous diseases necessitates additional imaging markers.
- Transverse relaxation time (T2) hypointensity in IgG4-ROD and granulomatous diseases versus intermediate T2 signal intensity in lymphomas provides complementary diagnostic value when ADC findings are equivocal.
- The combined ADC and T2 approach relies on routine magnetic resonance imaging sequences without requiring advanced protocols, offering a clinically practical solution to guide treatment decisions.

## Methods

### Patients

This retrospective study was conducted with approval from the Hacettepe University Health Sciences Research Ethics Board, with a waiver of informed consent (SBA 24/1115/2024, date: 27.11.2024). Orbital MRI reports dictated in the radiology department between January 2016 and January 2024 were searched for the keywords "lymphoma," "intraorbital mass," "inflammation," and "granulomatous". An additional retrospective search was conducted in the rheumatology department database to identify patients diagnosed with orbital IgG4-ROD, granulomatous polyangiitis (GPA), or sarcoidosis.

The inclusion criteria were as follows: 1) patients with an orbital mass or infiltration detected by MRI; 2) patients meeting established diagnostic criteria and/or having a histopathologic diagnosis of IgG4-related disease[14] or granulomatous diseases, including GPA[15] and sarcoidosis;[16] 3) patients with a histopathologic diagnosis of orbital lymphoma; 4) patients who underwent orbital MRI at the onset of orbital disease. A total of 36 patients with orbital lymphoma, 26 patients with IgG4-ROD, 10 patients with GPA, 5 patients with sarcoidosis, and 1 patient with fungal granulomatous angiitis were identified.

The exclusion criteria were as follows: 1) MRI performed after biopsy and/or corticosteroid use, 2) secondary orbital lymphoma, 3) technically inadequate scans, and 4) scans without DWI. A total of 15 patients with orbital lymphoma (14 with secondary lymphoma, 1 with a technically inadequate scan) and 5 patients with IgG4-ROD (3 without DWI, 1 with corticosteroid use, 1 with a technically inadequate scan) were excluded. The final study group (Figure 1) comprised 21 patients with primary orbital lymphoma (9 women and 12 men; age range 9-86 years;

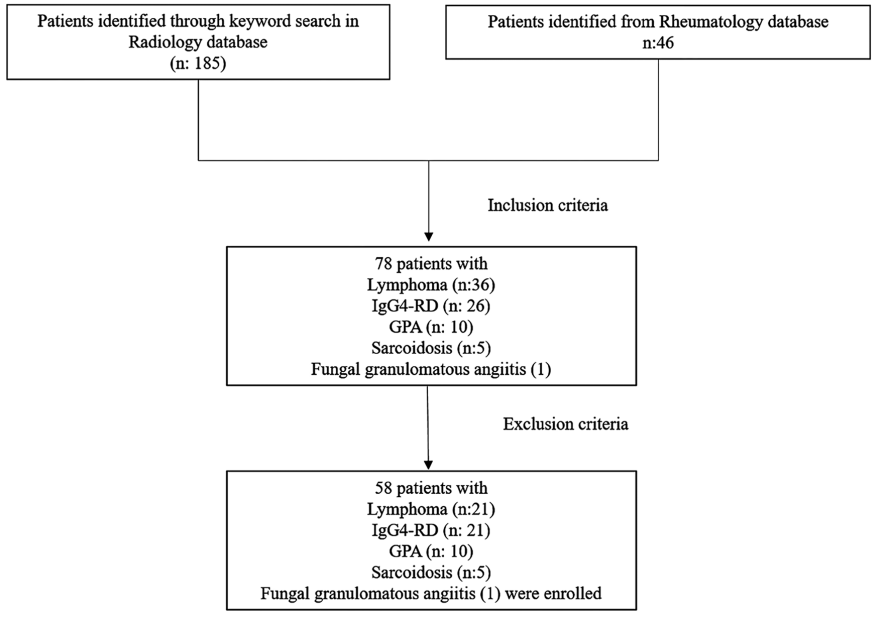



**Figure 1.** Flowchart of patient selection. IgG4-ROD, immunoglobulin G4-related orbital disease; GPA, granulomatous polyangiitis.

median 58.4 years), 21 patients with IgG4-ROD (13 women and 8 men; age range 4-65 years; median 37 years), 10 patients with orbital GPA (7 women and 3 men; age range 11-64 years; median 41.5 years), 5 patients with sarcoidosis (5 women; age range 26-59 years; median 44 years), and 1 patient with fungal granulomatous angiitis (78-year-old woman).

### Magnetic resonance imaging examination

MR studies were performed using either 1.5 T or 3 T MR scanners (Philips Healthcare, Best, the Netherlands; GE Healthcare, Milwaukee, WI, USA; and Siemens, Erlangen, Germany). The orbital MRI protocol included coronal and axial T1-weighted images, coronal and axial T2-weighted images with fat saturation, post-contrast axial and coronal T1-weighted images with fat saturation, and axial DWI.

The parameters for fat-saturated T2-weighted images were as follows: repetition time (TR)/echo time (TE): 5,680–2,720/100–80 ms, field of view (FOV): 215 × 215–120 × 105 mm, slice thickness/gap: 3–4/3.3–4.6 mm, number of excitations (NEX): 2–3, and matrix: 320 × 302–612 × 768. DWI was acquired using single-shot spin-echo echo planar imaging with the following parameters: TR/TE: 6,188–2,400/116–73 ms; slice thickness: 3–5 mm; flip angle: 90°; FOV: 267 × 227–160 × 100 mm; NEX: 2–12; matrix 256 × 256–128 × 128; and b values of 0 and 1,000 sec/mm$^2$.

### Image analysis

The images were reviewed by two head and neck radiologists (Reviewer 1 and Reviewer 2, with 11 and 7 years of experience, respectively), both blinded to the clinical diagnosis. Descriptive features, including laterality and location, were assessed by consensus. When both sides were affected, the larger area of involvement was selected for quantitative analysis. Quantitative analysis included T2 signal and ADC measurements. Measurements were made independently on axial planes at the level of maximal lesion diameter. Regions of interest (ROIs) were drawn manually to cover the entire lesion, avoiding cystic or necrotic areas by referencing post-contrast T1-weighted images. Lesion edges were excluded to reduce the effects of partial volume and susceptibility artifacts. The mean ROI sizes drawn by Reviewers 1 and 2 were as follows: for Reviewer 1, the T2 signal ROI was 2.62 cm$^2$ (range: 0.3–9.38 cm$^2$), and the ADC ROI was 2.56 cm$^2$ (range: 0.16–9.32 cm$^2$). For Reviewer 2, the T2 signal ROI was 2.08 cm$^2$ (range: 0.25–7.36 cm$^2$), and the ADC ROI was 1.84 cm$^2$ (range: 0.23–5.81 cm$^2$).

The ratios of T2 signal (lesion to mean T2 signal cortex) and ADC (lesion to mean ADC cortex) were calculated for each patient. The mean T2 signal cortex and ADC cortex values were derived from three separate cerebral cortical measurements taken on the same axial plane where lesion T2 signal and ADC values were measured. For cortical reference values, three circular ROIs (0.04–0.15 cm$^2$) were placed in areas with optimal gray–white matter differentiation, carefully avoiding white matter and sulci (Figure 2).

### Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics 23.0 (IBM Corp., Armonk, NY, USA). Categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. The Kolmogorov–Smirnov test was used to assess the normality of continuous variables. Normally distributed variables were presented as mean ± standard deviation, and group comparisons were made using independent samples t-tests. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic performance of T2 signal ratios and ADC values/ratios in differentiating orbital lymphoma from orbital inflammation. Optimal cut-off values were determined using the Youden index, an objective criterion that balances diagnostic test performance by identifying the threshold that maximizes both sensitivity and specificity simultaneously, corresponding to the point on the ROC curve nearest the upper-left corner (representing perfect discrimination). Based on these cut-off points, the two variables were converted into binary outcomes and combined using: (1) the "OR rule" (positive if either parameter was positive) and (2) the "AND rule" (positive only if both parameters were positive). The sensitivity and specificity of each combined test were calculated. Inter-rater reliability was assessed using intraclass correlation coefficients (ICC). A $P$ value of <0.05 was considered statistically significant.

### Findings

A total of 86% of patients with primary orbital lymphoma had low-grade non-Hodgkin lymphoma, and MALT lymphoma constituted 57% of all cases. The mean age of patients with lymphoma (58.48 ± 21.18 years) was higher than that of patients with orbital inflammatory diseases (40.54 ± 19.06 years) ($P$ = 0.002). No significant difference in gender distribution was observed between the lymphoma and inflammatory cohorts. Comparative analyses of descriptive findings are presented in Table 1. The majority of patients with lymphoma (n = 19, 90.5%) showed unilateral involvement, demonstrating a statistically significant predominance over inflammatory etiologies ($P$ = 0.02). Although extraconal fat involvement was observed in both cohorts, it was more frequent in inflammatory cases ($P$ = 0.036).

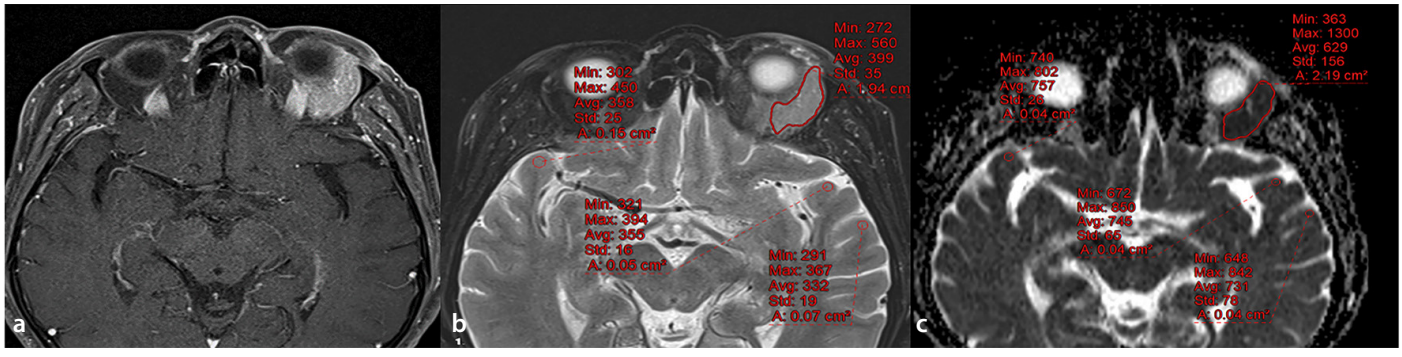



**Figure 2.** Representative example of lesion and reference cortical measurements in a patient with orbital lymphoma; using an axial fat-saturated post-contrast T1-weighted image **(a)** as a reference, manual regions of interest (ROIs) were placed to avoid cystic/necrotic components and lesion margins on the axial fat-saturated T2-weighted image **(b)** and apparent diffusion coefficient map **(c)**; reference cortical ROIs were positioned in normal-appearing cortex while carefully avoiding white matter and sulci.

MRI examinations were performed using 3 T scanners for 10 patients (3 lymphoma, 7 inflammatory) and 1.5 T scanners for 48 patients (18 lymphoma, 30 inflammatory). Mean T2 signal ratios and mean ADC values and ratios for each reviewer are presented in Table 2. All measurements showed statistically significant differences between the two cohorts ($P < 0.05$). Lymphomas demonstrated significantly lower ADC values/ratios and higher T2 signal ratios than inflammatory lesions. Interobserver variability in measurements showed excellent agreement, with ICC values ranging from 0.969 to 0.985.

ROC analysis using measurements from both reviewers is presented in Figure 3. The highest diagnostic performance was achieved by the ADC ratio measurements from Reviewer 1, with an area under the curve (AUC) of 0.920 (95% CI: 0.818–0.975, $P < 0.001$). Using the optimal cut-off (≤1), it demonstrated 85.7% sensitivity and 86.5% specificity, with an overall accuracy of 86.2%. The positive and negative predictive values were 78.3% and 91.4%, respectively.

The T2 signal ratio demonstrated moderate diagnostic performance, with an AUC of 0.726 (95% CI: 0.592–0.86, $P = 0.001$) for Reviewer 1. At the optimal cut-off (>0.88), the test achieved 80.95% sensitivity and 64.86% specificity, yielding an overall accuracy of 70.7%, with positive and negative predictive values of 56.7% and 85.7%, respectively.

The OR rule combination (applying either ADC ratio ≤1 or T2 signal ratio >0.88) showed a sensitivity of 95.2% and specificity of 56.8%. The AND rule combination (requiring both ADC ratio ≤1 and T2 signal ratio >0.88) resulted in a sensitivity of 71.4% and specificity of 94.6%.

**Table 1.** Comparison of descriptive characteristics

| | Lymphoma (%) | Inflammation (%) | *P* value |
|---|---|---|---|
| **Laterality** | | | |
| Unilateral | 19 (90.5%) | 23 (62.2%) | 0.02 |
| Bilateral | 2 (9.5%) | 14 (37.8%) | |
| **Location** | | | |
| Extraconal space only | 12 (57.1%) | 24 (64.9%) | 0.036 |
| Intraconal space only | 3 (14.3%) | 0 | |
| Both extra- and intraconal space | 5 (23.8%) | 13 (35.1%) | |
| Muscle cone only | 1 (4.8%) | 0 | |

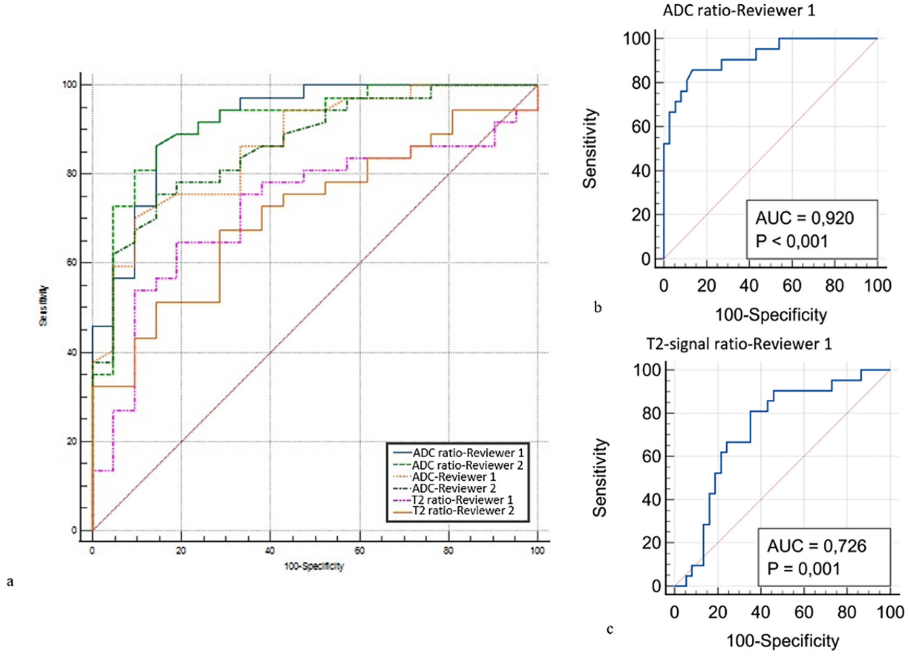



**Figure 3.** Receiver operating characteristic curve analysis evaluating the diagnostic performance of apparent diffusion coefficient (ADC) and transverse relaxation time (T2) signal ratio measurements between reviewers **(a)**; ADC ratio measurements by Reviewer 1 demonstrated superior diagnostic performance, with an area under the curve (AUC) of 0.920 ($P < 0.001$) **(b)**; among T2 signal ratio measurements, Reviewer 1's results showed moderate diagnostic value, with an AUC of 0.726 ($P = 0.001$) **(c)**.

## Discussion

This study demonstrates that combining the T2 signal ratio with ADC values considerably improves differentiation between orbital lymphoma and its challenging inflammatory mimics, particularly IgG4-ROD and granulomatous inflammation (Figure 4). The ADC ratio demonstrated superior diagnostic performance (AUC: 0.920; sensitivity: 85.7%; specificity: 86.5%), confirming the role of DWI as the primary discriminator for lymphoma detection. Although the T2 signal ratio alone exhibited only moderate discriminatory capability (AUC: 0.726; sensitivity: 80.95%; specificity: 64.86%), its integration with ADC values enhanced diagnostic power through complementary effects: the OR rule achieved 95.2% sensitivity, whereas the AND rule demonstrated 94.6% specificity.

Consistent with the known epidemiology of orbital lymphomas, which typically present in the 6th to 7th decades, our patients with lymphoma were significantly older than those with inflammatory disease ($P = 0.002$), as previously documented.[4,5] Our findings also align with earlier studies demonstrating the predominance of unilateral involvement in orbital lymphoma.[4,17] Although extraconal space involvement was frequent in the lymphoma group, supporting the findings of Priego et al.[17], we observed it to be more prevalent in inflammatory cases ($P = 0.036$).

Both reviewers demonstrated significantly lower mean ADC values and ratios in lymphoma ($P < 0.001$). These findings align with prior studies employing various DWI techniques, including single-shot echo-planar imaging (EPI), turbo spin-echo DWI, and diffusion-sensitized driven-equilibrium preparation.[4,6,7,18] Although non-EPI DWI techniques offer advantages over conventional EPI—such as reduced susceptibility artifacts and geometric distortion—all approaches consistently demonstrate lymphoma's characteristic ADC reduction compared with IgG4-ROD and other inflammatory conditions. These findings support the diagnostic utility of DWI independent of technique, although standardization of protocols could further improve its clinical application. Despite DWI's established role in differentiating orbital lymphoma from inflammatory mimickers, diagnostic challenges persist in certain cases. This challenge is especially evident in IgG4-ROD and granulomatous orbital diseases, where ADC values are frequently reduced compared with IOI. Notably, lacrimal gland involvement in IgG4-ROD often demonstrates ADC values below 600 × $10^{-6}$

**Table 2.** Inter-reviewer agreement and comparison of T2 signal ratios and ADC values/ratios between the lymphoma and inflammation groups

| | Lymphoma | Inflammation | *P* value | ICC |
|---|---|---|---|---|
| **Mean T2 signal ratio** | | | | |
| Reviewer 1 | 0.961 ± 0.134 | 0.822 ± 0.242 | 0.007 | 0.969 |
| Reviewer 2 | 0.980 ± 0.146 | 0.835 ± 0.267 | 0.01 | |
| **Mean ADC lesion ($\times 10^{-3}$ mm$^2$/s)** | | | | |
| Reviewer 1 | 0.731 ± 0.172 | 1.077 ± 0.261 | <0.001 | 0.985 |
| Reviewer 2 | 0.713 ± 0.173 | 1.058 ± 0.262 | <0.001 | |
| **Mean ADC ratio** | | | | |
| Reviewer 1 | 0.864 ± 0.184 | 1.319 ± 0.311 | <0.001 | 0.971 |
| Reviewer 2 | 0.852 ± 0.180 | 1.333 ± 0.336 | <0.001 | |

T2, *transverse relaxation time; ADC, apparent diffusion coefficient.*

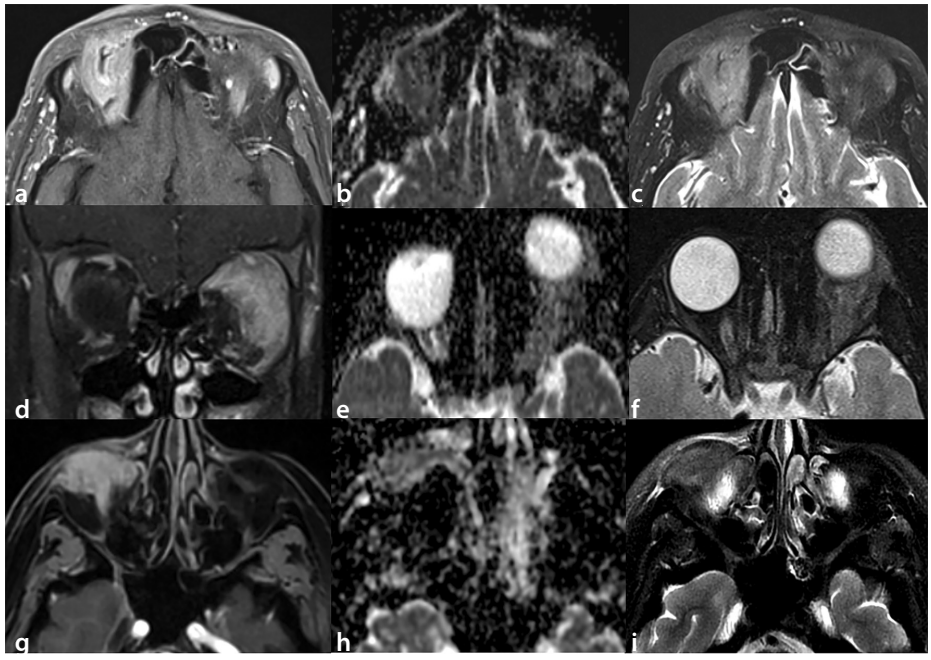



**Figure 4.** Space-occupying soft tissue masses with diffuse contrast enhancement on post-contrast T1-weighted images were observed in 3 patients diagnosed with primary orbital lymphoma (**a**), immunoglobulin G4-related orbital disease (IgG4-ROD) (**d**), and sarcoidosis (**g**); these lesions demonstrated overlapping apparent diffusion coefficient (ADC) values on ADC maps (**b**, **e**, and **h**, respectively); transverse relaxation time (T2)-weighted images helped differentiate lymphoma (**c**) from IgG4-ROD and sarcoidosis (**f** and **i**, respectively), with lymphoma appearing isointense to the cerebral cortex, whereas the inflammatory mimickers appeared hypointense.

mm$^2$/s.[10,11] Granulomatous inflammations may similarly exhibit ADC values overlapping with lymphoma.[19] For cases with indeterminate ADC measurements, additional discriminative imaging features become essential to accurately distinguish lymphoma from IgG4-ROD and granulomatous inflammation, ensuring appropriate clinical management.

Advanced MRI techniques, including dynamic contrast-enhanced (DCE) MRI and arterial spin labeling (ASL), have demonstrated diagnostic potential for differentiating orbital lymphoma from benign mimics. Hu et al.[5] reported that DCE-MRI parameters, particularly "$_{Ve}$" with 76.2% sensitivity and 94.9% specificity, could reliably distinguish malignant lymphoproliferative disorders from benign lesions based on their distinct microvascular characteristics. Eissa et al.[20] reported that total blood flow derived from ASL could discriminate lymphoma from idiopathic inflammatory pseudotumor ($P < 0.001$), although ADC provided superior diagnostic accuracy compared with ASL, and the combined use of both techniques yielded only minimal improvement over ADC alone, suggesting limited clinical added value for ASL. Although perfusion MRI parameters demonstrate promising discriminatory capability for differentiating orbital lymphoma from benign lymphoproliferative and inflammatory disorders, their clinical application has several limitations. Perfusion imaging is not currently part of standard orbital MRI protocols; its inclusion would prolong scan times, and the requirement for specialized post-processing further restricts its viability in daily practice. These limitations underscore the need for practical MRI biomarkers, notably those available through T2-weighted imaging.

Granulomatous inflammations and IgG4-related disease are well-recognized causes of T2-hypointense head and neck lesions. Characteristics of the T2 signal arise from granulomas in granulomatous diseases, including GPA and sarcoidosis, and from fibrosis in orbital IgG4-related disease.[9-12] In contrast, orbital lymphoma is mostly reported to present with an intermediate T2 signal.[13] Our results align well with the literature, as the mean T2 signal ratio of the inflammation cohort was considerably lower than that of the lymphoma cohort.

Although the T2 signal ratio yielded lower isolated diagnostic reliability than ADC values, it serves as a strong complementary tool that considerably enhances the discriminatory power of ADC. The combined application of both parameters using the OR rule (ADC ratio ≤1 or T2 signal ratio >0.88) demonstrated superior sensitivity (95.2%) compared with ADC alone (85.7%). More importantly, the AND rule combination (ADC ratio ≤1 and T2 signal ratio >0.88), achieved higher specificity (from 86.5% to 94.6%), effectively reducing false-positive cases while maintaining diagnostic reliability. These findings underscore that T2-weighted imaging contributes meaningfully to diagnostic confidence, particularly in equivocal cases where ADC values are borderline. Importantly, this approach has direct clinical utility, as it relies on standard MRI sequences without requiring specialized protocols, offering radiologists a practical method to guide patient management.

This study has several limitations, primarily inherent to its retrospective design. The multi-vendor nature of our MRI data (1.5 T and 3 T scanners from different manufactur-

ers) and variations in acquisition parameters may introduce technical variability in quantitative measurements, despite our standardization efforts using ratio-based analyses. Additionally, the small number of patients scanned at 3 T MRI (n = 10) limited our ability to evaluate field strength effects, warranting investigation in future larger studies. Another technical consideration is the use of EPI-DWI, which is particularly susceptible to artifacts and geometric distortion near bone–air interfaces, potentially compromising ADC measurements in small orbital lesions. Our inflammatory cohort demonstrated marked heterogeneity in subgroup distribution, with notably limited cases of sarcoidosis and fungal angiitis, which may affect the generalizability of our findings across all inflammatory subtypes. Furthermore, the retrospective design and modest sample size (n = 58) preclude advanced quantitative analyses such as radiomics, which typically require larger cohorts with standardized imaging protocols to ensure reproducible feature extraction.

In conclusion, the combined evaluation of ADC values and T2 signal ratio provides a clinically practical and effective method for differentiating orbital lymphoma from its challenging inflammatory mimics. This approach utilizes routine MRI sequences while demonstrating superior diagnostic performance compared with either parameter in isolation. Because accurate differentiation between these entities carries substantial therapeutic implications, our findings may help optimize patient management. Multicenter prospective studies with standardized protocols are needed to validate optimal cut-off values and guide clinical implementation.

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Gerbino G, Boffano P, Benech R, et al. Orbital lymphomas: clinical and radiological features. *J Craniomaxillofac Surg*. 2014;42(5):508-512. [Crossref]
2. Kharod SM, Herman MP, Morris CG, Lightsey J, Mendenhall WM, Mendenhall NP. Radiotherapy in the management of orbital lymphoma: a single institution's experience over 4 decades. *Am J Clin Oncol*. 2018;41(1):100-106. [Crossref]
3. Sun B, Song L, Wang X, et al. Lymphoma and inflammation in the orbit: diagnostic performance with diffusion-weighted imaging and dynamic contrast-enhanced. *J Magn Reson Imaging* 2017;45(5):1438-1445. [Crossref]
4. Haradome K, Haradome H, Usui Y, et al. Orbital lymphoproliferative disorders (OLPDs): value of MR imaging for differentiating orbital lymphoma from benign OPLDs. *AJNR Am J Neuroradiol*. 2014;35(10):1976-1982. [Crossref]
5. Hu H, Xu XQ, Liu H, et al. Orbital benign and malignant lymphoproliferative disorders: Differentiation using semi-quantitative and quantitative analysis of dynamic contrast-enhanced magnetic resonance imaging. *Eur J Radiol*. 2017;88:88-94. [Crossref]
6. Ren J, Yuan Y, Wu Y, Tao X. Differentiation of orbital lymphoma and idiopathic orbital inflammatory pseudotumor: combined diagnostic value of conventional MRI and histogram analysis of ADC maps. *BMC Med Imaging*. 2018;18(1):6. [Crossref]
7. Hiwatashi A, Yoshiura T, Togao O, et al. Diffusivity of intraorbital lymphoma vs. IgG4-related disease: 3D turbo field echo with diffusion-sensitised driven-equilibrium preparation technique. *Eur Radiol*. 2014;24(3):581-586. [Crossref]
8. Ferreira TA, Saraiva P, Genders SW, Buchem MV, Luyten GPM, Beenakker JW. CT and MR imaging of orbital inflammation. *Neuroradiology*. 2018;60(12):1253-1266. [Crossref]
9. Toyoda K, Oba H, Kutomi K, et al. MR imaging of IgG4-related disease in the head and neck and brain. *AJNR Am J Neuroradiol*. 2012;33(11):2136-2139. [Crossref]
10. Thompson A, Whyte A. Imaging of IgG4-related disease of the head and neck. *Clin Radiol*. 2018;73(1):106-120. [Crossref]
11. Dragan AD, Weller A, Lingam RK. Imaging of IgG4-related disease in the extracranial head and neck. *Eur J Radiol*. 2021;136:109560. [Crossref]
12. Kato H, Kawaguchi M, Ando T, Kaneko Y, Hyodo F, Matsuo M. Hypointense head and neck lesions on T2-weighted images: correlation with histopathologic findings. *Neuroradiology*. 2020;62(10):1207-1217. [Crossref]
13. Akansel G, Hendrix L, Erickson BA, et al. MRI patterns in orbital malignant lymphoma and atypical lymphocytic infiltrates. *Eur J Radiol*. 2005;53(2):175-181. [Crossref]
14. Umehara H, Okazaki K, Kawa S, et al. The 2020 revised comprehensive diagnostic (RCD) criteria for IgG4-RD. *Mod Rheumatol*. 2021;31(3):529-533. [Crossref]
15. Robson JC, Grayson PC, Ponte C, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for granulomatosis with polyangiitis. *Ann Rheum Dis*. 2022;81(3):315-320. [Crossref]
16. Crouser ED, Maier LA, Wilson KC, et al. Diagnosis and Detection of Sarcoidosis. An Official American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med*. 2020;201(8):e26-e51. [Crossref]
17. Priego G, Majos C, Climent F, Muntane A. Orbital lymphoma: imaging features and differential diagnosis. *Insights Imaging*. 2012;3(4):337-344. [Crossref]
18. Hiwatashi A, Togao O, Yamashita K, et al. Diffusivity of intraorbital lymphoma vs. inflammation: comparison of single shot turbo spin echo and multishot echo planar imaging techniques. *Eur Radiol*. 2018;28(1):325-330. [Crossref]
19. Bulut EG. Granulomatosis with polyangiitis: special focus on suggestive imaging findings in head and neck involvement. *Acta Medica*. 2019;50(3):9-16. [Crossref]
20. Eissa L, Abdel Razek AAK, Helmy E. Arterial spin labeling and diffusion-weighted MR imaging: utility in differentiating idiopathic orbital inflammatory pseudotumor from orbital lymphoma. *Clin Imaging*. 2021;71:63-68. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253652

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE



# Association between magnetic resonance imaging-based prostatic tissue morphology and changes in International Prostate Symptom Score and maximum urinary flow rate after prostatic artery embolization

Chen Wang[1]
Kun Liang[1]
Jiasheng Qin[1]
Xin Shu[1]
Lihua Yuan[1]
Jianfei Dong[2]

[1]Department of Interventional Radiology, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China

[2]Department of Otorhinolaryngology Head and Neck Surgery, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China

**PURPOSE**

To evaluate whether baseline multiparametric magnetic resonance imaging (mpMRI)-defined prostate morphological phenotypes are associated with changes in the International Prostate Symptom Score (ΔIPSS) and maximum urinary flow rate (ΔQmax) following prostatic artery embolization (PAE) in patients with benign prostatic hyperplasia (BPH).

**METHODS**

This retrospective single-center study included patients who underwent technically successful PAE with preprocedural mpMRI performed within 8 weeks of intervention. Prostate morphology was classified as glandular-dominant, stromal-dominant, or mixed phenotype based on the predominant nodule signal characteristics on T2-weighted imaging (T2WI), with supplemental assessment using diffusion-weighted and contrast-enhanced sequences when available. Symptomatic and functional outcomes—ΔIPSS and ΔQmax—were assessed at 3, 6, 12, and 24 months post-procedure. Associations between morphology and outcomes were evaluated using multivariable regression and subgroup analysis.

**RESULTS**

A total of 152 patients (mean age, 70.0 ± 9.9 years) who underwent technically successful PAE were included in the study, all of whom had preprocedural MRI performed within 8 weeks prior to the intervention. All MRI-defined morphological phenotypes demonstrated improvements in the Qmax and IPSS after PAE. The glandular-dominant phenotype exhibited the most pronounced and earliest response, with peak improvements at 6 months (ΔQmax: 10.45 mL/s; ΔIPSS: 14.11 points) and sustained benefits through 24 months (ΔQmax: 8.78 mL/s; ΔIPSS: 13.04 points). Stromal-dominant and mixed phenotypes showed smaller, delayed improvements, typically peaking at 12 months. Morphology-related phenotype differences remained statistically significant at 24 months in unadjusted comparisons, particularly between glandular and stromal phenotypes, although ΔIPSS differences were attenuated in multivariable models (24-month β: 0.104, $P = 0.547$). Glandular morphology was consistently associated with a greater ΔQmax across all timepoints (e.g., 24-month β: 0.450, $P < 0.001$) and significantly interacted with baseline symptom severity (IPSS ≥ 20) to predict enhanced 24-month ΔIPSS (interaction β: 0.349, $P = 0.045$). Subgroup analyses stratified by prostate volume (< 80 vs. ≥ 80 mL) corroborated these findings, with glandular morphology consistently outperforming that of stromal and mixed phenotypes.

**CONCLUSION**

Baseline MRI-defined prostate morphology, assessed primarily on T2WIs, was significantly associated with clinical outcomes after PAE; glandular-dominant morphology was linked to larger and more sustained improvements in both the Qmax and IPSS. Therefore, MRI-based phenotypes offer a practical imaging biomarker for patient stratification, warranting prospective validation.

**CLINICAL SIGNIFICANCE**

The MRI-based classification of prostate morphology may improve patient selection and predict clinical outcomes following PAE in BPH management.

**KEYWORDS**

Benign prostatic hyperplasia, lower urinary tract symptoms, prostatic artery embolization, magnetic resonance imaging, prostate morphology

**Handling editor:** İlkay İdilman

**Corresponding author:** Lihua Yuan

**E-mail:** lihua.yuan@njglyy.com

**Corresponding author:** Jianfei Dong

**E-mail:** alaterbloomer@sina.com

Received 22 September 2025; revision requested 01 November 2025; last revision requested 19 December 2025; accepted 29 December 2025.

Epub: 26.01.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253652

You may cite this article as: Wang C, Liang K, Qin J, Shu X, Yuan L, Dong J. Association between magnetic resonance imaging-based prostatic tissue morphology and changes in International Prostate Symptom Score and maximum urinary flow rate after prostatic artery embolization. *Diagn Interv Radiol*. 2026;32(4):463-472.

Benign prostatic hyperplasia (BPH) is a prevalent condition among aging men, characterized by non-malignant proliferation of epithelial and stromal components within the prostatic transition zone. This histologic expansion contributes to lower urinary tract symptoms (LUTS), including increased urinary frequency, urgency, and nocturia and a reduced urinary stream. Pharmacologic therapies—primarily α-adrenergic antagonists and 5α-reductase inhibitors—remain the first-line treatment; however, their efficacy is often limited and is associated with adverse effects such as hypotension, ejaculatory dysfunction, and poor long-term adherence. For patients with moderate-to-severe LUTS or an unsatisfactory response to medication, transurethral resection of the prostate (TURP) has long served as the gold-standard surgical intervention, providing rapid and durable symptom relief in most cases.[1,2] More recently, prostatic artery embolization (PAE) has emerged as a minimally invasive alternative that achieves prostate debulking through selective arterial occlusion, leading to ischemia and subsequent tissue necrosis. Unlike TURP, PAE does not require general anesthesia and is particularly suitable for elderly individuals, patients with substantial comorbidities, or those seeking nonsurgical options. Clinical studies have demonstrated that PAE effectively alleviates LUTS and improves quality of life; however, therapeutic responses remain heterogeneous, and reliable imaging biomarkers for outcome prediction are lacking.[3-6] Multiparametric magnetic resonance imaging (mpMRI), combining T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) sequences, has become the standard for comprehensive prostate evaluation. It provides a unified framework for lesion characterization across institutions. The Prostate Imaging Reporting and Data System PI-RADS v2.1 (American College of Radiology, Reston, VA, USA). standardizes the acquisition and interpretation of prostate MRI to detect clinically significant prostate cancer, particularly in the peripheral zone, where the majority of clinically significant prostate cancer occurs. Although PI-RADS assesses the transition zone, it is primarily designed to assess malignant lesions, and its utility in distinguishing benign changes, such as BPH, in the transition zone is limited. The transition zone, where stromal and glandular nodules exhibit heterogeneous morphology and signal patterns, is not sufficiently addressed by PI-RADS. Recent studies on BPH have emphasized the use of quantitative MRI-derived metrics, such as transition zone volume and transition zone index, which correlate with surgical enucleation volume and bladder outlet obstruction. These findings underscore the need for a dedicated classification system to assess benign changes in the transition zone.[7,8] Histopathologic investigations emphasize that BPH encompasses a spectrum of nodular hyperplasia with varying proportions of glandular and stromal elements, resulting in substantial tissue and imaging heterogeneity. Glandular-dominant nodules primarily consist of epithelial structures and glandular lumina, whereas stromal-dominant nodules contain abundant smooth muscle and fibrous tissue.[9] Clinically, patients with adenomatous-dominant BPH (AdBPH) have been shown to achieve greater prostate volume reduction and more pronounced symptomatic improvement following PAE, suggesting that imaging-based morphological subtyping may hold prognostic value.[10]

Building on these histologic insights, mpMRI has been explored as a non-invasive tool for preprocedural morphological characterization. In particular, T2-weighted and DWI features have been correlated with underlying tissue composition.[11] Glandular-dominant nodules typically exhibit high T2 signal intensity and correspond to improved post-PAE outcomes, whereas stromal-dominant patterns—characterized by a low T2 signal and dense fibromuscular stroma—are associated with less favorable responses.[12]

Despite these encouraging findings, a standardized and clinically applicable MRI-based classification system for prostate morphology has not been widely validated. In this study, we propose a T2-weighted MRI classification framework that categorizes transition zone nodules into three phenotypes: glandular dominant, stromal dominant, and mixed. We aimed to evaluate whether MRI-defined prostate morphological phenotypes are associated with changes in the International Prostate Symptom Score (IPSS) and maximum urinary flow rate (Qmax) following PAE. To explore potential effect modification, prespecified subgroup analyses stratified by the baseline prostate volume (< 80 vs. ≥ 80 mL) were also conducted.[13]

**Main points**

- Magnetic resonance imaging-based prostate morphology classification into glandular-dominant, stromal-dominant, and mixed phenotypes provides a practical imaging surrogate for underlying histologic composition in benign prostatic hyperplasia.
- Glandular-dominant prostates demonstrated the most pronounced and consistent improvements in both the International Prostate Symptom Score (IPSS) and maximum urinary flow rate (Qmax) after prostatic artery embolization, with the maximal response observed at 6 months and sustained benefits identified through long-term follow-up.
- Stromal-dominant and mixed phenotypes exhibited delayed or attenuated benefits, typically reaching their peak effect at around 12 months.
- Subgroup analysis revealed a morphology- and volume-dependent gradient of efficacy, with interphenotype differences in functional outcomes (ΔQmax) emerging earlier in larger prostates (≥ 80 mL).
- Magnetic resonance imaging-defined glandular-dominant morphology, particularly in patients with a high baseline IPSS, was independently associated with sustained symptomatic and functional benefits at 24 months.

## Methods

### Study design and patient selection

This study initially included 220 patients diagnosed with BPH-related LUTS. Of these, 68 were excluded because they did not meet the requirements of the inclusion criteria, with 152 patients ultimately included (as outlined in the flowchart, Figure 1, Table 1).

This retrospective single-center study was approved by the Ethics Committee of Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School (reference number: 2024-124-02, date: April 24, 2024), and the requirement for informed consent was waived due to the retrospective nature of the study. We screened consecutive patients who underwent PAE for symptomatic BPH between May 2020 and March 2022. A total of 152 patients (mean age, 70.0 ± 9.9 years) who underwent technically successful PAE were included in the study, all of whom had preprocedural MRI performed within 8 weeks prior to intervention.

Patients were eligible for inclusion if they had an IPSS of ≥ 15 at baseline, experienced persistent moderate-to-severe LUTS refractory to standard pharmacologic treatment, underwent multiparametric pelvic MRI within 8 weeks prior to embolization, and completed follow-up assessments at 3, 6, 12, and 24 months post-procedure. Exclusion criteria included prior prostate surgery, neurogenic bladder, known or suspected prostate malignancy, technical failure of embolization, or incomplete imaging or clinical data.

Baseline variables, including age, body mass index (BMI), total prostate volume (TPV), and MRI-defined morphology (glandular, stromal, mixed), were extracted from institutional electronic medical records. The prolate ellipsoid formula (TPV = 0.52 × transverse × anteroposterior × craniocaudal dimensions) was used to calculate TPV, measured on axial and sagittal T2WIs, which is the standard method recommended in contemporary prostate MRI protocols.

Prostate biopsy was not performed routinely and was only pursued when malignancy was clinically suspected (e.g., elevated or rising prostate-specific antigen levels, abnormal digital rectal examination, or MRI findings suggestive of malignancy). A small number of patients were diagnosed with prostate cancer during follow-up, either due to the absence of preprocedural suspicion warranting biopsy or as a result of false-negative preprocedural assessments. These cases were marked as exclusions in the flowchart (Figure 1). Although PI-RADS v2.1 scoring is included in standard institutional MRI reporting, it was not utilized in this study. Our focus was on evaluating benign transition zone morphology, a feature not adequately addressed by PI-RADS, which is primarily designed to detect malignant lesions in the peripheral zone. Although PI-RADS includes the transition zone in its evaluation, its primary application is to assess clinically significant prostate cancer, providing limited guidance for the assessment of BPH or other benign morphological changes in the transition zone. Therefore, we employed a morphology-based classification system specifically developed for assessing benign features in the transition zone—an aspect not covered by PI-RADS.

## Magnetic resonance imaging acquisition and morphological subtyping

### Magnetic resonance imaging acquisition

All MRI examinations were performed using either a 1.5-T or 3.0-T scanner (Ingenia CX or UMR platform) equipped with an 8–16-channel phased-array pelvic coil. The standard protocol included multiplanar T2-weighted turbo spin-echo sequences, which served as the primary basis for morphological classification. DWI was performed on all patients as part of the routine clinical protocol during the study period. However, DCE sequences were available only in a limited subset of patients (< 15%) according to the clinical routine at the time of scanning. When present, these sequences were used qualitatively as supportive references to improve tissue characterization or to exclude foci with suspected cancer. When available, the DCE sequences often showed patterns suggestive of early enhancement in gland-rich nodules and more prolonged enhancement in stromal-dominant nodules, consistent with the T2-weighted impressions of the respective tissue composition. However, no quantitative perfusion parameters were analyzed, and classification of nodule morphology was based solely on axial T2WI to maximize consistency across the cohort.[14,15] Acquisition parameters, including field strength, coil type, repetition and echo times, slice thickness, and pixel size, are summarized in Table 2.

Patient preparation procedures were standardized according to institutional protocols. Patients were instructed to fast for at least 6 hours prior to the MRI and PAE procedures. Additionally, for optimal prostate visualization during MRI, patients were asked to maintain a comfortably full bladder.

### Morphological subtyping

A phenotypic classification system was developed based on visual assessment of the dominant imaging characteristics of transition zone nodules, categorizing BPH into glandular-dominant, stromal-dominant, and mixed phenotypes. In the glandular-dominant phenotype, glandular nodules were larger and more prominent, predominantly composed of glandular tissue, and showed high signal intensity on T2WI, distinguishing them clearly from the surrounding stromal tissue. The mixed phenotype featured multiple glandular nodules, even when these nodules were not dominant in size. These nodules appeared alongside stromal elements, with alternating high and low signal intensities on T2WI, reflecting the heterogeneous nature of the transition zone. The stromal-dominant phenotype was characterized by minimal or absent glandular nodules, with the lesions predominantly composed of stromal tissue exhibiting low signal intensity, typical of be-

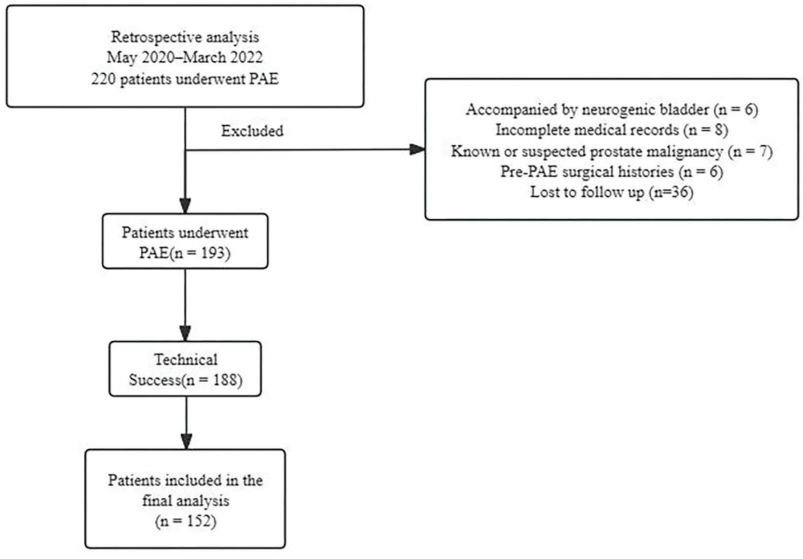



**Figure 1.** Flowchart illustrating the retrospective analysis design of this study.

PAE, prostatic artery embolization.

**Table 1.** Baseline characteristics by magnetic resonance imaging-defined prostate morphology

| Variable | Glandular (n = 53) | Mixed (n = 61) | Stromal (n = 38) | *P* |
|---|---|---|---|---|
| Age (years) | 70.83 ± 8.77 | 68.54 ± 10.32 | 66.63 ± 10.89 | 0.136 |
| BMI (kg/m$^2$) | 28.48 ± 2.79 | 27.06 ± 3.15 | 27.99 ± 2.28 | 0.027 |
| TPV (mL) | 96.53 ± 32.70 | 98.67 ± 32.46 | 91.52 ± 23.11 | 0.523 |
| Baseline IPSS | 24.98 ± 4.86 | 24.95 ± 6.10 | 22.66 ± 6.52 | 0.110 |
| Baseline Qmax (mL/s) | 6.57 ± 2.42 | 6.39 ± 1.99 | 6.76 ± 1.71 | 0.689 |

IPSS, International Prostate Symptom Score; Qmax, maximum urinary flow rate; TPV, total prostate volume; BMI, body mass index.

nign fibromuscular tissue. These lesions displayed a lamellar or fibrous texture, with no features suggestive of malignancy, such as irregular margins or a diffuse low signal, which are typical of tumor infiltration (Figure 2).[16]

### Interobserver agreement

All examinations were independently evaluated by two radiologists, each with more than 5 years of experience in prostate MRI, who were blinded to the clinical data. Morphologic phenotype (glandular dominant, stromal dominant, or mixed) and TPV (T2-weighted planimetry) were determined for each case. Interobserver agreement for phenotype classification was assessed using Cohen's kappa with 95% confidence intervals (CIs), whereas TPV reproducibility was evaluated using intraclass correlation coefficients (ICCs) from a two-way random-effects, absolute-agreement, single-measure model with 95% CIs. Discrepancies between readers were resolved by consensus, and the agreed results were used for subsequent analyses.[17]

### Embolization procedure and technical parameters

PAE was performed in all patients under local anesthesia using a transfemoral approach. Procedures were conducted by two experienced interventional radiologists using a cone-beam computed tomography (CT)-equipped C-arm system (Allura Xper FD20/XperCT, Philips). Selective catheterization of the internal iliac artery was achieved with a 4-Fr Simmons or Cobra catheter. When visualization of the prostatic artery was inadequate, tube angulation was adjusted accordingly. Super-selective catheterization of the prostatic artery was achieved using a 2.4–2.7-Fr microcatheter (e.g., Maestro, Merit Medical) over a 0.014-inch steerable microwire (e.g., Fathom, Boston Scientific). In all cases, intra-arterial nitroglycerin was administered to induce vasodilation, followed by cone-beam CT acquisition to delineate the prostatic arterial supply and exclude extra-prostatic collaterals before embolization. Bilaterally, embolization was performed using the PErFecTED technique. Larger microspheres (300–500 µm) were first injected proximally, followed by distal advancement of the microcatheter beyond potential collateral branches for administration of smaller microspheres (100–300 µm). Embolization was performed using calibrated tris-acryl gelatin microspheres (100–300 and 300–500 µm; Embosphere®, Merit Medical, South Jordan, UT, USA). Embolization was monitored under fluoroscopy and digital subtraction angiography, with the endpoint defined as slow antegrade flow or near-stasis in the prostatic artery without reflux, and decreased perfusion of the prostate on post-embolization angiography.[3,4]

### Follow-up and endpoints

Clinical assessments—including the IPSS and Qmax—were obtained at baseline and at 3, 6, 12, and 24 months following PAE. All follow-up evaluations were conducted by board-certified interventional radiologists who were blinded to the MRI-based prostate morphology classification. Longitudinal follow-up at each timepoint was used to assess changes from baseline, expressed as ΔIPSS and ΔQmax. These outcome measures were analyzed according to the baseline MRI-defined prostate morphology, categorized as glandular-dominant, stromal-dominant, and mixed phenotypes. Post-PAE complications and their management were systematically documented using information based on patient interviews, reports from referring physician reports, and a review of medical records.

### Subgroup analysis

To evaluate whether TPV influenced treatment outcomes, patients were stratified into two prespecified subgroups (< 80 vs. ≥ 80 mL), consistent with previous PAE studies that applied this threshold for both clinical and procedural considerations.[18-20] Baseline TPV was dichotomized at 80 mL for subgroup analyses, in accordance with the prior PAE-related literature and clinical convention distinguishing moderate-sized from large prostates. Within each subgroup, longitudinal changes in the IPSS and Qmax at 3, 6, 12, and 24 months were compared among the MRI-defined morphological phenotypes (glandular-dominant, stromal-dominant, and mixed types).

### Complications

All procedure-related complications were documented prospectively and categorized according to the Cardiovascular and Interventional Radiological Society of Europe (CIRSE) Standards of Practice for Reporting Complications (2018 update). Complications were defined as any deviation from the normal postoperative course and graded as minor (CIRSE grades 1–2) or major (grades 3–6) based on the need for therapy, hospitalization, or permanent sequelae. According to this classification, dysuria, mild hematuria, transient urinary retention, and self-limited pelvic pain were considered minor, whereas any event requiring prolonged hospitalization, repeat embolization, or surgical management were considered major.

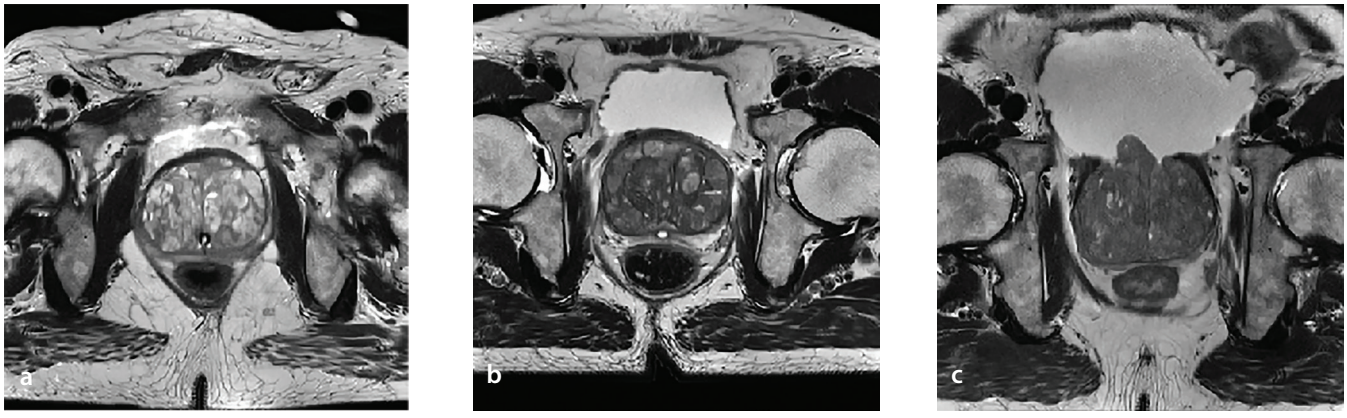



**Figure 2.** Representative T2-weighted magnetic resonance imaging patterns of transition zone phenotypes in benign prostatic hyperplasia. **(a)** Glandular-dominant phenotype: the transition zone is predominantly composed of glandular nodules, which present as high-signal structures on T2-weighted imaging (T2WI), exhibiting acinous or microcystic morphology. These nodules are easily distinguishable from the surrounding stromal tissue due to their high signal intensity. **(b)** Mixed phenotype: the transition zone shows a combination of glandular and stromal nodules. The glandular nodules are present but not dominant in size, exhibiting alternating high and low signal intensities on T2WI. This phenotype reflects the heterogeneous composition of the transition zone, with both glandular and stromal components clearly visible. **(c)** Stromal-dominant phenotype: the transition zone is primarily composed of stromal nodules, with fewer glandular nodules. The stromal tissue appears as low-signal structures on T2WI, exhibiting a lamellar or fibromuscular texture, typical of benign stromal hyperplasia.

### Statistical analysis

All statistical analyses were performed using SPSS (version 31.0; IBM Corp., Armonk, NY, USA) in accordance with the STrengthening the Reporting of OBservational Studies in Epidemiology reporting guidelines.[21] Continuous variables were summarized as mean ± standard deviation and categorical variables as frequencies and percentages. Continuous outcomes (ΔIPSS, ΔQmax) were compared using one-way Analysis of Variance or independent samples t-tests when normality was satisfied. Separate multivariable linear regression models were constructed at each follow-up interval to evaluate associations between baseline variables and clinical outcomes. The ΔIPSS or ΔQmax values served as the dependent variable, whereas prespecified covariates—including age, BMI, baseline TPV, and baseline symptom scores (IPSS for ΔIPSS models; Qmax for ΔQmax models)—were entered as independent variables. To explore potential effect modification, interaction terms between the baseline IPSS and prostate morphology phenotypes were included. Patients were stratified into two categories (IPSS < 20 and IPSS ≥ 20) based on the standard IPSS severity classification, consistent with previous studies that defined IPSS ≥ 20 as indicative of severe symptoms.[22] All statistical tests were two sided, and a *P* value < 0.05 was considered statistically significant.

## Results

### Patient characteristics

A total of 152 patients met the inclusion criteria (mean age 70.0 ± 9.9 years). Baseline TPV was comparable among the glandular-dominant, stromal-dominant, and mixed phenotypes. The cohort included 53 patients with glandular-dominant, 38 with stromal-dominant, and 61 with mixed-type morphology, representing 34.9%, 25.0%, and 40.1% of the study population, respectively. The baseline demographic and clinical characteristics summarized by MRI-defined morphology are shown in Table 1. Groups were comparable in age, TPV, baseline IPSS, and baseline Qmax (all $P > 0.05$); BMI differed modestly across phenotypes (overall $P = 0.027$). Across MRI-defined morphological phenotypes, both functional (ΔQmax) and symptomatic (ΔIPSS) outcomes improved significantly after PAE, although the magnitude and timing of improvement differed by phenotype. The glandular-dominant phenotype exhibited the most pronounced and earliest responses, reaching peak improvements at 6 months (ΔQmax: 10.45 mL/s; ΔIPSS: 14.11 points). Minor declines were observed thereafter at 12 and 24 months (ΔQmax: 9.76 and 8.78 mL/s; ΔIPSS: 13.43 and 13.04 points, respectively), yet treatment effects remained substantially above baseline throughout the 24-month follow-up. By contrast, mixed- and stromal-dominant phenotypes demonstrated slower and less marked improvements, with maximal changes occurring at around 12 months for both ΔIPSS and ΔQmax, followed by modest reductions at 24 months. At all follow-up intervals, the glandular-dominant phenotype consistently showed the greatest mean gains in both ΔQmax and ΔIPSS, followed by the mixed and stromal phenotypes. This morphology-dependent gradient in treatment response persisted over time, although the magnitude of intergroup difference slightly attenuated at later timepoints (Table 3).

### Multivariable linear regression

Relative to the stromal-dominant reference phenotype, the glandular-dominant phenotype showed greater improvement at 6 months—the timepoint of maximal separation—for both ΔIPSS (β: 0.510, $P < 0.001$) and ΔQmax (β: 0.567, $P < 0.001$). For ΔIPSS, this advantage was also evident at 3 and 12 months but was no longer significant at 24 months. No significant differences in ΔIPSS were observed between the mixed- and stromal-dominant phenotypes at any timepoint. For ΔQmax, the superiority of the glandular-dominant phenotype persisted across all follow-up intervals. Comparisons between mixed and stromal phenotypes revealed higher ΔQmax in the mixed group at 3 and 24 months, with borderline significance at 6 and 12 months. The full model coefficients and corresponding 95% CIs for each timepoint are presented in Table 4. An interaction

**Table 2.** Magnetic resonance imaging acquisition parameters for prostate imaging sequences

| Sequence | TR/TE | FA | ST | Res | Coil | Field | Availability |
|---|---|---|---|---|---|---|---|
| Axial T2 | ~4,000/100 | 120 | 3–4 | 0.8–1.2 | PA (8–16 ch) | 1.5/3.0 | All pts |
| DWI* | Vendor-spec. | — | 4–5 | 1.5–2.0 | same | 1.5/3.0 | All pts |
| DCE–MRI** | Short TR/TE | 10–15 | 2–4 | 1.0–1.5 | same | 1.5/3.0 | Subset |

*DWI parameters (including TR/TE, b-values, and fat-suppression technique) were vendor-dependent across scanners and were not used for phenotypic classification. **DCE–MRI was acquired using three-dimensional spoiled gradient-echo sequences. Temporal enhancement parameters were not analyzed and did not contribute to phenotypic classification.
T2, T2-weighted imaging; DWI, diffusion-weighted imaging; DCE, dynamic contrast-enhanced MRI; MRI, magnetic resonance imaging; TR, repetition time; TE, echo time; FA, flip angle; ST, slice thickness; Res, in-plane resolution; PA, pelvic array; ch, channels; pts, patients.

**Table 3.** Interphenotype comparison of ΔIPSS and ΔQmax in all patients

| Outcome | Glandular | Mixed | Stromal | *P | **P | ***P | ****P |
|---|---|---|---|---|---|---|---|
| ΔQmax 3 m | 6.64 ± 2.85 | 5.45 ± 2.69 | 4.71 ± 1.72 | 0.002 | 0.187 | <0.001 | 0.006 |
| ΔQmax 6 m | 10.45 ± 2.32 | 8.32 ± 2.38 | 7.36 ± 1.89 | < 0.001 | < 0.001 | <0.001 | 0.030 |
| ΔQmax 12 m | 9.76 ± 2.30 | 8.59 ± 2.22 | 7.73 ± 1.83 | < 0.001 | 0.007 | <0.001 | 0.035 |
| ΔQmax 24 m | 8.78 ± 2.45 | 7.46 ± 2.54 | 6.33 ± 1.97 | < 0.001 | 0.006 | <0.001 | 0.021 |
| ΔIPSS 3 m | 7.62 ± 2.96 | 6.62 ± 2.44 | 6.11 ± 2.74 | 0.024 | 0.051 | 0.015 | 0.330 |
| ΔIPSS 6 m | 14.11 ± 4.51 | 10.00 ± 2.86 | 9.42 ± 2.61 | <0.001 | <0.001 | <0.001 | 0.314 |
| ΔIPSS 12 m | 13.43 ± 3.43 | 12.36 ± 3.90 | 11.08 ± 3.77 | 0.013 | 0.124 | 0.003 | 0.111 |
| ΔIPSS 24 m | 13.04 ± 4.28 | 10.92 ± 3.57 | 9.16 ± 3.12 | <0.001 | 0.005 | <0.001 | 0.014 |

*P, overall three-phenotype comparison; **P, glandular vs. mixed; ***P, glandular vs. stromal; ****P, mixed vs. stromal. IPSS, International Prostate Symptom Score; ΔQmax, maximum urinary flow rate; ΔIPSS, change in International Prostate Symptom Score.

between baseline IPSS ≥ 20 and glandular-dominant morphology was identified for ΔIPSS at 24 months (β: 0.349, $P = 0.045$), indicating more pronounced long-term symptomatic improvement among patients with a higher baseline symptom burden and glandular-dominant features.

### Subgroup analysis

Subgroup analyses stratified by baseline TPV (< 80 vs. ≥ 80 mL) revealed clear volume-dependent differences in treatment response. In the < 80-mL subgroup, interphenotype variation in ΔQmax was not significant at 3 months but became significant at later timepoints and persisted thereafter. Glandular-dominant prostates showed consistently greater functional improvement than stromal-dominant prostates across all follow-up intervals and also exceeded mixed phenotypes at 6 and 24 months. Mixed and stromal phenotypes did not differ significantly. For ΔIPSS, interphenotype differences at 6 and 24 months were significant, with glandular-dominant morphology associated with greater symptomatic improvement than stromal morphology at all timepoints and with mixed morphology at 6 and 24 months (Table 5). In the ≥ 80 mL subgroup, interphenotype differences in ΔQmax were already significant at 3 months and remained so through 24 months. A clear hierarchical gradient (glandular > mixed > stromal) was observed, with mixed phenotypes separating earlier from stromal-dominant phenotypes in a ΔQmax-a pattern not seen in the smaller-volume subgroup. For ΔIPSS, significant interphenotype variation was again noted at 6 and 24 months, with glandular-dominant morphology showing the greatest improvement, whereas mixed and stromal-dominant phenotypes remained largely comparable (Table 6).

### Interobserver agreement

Interobserver agreement was assessed across all 152 patients for both TPV measurements and MRI-based phenotype classification. The ICC for TPV was 0.934, indicating excellent agreement. The Cohen's kappa coefficient for MRI-defined prostate type was 0.836, reflecting very good inter-rater reliability.

### Complications

A total of 61 patients (40.1%) experienced minor complications. These included dysuria (n = 19, 12.5%), hematuria (n = 22, 14.5%), urinary tract infections (n = 7, 4.6%), transient urinary retention (n = 14, 9.2%), and pelvic pain (n = 54, 35.5%). Hematospermia occurred in eight patients (5.3%). No major complications were reported in this phenotype.

**Table 4.** Multivariable regression results

| Timepoint | Variable | β coefficient | *P* value |
|---|---|---|---|
| ΔIPSS 3 m | Glandular | 0.193 | 0.048 |
| | Mixed | 0.033 | 0.738 |
| ΔIPSS 6 m | Glandular | 0.510 | < 0.001 |
| | Mixed | 0.024 | 0.781 |
| ΔIPSS 12 m | Glandular | 0.255 | 0.011 |
| | Mixed | 0.120 | 0.233 |
| ΔIPSS 24 m | Glandular | 0.104 | 0.547 |
| | Mixed | 0.269 | 0.602 |
| | IPSS ≥ 20 × glandular | 0.349 | 0.045 |
| ΔQmax 3 m | Glandular | 0.335 | < 0.001 |
| | Mixed | 0.208 | 0.038 |
| ΔQmax 6 m | Glandular | 0.567 | < 0.001 |
| | Mixed | 0.179 | 0.051 |
| ΔQmax 12 m | Glandular | 0.415 | < 0.001 |
| | Mixed | 0.189 | 0.055 |
| ΔQmax 24 m | Glandular | 0.450 | < 0.001 |
| | Mixed | 0.216 | 0.027 |

IPSS, International Prostate Symptom Score; ΔQmax, maximum urinary flow rate.

**Table 5.** Interphenotype comparison of ΔIPSS and ΔQmax in patients with total prostate volume <80 mL

| Outcome | Glandular | Mixed | Stromal | *P | **P | ***P | ****P |
|---|---|---|---|---|---|---|---|
| ΔQmax 3 m | 6.50 ± 3.03 | 5.58 ± 2.92 | 3.74 ± 1.65 | 0.064 | 0.359 | 0.006 | 0.092 |
| ΔQmax 6 m | 10.11 ± 2.13 | 8.17 ± 2.74 | 7.29 ± 2.66 | 0.017 | 0.025 | 0.007 | 0.429 |
| ΔQmax 12 m | 9.95 ± 1.93 | 8.36 ± 2.55 | 7.67 ± 2.56 | 0.041 | 0.043 | 0.017 | 0.513 |
| ΔQmax 24 m | 9.02 ± 2.30 | 7.15 ± 2.92 | 6.14 ± 2.74 | 0.025 | 0.041 | 0.009 | 0.392 |
| ΔIPSS 3 m | 8.00 ± 2.60 | 6.37 ± 2.22 | 5.56 ± 3.84 | 0.072 | 0.050 | 0.065 | 0.482 |
| ΔIPSS 6 m | 14.12 ± 3.39 | 9.47 ± 2.44 | 8.89 ± 3.66 | < 0.001 | <0.001 | 0.001 | 0.618 |
| ΔIPSS 12 m | 14.06 ± 3.44 | 12.63 ± 3.70 | 10.11 ± 4.34 | 0.047 | 0.240 | 0.018 | 0.123 |
| ΔIPSS 24 m | 13.82 ± 3.84 | 10.53 ± 3.44 | 9.89 ± 3.37 | 0.010 | 0.010 | 0.016 | 0.649 |

*P, overall three-phenotype comparison; **P, glandular vs. mixed; ***P, glandular vs. stromal; ****P, mixed vs. stromal. IPSS, International Prostate Symptom Score; ΔQmax, maximum urinary flow rate.

**Table 6.** Interphenotype comparison of ΔIPSS and ΔQmax in patients with total prostate volume ≥ 80 mL

| Outcome | Glandular | Mixed | Stromal | *P | **P | ***P | ****P |
|---|---|---|---|---|---|---|---|
| ΔQmax 3 m | 6.70 ± 2.81 | 6.11 ± 2.61 | 5.01 ± 1.66 | 0.025 | 0.342 | 0.004 | 0.033 |
| ΔQmax 6 m | 10.61 ± 2.42 | 8.39 ±2.23 | 7.39 ± 1.65 | <0.001 | <0.001 | <0.001 | 0.043 |
| ΔQmax 12 m | 9.68 ± 2.48 | 8.71 ± 2.08 | 7.74 ± 1.60 | 0.002 | 0.064 | <0.001 | 0.031 |
| ΔQmax 24 m | 8.67 ± 2.54 | 7.60 ± 2.38 | 6.39 ± 1.72 | <0.001 | 0.058 | 0.015 | 0.036 |
| ΔIPSS 3 m | 7.44 ± 3.13 | 6.74 ± 2.56 | 6.28 ± 2.36 | 0.218 | 0.276 | 0.101 | 0.443 |
| ΔIPSS 6 m | 14.11 ± 5.00 | 10.24 ±3.03 | 9.59 ± 2.24 | <0.001 | <0.001 | <0.001 | 0.328 |
| ΔIPSS 12 m | 13.14 ± 3.44 | 12.24 ± 4.03 | 11.38 ± 3.60 | 0.170 | 0.296 | 0.049 | 0.360 |
| ΔIPSS 24 m | 12.67 ± 4.48 | 11.10 ± 3.66 | 9.55 ± 2.98 | 0.006 | 0.092 | <0.001 | 0.064 |

*P, overall three-phenotype comparison; **P, glandular vs. mixed; ***P, glandular vs. stromal; ****P, mixed vs. stromal. IPSS, International Prostate Symptom Score; ΔQmax, maximum urinary flow rate.

## Discussion

This study demonstrates that MRI-defined prostatic morphological phenotypes are strongly associated with differential symptomatic and functional outcomes following PAE. Across the entire cohort, all morphology groups experienced significant improvements in ΔIPSS and ΔQmax, but the magnitude and timing of response varied markedly by phenotype. The glandular-dominant phenotype exhibited the most pronounced and earliest improvement, with maximal gains in both ΔIPSS and ΔQmax achieved by 6 months and sustained through 24 months. By contrast, stromal-dominant prostates exhibited smaller and delayed benefits, typically peaking at around 12 months, whereas mixed morphologies demonstrated intermediate responses. Multivariable regression confirmed that glandular-dominant morphology remained an independent predictor of greater ΔQmax after adjusting for age, baseline prostate volume, and baseline symptom severity, and interaction analysis revealed that this effect was especially pronounced among patients with severe baseline symptoms (IPSS ≥ 20). These findings collectively indicate that preprocedural MRI-based phenotyping can stratify functional recovery trajectories following PAE, providing non-invasive insight into the biological variability of the embolization response.

BPH demonstrates marked histological heterogeneity, with varying proportions of glandular epithelial and stromal components. This compositional variability correlates with differences in LUTS severity and therapeutic efficacy, including outcomes after PAE.[16,23] Considerable interpatient variation in post-PAE symptom relief and urinary flow was identified, which may be attributable to differences in nodular histologic architecture, underscoring the need to investigate morphology-specific treatment effects.

Prior studies have highlighted the prognostic relevance of AdBPH morphology. Patients with AdBPH have demonstrated significantly greater reductions in symptom scores and prostate volume following PAE than those exhibiting non-AdBPH phenotypes.[24] However, most existing studies have relied on binary classification frameworks (AdBPH vs. non-AdBPH) that overlook the histologic heterogeneity of hyperplastic nodules.

To address this limitation, we introduced a three-type MRI-based morphological classification—glandular, stromal, and mixed types—based on nodular composition and signal characteristics on T2-weighted sequences. Radiologic–pathologic correlation studies provide validation for this model. Notably, Dai et al.[16] demonstrated that stromal hyperplasia presents as low-signal, solid tissue on T2-weighted MRI, whereas glandular hyperplasia appears hyperintense with cystic components, facilitating non-invasive differentiation of tissue composition using MRI.

Histologically, glandular-dominant nodules exhibit a high epithelial-to-stromal ratio and a secretory glandular architecture, a histologic profile that demonstrates minimal variation across different prostate volumes.[25] By contrast, stromal-dominant nodules are characterized by an abundance of fibromuscular and fibroblastic elements, rich in smooth muscle and extracellular matrix, and minimal glandular content. Recent spatial and transcriptomic studies have confirmed this histologic phenotype, demonstrating increased expression of *ACTA2*, *VIM*, and collagen genes within stromal-dominant regions.[26,27]

This histological heterogeneity likely contributes to the variability in therapeutic efficacy observed among different nodule subtypes after PAE. Longitudinal analyses across the 24-month follow-up period revealed that prostates with glandular-dominant morphology exhibited significantly earlier and greater improvements in both the IPSS and Qmax than stromal-dominant or mixed types. Incorporating a tripartite MRI-based morphological classification thus allows a more refined interpretation of tissue-specific responses to embolization.

Descriptive and subgroup analyses revealed that glandular-dominant phenotypes reached peak improvements in ΔIPSS and ΔQmax at 6 months following PAE. By contrast, mixed and stromal-dominant phenotypes exhibited delayed peak responses at approximately 12 months. The slower improvement observed in stromal and mixed morphologies—characterized by a greater proportion of fibrous and smooth muscle elements—is likely attributable to reduced susceptibility to ischemic necrosis and delayed tissue remodeling following ischemic injury. Dense stromal tissue may resist ischemic degeneration and exhibit slower clearance of cellular debris, thereby postponing the onset of glandular atrophy and luminal decompression. Consequently, symptomatic relief and improvements in urinary flow emerge more gradually in these phenotypes than in glandular-dominant tissue, which undergoes more rapid ischemic shrinkage and volumetric regression. Preclinical canine models have demonstrated that prostatic artery occlusion induces acute hemorrhagic necrosis and inflammatory infiltration within 2 weeks, followed by progressive glandular atrophy and interstitial fibrotic remodeling by 6 months post-treatment.[28] Histological studies in human participants further support these findings, showing that ischemic necrosis, squamous metaplasia, and inflammation predominantly affect glandular regions, whereas stromal tissue displays relatively limited histopathologic alterations.[29]

At 24 months, both ΔIPSS and ΔQmax demonstrated a modest decline compared with earlier timepoints, indicating a gradual attenuation of the treatment effect across all morphology phenotypes. Despite this reduction in absolute effect size, interphenotype differences remained significant, with the glandular-dominant group consistently exhibiting the most pronounced and durable improvements in both symptomatic and functional outcomes. These findings reinforce the association between glandular-predominant morphology and greater, more sustained responsiveness to PAE. Although the overall treatment effect diminished slightly over time, clinically meaningful benefits relative to baseline were maintained in most patients. This longitudinal trend aligns with prior studies reporting that maximal symptomatic relief typically occurs within the first postoperative year, followed by mild attenuation while preserving overall therapeutic efficacy.[20]

Given that unadjusted analyses are inherently susceptible to confounding by baseline characteristics—such as age, baseline TPV, and initial IPSS or Qmax—multivariable regression modeling was employed to minimize these effects. These adjusted analyses confirmed that glandular-dominant morphology independently predicted greater symptomatic ΔIPSS and functional ΔQmax improvement, with the strongest effects observed during the first postoperative year. By contrast, the mixed-morphology phenotype showed only partial associations, displaying isolated gains in ΔQmax at certain timepoints but no consistent relationship with ΔIPSS. These findings suggest that mixed-type prostates provide variable and less durable functional benefits than the robust and sustained improvements observed in glandular-dominant cases. The predictive association between glandular morphology and ΔIPSS attenuated at 24 months, implying that long-term symptom relief may be influenced by additional pathophysiologic or behavioral factors beyond intrinsic nodule composition. This observation is consistent with findings from the STREAM trial, which reported that patients with stromal-predominant BPH experienced greater symptomatic improvement over prolonged follow-up despite the early advantages observed in adenomatous-dominant cases.[30]

The choice of 80 mL as the prostate volume threshold was driven by clinical convention rather than statistical optimization. From a surgical perspective, international LUTS/BPH guidelines recommend transurethral procedures (e.g., TURP and transurethral laser vaporization) primarily for prostate volumes in the range of 30–80 mL, whereas glands exceeding 80 mL generally require enucleation-based techniques or open/robotic simple prostatectomy, thereby defining ≥ 80 mL as "large-volume" BPH in routine practice. This surgical distinction has been increasingly included in the endovascular literature. Several prospective PAE studies have specifically evaluated cohorts with prostate volumes ≥ 80 mL, reporting meaningful symptom improvement and volume reduction in this population. Thus, using 80 mL as a stratification threshold allows our findings to be interpreted within both guideline-supported clinical boundaries and the established PAE evidence base. Importantly, this classification does not imply a biological transition at 80 mL but serves as a pragmatic cut-off that reflects current clinical decision-making.[18,31,32]

Subgroup analyses stratified by baseline TPV (< 80 vs. ≥ 80 mL) were directionally consistent with the overall results, confirming a morphology-dependent gradient in treatment efficacy. The most pronounced and consistent differences were observed between glandular- and stromal-dominant phenotypes, with glandular-dominant prostates showing superior functional improvements across follow-up intervals. Differences between glandular and mixed morphologies achieved statistical significance at selected timepoints, reflecting the inherent histologic heterogeneity and variable responsiveness characteristic of mixed-type tissue composition.

Consistent with the multivariable regression analyses, mixed and stromal-dominant phenotypes demonstrated no significant differences in most subgroup comparisons, particularly for ΔIPSS. Among patients with larger prostates (≥ 80 mL), significant intergroup variation in ΔQmax emerged as early as 3 months after PAE and persisted throughout the follow-up period. By contrast, patients with smaller prostates (< 80 mL) exhibited no early intergroup differences. These findings highlight a morphology- and volume-dependent gradient in treatment efficacy, supporting the value of combining imaging-based phenotyping with volumetric assessment for personalized therapeutic planning in BPH.

Interaction analysis revealed that only the coexistence of a baseline IPSS ≥ 20 and glandular-dominant morphology independently predicted greater symptomatic improvement at 24 months. Neither baseline symptom severity nor glandular morphology alone showed a significant association with long-term outcomes. These findings suggest that patients with both severe baseline symptoms and glandular-dominant nodules are most likely to achieve durable clinical benefit after PAE. In agreement with prior studies, individuals presenting with more severe LUTS demonstrated the largest long-term reductions in IPSS following embolization.[33,34]

From a clinical perspective, these results underscore the utility of MRI-based prostate morphology classification as a practical tool for patient stratification. Glandular-dominant phenotypes were associated with the most consistent improvements in both symptom scores and urinary flow, whereas mixed phenotypes showed intermediate responses, and stromal-dominant morphology was linked to the least favorable outcomes. Incorporating MRI-based tissue phenotyping into preprocedural evaluation may therefore enhance individualized treatment planning and improve the prediction of clinical response in the management of BPH.

This study has several limitations. First, the morphological classification of BPH nodules was based on a reader-dependent visual assessment of T2-weighted MRI signal characteristics without histopathological validation, introducing potential interobserver variability. Second, the retrospective and single-center design inherently increases the risk of selection bias and residual confounding despite statistical adjustment using multivariable regression modeling. Third, longitudinal measurements of post-void residual volume and prostate size were inconsistently obtained during follow-up, limiting evaluation of their temporal relationships with clinical outcomes. Fourth, other potentially relevant modifiers of treatment response—such as urodynamic parameters, intravesical prostatic protrusion, and systemic or local inflammatory markers—were not routinely assessed or included in the analysis. These unmeasured factors may contribute to interpatient heterogeneity in treatment outcomes and warrant systematic evaluation in future prospective research. Finally, the post-PAE outcome assessment in this retrospective cohort was limited to symptomatic and functional parameters (ΔIPSS and ΔQmax) without uniform post-contrast MRI to quantify ischemic necrosis. Although quantitative necrosis mapping could further elucidate the mechanistic link between baseline morphology and treatment response, post-emboliza-

tion contrast-enhanced imaging was inconsistently obtained in routine care, precluding robust volumetric analysis without selection bias. Future prospective studies with standardized post-PAE MRI (e.g., contrast-enhanced, perfusion, or quantitative mapping) are warranted to test whether morphology influences the extent and pattern of tissue infarction and mediates clinical benefits.

In conclusion, the present findings demonstrate that the histologic composition inferred from the dominant imaging characteristics of transition zone nodules influences both symptomatic and functional outcomes after PAE. Multiparametric MRI, assessed primarily on T2WI, offers a potential non-invasive method for characterizing prostate tissue composition and guiding treatment selection. Although further validation in prospective, multicenter cohorts is warranted, this imaging-based phenotyping framework may assist in identifying histologic phenotypes most likely to achieve favorable therapeutic responses.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Knight GM, Talwar A, Salem R, Mouli S. Systematic review and meta-analysis comparing prostatic artery embolization to gold-standard transurethral resection of the prostate for benign prostatic hyperplasia. *Cardiovasc Intervent Radiol*. 2021;44(2):183-193. [Crossref]
2. Alemu MH. Transurethral resection of the prostate (TURP)--in the treatment of benign prostatic hypertrophy (BPH) in Mekelle, Ethiopia. *Ethiop Med J*. 2009;47(1):65-69. [Crossref]
3. Ray AF, Powell J, Speakman MJ, et al. Efficacy and safety of prostate artery embolization for benign prostatic hyperplasia: an observational study and propensity-matched comparison with transurethral resection of the prostate (the UK-ROPE study). *BJU Int*. 2018;122(2):270-282. [Crossref]
4. Carnevale FC, Moreira AM, de Assis AM, et al. Prostatic artery embolization for the treatment of lower urinary tract symptoms due to benign prostatic hyperplasia: 10 years' experience. *Radiology*. 2020;296(2):444-451. [Crossref]
5. Müllhaupt G, Hechelhammer L, Graf N, et al. Prostatic artery embolisation versus transurethral resection of the prostate for benign prostatic obstruction: 5-year outcomes of a randomised, open-label, noninferiority trial. *Eur Urol Focus*. 2024;10(5):788-795. [Crossref]
6. Sun F, Lucas-Cava V, Sánchez-Margallo FM. Clinical predictive factors in prostatic artery embolization for symptomatic benign prostatic hyperplasia: a comprehensive review. *Transl Androl Urol*. 2020;9(4):1754-1768. [Crossref]
7. 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]
8. Gaudiano C, Rustici A, Corcioni B, et al. PI-RADS version 2.1 for the evaluation of transition zone lesions: a practical guide for radiologists. *Br J Radiol*. 2022;95(1131):20210916. [Crossref]
9. Li MJ, Hsu HS, Liang RC, Lin SY. Infrared microspectroscopic detection of epithelial and stromal growth in the human benign prostatic hyperplasia. *Ultrastruct Pathol*. 2002;26(6):365-370. [Crossref]
10. Little MW, Boardman P, Macdonald AC, et al. Adenomatous-dominant benign prostatic hyperplasia (AdBPH) as a predictor for clinical success following prostate artery embolization: an age-matched case-control study. *Cardiovasc Intervent Radiol*. 2017;40(5):682-689. [Crossref]
11. Boschheidgen M, Al-Monajjed R, Minko P, et al. Influence of benign prostatic hyperplasia patterns detected with MRI on the clinical outcome after prostatic artery embolization. *CVIR Endovasc*. 2023;6(1):9. [Crossref]
12. Chatterjee A, Gallan AJ, He D, et al. Revisiting quantitative multi-parametric MRI of benign prostatic hyperplasia and its differentiation from transition zone cancer. *Abdom Radiol (NY)*. 2019;44(6):2233-2243. [Crossref]
13. Zumstein V, Betschart P, Vetterlein MW, et al. Prostatic artery embolization versus standard surgical treatment for lower urinary tract symptoms secondary to benign prostatic hyperplasia: a systematic review and meta-analysis. *Eur Urol Focus*. 2019;5(6):1091-1100. [Crossref]
14. Noworolski SM, Vigneron DB, Chen AP, Kurhanewicz J. Dynamic contrast-enhanced MRI and MR diffusion imaging to distinguish between glandular and stromal prostatic tissues. *Magn Reson Imaging*. 2008;26(8):1071-1080. [Crossref]
15. Oto A, Kayhan A, Jiang Y, et al. Prostate cancer: differentiation of central gland cancer from benign prostatic hyperplasia by using diffusion-weighted and dynamic contrast-enhanced MR imaging. *Radiology*. 2010;257(3):715-723. [Crossref]
16. Dai JC, Morgan TN, Goueli R, et al. MRI features associated with histology of benign prostatic hyperplasia nodules: generation of a predictive model. *J Endourol*. 2022;36(3):381-386. [Crossref]
17. Li M, Gao Q, Yu T. Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC Cancer*. 2023;23(1):799. [Crossref]
18. Bhatia S, Sinha VK, Harward S, Gomez C, Kava BR, Parekh DJ. Prostate artery embolization in patients with prostate volumes of 80 ml or more: a single-institution retrospective experience of 93 patients. *J Vasc Interv Radiol*. 2018;29(10):1392-1398. [Crossref]
19. Veyg D, Mohanka R, Rumball IP, et al. Comparison of 24-month clinical outcomes after prostatic artery embolization in prostate glands larger versus smaller than 80 mL: a systematic review. *J Vasc Interv Radiol*. 2023;34(4):578-584.e1. [Crossref]
20. Neymark AI, Karpenko AA, Neymark BA, et al. Superselective prostatic artery embolization in the treatment of large benign prostatic hyperplasia. *Urologia*. 2021;88(4):374-381. [Crossref]
21. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349. [Crossref]
22. Yassin A, Kelly D, Nettleship J, et al. Testosterone treatment and change of categories of the International prostate symptom score (IPSS) in hypogonadal patients: 12 years prospective controlled registry study. *Aging Male*. 2023;26(1):2220567. [Crossref]
23. Diaz TA, Benson B, Clinkenbeard A, Long JR, Kawashima A, Yano M. MRI evaluation of patients before and after interventions for benign prostatic hyperplasia: an update. *AJR Am J Roentgenol*. 2022;218(1):88-99. [Crossref]
24. Strand DW, Costa DN, Francis F, Ricke WA, Roehrborn CG. Targeting phenotypic heterogeneity in benign prostatic hyperplasia. *Differentiation*. 2017;96:49-61. [Crossref]
25. Bauman TM, Nicholson TM, Abler LL, et al. Characterization of fibrillar collagens and extracellular matrix of glandular benign prostatic hyperplasia nodules. PLoS One. 2014;9(10):e109102. [Crossref]
26. Cao Y, Zhang H, Tu GL, et al. The symptoms of benign prostatic hyperplasia patients with stromal-dominated hyperplasia nodules may be associated with prostate fibrosis. Int J Gen Med. 2023;16:1181-1191. [Crossref]
27. Pollack AS, Kunder CA, Brazer N, et al. Spatial transcriptomics identifies candidate stromal drivers of benign prostatic hyperplasia. JCI Insight. 2024;9(2):e176479. [Crossref]
28. Lucas-Cava V, Sanchez-Margallo FM, Moreno-Lobato B, et al. Prostatic artery occlusion: initial findings on pathophysiological response in a canine prostate model. Transl Androl Urol. 2022;11(12):1655-1666. [Crossref]
29. Camara-Lopes G, Mattedi R, Antunes AA, et al. The histology of prostate tissue following

prostatic artery embolization for the treatment of benign prostatic hyperplasia. Int Braz J Urol. 2013;39(2):222-227. [Crossref]

30. Tapping CR, Little MW, Macdonald A, et al. The STREAM Trial (Prostatic Artery Embolization for the Treatment of Benign Prostatic Hyperplasia) 24-month clinical and radiological outcomes. Cardiovasc Intervent Radiol. 2021;44(3):436-442. [Crossref]
31. Wang M, Guo L, Duan F, et al. Prostatic arterial embolization for the treatment of lower urinary tract symptoms caused by benign prostatic hyperplasia: a comparative study of medium- and large-volume prostates. BJU Int. 2016;117(1):155-164. [Crossref]
32. Gravas S, Gacci M, Gratzke C, et al. Summary paper on the 2023 European Association of Urology Guidelines on the management of non-neurogenic male lower urinary tract symptoms. Eur Urol. 2023;84(2):207-222. [Crossref]
33. Thulasidasan N, Kok HK, Elhage O, Clovis S, Popert R, Sabharwal T. Prostate artery embolisation: an all-comers, single-operator experience in 159 patients with lower urinary tract symptoms, urinary retention, or haematuria with medium-term follow-up. Clin Radiol. 2019;74(7):569.e1-569.e8. [Crossref]
34. Moreira AM, de Assis AM, Carnevale FC, Oliveira DS, Antunes AA. Improvements in irritative versus obstructive symptoms of the international prostate symptom score after prostatic artery embolization in 174 patients, in a single center. Cardiovasc Intervent Radiol. 2020;43(4):613-619. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253585

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE



# The safety and efficacy of the percutaneous balloon-expandable biodegradable magnesium biliary stents in patients with liver transplant

Ramazan Kutlu
Nurullah Dağ
Eldiiar Saparbekov

İnönü University Faculty of Medicine, Department of Radiology and Liver Transplantation Institute, Malatya, Türkiye

**PURPOSE**

This study aims to present our institutional experience with the use of percutaneous balloon-expandable biodegradable magnesium biliary stents (MBS) in the treatment of benign, refractory anastomotic strictures (AS), where initial percutaneous or endoscopic approaches fail or are ineffective, after liver transplantation (LT).

**METHODS**

In this retrospective single-center study, 13 patients with clinically refractory AS who underwent MBS placement between July 2021 and August 2024 were evaluated. Statistical analyses included Kaplan–Meier survival analysis for patency and Spearman's correlation for recurrence risk. Primary outcomes included stricture recurrence and time to reintervention during follow-up.

**RESULTS**

The median age of patients was 35 years [interquartile range (IQR): 22–48], and 11 were male (85%). Living donor LT was performed in 12 (92%). The MBS were placed at a median of 8 months post-transplant (IQR: 5–44), with a technical success in all cases (100%). Before MBS placement, patients underwent a median of one endoscopic retrograde cholangiopancreatography (IQR: 0–3) and three percutaneous transhepatic biliary drainage procedures (IQR: 1–8). The median follow-up was 25 months (IQR: 15–33). The MBS patency rates were 93%, 85%, and 67% at 6, 12, and 24 months, respectively. Stricture recurrence occurred at a median of 30 months post-placement (95% confidence interval: 23.6–36.3). A moderate positive correlation was observed between the number of pre-stent interventions and recurrence risk (rho: 0.582, $P = 0.023$). Post-procedural complications (cholangitis) occurred in 1 patient.

**CONCLUSION**

Balloon-expandable biodegradable MBS may provide a safe and effective treatment for refractory AS following LT. Early placement of MBS, particularly after fewer prior interventions, appears to be associated with improved patency and longer stricture-free survival.

**CLINICAL SIGNIFICANCE**

Early use of biodegradable magnesium stents after LT may reduce the need for multiple interventions and improve long-term biliary patency.

**KEYWORDS**

Biodegradable stent, magnesium, benign biliary anastomotic stricture, biliary stent, percutaneous transhepatic biliary drainage, liver transplantation

**Handling editor:** Bahri Üstünsöz

**Corresponding author:** Ramazan Kutlu

**E-mail:** ramazan.kutlu@inonu.edu.tr

Received 28 July 2025; revision requested 21 August 2025; last revision received 21 September 2025; accepted 25 September 2025.

Epub: 20.10.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253585

You may cite this article as: Kutlu R, Dağ N, Saparbekov E. The safety and efficacy of the percutaneous balloon-expandable biodegradable magnesium biliary stents in patients with liver transplant. *Diagn Interv Radiol*. 2026;32(4):473-478.

Biliary anastomotic strictures (AS) represent one of the most frequent and clinically significant complications following liver transplantation (LT).[1] These strictures typically arise from factors such as surgical technique, graft-recipient bile duct size mismatch, or ischemic injury to the biliary epithelium, and they contribute substantially to post-transplant morbidity and mortality.[2] The reported incidence of AS varies by transplant type, occurring in approximately 15% of patients after deceased donor LT (DDLT) and ranging from 19% to 40% following living donor LT (LDLT), where biliary reconstruction is often more technically challenging.[3]

In patients with AS following LT, the first-line treatment is typically endoscopic retrograde cholangiopancreatography (ERCP) with balloon dilatation and plastic stent placement.[4,5] When ERCP is unsuccessful or not feasible—such as in patients with altered anatomy or inaccessible anastomoses—percutaneous transhepatic biliary drainage (PTBD) combined with bilioplasty and stent insertion is employed as a more invasive but effective alternative.[6,7] Although these interventions are generally efficacious, they often require multiple sessions or prolonged catheterization, which can negatively impact the patient's quality of life and increase the risk of complications.[8,9]

Biodegradable stents (BS) have been successfully employed in the management of benign gastrointestinal and vascular conditions, and their application in the biliary system represents a relatively recent and innovative therapeutic approach.[10,11] Compared with conventional non-degradable stents, BS are associated with lower complication rates, including reduced risks of infection, stent occlusion, migration, and chronic inflammation. Moreover, they eliminate the need for a secondary intervention for stent removal, as the material gradually degrades *in situ*.[12] Given the high incidence of AS and the frequent need for repeated biliary interventions following LT, BS represent a promising alternative to traditional endobiliary devices. Magnesium biliary stents (MBS) biodegrade naturally, eliminating the need for retrieval while maintaining biocompatibility and optimal mechanical properties.[13-17] Their degradation products are safely excreted and may provide anti-inflammatory, antimicrobial, and healing benefits.[13,15,18] This study presents our institutional experience with magnesium-based BS in the treatment of post-transplant AS and aims to evaluate their safety and efficacy.

**Main points**

- Percutaneous balloon-expandable biodegradable magnesium biliary stents (MBS) are effective and safe for treating biliary strictures after liver transplantation.
- Biodegradable MBS placed after fewer interventional procedures demonstrate higher patency rates and longer stricture-free intervals.
- Early placement of biodegradable magnesium stents in refractory anastomotic strictures is associated with improved clinical outcomes and reduced recurrence risk.

## Methods

### Study design

This single-center retrospective study was approved by İnönü University Ethics Committee (approval number 2025/6784, date: 03.01.2025). Written informed consent was obtained from all patients prior to interventional procedures. Thirteen patients treated for AS following LT between July 2021 and August 2024 were included. The inclusion criteria were as follows: patients with clinically and radiologically confirmed benign AS following LT; failure of previous endoscopic or percutaneous management to maintain long-term patency. The exclusion criteria were as follows: evidence of hepatic artery thrombosis; malignant strictures; or inadequate follow-up data. No patients were lost to follow-up. All percutaneous procedures were performed by the same interventional radiologist.

### Patient management and intervention protocol

Management decisions were determined through a multidisciplinary approach involving interventional radiologists, transplant surgeons, and gastroenterologists. AS were confirmed by ultrasonography (US) and magnetic resonance cholangiopancreatography (MRCP) in patients with clinical or laboratory suspicion of biliary complications.

When feasible, ERCP was attempted as the first-line intervention. If ERCP failed or clinical improvement was not achieved after repeated sessions, PTBD was performed; furthermore, PTBD was the first-line treatment in patients with hepaticojejunostomy (HJ)-type biliary anastomosis. Refractory AS was defined as persistent stricture after multiple PTBD sessions, and BS placement was considered in these cases.

All PTBD procedures were performed under sterile conditions with intravenous sedation and analgesia, with the patient in the supine position. Access to the biliary system was obtained via a transhepatic approach using a 0.018-inch guidewire and a coaxial access system (Accustick, Boston Scientific, Marlborough, MA, USA) through an appropriate intercostal space, under US and fluoroscopic guidance.

After access, contrast cholangiography was used to delineate biliary anatomy and stricture severity. In the initial session, an internal-external biliary drainage catheter (Flexima, Boston Scientific) was placed for decompression. Drainage output was monitored daily, with routine catheter flushing and maintenance.

In subsequent sessions, under anesthesia, balloon dilatation of the stricture was performed using angioplasty balloons inflated to 18–20 atmospheres for approximately 2 minutes. A multi-hole internal–external catheter was then reinserted. The balloon diameter was selected to match the normal proximal (upstream) bile duct caliber, and drainage catheter size was chosen accordingly to avoid overdistension while ensuring sufficient scaffolding.

If the stricture persisted during follow-up, a balloon-expandable biodegradable MBS (UNITY-B, Q3 Medical, Dublin, Ireland; CE certified) was deployed. One patient with two HJ anastomoses and dual strictures received two stents.

Technical success was defined as the absence of residual stricture on completion of cholangiography after stent placement.

The percutaneous catheter was not immediately removed post-stent placement, allowing early complication management. It was withdrawn within 24–48 hours after confirming normal findings on follow-up cholangiography and clinical assessment.

### Follow-up and outcome measures

All patients were followed at 1 month post-procedure, then every 3 months during the 1st year in the absence of symptoms. After the 1st year, visits were scheduled every 6 months. Routine follow-up included liver function tests and hepatobiliary US. In the presence of clinical signs of cholangitis, abnormal laboratory values, or biliary dilatation on US, MRCP was performed to assess for stricture recurrence.

Recurrence of AS was defined as a new anastomotic-level stricture on MRCP requiring intervention. Clinical success was defined as BS patency without AS recurrence during follow-up.

### Statistical analysis

Statistical analyses were performed using the SPSS 22.0 software (IBM Corp., Armonk, NY, USA) package. The Kolmogorov–Smirnov test was used to assess the normality of the data distribution. As the data did not follow a normal distribution, quantitative variables were presented as median [interquartile range (IQR)]. Due to the violation of parametric test assumptions, the Mann–Whitney U test or the Wilcoxon signed-rank test was used for comparisons of quantitative variables within dependent groups. Correlations between two quantitative variables were analyzed using Spearman's rho coefficient. Categorical variables were expressed as frequencies and percentages (%). The recurrence rate of AS was assessed using the Kaplan–Meier method. A *P* value of <0.05 was considered indicative of statistical significance in all analyses.

## Results

A total of 13 patients (11 men, 2 women; ratio 5.5:1) were included. The median age at the time of LT was 35 years (IQR: 22–48). LDLT was performed in 12 patients (92%), and DDLT in 1 patient (8%). Indications for LT included hepatitis B (n = 6), hepatitis C (n = 3), fulminant liver failure (n = 2), and hepatocellular carcinoma (n = 2). The type of biliary anastomosis was duct-to-duct in 7 patients and HJ in seven anastomoses (1 patient had two HJ anastomoses). Table 1 presents the demographic and clinical characteristics of the study population. Stent placement occurred at a median of 8 months after LT (IQR: 5–44). The technical success rate was 100%. Figures 1–3 demonstrate representative cholangiographic images before and after MBS placement.

Pre-stent interventions included a median of one ERCP session (IQR: 0–3) and three PTBD procedures (IQR: 1–8). The median follow-up period after MBS placement was 25 months (IQR: 15–33) (Figure 4). Stricture recurrence occurred at a median of 30 months post-stenting (standard error: 3.2; 95% confidence interval: 23.6–36.3) (Figure 5). The BS patency rates were 93% at 6 months, 85% at 12 months, and 67% at 24 months. Among patients with recurrence, the shortest stent patency was 6 months, and the longest was 35 months. Three patients remained recurrence-free throughout follow-up (7, 36, and 45 months) (Figure 6, Table 2). There was a moderate positive correlation between the total number of interventional procedures prior to BS placement and stricture recurrence (Spearman's rho: 0.582, $P = 0.023$).

**Table 1.** Demographic characteristics and clinical information of the patients

| Variables | |
|---|---|
| Age (median, IQR) (years) | 35 (22–48) |
| Sex (male/female) | 11/2 |
| Liver transplantation indication | Hepatitis B (n = 6),<br>Hepatitis C (n = 3),<br>Fulminant liver failure (n = 2),<br>Hepatocellular carcinoma (n = 2) |
| Type of transplant<br>LDLT<br>DDLT | <br>12<br>1 |
| Biliary anastomosis<br>DD<br>HJ | <br>7<br>7 |

Continuous data are expressed as median. IQR, interquartile range; LDLT, living donor liver transplantation; DDLT, deceased donor liver transplantation, DD: duct-to-duct; HJ, hepaticojejunostomy.

**Table 2.** Procedural details and clinical outcomes of the patients

| Variables | |
|---|---|
| Time from LT to BS (median, IQR) (months) | 8 (5–44) |
| Number of ERCP(s) before BS (median, IQR) | 1 (0–3) |
| Number of PTBD(s) before BS (median, IQR) | 3 (1–8) |
| Clinical success rate (%)<br>6 months<br>12 months<br>24 months | <br>92%<br>85%<br>67% |
| Follow-up without recurrence after BS | 3 patients (7, 36, 45 months) |
| Complications (n) | 1 (cholangitis) |

Continuous data are expressed as median. IQR, interquartile range; LT, liver transplantation; BS, biodegradable stent; ERCP, endoscopic retrograde cholangiopancreatography; PTBD, percutaneous transhepatic biliary drainage.

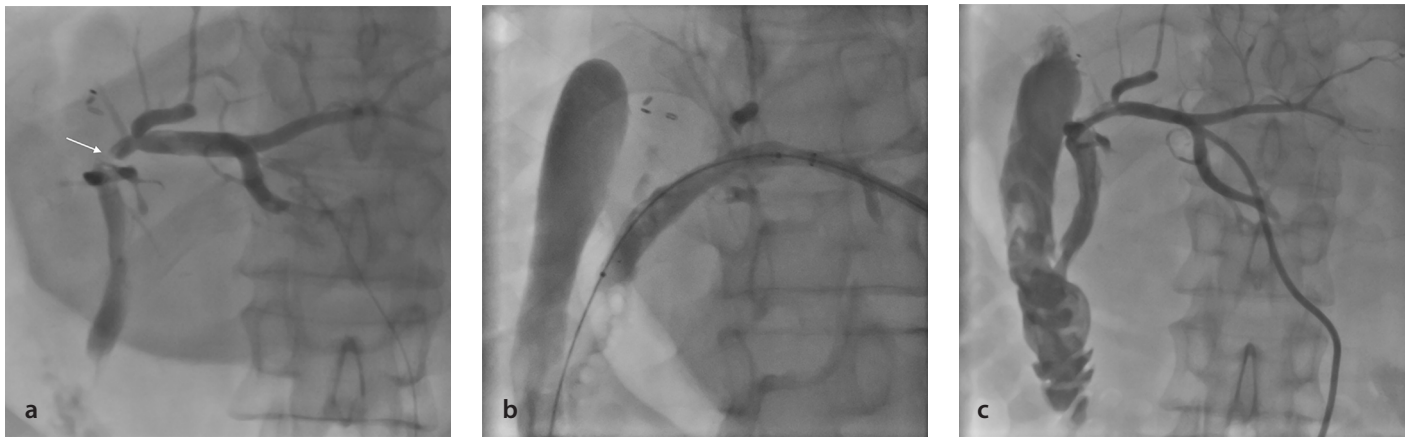



**Figure 1.** Cholangiogram of a left lobe recipient living donor liver transplantation patient demonstrating a significant duct-to-duct anastomosis stenosis (white arrow) **(a)**. Implantation cholangiogram showing successful deployment of balloon-expandable biodegradable magnesium biliary stent **(b)**. Cholangiogram after placement of an internal-external biliary drainage catheter for safety, demonstrating a reduction in the biliary dilatation **(c)**.

Among a predefined list of potential complications (pancreatitis, bleeding, perforation, stent migration, bile duct or duodenal abrasion, misplacement, severe pain, bile occlusion, and cholangitis), only one case of cholangitis was observed. No major complications, such as stent migration, bleeding, or ductal injury, occurred.

There was a statistically significant decrease in both total bilirubin and direct bilirubin levels from the time of BS placement to 1 month post-placement ($P < 0.05$). Although alanine transaminase, aspartate transaminase, alkaline phosphatase, and gamma-glutamyl-transpeptidase levels also decreased, these changes were not statistically significant ($P > 0.05$) (Table 3).

## Discussion

BS have shown promise in the treatment of benign gastrointestinal, coronary, and peripheral vascular stenoses.[16] Their application in the biliary tract, however, is relatively recent. BS can be categorized into two groups: organic stents, which are made from materials such as polycaprolactone and polylactic acid, offering benefits such as biocompatibility, elasticity, and flexibility; and metallic stents, built from magnesium, iron, and zinc, providing advantages in biocompatibility and superior mechanical properties.[13,16]

Currently, three Conformité Européene-marked biodegradable biliary stents are commercially available: Ella (Ella-CS, Czech Republic), Archimedes (Q3 Medical, Ireland), and UNITY-B (Q3 Medical, Ireland). The Ella stent has been evaluated in several small studies for benign biliary strictures and shown to have good clinical outcomes, with low rates of epithelial proliferation and stent-related complications.[19-22] However, evidence remains limited regarding the use of BS—particularly magnesium-based BS—in the setting of AS following LT. The mid- and long-term performance of these devices in this specific population remains underexplored.

There is currently no standardized protocol for managing AS after LT. The primary goal of treatment is to achieve long-term biliary patency while minimizing reintervention, prolonged catheterization, and the need for surgical revision.[3] Although both endoscopic and percutaneous interventions are effective, multiple sessions for catheter

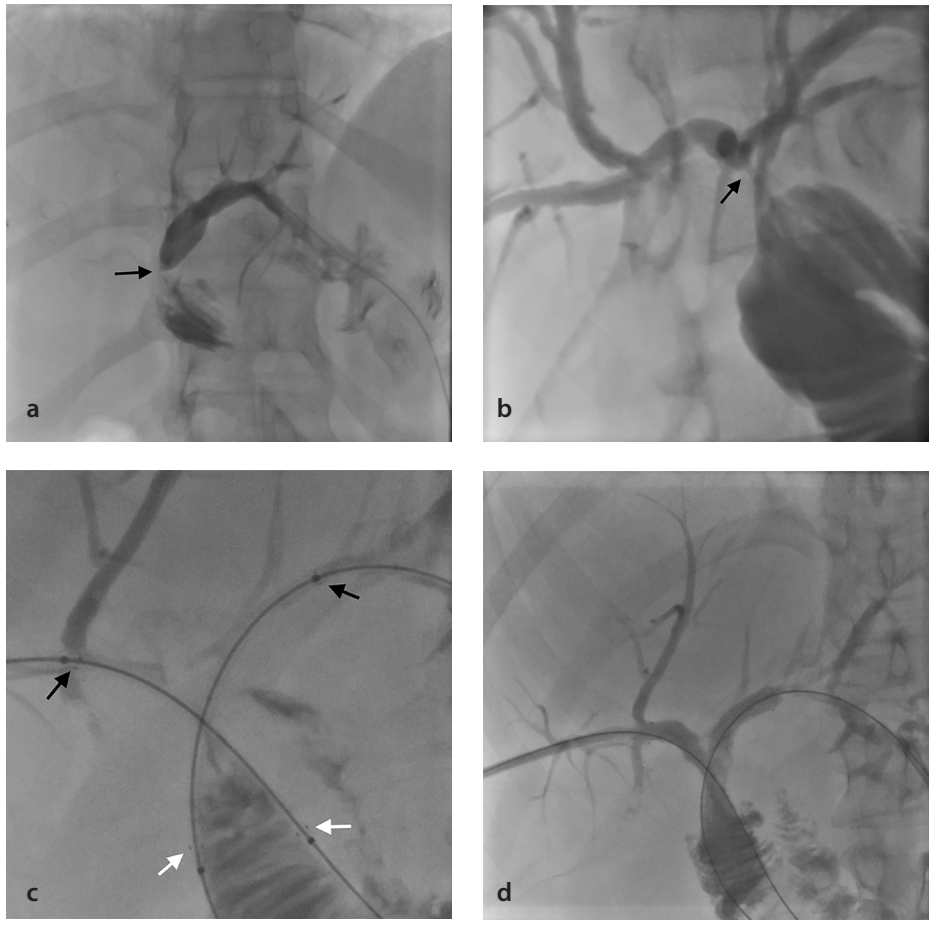



**Figure 2.** Cholangiograms showing significant stenoses (black arrows) in the left **(a)** and right **(b)** main hepatic duct to hepaticojejunostomy anastomosis in a cadaveric liver transplant recipient. Simultaneous implantation of two stents was performed in the patient (black arrows point to the proximal and white arrows point to the distal ends of the stents) **(c)**. Post-implantation control cholangiogram demonstrates complete resolution of stenoses and patency of implanted stents **(d)**.

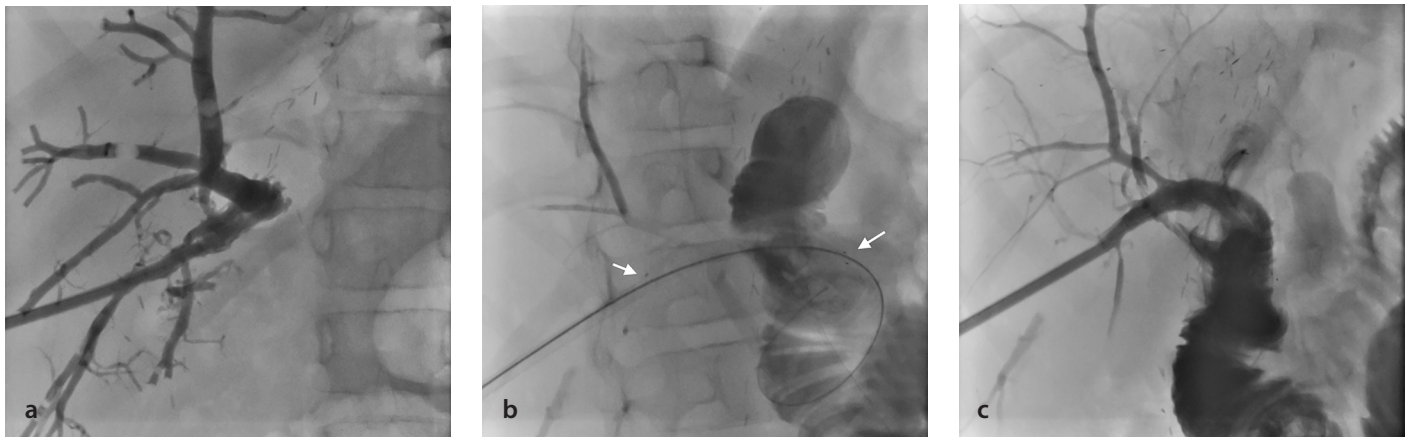



**Figure 3.** Cholangiogram showing marked dilatation and stenosis of hepaticojejunostomy (HJ) anastomosis in a pediatric patient who underwent a right lobe living donor liver transplantation **(a)**. A biodegradable magnesium biliary stent (white arrows) was implanted through the HJ anastomosis **(b)**. Post-implantation control cholangiogram showing resolution of stenosis and normal caliber of intrahepatic bile ducts **(c)**.

exchanges, stent removals, or repeat dilatations negatively affect quality of life and increase the risk of procedure-related complications.[23,24] In this context, BS offer a significant advantage, especially in refractory cases.

Our study supports the feasibility and effectiveness of BS in this setting, with a 100% technical success rate, 85% clinical success within 1 year, and a median stent patency of 30 months. These outcomes are consistent with prior reports. For example, Dopazo et al.[26] reported a 40% recurrence over a median 23-month follow-up in 10 patients with LT treated with Ella stents, and Battistel et al.[25] reported a 72% success rate with minimal complications in 18 patients with LT with percutaneously placed BS. A large multicenter study (n = 159) demonstrated stent patency rates of 86%, 79%, and 78% at 12, 36, and 60 months, respectively, though only 4.6% of patients in that cohort had undergone LT.[27]

Unlike previous studies, we used the UNITY-B stent, a balloon-expandable, hybrid BS consisting of a magnesium alloy core (MgNdMn21) with a polymer coating, designed for percutaneous and endoscopic implantation. The percutaneous version of the stent was used in the study. MBS degrade naturally *in vivo*, eliminating the need for retrieval and showing high biocompatibility, minimal inflammatory reaction, and low migration risk.[13,17] They also exhibit favorable mechanical properties, including adequate radial force, thin strut profiles, and enhanced inner lumen area. Additionally, magnesium degradation products are safely excreted and may confer anti-inflammatory, antibacterial, and tissue healing properties.[14,16] However, rapid corrosion in bile remains a significant challenge. The alkaline, enzyme-rich environment of bile accelerates magnesium degradation, potentially limiting stent durability. Advances in surface coatings and alloying elements (e.g., zinc, calcium) are necessary to enhance corrosion resistance and extend functional life in the biliary system.[13,15,16]

Our study population was predominantly LDLT recipients (92%), a group known to have more complex biliary anatomy and stricter anastomoses than DDLT recipients. The finding in our study that there was a patient who developed AS recurrence 6 months after BS, as well as a patient who was followed up for 45 months without recurrence, supports that this situation is due to multifactorial factors, which may partially explain the reduced mid-term patency rates. The tim-

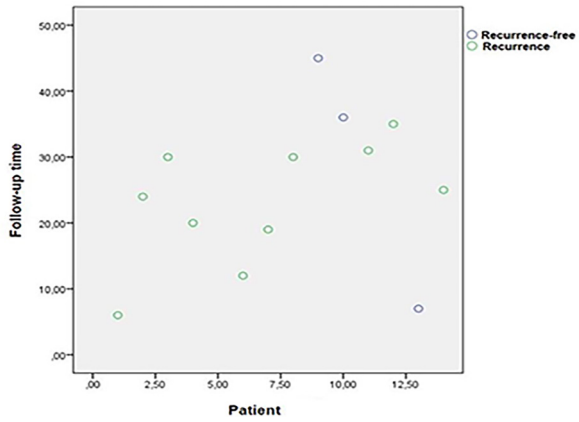



**Figure 4.** Scatter plot depicting individual patient follow-up time. Each data point represents a patient, categorized as either recurrence (green) or recurrence-free (blue).

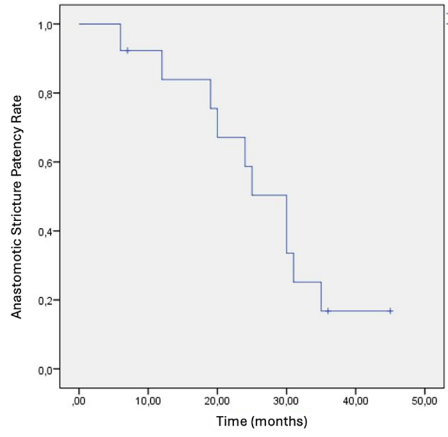



**Figure 5.** Graphical representation of anastomotic stricture patency rates after biodegradable stenting as estimated by Kaplan–Meier survival analysis.

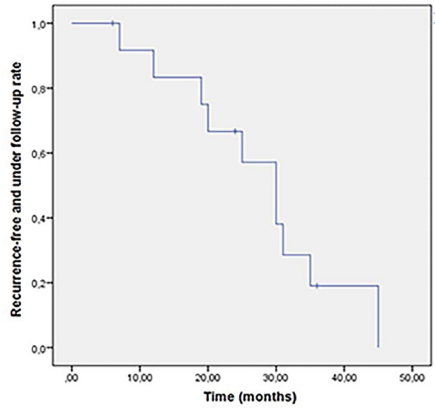



**Figure 6.** Graphical representation of recurrence-free and under follow-up rate after biodegradable stenting as estimated by reverse Kaplan–Meier survival analysis.

**Table 3.** Pre- and post-stent laboratory values of the patients

| Laboratory tests | Time of BS placement | First month after BS placement | *P* |
|---|---|---|---|
| Total bilirubin, mg/dL | 2.1 (0.8–2.3) | 1.0 (0.7–1.3) | 0.019 |
| Conjugated bilirubin, lmol/L | 0.7 (0.4–1.9) | 0.3 (0.2–0.6) | 0.034 |
| AST, U/L | 53 (38–78) | 31 (24–64) | 0.333 |
| ALT, U/L | 65 (45–122) | 44 (29–84) | 0.268 |
| ALP, U/L | 187 (131–268) | 148 (111–259) | 0.125 |
| GGT, U/L | 164 (96–237) | 105 (42–187) | 0.112 |

Continuous data are expressed as median interquartile range. BS, biodegradable stent; AST, aspartate transaminase; ALT, alanine transaminase; ALP, alkaline phosphatase; GGT, gamma-glutamyl-transpeptidase.

ing of BS placement also appears to influence outcomes; patients with earlier stent deployment after fewer failed interventions had longer stricture-free intervals. These findings suggest that early BS placement, rather than prolonged attempts at repeated ERCP or PTBD, may optimize clinical outcomes.

Complication rates were low and comparable with existing literature.[2] Only one patient developed cholangitis, and no serious adverse events, such as stent migration or bile duct perforation, were observed in the present study.

This study has several limitations. First, the small sample size limits the statistical power and generalizability of the results. In addition, variability in time from LT to BS placement and the number of prior procedures complicates analysis and underscores the need for standardized treatment pathways. Further analysis of risk factors for AS recurrence after BS placement was not possible with the available data set. As stent visibility was not the main focus of our study due to the retrospective design, we do not have precise information on stent visibility during the resorption process. In this context, future prospective studies may provide more comprehensive data to evaluate the biological lifecycle of BS. Finally, our study did not directly compare BS with other treatment strategies, such as repeat bilioplasty or surgery. Prospective, multicenter studies with larger sample sizes and longer follow-up are needed to fully establish the role of MBS in the treatment of AS after LT.

Magnesium-based biodegradable biliary stents appear to be a safe and effective treatment option for refractory AS following LT. Our findings suggest that earlier stent placement, particularly in patients with fewer prior interventions, may result in prolonged stent patency and extended stricture-free survival. Although further comparative studies are warranted, MBS may offer a promising minimally invasive alternative in the evolving management of post-transplant biliary complications.

### Footnotes

### Conflict of interest disclosure

Nurullah Dağ was not serving in any editorial role at the time of the manuscript's submission, evaluation, acceptance, or ahead of print publication. At the time of inclusion of the article in the regular issue, he is serving as Section Editor of the journal. Nurullah Dağ was not involved in the editorial evaluation, peer-review process, or publication decision of this manuscript. The other authors declare no conflict of interest.

## References


1. Brookmeyer CE, Bhatt S, Fishman EK, Sheth S. Multimodality imaging after liver transplant: top 10 important complications. *Radiographics*. 2022;42(3):702-721. [Crossref]
2. Senter-Zapata M, Khan AS, Subramanian T, et al. Patient and graft survival: biliary complications after liver transplantation. *J Am Coll Surg*. 2018;226(4):484-494. [Crossref]
3. Akamatsu N, Sugawara Y, Hashimoto D. Biliary reconstruction, its complications and management of biliary complications after adult liver transplantation: a systematic review of the incidence, risk factors and outcome. *Transpl Int*. 2011;24(4):379-392. [Crossref]
4. Boeva I, Karagyozov PI, Tishkov I. Post-liver transplant biliary complications: current knowledge and therapeutic advances. *World J Hepatol*. 2021;13(1):66-79. [Crossref]
5. Bofill A, Cárdenas A. A practical approach to the endoscopic management of biliary strictures after liver transplantation. *Ann Hepatol*. 2024;29(2):101186. [Crossref]
6. Lee IJ, Lee JH, Kim SH, et al. Percutaneous transhepatic treatment for biliary stricture after duct-to-duct biliary anastomosis in living donor liver transplantation: a 9-year single-center experience. *Eur Radiol*. 2022;32(4):2414-2425. [Crossref]
7. Karatoprak S, Kutlu R, Karatoprak NB, Dağ N, Yılmaz S. Percutaneous radiological biliary interventions after failed endoscopic treatment in living liver donors: experience of a high-volume transplantation center. *Transpl Int*. 2021;34(12):2846-2855. [Crossref]
8. Koksal AS, Eminler AT, Parlak E, Gurakar A. Management of biliary anastomotic strictures after liver transplantation. *Transplant Rev (Orlando)*. 2017;31(3):207-217. [Crossref]
9. Villa NA, Harrison ME. Management of biliary strictures after liver transplantation. *Gastroenterol Hepatol (N Y)*. 2015;11(5):316-328. [Crossref]
10. Shen Y, Yu X, Cui J, et al. Development of biodegradable polymeric stents for the treatment of cardiovascular diseases. *Biomolecules*. 2022;12(9):1245. [Crossref]
11. Vermeulen BD, Siersema PD. Esophageal stenting in clinical practice: an overview. *Curr Treat Options Gastroenterol*. 2018;16(2):260-273. [Crossref]
12. Mauri G, Michelozzi C, Melchiorre F, et al. Benign biliary strictures refractory to standard bilioplasty treated using polydoxanone biodegradable biliary stents: retrospective multicentric data analysis on 107 patients. *Eur Radiol*. 2016;26(11):4057-4063. [Crossref]
13. Chen X, Xia Y, Shen S, et al. Research on the current application status of magnesium metal stents in human luminal cavities. *J Funct Biomater*. 2023;14(9):462. [Crossref]
14. Kwon CI, Son JS, Kim KS, et al. Mechanical properties and degradation process of biliary self-expandable biodegradable stents. *Dig Endosc*. 2021;33(7):1158-1169. [Crossref]
15. Li Y, Yuan K, Deng C, et al. Biliary stents for active materials and surface modification: recent advances and future perspectives. *Bioact Mater*. 2024;42:587-612. [Crossref]
16. Song G, Zhao HQ, Liu Q, Fan Z. A review on biodegradable biliary stents: materials and future trends. *Bioact Mater*. 2022;17:488-495. [Crossref]
17. Song Y, Qin G, Du L, Hu H, Han Y. *In vitro* and *in vivo* assessment of biocompatibility of AZ31 alloy as biliary stents: a preclinical approach. *Arch Med Sci*. 2020;18(1):195-205. [Crossref]
18. Peng H, Fan K, Zan R, et al. Degradable magnesium implants inhibit gallbladder cancer. *Acta Biomater*. 2021;128:514-522. [Crossref]
19. Chen A, Tey K, Verhage R, Maan R, Pieterman K. Percutaneous biodegradable stent placement for treatment of benign biliary strictures: a systematic review. *J Vasc Interv Radiol*. 2025;36(4):556-563. [Crossref]
20. G Almeida G, Donato P. Biodegradable versus multiple plastic stent implantation in benign biliary strictures: a systematic review and meta-analysis. *Eur J Radiol*. 2020;125:108899. [Crossref]
21. Vaz OP, Al-Islam S, Khan ZA, et al. Biodegradable stents: primary experience in a tertiary hepatopancreaticobiliary center in the United Kingdom. *Cureus*. 2021;13(10):e19075. [Crossref]
22. Yamamoto K, Yoshioka T, Furuichi K, et al. Experimental study of poly-L-lactic acid biodegradable stents in normal canine bile ducts. *Cardiovasc Intervent Radiol*. 2011;34(3):601-608. [Crossref]
23. Magro B, Tacelli M, Mazzola A, Conti F, Celsa C. Biliary complications after liver transplantation: current perspectives and future strategies. *Hepatobiliary Surg Nutr*. 2021;10(1):76-92. [Crossref]
24. Öztürk M, Dağ N, Sığırcı A, Yılmaz S. Evaluation of early and late complications of pediatric liver transplantation with multi-slice computed tomography: a high-volume transplant single-center study. *Turk J Gastroenterol*. 2021;32(7):586-592. [Crossref]
25. Battistel M, Senzolo M, Ferrarese A, et al. Biodegradable biliary stents for percutaneous treatment of post-liver transplantation refractory benign biliary anastomotic strictures. *Cardiovasc Intervent Radiol*. 2020;43(5):749-755. [Crossref]
26. Dopazo C, Diez I, Quintero J, et al. Role of biodegradable stents as part of treatment of biliary strictures after pediatric and adult liver transplantation: an observational single-center study. *J Vasc Interv Radiol*. 2018;29(6):899-904. [Crossref]
27. De Gregorio MA, Criado E, Guirola JA, et al. Absorbable stents for treatment of benign biliary strictures: long-term follow-up in the prospective Spanish registry. *Eur Radiol*. 2020;30(8):4486-4495. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253699

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE



# Impact of biopsy route, muscle pathway, and cortex target on safety and diagnostic yield in ultrasound-guided renal parenchymal biopsy

Kadir Han Alver[1]
Halil Serdar Aslan[1]
Muhammet Arslan[1]
Muhammed Tekinhatun[2]
Mahmut Demirci[3]
Nagihan Yalçın[4]
Ahmet Baki Yağcı[1]

[1]Pamukkale University Faculty of Medicine, Department of Radiology, Denizli, Türkiye

[2]Dicle University Faculty of Medicine, Department of Radiology, Diyarbakır, Türkiye

[3]Adnan Menderes University Faculty of Medicine, Department of Radiology, Aydın, Türkiye

[4]Pamukkale University Faculty of Medicine, Department of Pathology, Denizli, Türkiye

**PURPOSE**

To compare the safety and diagnostic yield of two ultrasound (US)-guided percutaneous renal biopsy (PRB) approaches, lateral to medial and medial to lateral, which differ in access route, muscle groups traversed, and cortical targets.

**METHODS**

This retrospective study included 490 patients (mean age: 38.2 ± 21.2 years; 267 men, 223 women) who underwent US-guided PRB between 2019 and 2024 and had abdominal computed tomography (CT)/magnetic resonance imaging (MRI) within 1 year. At the left kidney lower pole level (L3–L4), anterior–posterior thicknesses of the traversed muscle groups were measured on CT/MRI. Complications were classified according to the Society of Interventional Radiology guidelines. Diagnostic yield was categorized as optimal (≥ 12 glomeruli), suboptimal (≥ 3 glomeruli), and pathologist based (diagnostic according to final pathology assessment). Group comparisons were performed using the chi-square test, Fisher's exact test, and t-test.

**RESULTS**

In 490 PRBs (237 lateral to medial, 253 medial to lateral), the medial-to-lateral approach, despite traversing thicker muscles (35.7 vs. 11.5 mm, $P = 0.001$), produced smaller hematomas (8.6 vs. 17.3 mm, $P = 0.001$) with similar complication rates (major: 3.6% vs. 3.4%, $P = 0.913$; minor: 36% vs. 33.8%, $P = 0.608$). Diagnostic adequacy was comparable, but optimal yield was higher with the medial-to-lateral route (85.0% vs. 73.0%, $P = 0.001$).

**CONCLUSION**

Both approaches demonstrated comparable safety. However, the medial-to-lateral route was associated with smaller hematomas and a higher proportion of optimal biopsies from the lateral cortex, but suboptimal and pathologist-based adequacy remained high in both techniques.

**CLINICAL SIGNIFICANCE**

When standard lower pole lateral cortex biopsy is not feasible due to cortical scarring, cysts, overlying skin lesions, or anatomic limitations—especially in patients for whom contralateral biopsy is not possible (e.g., solitary or ectopic pelvic kidney, severe unilateral hydronephrosis)—alternative cortical targets must be used. Understanding how different access routes and muscle pathways influence hemorrhage control and diagnostic yield helps operators choose the safest and most effective technique in these situations.

**KEYWORDS**

Percutaneous renal biopsy, diagnostic yield, complications, muscle thickness, cortical targeting

**Handling editor:** Barbaros Çil

**Corresponding author:** Kadir Han Alver

**E-mail:** kadirhanalver@gmail.com

Received 16 October 2025; revision requested 27 October 2025; accepted 21 November 2025.

Epub: 08.12.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253699

Percutaneous renal biopsy (PRB) is a key procedure for diagnosing kidney disease, assessing prognosis, and guiding treatment. It is typically performed in patients with substantial proteinuria, hematuria, declining renal function, or suspected systemic disease involvement such as lupus or amyloidosis.[1] Although computed tomography (CT) guidance may be used in selected situations, ultrasound (US) is the preferred modality due to its re-

You may cite this article as: Alver KH, Aslan HS, Arslan M, et al. Impact of biopsy route, muscle pathway, and cortex target on safety and diagnostic yield in ultrasound-guided renal parenchymal biopsy. *Diagn Interv Radiol*. 2026;32(4):479-486.

al-time visualization, lack of radiation, lower cost, and relatively short procedure time. When combined with spring-loaded core biopsy systems, US-guided PRB provides high diagnostic yield with an established safety profile.[2]

According to the Society of Interventional Radiology (SIR) and the Cardiovascular and Interventional Radiological Society of Europe (CIRSE), the standard technique for native kidney biopsy is a tangential medial-to-lateral approach targeting the lower pole lateral cortex, corresponding to Brödel's avascular zone, which provides a favorable balance between diagnostic yield and safety.[3,4] However, this pathway may not be feasible in cases such as asymmetric cortical thinning, large or multiple cysts, overlying skin lesions, or difficulty achieving optimal patient positioning. In these situations, particularly when contralateral biopsy is not possible (e.g., solitary or ectopic pelvic kidney, severe unilateral hydronephrosis, or extensive nephrolithiasis), biopsy may need to target alternative cortical regions via different anatomical routes. Although recent technical perspectives suggest that any cortical region can be safely sampled if a tangential trajectory is maintained to avoid medullary and sinus injury, this assumption has not been systematically evaluated.[5] To date, no study has assessed whether differences in cortical target or the muscle groups traversed affect the safety or diagnostic performance of PRB.

At Pamukkale University Faculty of Medicine, Department of Interventional Radiology, two interventional radiologists, each with over 10 years of experience, routinely performed PRBs using different but consistent techniques, reflecting their procedural training and background. One radiologist used the standard medial-to-lateral approach, targeting the lower pole lateral cortex, whereas the other employed a lateral-to-medial approach, targeting the medial cortex. This systematic variation in cortical target, muscle pathway, and access route created a natural and unique comparison setting, enabling us to retrospectively evaluate how these technical differences influence safety and diagnostic yield, addressing a previously unexplored gap in the literature.

**Main points**

- This study directly compared two ultrasound-guided percutaneous renal biopsy techniques—lateral to medial and medial to lateral—that differ in cortical targets, routes, and muscle pathways. The medial-to-lateral approach traverses the paravertebral muscles (multifidus, erector spinae, quadratus lumborum) to sample the lower pole lateral cortex, whereas the lateral-to-medial approach passes through the posterolateral abdominal wall muscles (latissimus dorsi, external and internal obliques, transversus abdominis) to reach the lower pole medial cortex.
- Safety was comparable between approaches, with similar rates of minor and major complications within Society of Interventional Radiology guideline thresholds.
- The medial-to-lateral approach, which traverses thicker paravertebral muscles, resulted in smaller non-transfusion-requiring perirenal hematomas.
- Optimal diagnostic yield (≥ 12 glomeruli) was higher with the medial-to-lateral approach, whereas suboptimal yield (≥ 3 glomeruli) and pathologist-based adequacy remained high in both techniques.
- This is the first study to systematically evaluate how cortical targets and muscle pathways affect biopsy outcomes, providing practical guidance for cases in which standard lower pole lateral cortex access cannot be used.

## Methods

### Patient selection

This retrospective study was approved by the Pamukkale University Ethical Committee of Non-Invasive Clinical Research (approval number: E-60116787-020-703301, date: 10.06.2025), and informed consent was waived. Patients were eligible if they underwent US-guided left PRB and had abdominal CT or magnetic resonance imaging (MRI) within 1 year of the procedure. Among the 716 PRBs performed between January 2019 and December 2024, patients were excluded for renal mass biopsy (n = 32), CT-guided biopsy (n = 7), transplanted kidney biopsy (n = 34), absence of CT/MRI within 1 year (n = 84), and renal orientation anomalies impairing cortical or trajectory assessment (n = 9). Additional exclusions included resident-performed procedures (n = 13), 16-gauge needle use (n = 10), non-coaxial technique (n = 14), incomplete 24-hour follow-up data (n = 11), and right kidney biopsies (n = 12). The final cohort consisted of 490 patients, comprising Group 1 (lateral-to-medial approach targeting the medial cortex; n = 237; mean age: 38 ± 20.9 years; 129 men) and Group 2 (medial-to-lateral approach targeting the lateral cortex; n = 253; mean age: 37.9 ± 21.4 years; 138 men) (Figure 1).

### Prebiopsy preparation

On the day of the procedure, all patients underwent laboratory evaluation. Biopsy was performed only if the platelet count was > 50,000/mL, the international normalized ratio was < 1.5, and hemoglobin was > 8 g/dL.[5] Antiplatelet and anticoagulant medications were reviewed and adjusted (temporary discontinuation, bridging when indicated, and post-biopsy resumption) in accordance with SIR and CIRSE guidelines.[4,6] Because uncontrolled hypertension increases bleeding risk, biopsy proceeded only when blood pressure was < 160/90 mmHg; otherwise, antihypertensive optimization and rescheduling were performed.[1,7]

### Biopsy procedure

All biopsies were performed under real-time US guidance (ACUSON Sequoia, Siemens Healthineers, Mountain View, CA, USA; Aplio 500, Toshiba Medical Systems, Tokyo, Japan) using 1.5–6.0-MHz convex probes, with patients in the prone position under sterile conditions. Due to its more accessible position, the left kidney was consistently targeted, with all procedures guided sonographically in the transverse plane. Local anesthesia was provided using 10–20 mL of 1%–2% lidocaine. Two interventional radiologists, each with over 10 years of experience and more than 1,000 prior PRBs, performed the procedures using different but consistent approaches based on training preference. In Group 1, the biopsy was performed via a lateral skin entry, traversing the posterolateral abdominal wall muscles (latissimus dorsi, external oblique, internal oblique, transversus abdominis) to target the lower pole medial cortex (Figure 2). In Group 2, a medial skin entry was used, traversing the paravertebral muscles (multifidus, erector spinae, quadratus lumborum) to reach the lower pole lateral cortex (Figure 3). Patient allocation between the two groups was entirely random, as biopsies were performed by whichever radiologist was on duty on the day of the procedure. A coaxial technique was used in all cases: a 17-gauge introducer needle was advanced to the renal capsule, through which a 15-cm, 18-gauge spring-loaded core needle (Maxcore, Bard Biopsy Systems, Tempe, Arizona) with a penetration depth of 22 mm and a sample notch of 18 mm was used to obtain two tissue samples with a tangential cortical trajectory. If bleeding occurred through the coaxial introducer, temporary tract tamponade was performed by reinserting the stylet, and persistent bleeding was managed with autologous blood clot or absorbable

gelatin sponge (Spongostan, Ethicon, NJ, USA). Specimens were transported in phosphate-buffered saline for light microscopy (LM), immunofluorescence (IF), and electron microscopy (EM). On-site adequacy assessment was not performed due to the lack of cytotechnologist support. Local anesthesia was standard, and conscious sedation was used selectively in pediatric or anxious patients with anesthesiology support.[8]

## Post-biopsy monitoring

After biopsy and dressing, patients were kept supine with a sandbag under the left kidney for 6 hours and observed for 24 hours. At the 6th hour, follow-up renal US was performed in the interventional radiology suite, and the presence and maximum size of biopsy-related hemorrhage were recorded. Hemoglobin was assessed at the 6th hour and repeated at 24 hours before discharge. Vital signs were monitored every 30 minutes for 2 hours, hourly for 4 hours, and every 2 hours thereafter. Patients were observed for hematuria and bleeding-related symptoms (flank pain, dizziness, vomiting, presyncope). If major bleeding was suspected—macrohematuria, hypotension, tachycardia, or large perirenal/retroperitoneal hematoma—non-contrast CT was performed, followed by contrast-enhanced CT or angiography when needed. Patients who were hemodynamically stable and asymptomatic without ongoing hemoglobin decline were discharged after 24 hours with return precautions.[1,5,7]

In accordance with the SIR guidelines, complications were classified as major (requiring transfusion, radiologic/surgical intervention, or associated with renal obstruction/failure, sepsis, permanent sequelae, or death) or minor (transient hematuria or small perirenal/retroperitoneal hematomas resolving spontaneously without transfusion or extended hospitalization).[9]

## Radiologic and pathologic evaluation

During pre-procedural US, the long-axis length and parenchymal thickness of the target kidney were measured. At the lower pole level (typically L3–L4), the anterior–posterior thickness of the muscle groups along the biopsy path was measured on CT/MRI. In the lateral-to-medial approach (Group 1), this included the posterolateral abdominal wall muscles (external oblique, internal oblique, transversus abdominis, latissimus dorsi), whereas in the medial-to-lateral approach (Group 2), it included the paravertebral muscles (multifidus, erector spinae, quadratus lumborum). Multiple measurements were

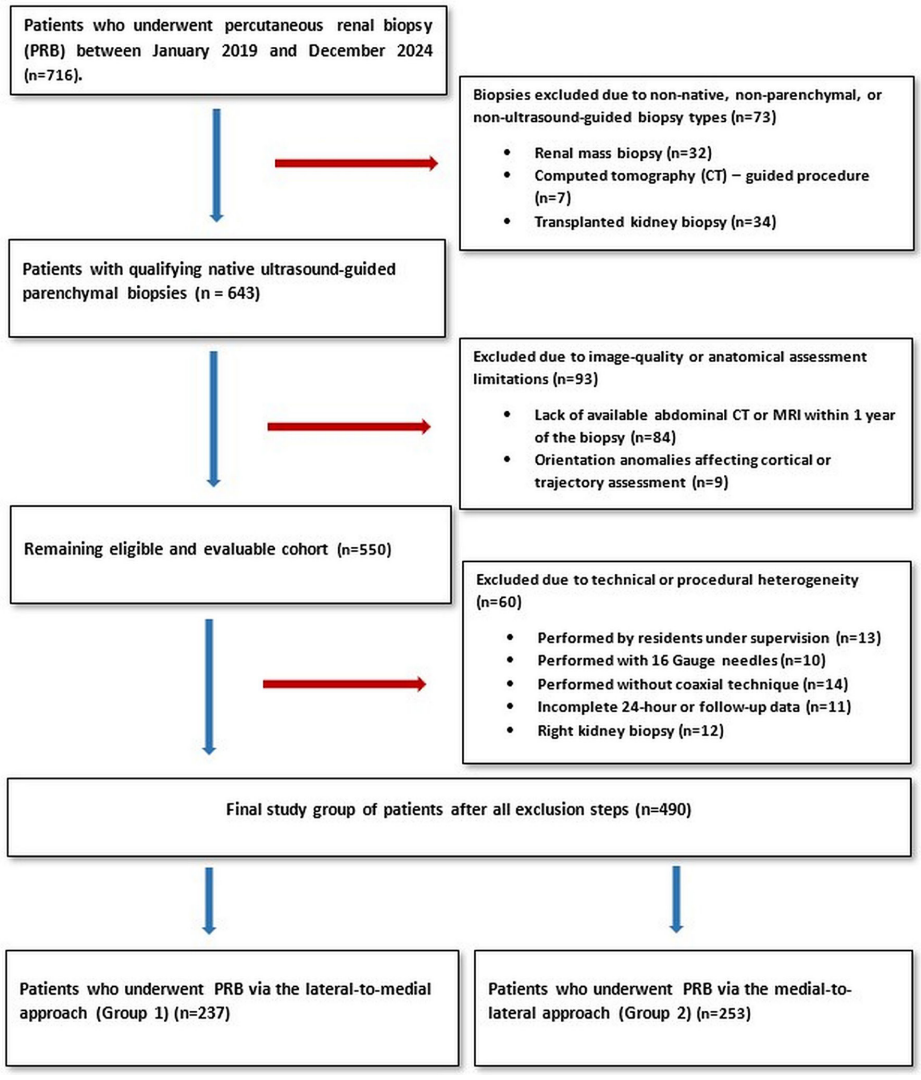



**Figure 1.** Flowchart of study population. MRI, magnetic resonance imaging.

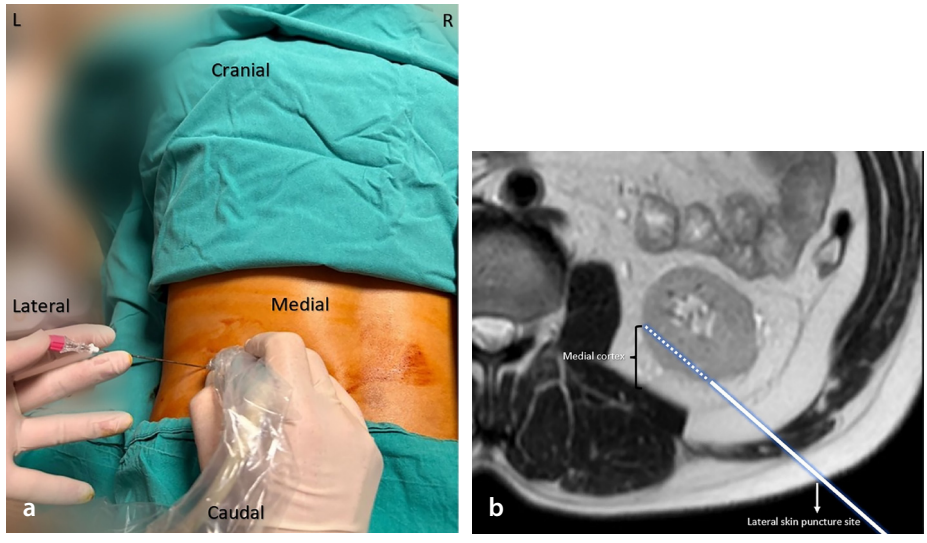



**Figure 2.** Lateral-to-medial approach for left kidney biopsy. **(a)** Intra-procedural photograph, obtained under sterile conditions with the patient in the prone position, demonstrating lateral skin entry of a 17G introducer needle under ultrasound guidance to target the medial cortex. **(b)** Illustration of the renal parenchymal biopsy performed using the cortical–tangential technique, with traversal of the posterolateral abdominal wall muscles and targeting of the medial cortex (L, left; R, right; G, gauge).

obtained across each muscle group at this level, and mean values were used for analysis. Demographics, biopsy approach and operator, hemoglobin values, sonographic findings, histopathology, glomerular counts, and complications were retrieved from electronic records and Picture Archiving and Communication Systems.

All specimens were evaluated by a nephropathologist with over 15 years of experience. Consistent with the literature, biopsies were categorized based on total glomerular count across LM, IF, and EM:[10,11]

• **Group A (optimal):** ≥ 12 glomeruli total (≥ 10 for LM + ≥ 1 for IF + ≥ 1 for EM),

• **Group B (suboptimal):** ≥ 3 glomeruli total, with at least one evaluable glomerulus in each modality but not meeting the Group A criteria, and

• **Group C (inadequate):** < 3 glomeruli or missing required modalities.

Diagnostic yield was defined as follows:

**1. Optimal yield:** Proportion of Group A biopsies (≥ 12 glomeruli),

**2. Suboptimal yield:** Proportion of biopsies with ≥ 3 glomeruli (Groups A + B), and

**3. Pathologist-based yield:** Biopsies deemed diagnostic in the final report, regardless of glomerular count.

Yield ratios were calculated as diagnostic biopsies divided by total biopsies in each group.

### Statistical analysis

Statistical analyses were conducted using SPSS version 25.0 (IBM Corp., Armonk, NY, USA). Categorical variables were reported as counts and percentages and continuous variables as mean ± standard deviation. Normality was assessed using the Shapiro–Wilk test.

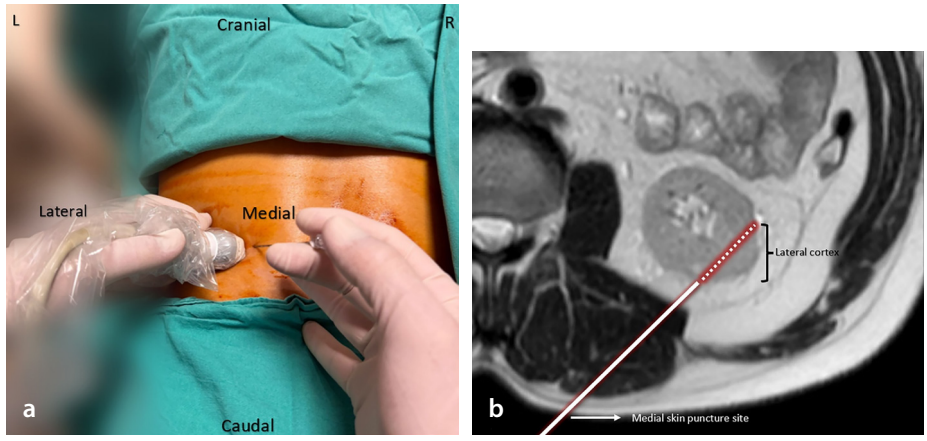



**Figure 3.** Medial-to-lateral approach for left kidney biopsy. **(a)** Intra-procedural photograph, obtained under sterile conditions with the patient in the prone position, demonstrating medial skin entry of a 17G introducer needle under ultrasound guidance to target the lateral cortex. **(b)** Illustration of the renal parenchymal biopsy performed using the cortical–tangential technique, with traversal of the paravertebral muscles and targeting of the lateral cortex (L, left; R, right; G, gauge).

Patients were grouped according to the biopsy approach (Group 1: lateral to medial; Group 2: medial to lateral). Diagnostic yield and major/minor complications were compared using chi-square or Fisher's exact tests. Total muscle thickness, total glomerular count, hemoglobin change (baseline vs. 6 and 24 hours), and hematoma size were compared using the independent-samples t-test or Mann–Whitney U test, depending on the distribution. A $P$ value < 0.05 was considered statistically significant.

## Results

A total of 490 patients were included (mean age: 38.2 ± 21.2 years; 267 men, 223 women), comprising 237 who underwent the lateral-to-medial approach (Group 1) and 253 who underwent the medial-to-lateral approach (Group 2). The most common histopathological diagnosis was focal segmental glomerulosclerosis (79/490, 16.1%), followed by membranous nephropathy (48/490, 9.8%) and lupus nephritis (39/490, 8.0%). The full diagnostic distribution is summarized in Table 1.

Baseline demographic characteristics, renal long-axis length, parenchymal thickness, and mean hemoglobin decline at both the 6$^{th}$ and 24$^{th}$ hours were comparable between the two groups (all $P > 0.05$). However, muscle thickness along the biopsy tract differed substantially, being significantly lower in Group 1 than in Group 2 (11.5 ± 4.2 vs. 35.7 ± 10.9 mm, $P = 0.001$) (Table 2).

The suboptimal diagnostic yield (≥ 3 glomeruli) was similarly high in Group 1 and Group 2 (95.4% vs. 95.3%, $P = 0.958$), and pathologist-based adequacy was also comparable (96.6% vs. 98.0%, $P = 0.335$). However, the proportion of biopsies achieving ≥ 12 glomeruli (optimal diagnostic yield) was significantly higher in Group 2 (85.0% vs. 73.0%, $P = 0.001$) (Table 2).

**Table 1.** Distribution of histopathological diagnoses by biopsy approach

| | | Lateral-to-medial approach (Group 1) (n = 237) | Medial-to-lateral approach (Group 2) (n = 253) | Total | Percentage (%) |
|---|---|---|---|---|---|
| 1 | Focal segmental glomerulosclerosis | 38 | 41 | 79 | 16.1% |
| 2 | Membranous nephropathy | 23 | 25 | 48 | 9.8% |
| 3 | Lupus nephritis | 16 | 23 | 39 | 8.0% |
| 4 | Crescentic glomerulonephritis | 15 | 18 | 33 | 6.7% |
| 5 | Immunoglobulin A nephropathy | 15 | 14 | 29 | 5.9% |
| 6 | Acute interstitial nephritis | 14 | 13 | 27 | 5.5% |
| 7 | Chronic glomerulonephritis | 12 | 14 | 26 | 5.3% |
| 8 | Henoch–Schönlein purpura nephritis | 13 | 12 | 25 | 5.1% |

**Table 1.** Continued

| | | Lateral-to-medial approach (Group 1) (n = 237) | Medial-to-lateral approach (Group 2) (n = 253) | Total | Percentage (%) |
|---|---|---|---|---|---|
| 9 | Normal histology/no specific pathology | 12 | 13 | 25 | 5.1% |
| 10 | Membranoproliferative glomerulonephritis type I | 10 | 7 | 17 | 3.5% |
| 11 | Diabetic nephropathy | 7 | 8 | 15 | 3.1% |
| 12 | Serum amyloid A protein amyloidosis | 5 | 9 | 14 | 2.9% |
| 13 | C3 glomerulopathy | 6 | 7 | 13 | 2.7% |
| 14 | Non-diagnostic | 8 | 5 | 13 | 2.7% |
| 15 | Mesangial hypercellularity ± glomerulosclerosis, not otherwise specified | 5 | 6 | 11 | 2.2% |
| 16 | Chronic interstitial nephritis | 5 | 6 | 11 | 2.2% |
| 17 | Minimal change disease | 7 | 4 | 11 | 2.2% |
| 18 | Vasculitis | 6 | 1 | 7 | 1.4% |
| 19 | Acute tubular necrosis | 1 | 5 | 6 | 1.2% |
| 20 | Hypertensive nephropathy | 2 | 4 | 6 | 1.2% |
| 21 | Small-vessel vasculitis | 4 | 1 | 5 | 1.0% |
| 22 | Postinfectious glomerulonephritis | 2 | 3 | 5 | 1.0% |
| 23 | Diffuse proliferative glomerulonephritis | 3 | 1 | 4 | 0.8% |
| 24 | Chronic pyelonephritis | 1 | 2 | 3 | 0.6% |
| 25 | Myeloma cast nephropathy | 0 | 3 | 3 | 0.6% |
| 26 | Immunoglobulin light chain amyloidosis | 0 | 2 | 2 | 0.4% |
| 27 | Microscopic polyangiitis | 1 | 1 | 2 | 0.4% |
| 28 | Diffuse large B-cell lymphoma | 1 | 1 | 2 | 0.4% |
| 29 | Thrombotic microangiopathy | 1 | 1 | 2 | 0.4% |
| 30 | Isolated Bowman's capsule thickening | 0 | 1 | 1 | 0.2% |
| 31 | Fibrillary glomerulonephritis | 1 | 0 | 1 | 0.2% |
| 32 | Isolated glomerulomegaly | 1 | 0 | 1 | 0.2% |
| 33 | Granulomatosis with polyangiitis | 0 | 1 | 1 | 0.2% |
| 34 | Lymphoplasmacytic lymphoma | 1 | 0 | 1 | 0.2% |
| 35 | Nephrocalcinosis | 1 | 0 | 1 | 0.2% |
| 36 | Necrotizing granulomatous inflammation | 0 | 1 | 1 | 0.2% |
| | Total | 237 | 253 | 490 | 100.0% |

**Table 2.** Procedural characteristics, diagnostic yield, and complications in Group 1 (lateral to medial) and Group 2 (medial to lateral) approaches

| | Lateral-to-medial approach (Group 1) (n = 237) | Medial-to-lateral approach (Group 2) (n = 253) | *P* value | Test |
|---|---|---|---|---|
| Age (years) | 38 ± 20.9 | 37.9 ± 21.4 | 0.693 | t |
| Gender (men/women) | (129/108) | (138/115) | 0.980 | $\chi^2$ |
| Renal long-axis length (mm) | 109.3 ± 13 | 109.5 ± 12.9 | 0.830 | t |
| Renal parenchymal thickness (mm) | 16.2 ± 2.8 | 16.3 ± 3.1 | 0.801 | t |
| Total muscle thickness traversed (mm) | 11.5 ± 4.2 | 35.7 ± 10.9 | **0.001** | t |
| Mean hemoglobin decline (Prebiopsy - 6$^{th}$ hour post-biopsy) (g/dL) | 0.44 ± 0.79 | 0.37 ± 0.63 | 0.297 | t |
| Mean hemoglobin decline (Prebiopsy - 24$^{th}$ hour post-biopsy) (g/dL) | 0.34 ± 0.84 | 0.33 ± 0.78 | 0.980 | t |
| Mean number of glomeruli in biopsies | 17.3 ± 10.9 (n = 237) | 18.4 ± 10.3 (n = 253) | 0.243 | t |

**Table 2.** Continued

| | Lateral-to-medial approach (Group 1) (n = 237) | Medial-to-lateral approach (Group 2) (n = 253) | *P* value | Test |
|---|---|---|---|---|
| Optimal (≥ 12 glomeruli) | 21.5 ± 9.8 (n = 173) | 20.9 ± 9.2 (n = 215) | 0.511 | t |
| Suboptimal (≥ 3 and < 12 glomeruli) | 7.0 ± 1.9 (n = 53) | 6.2 ± 1.6 (n = 26) | 0.058 | t |
| Inadequate (< 3 glomeruli) | 1.5 ± 1.4 (n = 11) | 1.7 ± 1.2 (n = 12) | 0.724 | f |
| Optimal diagnostic yield* | 73.0% | 85.0% | **0.001** | $\chi^2$ |
| Suboptimal diagnostic yield* | 95.4% | 95.3% | 0.958 | $\chi^2$ |
| Diagnostic yield with pathology opinion | 96.6% | 98.0% | 0.335 | $\chi^2$ |
| Minor complications | 80/237 (33.8%) | 91/253 (36%) | 0.608 | $\chi^2$ |
| Maximum perirenal hematoma size (minor, non-transfusion requiring) (mm) | 17.3 ± 12.6 | 8.6 ± 8.9 | **0.001** | t |
| Major complications | 8/237 (3.4%) | 9/253 (3.6%) | 0.913 | $\chi^2$ |
| Total complications | 88/237 (37.1%) | 100/253 (39.5%) | 0.586 | $\chi^2$ |

t: Student's t-test for independent samples; $\chi^2$, chi-square test; f, Fisher's exact test; *, optimal diagnostic yield was calculated as the proportion of biopsies with ≥ 12 glomeruli (O) among all biopsies (O + SO + IA). Suboptimal diagnostic yield was calculated as the proportion of biopsies with ≥ 3 glomeruli (O + SO) among all biopsies (O + SO + IA). O: Optimal (≥ 12 glomeruli); SO, suboptimal (3–11 glomeruli); IA, inadequate (< 3 glomeruli).

Overall, minor and major complication rates were also similar between Group 1 and Group 2 (37.1% vs. 39.5%, *P* = 0.586; 33.8% vs. 36.0%, *P* = 0.608; and 3.4% vs. 3.6%, *P* = 0.913, respectively). The most common minor and major complications were non-transfusion-requiring perirenal hematoma (29.5% vs. 30.4%) and hemorrhage requiring transfusion (1.3% vs. 2.0%), respectively (Figures 4 and 5). Although the frequency of perirenal hematoma was similar, hematoma size was significantly smaller in Group 2 (8.6 ± 8.9 vs. 17.3 ± 12.6 mm, *P* = 0.001). Detailed complication subtypes are presented in Table 3.

**Table 3.** Minor and major complications following percutaneous renal biopsy in Group 1 (lateral to medial) and Group 2 (medial to lateral) approaches

| Complication | Lateral-to-medial approach (Group 1) | | Medial-to-lateral approach (Group 2) | | Total | |
|---|---|---|---|---|---|---|
| | (n = 237) | | (n = 253) | | (n = 490) | |
| | (n) | (%) | (n) | (%) | (n) | (%) |
| Perirenal hematoma (non-transfusion requiring) | 70 | 29.5% | 77 | 30.4% | 147 | 30.0% |
| Self-limiting discomfort symptoms | 4 | 1.7% | 6 | 2.4% | 10 | 2.0% |
| Macroscopic hematuria | 5 | 2.1% | 6 | 2.4% | 11 | 2.2% |
| Vasovagal reaction | 1 | 0.4% | 1 | 0.4% | 2 | 0.4% |
| Bladder hematoma | 0 | 0.0% | 1 | 0.4% | 1 | 0.2% |
| **Minor complications** | **80** | **33.8%** | **91** | **36.0%** | **171** | **34.9%** |
| Hemorrhage requiring transfusion (perirenal or retroperitoneal) | 3 | 1.3% | 5 | 2.0% | 8 | 1.6% |
| Subcapsular hemorrhage requiring transfusion | 1 | 0.4% | 0 | 0.0% | 1 | 0.2% |
| Arteriovenous fistula | 1 | 0.4% | 2 | 0.8% | 3 | 0.6% |
| Arteriovenous fistula and pseudoaneurysm | 2 | 0.8% | 2 | 0.8% | 4 | 0.8% |
| Death | 1 | 0.4% | 0 | 0.0% | 1 | 0.2% |
| **Major complications** | **8** | **3.4%** | **9** | **3.6%** | **17** | 3.5% |
| Total complications | 88 | 37.1% | 100 | 39.5% | 188 | 38.4% |

## Discussion

This study compared medial-to-lateral and lateral-to-medial PRB approaches, showing similar overall safety but notable differences related to the muscle thickness and cortical region sampled. The medial-to-lateral approach traversed thicker paravertebral muscles (35.7 ± 10.9 vs. 11.5 ± 4.2 mm, *P* = 0.001) and was associated with smaller non-transfusion-requiring perirenal hematomas (8.6 ± 8.9 vs. 17.3 ± 12.6 mm, *P* = 0.001). This suggests that thicker, more compact muscle groups and a shorter retroperitoneal course may provide a tamponade effect that limits hematoma expansion, whereas thinner posterolateral muscles may permit greater spread. Although bleeding originates in the kidney, similar principles have been described in other contexts. In subcapsular liver lesion biopsies, traversing a short segment of normal parenchyma is recommended so that the surrounding tissue can help compress and limit bleeding.[12] Likewise, rectus sheath and iliopsoas hematomas often remain self-contained due to the natural pressure of muscular compartments.[13,14] Additionally, greater muscle thickness along nephrostomy catheter tracts has been associated with reduced catheter displacement, suggesting that deeper musculature has a stabilizing effect.[15] Although no study has directly evaluated the influence of muscle layers on bleeding after renal biopsy, these parallels support the plausibility that thicker muscle groups may similarly help constrain post-biopsy hemorrhage.

Major complications occurred at similar rates in the lateral-to-medial and medial-to-lateral groups (3.4% vs. 3.6%, *P* = 0.913), remaining below the < 5% threshold recommended by SIR for PRB.[3] These low rates likely reflect strict adherence to guideline-based anticoagulation management,[4-6] optimization of blood pressure,[16] performance in a high-volume tertiary center,[17] use of 18-gauge needles[18] and a coaxial technique,[19] tract tamponade/embolization when indicated,[20] and maintenance of a cortical–tangential trajectory.[11,21,22] Minor complications were also comparable between groups (33.8% vs. 36.0%, *P* = 0.608), with

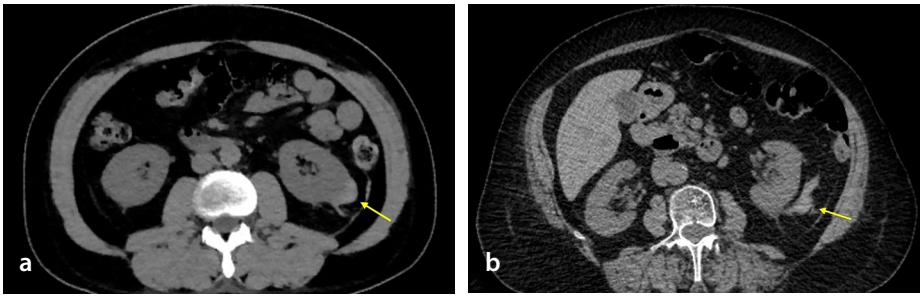


**Figure 4.** Examples of perirenal hemorrhage following biopsy using the lateral-to-medial approach. **(a)** Transverse non-contrast abdominal computed tomography (CT) images of a 45-year-old male patient who underwent percutaneous renal biopsy (PRB) for proteinuria, showing perirenal hemorrhage (yellow arrow) at the level corresponding to the needle entry site into the left kidney. **(b)** Axial non-contrast abdominal CT images obtained after PRB in a 51-year-old female patient being followed for vasculitis to assess renal involvement, demonstrating minimal perirenal hemorrhage (yellow arrow).

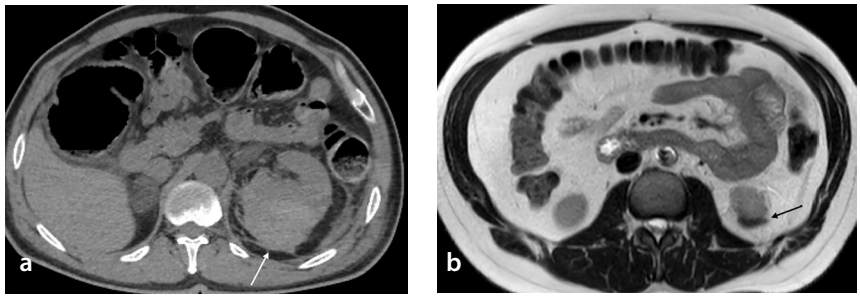


**Figure 5.** Examples of perirenal hemorrhage following biopsy using the medial-to-lateral approach. **(a)** Axial non-contrast abdominal computed tomography images of a 37-year-old male patient who underwent parenchymal renal biopsy (PRB) for nephrotic syndrome, showing perirenal hemorrhage (white arrow) in continuity with the kidney at its posterior aspect. **(b)** Axial non-fat-suppressed T2-weighted magnetic resonance imaging of a 45-year-old male patient who underwent PRB for acute kidney injury, demonstrating minimal hypointense perirenal hemorrhage (black arrow).

non-transfusion-requiring perirenal hematomas being the most frequent (29.5% vs. 30.4%, $P = 0.792$). Although these rates may appear high relative to the broad literature estimates (1.3%–33.3%),[7] they align with studies performing routine post-biopsy US in all patients, which report hematoma detection rates of 33%–58%.[23-25] This suggests that perirenal hematomas are likely under-recognized in studies that image only symptomatic patients.

Most studies describe targeting the lower pole in PRB without specifying cortical orientation, with only a few explicitly mentioning the lower pole lateral cortex.[23,24] However, none evaluate whether sampling the medial cortex affects diagnostic outcomes, making this the first study to systematically assess this distinction in routine practice. Suboptimal diagnostic yield (≥ 3 glomeruli) (95.4% vs. 95.3%, $P = 0.958$) and pathologist-based diagnostic adequacy (96.6% vs. 98.0%, $P = 0.335$) were similar between groups, as were mean glomerular counts across optimal (21.5 ± 9.8 vs. 20.9 ± 9.2, $P = 0.511$), suboptimal (7.0 ± 1.9 vs. 6.2 ± 1.6, $P = 0.058$), and inadequate (1.5 ± 1.4 vs. 1.7 ± 1.2, $P = 0.724$) biopsy categories. However, the proportion of optimal biopsies (≥ 12 glomeruli) was lower in the lateral-to-medial approach (73.0% vs. 85.0%, $P = 0.001$), below the 88% technical success benchmark recommended by SIR.[3] Although this difference may suggest that the cortical region targeted could influence tissue yield, we emphasize that this remains speculative, as no histopathologic study has compared glomerular density between the medial and lateral cortices. Although SIR recommends obtaining ≥ 10 glomeruli in diffuse and ≥ 20 in focal diseases, no universally accepted adequacy threshold exists for native kidney biopsies.[3,25-30] Adequacy assessment varies among nephropathologists, and the required number of glomeruli depends on disease distribution (diffuse vs. focal); the necessity of LM, IF, and EM for all cases also remains debated.[25,31,32] Classification systems such as that developed by Geldenhuys et al.,[10] although useful for quality benchmarking, may not fully reflect real-world diagnostic decision-making. In our study, although an optimal yield difference was observed, both approaches achieved high suboptimal and pathologist-based diagnostic adequacy, suggesting that overall diagnostic reliability remained clinically acceptable. Nevertheless, when medial cortical sampling is required, operators may consider yield-enhancing strategies such as using a needle with a larger gauge and obtaining ≥ 3 cores[33] or coordinating with an on-site or bedside pathologist.[5] Finally, although one autopsy study compared glomerular distribution between juxtamedullary and subcapsular regions,[34] no nephrectomy or autopsy study has compared glomerular density between the medial and lateral cortices. Future comparative studies may clarify whether intrinsic cortical differences contribute to the yield patterns observed here.

Our study has limitations. First, it was a single-center, retrospective analysis, and selection and information bias may have been introduced. Second, muscle thickness measurements were performed at the lower pole level on CT/MRI, but the biopsy needle trajectory may not have perfectly matched the measurement plane. However, the marked difference in muscle thickness between approaches and the exclusion of cases with orientation anomalies likely minimized the impact of this limitation. Because the traversed muscle groups differed substantially, procedure-related pain may also have differed between approaches; however, pain assessment (e.g., visual analog scale scoring) was not available due to the retrospective design. Additionally, procedures were performed by two interventional radiologists, introducing potential operator-related variability. Nevertheless, both had > 10 years of experience and had performed > 1,000 biopsies each, consistently using their respective techniques, which likely reduced skill-based differences. Prospective, multicenter, randomized studies are needed to validate these findings and further clarify the influence of approach-related factors on biopsy outcomes.

In conclusion, this study—the first to compare how different access routes, muscle pathways, and cortical targets influence PRB performance—demonstrated that both approaches are safe with acceptable complication rates. The medial-to-lateral approach, which traverses thicker paravertebral muscles to sample the lateral cortex, resulted in smaller hematomas and a higher optimal diagnostic yield. Although suboptimal and pathologist-based adequacy were similarly high in both groups, operators should recognize the slightly lower optimal yield when sampling the medial cortex and may adjust techniques accordingly.

## Footnotes

### Conflict of interest disclosure

The authors declared that there is no conflict of interest.

## References


1. Granata A, Distefano G, Pesce F, et al. Performing an ultrasound-guided percutaneous needle kidney biopsy: an up-to-date procedural review. *Diagnostics (Basel)*. 2021;11(12):2186. [Crossref]
2. MacGinley R, Champion De Crespigny PJ, Gutman T, et al. KHA-CARI Guideline recommendations for renal biopsy. *Nephrology (Carlton)*. 2019;24(12):1205-1213. [Crossref]
3. Sheth RA, Baerlocher MO, Connolly BL, et al. Society of Interventional Radiology Quality Improvement Standards on percutaneous needle biopsy in adult and pediatric patients. *J Vasc Interv Radiol.* 2020;31(11):1840-1848. [Crossref]
4. Veltri A, Bargellini I, Giorgi L, Almeida PAMS, Akhan O. CIRSE Guidelines on percutaneous needle biopsy (PNB). *Cardiovasc Intervent Radiol*. 2017;40(10):1501-1513. [Crossref]
5. Shyn PB, Patel MD, Itani M, et al. Image-guided renal parenchymal biopsies- how we do it. *Abdom Radiol (NY)*. 2025;50(6):2595-2605. [Crossref]
6. Davidson JC, Rahim S, Hanks SE, et al. Society of Interventional Radiology Consensus Guidelines for the periprocedural management of thrombotic and bleeding risk in patients undergoing percutaneous image-guided interventions-part I: review of anticoagulation agents and clinical considerations: endorsed by the Canadian Association for Interventional Radiology and the Cardiovascular and Interventional Radiological Society of Europe. *J Vasc Interv Radiol*. 2019;30(8):1155-1167. [Crossref]
7. Schnuelle P. Renal biopsy for diagnosis in kidney disease: indication, technique, and safety. *J Clin Med*. 2023;12(19):6424. [Crossref]
8. Patatas K, Koukkoulli A. The use of sedation in the radiology department. *Clin Radiol*. 2009;64(7):655-663. [Crossref]
9. Khalilzadeh O, Baerlocher MO, Shyn PB, et al. Proposal of a new adverse event classification by the Society of Interventional Radiology Standards of Practice Committee. *J Vasc Interv Radiol.* 2017;28(10):1432-1437. [Crossref]
10. Geldenhuys L, Nicholson P, Sinha N, et al. Percutaneous native renal biopsy adequacy: a successful interdepartmental quality improvement activity. *Can J Kidney Health Dis*. 2015;2:8. [Crossref]
11. Liu B, O'Dell M, Flores M, et al. CT-guided native medical renal biopsy: cortical tangential versus non-tangential approaches-a comparison of efficacy and safety. *Radiology*. 2017;283(1):293-299. [Crossref]
12. Potretzke TA, Saling LJ, Middleton WD, Robinson KA. Bleeding complications after percutaneous liver biopsy: do subcapsular lesions pose a higher risk? *AJR Am J Roentgenol*. 2018;211(1):204-210. [Crossref]
13. Allen M, Sevensma KE. Rectus Sheath Hematoma. In: *StatPearls*. StatPearls Publishing; 2025. Accessed July 28, 2025. [Crossref]
14. Kameda T, Fujita M, Takahashi I. Diagnosis of traumatic iliopsoas hematoma using point-of-care ultrasound. *Crit Ultrasound J*. 2011;3(1):59-61. [Crossref]
15. Ütebey AR, Aslan HS, Arslan M, et al. Predictive factors for spontaneous dislodgement of percutaneous nephrostomies for malignant ureteral obstruction. *Abdom Radiol (NY)*. 2025;50(9):4268-4282. [Crossref]
16. Kriegshauser JS, Patel MD, Young SW, Chen F, Eversman WG, Chang YH. Risk of bleeding after native renal biopsy as a function of preprocedural systolic and diastolic blood pressure. *J Vasc Interv Radiol*. 2015;26(2):206-212. [Crossref]
17. Tøndel C, Vikse BE, Bostad L, Svarstad E. Safety and complications of percutaneous kidney biopsies in 715 children and 8573 adults in Norway 1988-2010. *Clin J Am Soc Nephrol*. 2012;7(10):1591-1597. [Crossref]
18. Chunduri S, Whittier WL, Korbet SM. Adequacy and complication rates with 14- vs. 16-gauge automated needles in percutaneous renal biopsy of native kidneys. *Semin Dial*. 2015;28(2):E11-4. [Crossref]
19. Fung KFK, Cheng KK, Chan EY, Ma LTA, Cho HYD, Kan YLE. Percutaneous ultrasound-guided renal biopsies in a paediatric population: comparison of coaxial and non-coaxial techniques using 18-gauge core biopsy needles. *Pediatr Radiol*. 2022;52(12):2431-2437. [Crossref]
20. Strnad BS, Itani M, Middleton WD. Detection and management of bleeding in the setting of image-guided percutaneous needle biopsy. *Abdom Radiol (NY)*. 2022;47(8):2681-2696. [Crossref]
21. Patel MD, Phillips CJ, Young SW, et al. US-guided renal transplant biopsy: efficacy of a cortical tangential approach. *Radiology*. 2010;256(1):290-296. [Crossref]
22. Li Q, Lin X, Zhang X, Samir AE, Arellano RS. Imaging-related risk factors for bleeding complications of US-guided native renal biopsy: a propensity score matching analysis. *J Vasc Interv Radiol*. 2019;30(1):87-94. [Crossref]
23. Xu S, Xiong B, Lin S, Li Q, Wang L, Zhao W. Predictors of perirenal haematoma post-percutaneous ultrasound-guided renal biopsy. *J Int Med Res*. 2021;49(11):3000605211058377. [Crossref]
24. Li FF, Guan YX, Li TX, et al. Analysis of hemorrhage upon ultrasound-guided percutaneous renal biopsy in China: a retrospective study. *Int Urol Nephrol*. 2024;56(5):1713-1720. [Crossref]
25. Manno C, Strippoli GF, Arnesano L, et al. Predictors of bleeding complications in percutaneous ultrasound-guided renal biopsy. *Kidney Int*. 2004;66(4):1570-1577. [Crossref]
26. Loupy A, Haas M, Roufosse C, et al. The Banff 2019 Kidney Meeting Report (I): updates on and clarification of criteria for T cell- and antibody-mediated rejection. *Am J Transplant.* 2020;20(9):2318-2331. [Crossref]
27. Constantin A, Brisson ML, Kwan J, Proulx F. Percutaneous US-guided renal biopsy: a retrospective study comparing the 16-gauge end-cut and 14-gauge side-notch needles. *J Vasc Interv Radiol.* 2010;21(3):357-61. [Crossref]
28. Ferguson C, Winters S, Jackson S, McToal M, Low G. A retrospective analysis of complication and adequacy rates of ultrasound-guided native and transplant non-focal renal biopsies. *Abdom Radiol (NY)*. 2018;43(8):2183-2189. [Crossref]
29. Goldstein MA, Atri M, O'Malley M, et al. Nonfocal renal biopsies: adequacy and factors affecting a successful outcome. *J Comput Assist Tomogr*. 2013;37(2):176-182. [Crossref]
30. Wooldridge JT, Davis A, Fischer WG, Khalil MF, Zhang M, Afrouzian M. The impact of renal tissue procurement at bedside on specimen adequacy and best practices. *Am J Clin Pathol.* 2019;151(2):205-208. [Crossref]
31. Chang A, Gibson IW, Cohen AH, et al. A position paper on standardizing the nonneoplastic kidney biopsy report. *Clin J Am Soc Nephrol.* 2012;7(8):1365-1368. [Crossref]
32. Walker PD, Cavallo T, Bonsib SM; Ad Hoc Committee on Renal Biopsy Guidelines of the Renal Pathology Society. Practice guidelines for the renal biopsy. *Mod Pathol.* 2004;17(12):1555-1563. [Crossref]
33. Corapi KM, Chen JLT, Balk EM, Gordon CE. Bleeding complications of native kidney biopsy: a systematic review and meta-analysis. *Am J Kidney Dis.* 2012;60(1):62-73. [Crossref]
34. Kanzaki G, Tsuboi N, Utsunomiya Y, Ikegami M, Shimizu A, Hosoya T. Distribution of glomerular density in different cortical zones of the human kidney. *Pathol Int*. 2013;63(3):169-175. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253623

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE



# Minimizing radiation exposure in children: the role of spot region of interest imaging in venous access procedures

Özhan Özgür
Cemil Oktay
Demet Gündüz
Erhan Heperenler
Hakkı Timur Sindel

Akdeniz University Faculty of Medicine, Department of Radiology, Antalya, Türkiye

**PURPOSE**

To evaluate the effectiveness of the spot region of interest (ROI) technique in reducing radiation exposure during fluoroscopically guided venous access procedures in pediatric patients.

**METHODS**

This retrospective study included pediatric patients who underwent central venous access procedures in an interventional radiology unit of a tertiary care center. Data collected included patient demographics, procedure type, target vein, dose area product (DAP), cumulative dose, fluoroscopy time, and the DAP/fluoroscopy time ratio.

**RESULTS**

A total of 131 patients (mean age: 8 ± 4.91; 48.9% women) were included, of whom 44 (33.6%) underwent procedures using the spot ROI technique. The spot ROI group demonstrated significantly lower DAP and cumulative dose than the non-ROI group (reduction ratios: 63.8% and 67.2%, respectively, $P < 0.001$ for all). When normalized to fluoroscopy time, the DAP/fluoroscopy time ratio was also significantly reduced in the spot ROI group [15.34 (7.18–23.57) vs. 25.17 (18.49–42.03); $P < 0.001$].

**CONCLUSION**

Spot ROI is an effective and safe technique for reducing radiation exposure during pediatric venous access procedures without compromising procedural success. Given the high radiation sensitivity of pediatric patients and the potential need for repeated interventions in those with chronic conditions, spot ROI represents a valuable tool for dose optimization and aligns with the As Low As Reasonably Achievable (ALARA) principle.

**CLINICAL SIGNIFICANCE**

Spot ROI-based fluoroscopy significantly reduces radiation exposure in pediatric central venous catheterization without compromising procedural success. This hardware-based dose reduction technique complements the traditional ALARA principle and is particularly valuable for children requiring repeated interventions. These findings support broader clinical adoption and warrant validation in future prospective multicenter studies.

**KEYWORDS**

Catheter, pediatric, radiation, spot ROI, venous access

**Handling editor:** Mehmet Ruhi Onur

**Corresponding author:** Özhan Özgür

**E-mail:** ozhanozgur@gmail.com

Received 03 October 2025; revision requested 03 November 2025; last revision received 24 November 2025; accepted 01 December 2025.

Epub: 16.12.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253623

You may cite this article as: Özgür Ö, Oktay C, Gündüz D, Heperenler E, Sindel HT. Minimizing radiation exposure in children: the role of spot region of interest imaging in venous access procedures. *Diagn Interv Radiol*. 2026;32(4):487-491.

Reducing radiation exposure is essential for ensuring the safety of all interventional radiology procedures, especially in the pediatric population. Pediatric patients are more sensitive to ionizing radiation than adults, and the lifetime risk of radiation-induced malignancy is considerably higher in this population.[1-3] In interventional procedures requiring prolonged fluoroscopy, radiation dose optimization is of paramount importance for patient safety and fulfilling ethical responsibility. This is embodied in the concept of As Low As Reasonably Achievable (ALARA), which emphasizes minimizing radiation dose while maintaining adequate image quality for clinical decision-making.[2] Incorporating the ALARA principle into clinical practice is particularly critical in children, who are more vulnerable to the cumulative effects of repeated imaging and interventions.

Various systems have been introduced to provide either direct or indirect protection against X-ray exposure. Among the earliest of these are beam collimators, which have been used in different designs and configurations since the early days of X-ray imaging to limit radiation dose.[4,5] More recently, the spot region of interest (ROI) function has been developed as an innovative technique to reduce radiation exposure, particularly during neurointerventional procedures.[6]

ROI-based fluoroscopic dose reduction technology delivers full-dose X-rays exclusively to the targeted anatomical region while reducing the dose to surrounding areas by 65%–85% through a motorized attenuation filter.[6] The frosted appearance of the peripheral area allows the operator to visualize the ROI in high resolution while maintaining anatomical orientation.[6,7] Additionally, unlike conventional collimation, the ROI can be moved freely by the table in any X–Y direction.

In pediatric patients, radiation exposure associated with fluoroscopically guided venous access procedures has not been extensively investigated, largely due to the prevailing assumption that the doses involved are relatively low.[8,9]

**Main points**

- Spot region of interest technology provides significant radiation dose reduction during pediatric central venous catheter placement.
- Dose reduction is achieved without compromising procedural success, image quality, or safety.
- This technique offers an effective complement to the As Low As Reasonably Achievable principle and supports safer pediatric interventional radiology practices.

The present study aims to compare radiation doses in pediatric central venous catheterization procedures performed with and without the spot ROI technique on the same angiography system.

## Methods

This retrospective study was conducted at the Interventional Radiology Unit of Akdeniz University Hospital between January 2019 and December 2024. Institutional review board approval was obtained, and the requirement for informed consent was waived due to the retrospective design. Ethical approval for this retrospective study was obtained from the Akdeniz University Ethics Committee (decision number: TBAEK-786; date of approval: 28.08.2025).

The inclusion criteria were as follows: patients aged 0–18 years who underwent venous catheter placement, including temporary or permanent hemodialysis catheters, Hickman and Broviac catheters, and central venous lines (Figure 1). Procedures were performed on an angiography system equipped with spot ROI technology (Alphenix biplane angiographic machine, Canon Medical Systems, Canon, Tochigi, Japan, 2019). Patients with missing dose data were excluded from the study.

For the ROI application, the target anatomical region was aligned with the ROI field on the fluoroscopy monitor (Figure 2). The operator controlled the motorized attenuation filter dynamically via a joystick. The peripheral area appeared frosted but retained low-dose visibility for anatomical orientation. The ROI position could be readjusted as necessary during the procedure.

Procedures were initiated with ultrasound-guided puncture of the vein, followed by ROI-guided fluoroscopy. In cases of vascular variation or procedural difficulty, additional ROI-guided fluoroscopy with contrast injection was used to visualize the venous anatomy before catheter placement.

### Statistical analysis

Data collected included patient demographics, procedure type, target vein, dose area product (DAP), cumulative dose, fluoroscopy time, and the DAP/fluoroscopy time ratio. Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp., Armonk, NY, USA). The Shapiro–Wilk test was applied to assess the normality of continuous variables. Data not conforming to a normal distribution were reported as median values with interquartile ranges. Comparisons between two groups were conducted using the Mann–Whitney U test. Categorical variables were summarized as counts (n) and percentages (%), with differences between groups evaluated using the chi-square test. A $P$ value < 0.05 was considered statistically significant for all analyses.

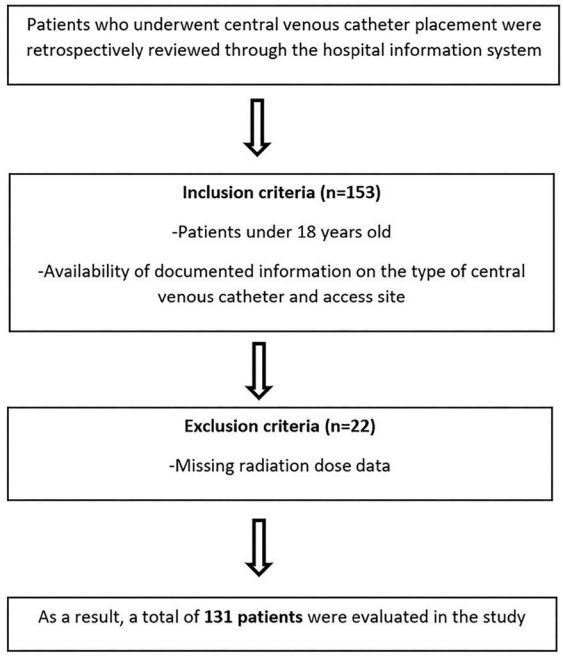



**Figure 1.** Flowchart illustrating the data collection process with the inclusion and exclusion criteria.

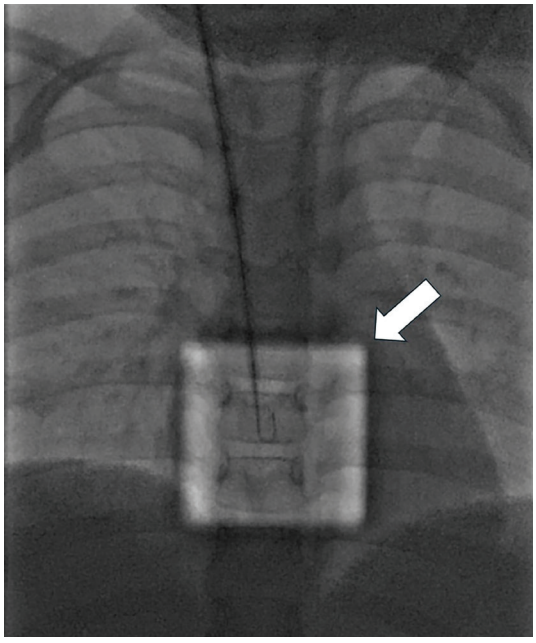


**Figure 2.** The bright central square (white arrow) represents the area imaged with the standard dose, whereas the darker peripheral zone reflects the application of an additional 0.7 mm copper filter for beam attenuation.

## Results

A total of 153 pediatric patients underwent venous catheter placement during the study period. Twenty-two patients were excluded due to missing radiation dose data, leaving 131 patients [mean age: 8 ± 4.91; 48.9% women (n = 64)] for analysis. Of these, 44 (33.6%) procedures were performed using the spot ROI technique, and 87 (66.4%) without it.

There were no significant differences between the groups in terms of age or sex distribution. In both groups, tunneled hemodialysis catheter placement was the most frequent procedure, with the internal jugular vein being the most common access site. In the ROI group, DAP, cumulative dose, and fluoroscopy time were all significantly lower than in the non-ROI group (Table 1). To account for potential differences in fluoroscopy time, the DAP-to-fluoroscopy time ratio was calculated as an indicator of radiation exposure per unit time. This ratio was also significantly reduced in the ROI group compared with the non-ROI group [15.34 (7.18–23.57) vs. 25.17 (18.49–42.03), $P < 0.001$]. Based on median values, the use of spot ROI resulted in reductions of 63.8% in DAP, 67.2% in cumulative dose, and 39% in DAP/fluoroscopy time.

## Discussion

Our study findings indicate that the application of spot ROI resulted in a significant reduction in both total and time-normalized radiation doses. The spot ROI-based fluoroscopic dose reduction technique was initially developed for neurointerventional procedures to reduce radiation exposure for both patients and operators. Phantom studies by Borota and Patz[6] demonstrated that ROI provides significantly lower DAP, air kerma, and skin dose than conventional collimation and spot fluoroscopy methods. This is achieved without compromising the visibility of the target anatomy.[6,7]

To our knowledge, no previous studies have specifically evaluated ROI-based dose reduction in pediatric central venous access procedures. Pediatric dose optimization strategies have traditionally relied on the ALARA principle, frame rate reduction, a narrow field of view, or the air gap technique.[2,10] These methods do not incorporate hardware-based ROI filtering mechanisms.

**Table 1.** Demographic characteristics of the patients and dose-related data of the procedures

| | ROI group (n = 44) | Non-ROI group (n = 87) | *P* value |
|---|---|---|---|
| Sex, n (%)<br>• Female<br>• Male | <br>20 (45.5)<br>24 (54.5) | <br>44 (50.6)<br>43 (49.4) | 0.580 |
| Age* | 7 (5–12) | 9 (4.5–14) | 0.071 |
| Types of venous access, n (%)<br>• Chest wall port<br>• Non-tunneled catheter<br>• Tunneled catheter | <br>4 (9.1)<br>4 (9.1)<br>36 (81.8) | <br>0 (0)<br>0 (0)<br>87 (100) | < 0.001 |
| Venous access site, n (%)<br>• Jugular<br>• Subclavian<br>• Femoral<br>• Transhepatic | <br>40 (90.9)<br>0 (0)<br>2 (4.5)<br>2 (4.5) | <br>86 (98.9)<br>1 (1.1)<br>0 (0)<br>0 (0) | 0.035 |
| Side, n (%)<br>• Right<br>• Left | <br>38 (86.4)<br>6 (13.6) | <br>68 (78.2)<br>19 (21.8) | 0.259 |
| Fluoroscopy time (minute)* | 1.9 (1.52–2.53) | 2.5 (1.99–3.44) | 0.003 |
| DAP (cGy·cm$^2$)* | 24.6 (14.36–43.51) | 68.03 (38.55–123.44) | < 0.001 |
| Cumulative dose (mGy)* | 1.09 (0.61–1.43) | 3.32 (2.06–4.95) | < 0.001 |
| DAP/fluoroscopy time* | 15.34 (7.18–23.57) | 25.17 (18.49–42.03) | < 0.001 |

*Values are presented as median (interquartile range). ROI, region of interest; DAP, dose area product

Borota and Patz[6] reported that neuroangiographic interventions performed with a biplane angiography system using spot ROI and peripheral-field shielding with a physical attenuation filter resulted in up to a 68% dose reduction. The findings of our pediatric central venous catheter cohort (n = 131) are consistent with these results: in procedures performed with spot ROI, the median DAP was 24.6 cGy·cm$^2$ compared with 68.03 cGy·cm$^2$ without ROI, representing an approximate 63.8% reduction ($P < 0.001$); the cumulative dose was 1.09 mGy vs. 3.32 mGy, indicating about a 67.2% reduction ($P < 0.001$); and the DAP/fluoroscopy time ratio, reflecting exposure per unit time, was 15.34 vs. 25.17, approximately 39% lower ($P < 0.001$). Although image quality was not formally scored in this cohort, as reported in the studies by Borota and Patz,[6] the ability to follow surrounding anatomy and the clinical applicability of spot ROI were preserved.

Beyond neurointerventional procedures, this technique has also been applied in percutaneous coronary interventions. Yoshinaga et al.[7] reported that the use of spot ROI markedly decreases radiation dose rates compared with conventional collimation. In their findings, the dose rate with spot ROI was typically reduced to approximately one-third to one-half of that observed without its use. Moreover, this approach may help avoid unintentional dose escalation, which can occur when aggressive collimation triggers automatic dose compensation by the imaging system.

Compared with conventional collimation, the spot ROI functionality provides a more precise and efficient method for radiation dose reduction. Although classical collimation can decrease radiation exposure, it often leads to unintended dose increases due to automatic exposure compensation, and it may also compromise image quality in peripheral regions.[11,12] Spot Fluoroscopy was introduced as an advancement over standard collimation and has been shown to reduce both patient and operator doses while maintaining sufficient image quality. However, it is limited by reduced flexibility and a smaller visual field.[12] In contrast, spot ROI offers a distinct advantage by allowing selective visualization of the target region with continuous dose reduction while preserving peripheral anatomical context.[6] Borota and Patz[6] reported that spot ROI achieves approximately 30%–50% dose reduction compared with conventional collimation, without the drawback of dose compensation effects seen in standard methods. Thus, spot ROI represents a superior evolution of dose-optimization strategies, combining the benefits of spot fluoroscopy with enhanced safety and image fidelity.

The spot ROI technique is applied in neurointerventional and cardiac procedures, where intervention times are typically prolonged. However, its potential benefits in pediatric patients are particularly major, regardless of procedure time. Children are more radiosensitive than adults due to higher cell proliferation rates and longer life expectancy, making any reduction in dose clinically meaningful, especially in those with chronic conditions who require repeated central venous access and are at risk of cumulative radiation exposure due to their increased radiosensitivity.[2] By demonstrating the effectiveness and advantages of spot ROI in this vulnerable group, our study provides a meaningful contribution to the existing literature.

Importantly, ROI use did not adversely affect procedural duration, technical success rates, or image quality, supporting its feasibility and safety in pediatric interventional radiology. This study helps fill a gap in the literature and provides a basis for larger multicenter prospective trials to validate the benefits of spot ROI in this patient population for other interventional radiology procedures.

This study has several limitations. First, its retrospective design may carry inherent risks of selection and referral bias, given that the cohort was derived from a tertiary care institution. Pediatric interventional radiology procedures present specific challenges related to body size, which in this population can range from infants with extremely low birth weight to adolescents and young adults. There was no information regarding the patients' weight and height at the time of the procedure, which may affect the study's dose analysis. Although the case distribution in tertiary centers is typically skewed toward smaller patients, no significant age differences were identified between the groups in this study. Second, the calculated radiation dose estimates are directly relevant only to operators employing comparable equipment, techniques, and training. As our institution is a teaching hospital, a proportion of venous access procedures are performed by radiology residents under faculty supervision. Procedures conducted exclusively by highly experienced operators would likely result in lower radiation exposure. Lastly, the spot ROI technique is currently commercially available only on a single angiography system, which may restrict the generalizability of our findings. Additionally, the implementation of this technology in other centers may be limited by the need for equipment upgrades or associated costs.

In conclusion, the spot ROI-based fluoroscopic dose reduction technique effectively reduces patient radiation exposure during pediatric central venous catheter placement without compromising procedural success. This emerging technology has the potential to enhance both the efficiency and safety of interventional radiology practice. However, its effectiveness and safety should be further validated through prospective multicenter studies.

### Footnotes

### Conflict of interest disclosure

The authors declared that there is no conflict of interest.

## References


1. Strauss KJ. Pediatric interventional radiography equipment: safety considerations. *Pediatr Radiol*. 2006;36 Suppl 2:126-135. [Crossref]
2. Connolly B, Racadio J, Towbin R. Practice of ALARA in the pediatric interventional suite. *Pediatr Radiol*. 2006;36 Suppl 2:163-167. [Crossref]
3. Osei FA, Hayman J, Sutton NJ, Pass RH. Radiation dosage during pediatric diagnostic or interventional cardiac catheterizations using the "air gap technique" and an aggressive "as low as reasonably achievable" radiation reduction protocol in patients weighing <20 kg. *Ann Pediatr Cardiol*. 2016;9:16-21. [Crossref]
4. Bednarek DR. Collimation. In: Karellas A, Thomadsen BR, editors. Cardiovascular and neurovascular imaging. Boca Raton (FL): CRC Press; 2016. p. 427-429. [Crossref]
5. Stuchebrov SG, Miloichikova IA, Danilova IB. The X-ray beam passage through the collimator made of different materials: numerical simulation. *J Phys Conf Ser*. 2016;671:012012. [Crossref]
6. Borota L, Patz A. Spot region of interest imaging: a novel functionality aimed at X-ray dose reduction in neurointerventional procedures. *Radiat Prot Dosimetry*. 2020;188:322-331. [Crossref]
7. Yoshinaga M, Muramatsu T, Kondo Y, Miyazaki A, Fukushima T, Sobue Y, et al. Using a spot ROI function to minimize radiation during chronic total occlusion percutaneous coronary interventions. *JACC Case Rep*. 2025;30:103526. [Crossref]

8. Storm ES, Miller DL, Hoover LJ, Georgia JD, Bivens T. Radiation doses from venous access procedures. *Radiology*. 2006;238:1044-1050. [Crossref]
9. Miller DL, Balter S, Cole PE, Lu HT, Schueler BA, Geisinger M, et al. Radiation doses in interventional radiology procedures: the RAD-IR study: Part I: Overall measures of dose. *J Vasc Interv Radiol.* 2003;14:711-727. [Crossref]
10. Bacher K, Bogaert E, Lapere R, De Wolf D, Thierens H. Patient-specific dose and radiation risk estimation in pediatric cardiac catheterization. *Circulation*. 2005;111:83-89. [Crossref]
11. Kim DJ, Park MK, Jung DE, Kang JH, Kim BM. Radiation dose reduction without compromise to image quality by alterations of filtration and focal spot size in cerebral angiography. *Korean J Radiol.* 2017;18:722-728. [Crossref]
12. Borota L, Jangland L, Åslund PE, Ronne-Engström E, Nyberg C, Mahmoud E, et al. Spot fluoroscopy: a novel innovative approach to reduce radiation dose in neurointerventional procedures. *Acta Radiol*. 2017;58:600-608. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253386

MUSCULOSKELETAL IMAGING

ORIGINAL ARTICLE



# Association of patellofemoral malalignment with early trochlear and patellar chondromalacia: a prospective T2* mapping study

Uğurcan Süner[1]
Atilla Hikmet Çilengir[2]
Tuğrul Bulut[3]
Merve Gürsoy[4]
Yılmaz Önder[3]
Berna Dirim Mete[2]

[1]University of Health Sciences Türkiye, Van Training and Research Hospital, Department of Orthopedics and Traumatology, Van, Türkiye

[2]İzmir Democracy University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[3]İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Orthopedics and Traumatology, İzmir, Türkiye

[4]İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Radiology, İzmir, Türkiye

**PURPOSE**

To investigate the association between patellofemoral malalignment and early-stage trochlear and patellar chondromalacia using the T2* mapping method.

**METHODS**

Seventy-five patients were included in the study and divided into two groups based on the presence (patient group) or absence (control group) of patellofemoral malalignment on magnetic resonance imaging. The T2* mapping measurements were evaluated by dividing patellar and trochlear cartilage into 12 quadrants on sagittal slices. The groups were first compared based on the mean T2* relaxation times of the cartilage. Subsequently, the 12 quadrants were compared individually between the two groups. Cut-off values were calculated for the quadrants, with significant differences observed.

**RESULTS**

The patient group included 39 patients, and the control group included 36 patients. There was no significant difference between the groups in terms of mean T2* relaxation values for the trochlear and patellar cartilage. However, in the separate comparison of the 12 quadrants, T2* relaxation values in the upper-outer–outer (P1, T1) and upper-outer–inner (P2, T2) quadrants of both the trochlear and patellar cartilage were found to be statistically significantly higher in the patient group. Similarly, significant cut-off values were identified for the T1, P1, and P2 quadrants.

**CONCLUSION**

Early chondromalacia can be quantitatively detected using T2* mapping. In patients with elevated T2* relaxation values in the superior–lateral regions of the patellar and trochlear cartilage, patellofemoral malalignment should be considered in the etiology.

**CLINICAL SIGNIFICANCE**

Chondromalacia caused by patellofemoral malalignment may exhibit an asymmetric onset, with the superior–lateral quadrant as the initial site of cartilage damage in both trochlear and patellar cartilage.

**KEYWORDS**

Cartilage, malalignment, mapping, patella, trochlea

**Handling editor:** Zeynep Maraş Özdemir

**Corresponding author:** Atilla Hikmet Çilengir

**E-mail:** acilengir@gmail.com

Received 08 April 2025; revision requested 24 May 2025; accepted 14 July 2025.

Epub: 18.08.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253386

Patellofemoral malalignment is a common cause of anterior knee pain in the young population. The normal alignment of the patella with the trochlear groove (TG) is a critical factor in load bearing, and malalignment (displacement of the patella from its expected trajectory within the knee joint) can lead to chondromalacia, synovial proliferation, and subchondral bone changes from a young age.[1,2] Early diagnosis and preventive interventions play a key role in avoiding irreversible cartilage damage. Therefore, it is important to monitor and thoroughly examine patients with patellofemoral malalignment from an early stage for potential cartilage damage.

You may cite this article as: Süner U, Çilengir AH, Bulut T, Gürsoy M, Önder Y, Dirim Mete B. Association of patellofemoral malalignment with early trochlear and patellar chondromalacia: a prospective T2* mapping study. *Diagn Interv Radiol*. 2026;32(4):492-499.

It is accepted that cartilage damage can be reversible in the early stages or that its progression can be halted with preventive interventions.[3] Thus, for disease prevention or effective treatment, cartilage degeneration must be reliably detected at the earliest stage. However, with conventional magnetic resonance imaging (MRI), early detection is generally not possible. Standard sequences provide only morphological information about cartilage. This limitation exists because changes in water content within degenerated cartilage are minimal in the early stages, and the sensitivity of standard sequences to detect these changes is low.[4,5] Furthermore, due to their macromolecular structure, both proteoglycan and collagen protons have very short T2 relaxation times, making direct MRI measurement difficult.[5-7] As a result, quantitative MRI techniques such as T2 and T2* mapping have gained importance in this field.[8,9] These mapping methods provide quantitative information about cartilage composition by evaluating changes in extracellular matrix components.[10]

In recent years, most studies examining the association between patellofemoral malalignment and chondromalacia have focused on patellar cartilage, often overlooking trochlear cartilage.[11-13] Similarly, studies using T2 and T2* mapping have generally evaluated tibiofemoral and patellar cartilage, with limited attention to trochlear cartilage.[14-21] However, in cases of patellofemoral malalignment, the trochlear cartilage can also be affected and may contribute to anterior knee pain even in the absence of patellar cartilage damage. To the best of our knowledge, no previous study has examined the association between patellofemoral malalignment and early-stage damage to both patellar and trochlear cartilage using T2* mapping.

The main hypothesis of this study is that patellofemoral malalignment is an etiological factor in both patellar and trochlear chondromalacia and that the chondromalacia caused by this malalignment exhibits an asymmetric onset on both joint surfaces from an early age. Another hypothesis is that this asymmetric chondromalacia can be detected at an early stage in both the patella and trochlea using T2* mapping.

**Main points**

- Patellofemoral malalignment is a key cause of chondromalacia.
- Early chondromalacia in patellofemoral malalignment typically involves the superolateral portions of patellar and trochlear cartilage.
- Early detection of chondromalacia using T2* mapping may improve treatment success.

In line with these hypotheses, one of the aims of the study is to investigate the association between patellofemoral malalignment and early-stage chondromalacia of trochlear and patellar cartilage using T2* mapping in young adults. Another aim is to determine which regions of the cartilage is first affected by chondromalacia.

## Methods

### Participants

This prospective study was approved by the İzmir Katip Çelebi University Faculty of Medicine Clinical Researches Ethics Committee (date: 12.09.2019, number: 88). Between January 2023 and January 2024, 524 patients who presented to the orthopedics and traumatology department with complaints of anterior knee pain and were suspected of having patellar malalignment based on physical examination by two orthopedists were included in the study (Figure 1). Exclusion criteria included being under 18 or over 40 years of age, a history of trauma or surgery in the knee region, any rheumatologic disease, or evidence of osteoarthritis on knee radiographs. A power analysis was conducted to determine the sample size, calculating that at least 36 patients per group (72 in total) would be required. The study was planned to conclude once this minimum number was reached in each group.

From the 524 patients prospectively evaluated, the following were excluded: those under 18 years (n = 62), over 40 years (n = 308), with a history of trauma (n = 23) or surgery (n = 20) in the knee region, with any rheumatologic disease (n = 12), and with osteoarthritic findings on knee radiographs (n = 12). The age range of 18–40 years was selected to eliminate the influence of pediatric cartilage and age-related degenerative changes on the results. Additionally, cases with traumatic, postoperative, rheumatologic, or osteoarthritic cartilage changes were excluded to avoid confounding effects.

A routine knee MRI and T2* mapping protocol was performed using a 3T MRI scanner (Magnetom Lumina, Siemens Healthineers, Erlangen, Germany) on the 87 patients who met the inclusion criteria (Supplementary Table 1). Four patients were excluded due to motion artifacts that rendered the MRI scans non-diagnostic. Furthermore, since the primary aim of the study was to detect early-stage chondromalacia in young adults and given that the optimal T2* relaxation time measurement is not feasible in the presence of full-thickness or near full-thickness cartilage defects due to the partial volume effect of synovial fluid within the mapping area, eight patients with high-grade (stage 3 or 4) chondromalacia were excluded.[22] This ensured a homogeneous study cohort and alignment with the study's objectives.

### Patellofemoral instability magnetic resonance imaging parameters

A total of 75 patients underwent morphological evaluation using standard MRI sequences. Parameters identified in the literature for diagnosing patellofemoral malalignment–including the Insall–Salvati index, lateral patellofemoral angle, lateral trochlear inclination angle, trochlear sulcus depth, trochlear sulcus angle, tibial tubercle (TT)–TG distance, and medial trochlea/lateral trochlea length ratio–were measured. Patients were divided into two groups, those with patellofemoral malalignment (patient group) and those without (control group), based on the cut-off values established in the literature (Table 1). These measurements were performed using validated techniques described in the Supplementary Material 1.[23-28] Patients presenting with anterior knee pain and physical examination findings consistent with patellofemoral malalignment were classified into the malalignment group if one or more of these MRI-based morphological parameters were present.

### Cartilage T2* relaxation measurement

Next, T2* relaxation measurements were performed on sagittal slices with a thickness of 3 mm. In the T2* mapping sequences, the trochlear and patellar cartilage was manually divided on sagittal images into superior, middle, and inferior thirds along the vertical axis and into medial–medial, medial–lateral, lateral–medial, and lateral–lateral quadrants along the horizontal axis, yielding a total of 12 quadrants for each structure (Table 2). This approach allowed a comprehensive sampling of both the patellar and trochlear cartilage.

After identifying the four main quadrants along the horizontal axis, a sagittal slice passing through the center of each quadrant was selected separately for the patellar and trochlear cartilage (yielding eight sagittal slices in total). A region of interest (ROI) was then manually placed at the center of the superior, middle, and inferior thirds of each

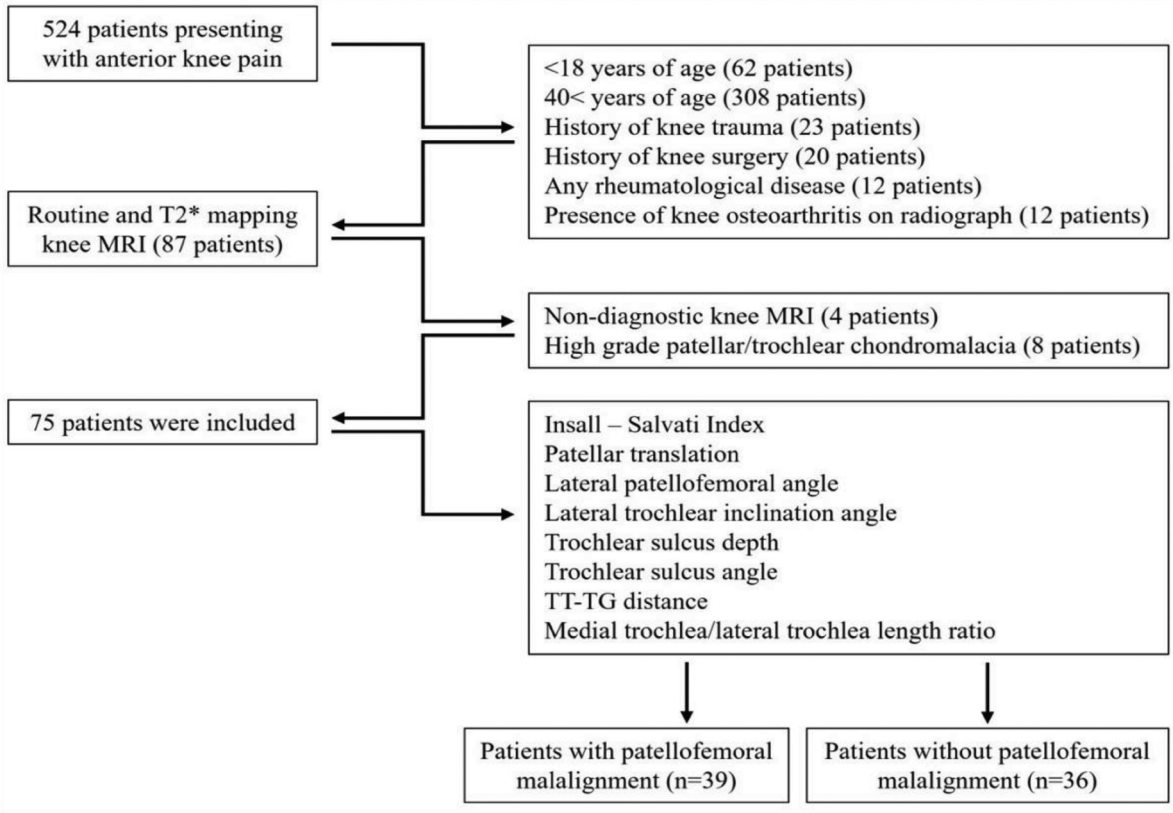



**Figure 1.** Study flowchart. MRI, magnetic resonance imaging; TT-TG, tibial tubercle-trochlear groove.

**Table 1.** Patellofemoral malalignment magnetic resonance imaging measurements and values associated with patellofemoral malalignment

| Patellofemoral malalignment MRI measurements | Values associated with patellofemoral malalignment |
|---|---|
| Insall–Salvati index | >1.2 |
| Lateral patellofemoral angle | <8° |
| Lateral trochlear inclination angle | <11° |
| Trochlear sulcus depth | <3 mm |
| Trochlear sulcus angle | >144° |
| TT–TG distance | >15 mm |
| Medial trochlea/lateral trochlea ratio | <40% |

MRI, magnetic resonance imaging; TT–TG: tibial tubercle–trochlear groove.

selected sagittal slice, and T2* relaxation values were recorded (Figure 2). Measurements were performed using the Siemens syngo. via (Siemens Healthcare, Erlangen, Germany) software on the MRI workstation. To minimize sampling error, small and similarly sized ROIs were used whenever possible, and high magnification was employed to avoid interfaces with synovial effusion and subchondral bone.[29]

All measurements were performed in consensus by two radiologists, one with 10 years of experience and another with 24 years of experience in musculoskeletal radiology. To reduce measurement errors and potential bias, morphological assessments and T2* relaxation measurements were conducted in separate sessions, spaced 1 month apart.

### Statistical analysis

To calculate the sample size, G*Power 3.1 (Heinrich Heine University, Düsseldorf, Germany) software was used. The study by Subhawong et al.[30] served as the basis for this calculation. Based on the patient and control group average values reported in that study, with an α value of 0.05 and power of 0.8, the total sample size was calculated as 72, with 36 in each group.

Statistical analyses were performed using IBM SPSS Statistics Standard Concurrent User V27 (IBM Corp., Armonk, New York, USA). Descriptive statistics were presented as count (n), percentage (%), mean ± standard deviation (SD), median, minimum, and maximum values. The normality of numerical variables was assessed using the Shapiro–Wilk test, and homogeneity of variances was evaluated using the Levene test.

Group comparisons were performed using the independent two-sample t-test, as the data met the assumptions for parametric testing. The receiver operating characteristic (ROC) analysis method was used to evaluate the area under the curve (AUC). Repeated measures analysis of variance (ANOVA) was used to compare measurement values, mixed design ANOVA was used to compare values between groups, and Bonferroni correction was applied for main effect comparisons in the mixed design ANOVA. A *P* value of <0.05 was considered statistically significant.

## Results

Among the 75 patients included in the study, 39 had MRI findings of patellofemoral malalignment (patient group), whereas the remaining 36 did not show such findings (control group). The patient group comprised 17 men and 22 women, and the control group contained 16 men and 20 women. The mean age of the patient group

**Table 2.** T2* relaxation measurement areas of the trochlear and patellar cartilage

| Trochlea quadrant | Region description | Patella quadrant | Region description |
|---|---|---|---|
| T1 | Upper-lateral–lateral quadrant | P1 | Upper-lateral–lateral quadrant |
| T2 | Upper-lateral–medial quadrant | P2 | Upper-lateral–medial quadrant |
| T3 | Upper-medial–lateral quadrant | P3 | Upper-medial–lateral quadrant |
| T4 | Upper-medial–medial quadrant | P4 | Upper-medial–medial quadrant |
| T5 | Mid-lateral–lateral quadrant | P5 | Mid-lateral–lateral quadrant |
| T6 | Mid-lateral–medial quadrant | P6 | Mid-lateral–medial quadrant |
| T7 | Mid-medial–lateral quadrant | P7 | Mid-medial–lateral quadrant |
| T8 | Mid-medial–medial quadrant | P8 | Mid-medial–medial quadrant |
| T9 | Lower-lateral–lateral quadrant | P9 | Lower-lateral–lateral quadrant |
| T10 | Lower-lateral–medial quadrant | P10 | Lower-lateral–medial quadrant |
| T11 | Lower-medial–lateral quadrant | P11 | Lower-medial–lateral quadrant |
| T12 | Lower-medial–medial quadrant | P12 | Lower-medial–medial quadrant |

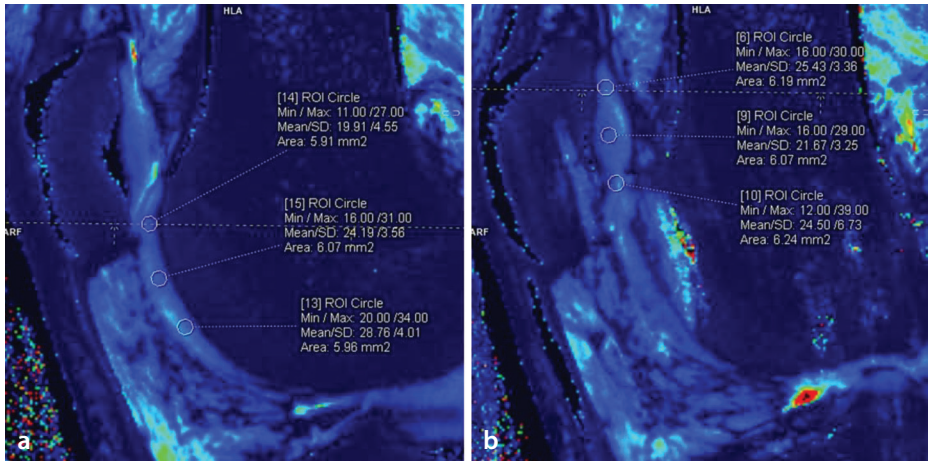



**Figure 2.** Trochlear **(a)** and patellar **(b)** cartilage T2* relaxation time measurements. ROI, region of interest; SD, standard deviation.

was 27.69 years (range: 19–40; SD: ±6.39), whereas the control group had a mean age of 28.17 years (range: 19–39; SD: ±6.58). The mean body mass index (BMI) was 24.38 (SD: ±3.01) in the patient group and 24.68 (SD: ±3.73) in the control group. No statistically significant differences were found between the groups regarding gender, age, or BMI ($P = 0.949$, $P = 0.680$, and $P = 0.725$, respectively).

An overall average T2* relaxation value for both the trochlea and patella was calculated by averaging the detailed T2* relaxation values from the 12 quadrants. When comparing these average T2* relaxation values without quadrant differentiation, no statistically significant differences were observed between the groups (Table 3).

Next, T2* relaxation values obtained by dividing the trochlear and patellar cartilage into 12 quadrants in the sagittal plane were compared. According to the results, T2* relaxation values in the superior-lateral–lateral (T1, P1) and superior-lateral–medial (T2, P2) quadrants of both the trochlear and patellar cartilage were significantly higher in the patient group ($P = 0.013$, $P = 0.001$, $P = 0.017$, and $P = 0.044$, respectively) (Tables 4 and 5).

Additionally, a ROC analysis was conducted to determine cut-off values for T2* relaxation in the 12 defined quadrants (Supplementary Tables 2 and 3). The evaluation revealed that the AUC was statistically significant for the P1, P2, and T1 quadrants. When assessing T2* relaxation values in these quadrants, values in the T1, P1, and P2 quadrants were significantly higher in the patient group. It was also possible to determine a diagnostic cut-off value for each of these regions. The threshold values were calculated as above 21.7 milliseconds for T1, above 24.1 milliseconds for P1, and above 24.3 milliseconds for P2 (Supplementary Table 4).

## Discussion

In this study, we evaluated the association between patellofemoral malalignment and early-stage trochlear and patellar chondromalacia using T2* mapping. There was no difference in the average T2* relaxation values of the trochlear and patellar cartilage between the two groups, with and without MRI findings of patellofemoral malalignment. However, when the T2* relaxation values of the 12 quadrants were compared individually, values in the superior-lateral–lateral quadrant (P1, T1) and superior-lateral–medial quadrant (P2, T2) of both the trochlear and patellar cartilage were significantly higher in the patient group. These results suggest that chondromalacia caused by patellofemoral malalignment may exhibit an asymmetric onset, with the superior-lateral quadrant as the initial site of cartilage damage in both the trochlear and patellar cartilage.

It is well established that early-stage cartilage degeneration can be reversible or its progression halted through various interventions, including pharmacotherapy, lifestyle changes, or realignment surgery.[3] Therefore, reliable early detection of cartilage degeneration is critical for successful treatment. Standard sequences (T1 weighted, T2 weighted, and proton density) only provide information about cartilage morphology and have relatively low sensitivity for detecting early degeneration.[5] Quantitative MRI techniques, such as T2 and T2* mapping, have gained increasing importance for early cartilage assessment.[8,9] In particular, T2 mapping is used to evaluate water content and collagen fiber orientation within cartilage and is widely applied in clinical practice.[31] By contrast, T2* mapping similarly assesses cartilage water content but has the advantage of

**Table 3.** Comparison of mean T2* relaxation values (milliseconds) of the trochlear and patellar cartilage in groups with patellofemoral malalignment

| Patellofemoral malalignment | Trochlear cartilage (mean ± SD) | Patellar cartilage (mean ± SD) |
|---|---|---|
| Yes | 22.18 ± 3.59 | 22.52 ± 3.15 |
| No | 20.78 ± 3.47 | 21.63 ± 2.06 |
| *P* value | 0.094 | 0.163 |

SD, standard deviation.

**Table 4.** Comparison of trochlear cartilage T2* relaxation values (milliseconds) between groups

| Quadrant | Patellofemoral malalignment | | *P* value |
|---|---|---|---|
| | No | Yes | |
| T1 | 17.60 ± 4.17 | 20.53 ± 5.54 | **0.013** |
| T2 | 19.38 ± 5.99 | 22.60 ± 6.04 | **0.017** |
| T3 | 23.31 ± 6.18 | 25.60 ± 6.68 | 0.132 |
| T4 | 16.25 ± 4.92 | 18.54 ± 5.86 | 0.074 |
| T5 | 24.88 ± 4.72 | 26.52 ± 6.91 | 0.242 |
| T6 | 25.59 ± 5.34 | 25.88 ± 5.38 | 0.818 |
| T7 | 24.80 ± 7.92 | 27.31 ± 5.77 | 0.122 |
| T8 | 20.73 ± 5.47 | 22.20 ± 5.32 | 0.247 |
| T9 | 18.18 ± 5.96 | 19.89 ± 6.88 | 0.260 |
| T10 | 19.94 ± 6.46 | 19.09 ± 5.66 | 0.548 |
| T11 | 18.36 ± 5.54 | 19.23 ± 5.41 | 0.496 |
| T12 | 18.38 ± 5.94 | 18.77 ± 5.16 | 0.762 |

Values are presented as mean ± standard deviation. The values written in bold are statistically significant.

**Table 5.** Comparison of patellar cartilage T2* relaxation values (milliseconds) between groups

| Quadrant | Patellofemoral malalignment | | *P* value |
|---|---|---|---|
| | No | Yes | |
| P1 | 23.22 ± 4.37 | 27.03 ± 4.91 | **0.001** |
| P2 | 23.96 ± 4.51 | 26.49 ± 5.92 | **0.044** |
| P3 | 23.11 ± 5.32 | 23.34 ± 4.30 | 0.840 |
| P4 | 19.46 ± 3.04 | 20.32 ± 4.37 | 0.322 |
| P5 | 22.52 ± 5.02 | 23.14 ± 7.27 | 0.674 |
| P6 | 23.33 ± 6.10 | 24.28 ± 5.89 | 0.498 |
| P7 | 22.31 ± 4.12 | 21.83 ± 4.42 | 0.629 |
| P8 | 19.64 ± 4.1 | 19.98 ± 5.87 | 0.775 |
| P9 | 20.29 ± 3.68 | 19.58 ± 5.12 | 0.499 |
| P10 | 22.14 ± 3.39 | 23.22 ± 5.57 | 0.323 |
| P11 | 20.40 ± 4.04 | 21.91 ± 4.74 | 0.149 |
| P12 | 19.29 ± 4.16 | 19.20 ± 4.89 | 0.930 |

Values are presented as mean ± standard deviation. The values written in bold are statistically significant.

a shorter acquisition time compared with T2 mapping. Additionally, T2* mapping is more sensitive to local magnetic field inhomogeneities, making it more effective for detecting early-stage changes.[14,32]

In addition, T1 rho mapping, which is sensitive to proteoglycan content, can indicate early degeneration but is limited by longer acquisition times and reduced scanner compatibility.[33] Delayed gadolinium-enhanced MRI of cartilage measures glycosaminoglycan concentrations using a contrast agent and can detect early glycosaminoglycan loss. However, its use is limited by the need for contrast injection and prolonged imaging duration.[33]

Quantitative T2* mapping has been reported as an effective technique for assessing cartilage compositional integrity.[14] Compared with T2 mapping, T2* mapping is more sensitive to subtle changes in tissue composition due to its sensitivity to local field inhomogeneities, which have dephasing effects.[14,32] It also provides a high signal-to-noise ratio and high spatial resolution, with relatively short scan times.[16] Mars et al.[34] concluded that T2* mapping was superior to T2 mapping for detecting cartilage injury.

Therefore, T2* mapping is considered more sensitive for detecting changes within cartilage tissue. However, its role across different stages of cartilage degeneration remains incompletely defined. There is evidence that T2* mapping is useful for identifying early-stage degeneration in cartilage areas that exhibit elevated T2* relaxation times compared with normal cartilage.[14] However, no previous studies have identified the specific anatomical regions where early cartilage damage begins in the patella and trochlea.

The most significant difference between our study and previous studies in the literature is that we performed detailed measurements by dividing the trochlear and patellar cartilage into 12 quadrants to detect heterogeneous changes. This aspect of our study is superior to others in the literature. Only Ruiz Santiago et al.[9] used a similar quadrant division method; however, the main purpose of their study was to determine the relationship between cartilage T2 values and the current staging of chondromalacia, which is not the subject of our research.

In our study, when comparing the average T2* relaxation values of the trochlea and patella in the sagittal plane between the patient and control groups, no statistically significant difference was found. However, when detailed T2* relaxation measurements were made by dividing the trochlea and patella into 12 quadrants, statistically significant results were obtained in specific quadrants. Similar to our findings, Kim et al.[35] reported that the involvement of lateral patellar facet cartilage was statistically significantly higher in the patellofemoral malalignment group. Additionally, in our study, T2* relaxation values in the lateral quadrants (P1, P2) of the patellar cartilage were higher in the patellofemoral malalignment group. Furthermore, T2* relaxation values in the lateral quadrants (T1, T2) of the trochlear cartilage, similar to the patella, were also significantly higher in our study. Various mechanisms may explain this. One explanation is that in patients with patellofemoral malalignment, the contact area between the patella and the TG decreases during flexion, leading to increased maximum patellofemoral contact pressure.[36] Additionally, when the TT–TG distance increases, the patella shifts laterally, which especially increases patellofemoral joint stress on the lateral side.[37] These changes in load transfer may explain the greater occurrence of chondromalacia on the lateral side.

Consistent with these mechanisms, our study also found statistically significant T2* relaxation values in the lateral quadrants and statistically significant results in the ROC analysis. These findings indicate that chondromalacia caused by patellofemoral malalignment exhibits an asymmetric onset and is more prevalent on the lateral joint surfaces of both the patella and trochlea. Therefore, evaluating the trochlear and patellar cartilage by dividing it into quadrants, rather than relying on average values, using the T2* mapping method is more sensitive and meaningful for assessing early-stage chondromalacia.

Another statistical analysis in our study involved conducting ROC analysis to determine cut-off values for the presence of chondromalacia in all quadrants of both the trochlea and patella. Accordingly, it was found that cut-off values could be determined for specific quadrants in the sagittal plane. In particular, T2* mapping measurements in the T1, P1, and P2 quadrants revealed statistically significant relaxation values and allowed for the determination of cut-off values in the ROC analysis.

Therefore, we believe it is meaningful to focus on the T1 and P1 quadrants for measurements and to consider the results obtained from these regions. Although a cut-off value could also be determined for the P2 quadrant in the ROC analysis, the practical and memorable nature of obtaining significant results in both the T2* relaxation values and ROC analysis for the T1 and P1 areas may facilitate easier measurement and clinical application. For cut-off values to be practically applicable in clinical settings, they need to be verified through further studies. Thus, rather than proposing the cut-off values identified in our study as definitive criteria, we recommend that quadrants where cut-off values can be determined be evaluated more carefully when assessing early-stage chondromalacia. This approach will help alert clinicians to areas that require focused attention and support more effective follow-up.

Similarly, Fulkerson et al.[38] noted that different regions of the patellar cartilage are not equally affected by chondromalacia and that different mechanisms are responsible for damage in different regions. He reported that lateral patellar facet cartilage damage is associated with chronic patellar tilt, whereas medial facet cartilage damage is linked to patellar dislocation. Therefore, in clinical evaluations, both T2* relaxation values and the patient's morphological MRI findings and clinical history should be considered.

Our study has some limitations. The main limitation is that, despite dividing the trochlea and patella into 12 quadrants and measuring from the midpoint of each quadrant to detect heterogeneous chondromalacia distribution, exact standardization may not be achieved for measurement localization in each patient. Chondromalacia in each quadrant may not occur precisely in the center but may develop at the edges.

Second, in our study, T2* relaxation value measurements were performed only on sagittal slices for both the patella and trochlea, without including axial or coronal slices. Given the imaging planes, the rationale for performing T2* relaxation measurements in the sagittal plane for both trochlear and patellar chondromalacia was to ensure more accurate assessment, particularly for the trochlea, which has a steep orientation. Sagittal plane acquisition allows evaluation of the articular cartilage perpendicular to the shearing forces acting on the joint. Anatomically, measurements taken in other planes may not yield accurate results for the trochlea. Additionally, the relatively thick slice thickness of T2* mapping compared with cartilage thickness in other planes may contribute to inaccuracies. These factors may lead to the incorrect calculation of T2* relaxation values due to the partial volume effect from non-cartilage structures.[22] Therefore, we believe sagittal plane measurements are more reliable than those taken in the axial or coronal planes, especially for the trochlea. However, evaluation and comparison across slices in axial and coronal planes could contribute to standardizing the most appropriate measurement plane.

All patients in our study were symptomatic with anterior knee pain; however, the absence of an objective scoring system to quantify symptom severity may be considered a third limitation. As a general limitation of imaging-based studies, there was no histopathological verification of chondromalacia. Nonetheless, given the presence of well-defined and widely accepted diagnostic criteria, invasive procedures are no longer preferred or routinely used in current clinical practice and research. Although T2* value measurements in our study were performed by consensus between two radiologists experienced in musculoskeletal radiology, intra-observer and inter-observer agreement were not assessed.

In conclusion, in the presence of patellofemoral malalignment, both patellar and trochlear cartilage can be examined more specifically by dividing them into quadrants

rather than using a general assessment. Early chondromalacia can be detected by quantitatively assessing the superior-lateral sections of the patellar and trochlear cartilage using T2* mapping. Additionally, in patients with high T2* relaxation values identified in the superior-lateral sections of the patellar and trochlear cartilage, patellofemoral malalignment should be primarily considered as the etiology of chondromalacia. Thus, T2* mapping may serve as a valuable non-invasive imaging tool for the early diagnosis and monitoring of cartilage degeneration in patients with patellofemoral malalignment.

### Footnotes

### Conflict of interest disclosure

The authors declared no conflicts of interest.


### Funding

İzmir Katip Çelebi University Scientific Research Projects Coordination (project number: 2022-TDU-TIPF-0012).


**Supplementary Material 1:** https://d2v96fxpocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/23f2337d-f1ce-4ce6-916a-67e9a5f-4f3bd.pdf

**Supplementary Tables:** https://d2v96fx-pocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/f7d5e54f-deaf-4a03-b524-6754dba51474.pdf

## References


1. Tsavalas N, Katonis P, Karantanas AH. Knee joint anterior malalignment and patellofemoral osteoarthritis: an MRI study. *Eur Radiol*. 2012;22(2):418-428. [Crossref]
2. Liao TC, Pedoia V, Link TM, Majumdar S, Souza RB. Association of patella alignment with cartilage relaxation times and self-reported symptoms in individuals with patellofemoral degeneration. *J Orthop Res*. 2023;41(3):562-569. [Crossref]
3. Feucht MJ, Izadpanah K, Vogt S, Mehl J. Cartilage repair procedures for early osteoarthritis. *Orthopade*. 2021;50(5):356-365. [Crossref]
4. Quatman CE, Hettrich CM, Schmitt LC, Spindler KP. The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review. *Am J Sports Med*. 2011;39(7):1557-1568. [Crossref]
5. Zibetti MVW, Menon RG, de Moura HL, Zhang X, Kijowski R, Regatte RR. Updates on compositional MRI mapping of the cartilage: emerging techniques and applications. *J Magn Reson Imaging*. 2023;58(1):44-60. [Crossref]
6. Truhn D, Sondern B, Oehrl S, et al. Differentiation of human cartilage degeneration by functional MRI mapping-an ex vivo study. *Eur Radiol*. 2019;29(12):6671-6681. [Crossref]
7. Souza RB, Kumar D, Calixto N, et al. Response of knee cartilage T1rho and T2 relaxation times to in vivo mechanical loading in individuals with and without knee osteoarthritis. *Osteoarthritis Cartilage*. 2014;22(10):1367-1376. [Crossref]
8. Apprich S, Mamisch TC, Welsch GH, et al. Quantitative T2 mapping of the patella at 3.0T is sensitive to early cartilage degeneration, but also to loading of the knee. *Eur J Radiol*. 2012;81(4):438-443. [Crossref]
9. Ruiz Santiago F, Pozuelo Calvo R, Almansa López J, Guzmán Álvarez L, Castellano García MDM. T2 mapping in patellar chondromalacia. *Eur J Radiol*. 2014;83(6):984-988. [Crossref]
10. Kijowski R, Blankenbaker DG, Munoz Del Rio A, Baer GS, Graf BK. Evaluation of the articular cartilage of the knee joint: value of adding a T2 mapping sequence to a routine MR imaging protocol. *Radiology*. 2013;267(2):503-513. [Crossref]
11. van Middelkoop M, Macri EM, Eijkenboom JF, et al. Are patellofemoral joint alignment and shape associated with structural magnetic resonance imaging abnormalities and symptoms among people with patellofemoral pain? *Am J Sports Med*. 2018;46(13):3217-3226. [Crossref]
12. Gürsoy M, Dirim Mete B, Oyar O, et al. The association of patellar maltracking with infrapatellar fat pad edema and chondromalacia patella: a quantitative morphological magnetic resonance imaging analysis. *Turk J Phys Med Rehab*. 2018;64(3):246-252. [Crossref]
13. Li J, Sheng B, Yu F, et al. Quantitative magnetic resonance imaging in patellar tendon-lateral femoral condyle friction syndrome: relationship with subtle patellofemoral instability. *Skeletal Radiol*. 2019;48(8):1251-1259. [Crossref]
14. Khandelwal R, Kharat A, Botchu R, Koganti D, Shah VP. High resolution T2* mapping in assessment of knee articular cartilage on 3T MRI. *J Clin Orthop Trauma*. 2022;27:101823. [Crossref]
15. Bittersohl B, Hosalkar HS, Sondern M, et al. Spectrum of T2* values in knee joint cartilage at 3 T: a cross-sectional analysis in asymptomatic young adult volunteers. *Skeletal Radiol*. 2014;43(4):443-452. [Crossref]
16. Welsch GH, Trattnig S, Hughes T, et al. T2 and T2* mapping in patients after matrix-associated autologous chondrocyte transplantation: initial results on clinical use with 3.0-Tesla MRI. *Eur Radiol*. 2010;20(6):1515-1523. [Crossref]
17. Newbould RD, Miller SR, Toms LD, et al. T2* measurement of the knee articular cartilage in osteoarthritis at 3T. *J Magn Reson Imaging*. 2012;35(6):1422-1429. [Crossref]
18. Welsch GH, Mamisch TC, Hughes T, et al. In vivo biochemical 7.0 Tesla magnetic resonance: preliminary results of dGEMRIC, zonal T2, and T2* mapping of articular cartilage. *Invest Radiol*. 2008;43(9):619-626. [Crossref]
19. Tsai PH, Wong CC, Chan WP, Lu TW. The value of MR T2* measurements in normal and osteoarthritic knee cartilage: effects of age, sex, and location. *Eur Radiol*. 2019;29(8):4514-4522. [Crossref]
20. Mamisch TC, Hughes T, Mosher TJ, et al. T2 star relaxation times for assessment of articular cartilage at 3 T: a feasibility study. *Skeletal Radiol*. 2012;41(3):287-292. [Crossref]
21. Pai A, Li X, Majumdar S. A comparative study at 3 T of sequence dependence of T2 quantitation in the knee. *Magn Reson Imaging*. 2008;26(9):1215-1220. [Crossref]
22. Liu F, Chaudhary R, Block WF, Samsonov A, Kijowski R. Multicomponent T2 analysis of articular cartilage with synovial fluid partial volume correction. *J Magn Reson Imaging*. 2016;43(5):1140-1147. [Crossref]
23. Giovagnorio F, Olive M, Casinelli A, et al. Comparative US-MRI evaluation of the Insall–Salvati index. *Radiol Med*. 2017;122(10):761-765. [Crossref]
24. Jibri Z, Jamieson P, Rakhra KS, Sampaio ML, Dervin G. Patellar maltracking: an update on the diagnosis and treatment strategies. *Insights Imaging*. 2019;10(1):65. [Crossref]
25. Carrillon Y, Abidi H, Dejour D, Fantino O, Moyen B, Tran-Minh VA. Patellar instability: assessment on MR images by measuring the lateral trochlear inclination-initial experience. *Radiology*. 2000;216(2):582-585. [Crossref]
26. Walker C, Cassar-Pullicino VN, Vaisha R, McCall IW. The patello-femoral joint a critical appraisal of its geometric assessment utilizing conventional axial radiography and computed arthro-tomography. *Br J Radiol*. 1993;66(789):755-761. [Crossref]
27. Jibri Z, Martin D, Mansour R, Kamath S. The association of infrapatellar fat pad oedema with patellar maltracking: a case-control study. *Skeletal Radiol*. 2012;41(8):925-931. [Crossref]
28. Wolfe S, Varacallo M, Thomas JD, Carroll JJ, Kahwaji CI. Patellar instability. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024. [Crossref]
29. Cobianchi Bellisari F, De Marino L, Arrigoni F, et al. T2-mapping MRI evaluation of

patellofemoral cartilage in patients submitted to intra-articular platelet-rich plasma (PRP) injections. *Radiol Med*. 2021;126(8):1085-1094. [Crossref]

30. Subhawong TK, Thakkar RS, Padua A, Flammang A, Chhabra A, Carrino JA. Patellofemoral friction syndrome: magnetic resonance imaging correlation of morphologic and T2 cartilage imaging. *J Comput Assist Tomogr*. 2014;38(2):308-312. [Crossref]
31. Gao S, Peng C, Wang G, Deng C, Zhang Z, Liu X. Cartilage T2 mapping-based radiomics in knee osteoarthritis research: Status, progress and future outlook. *Eur J Radiol*. 2024;181:111826. [Crossref]
32. Bittersohl B, Hosalkar HS, Miese FR, et al. Zonal T2* and T1Gd assessment of knee joint cartilage in various histological grades of cartilage degeneration: an observational in vitro study. *BMJ Open*. 2015;5(2):e006895. [Crossref]
33. Li X, Kim J, Yang M, et al. Cartilage compositional MRI-a narrative review of technical development and clinical applications over the past three decades. *Skeletal Radiol*. 2024;53(9):1761-1781. [Crossref]
34. Mars M, Tbini Z, Gharbi S, Bouaziz MC, Ladep F. T2 Versus T2* MRI mapping in the Knee Articular Cartilage at 1.5 Tesla and 3 Tesla. *Open Med*. 2018;5:119-129. [Crossref]
35. Kim JH, Lee SK, Jung JY. Superolateral Hoffa's fat pad oedema: relationship with cartilage T2* value and patellofemoral maltracking. *Eur J Radiol*. 2019;118:122-129. [Crossref]
36. Elias JJ, Jones KC, Lalonde MK, Gabra JN, Rezvanifar SC, Cosgarea AJ. Allowing one quadrant of patellar lateral translation during medial patellofemoral ligament reconstruction successfully limits maltracking without overconstraining the patella. *Knee Surg Sports Traumatol Arthrosc*. 2018;26(10):2883-2890. [Crossref]
37. Ghany JF, Kamel S, Zoga A, et al. Extensor mechanism tendinopathy in patients with lateral patellar maltracking. *Skeletal Radiol*. 2021;50(11):2205-2212. [Crossref]
38. Fulkerson JP, Buuck DA, Dye SF, Farr JII, Post WR. Disorders of the patellofemoral joint. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004:239-253. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263921

MUSCULOSKELETAL IMAGING

ORIGINAL ARTICLE



# Magnetic resonance imaging-based morphologic features associated with ankle sprain and increased risk of ligament tear

Hande Özen Atalay[1]
Gizem Timoçin Yığman[2]
Erol Erinç Dokuyucu[3]
Evrim Özmen[2]

[1]Hakkari State Hospital, Clinic of Radiology, Hakkari, Türkiye

[2]Koç University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

[3]Koç University Faculty of Medicine, İstanbul, Türkiye

**PURPOSE**

To identify magnetic resonance imaging (MRI)-based morphologic features of the distal tibiofibular syndesmosis and talocrural joint associated with ankle sprains and to determine which parameters are specifically linked to an increased risk of ligament tears after sprains.

**METHODS**

This retrospective study included ankle MRI examinations performed between January 2022 and November 2025. Two analytic datasets were constructed: Dataset 1 compared patients with ankle sprains and MRI-confirmed ligament tears to healthy controls with completely normal ankle MRIs; Dataset 2 compared controls, patients with sprains but no ligament tears, and patients with sprains and ligament tears. Standardized morphometric measurements of the distal tibiofibular syndesmosis and talocrural joint were obtained, including absolute and ratio-based parameters. Group comparisons were performed using appropriate univariable tests with multiple-comparison correction. Multivariable logistic regression was used to identify independent predictors of ligament tears. Inter-reader reliability was assessed using intraclass correlation coefficients and Bland–Altman analysis.

**RESULTS**

In Dataset 1, compared with healthy controls, tibiofibular clear space and the fibular notch depth–to–tibial thickness ratio were significantly higher in patients with ankle sprains and ligament tears, whereas the lateral malleolar height–to–talar articular width ratio was significantly lower (all $P < 0.05$). Multivariable analysis demonstrated that tibiofibular clear space and the fibular notch depth–to–tibial thickness ratio were independent predictors. In Dataset 2, tibiofibular clear space, fibular notch depth–to–tibial thickness ratio, and lateral malleolar height–to–talar articular width ratio differed significantly across three groups (all $P < 0.05$). Notably, only the lateral malleolar height–to–talar articular width ratio independently differentiated patients with sprains and ligament tears from those without tears ($P = 0.025$). Model discrimination was moderate to good (area under the curve: 0.699 and 0.785).

**CONCLUSION**

Specific MRI-based morphologic features are associated with both ankle sprain susceptibility and an increased risk of ligament tears. Among these, the lateral malleolar height–to–talar articular width ratio appears to be a morphometric parameter associated with an increased likelihood of ligament tears.

**CLINICAL SIGNIFICANCE**

MRI-based morphometric assessment, particularly the lateral malleolar height–to–talar articular width ratio, may help identify patients with ankle sprains who are at an increased risk for ligament tears and support more individualized clinical management.

**KEYWORDS**

Ankle joint, ankle sprain, ligament injury, magnetic resonance imaging, trauma

**Handling editor:** Zeynep Maraş Özdemir

**Corresponding author:** Hande Özen Atalay

**E-mail:** handeozen15@gmail.com

Received 05 February 2026; revision requested 04 March 2026; last revision requested 04 April 2026; accepted 13 April 2026.

Epub: 05.05.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263921

You may cite this article as: Özen Atalay H, Timoçin Yığman G, Dokuyucu EE, Özmen E. Magnetic resonance imaging-based morphologic features associated with ankle sprain and increased risk of ligament tear. *Diagn Interv Radiol*. 2026;32(4):500-509.

Lateral ankle sprain is among the most common musculoskeletal injuries in athletes, as well as in general populations, and represents a major source of morbidity worldwide.[1,2] The typical mechanism is forced inversion combined with internal rotation. It frequently occurs during sudden cutting or landing maneuvers, particularly in sports activities.[3,4] Despite being considered a relatively benign trauma, it is associated with high recurrence, chronic symptoms, and instability, eventually leading to long-term consequences such as post-traumatic osteoarthritis.[5]

Anatomical and biomechanical factors play an important role in the mechanism of ankle sprains. The ankle joint has a complex structure and limited osseous stability; therefore, its integrity depends on ligamentous structures.[6] The distal tibiofibular syndesmosis, in particular, contributes substantially to mortise stability; even minor widening of the syndesmosis can change joint mechanics and increase the risk of instability.[7] Additionally, variations in the talocrural joint have been shown to influence the susceptibility to lateral sprains, lateral malleolar fractures, or chronic instability.[8,9] In parallel, recent studies have demonstrated that chronic ankle instability is accompanied by functional changes that may be associated with underlying osseous morphology.[10]

Although several investigations have explored structural risk factors for ankle sprains, most prior studies have focused on either sprain history with ligament injury vs. non-sprain group (not a completely healthy control group) comparisons or on differentiating sprains from fractures.[8,11-16] Importantly, very few have incorporated a well-defined control group with completely normal magnetic resonance imaging (MRI), and previous research has rarely attempted to examine which morphological features increase the risk of sprains vs. those that contribute to the presence of ligament tears.

Thus, the specific MRI-based anatomical characteristics that either promote initial injuries or determine whether sprains evolve into ligament tears remain insufficiently understood. Identifying anatomical features associated with the presence of tears may have important clinical implications, including earlier risk stratification, targeted prevention strategies, and more individualized rehabilitation or treatment planning. To address this gap, this study aims to identify MRI-based morphological parameters of the distal tibiofibular syndesmosis and talocrural joint that are associated with ankle sprains and to determine which anatomical features are specifically linked to ligament tears.

**Main points**

- Distal tibiofibular syndesmosis and talocrural joint morphology show measurable differences on magnetic resonance imaging (MRI) in patients with ankle sprains and ligament tears.
- Tibiofibular clear space and the fibular notch depth–to–tibial thickness ratio is increased in patients with lateral ankle sprains.
- The lateral malleolar height–to–talar articular width ratio is the parameter that independently differentiates patients with sprains and ligament tears from those without tears.
- Detailed MRI-based morphometric assessment may improve anatomical risk stratification in patients with ankle sprains.

## Methods

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Koç University Biomedical Research Ethics Committee (approval number: 2025.503.IRB2.226, date: 10.11.2025). Written informed consent, including permission for the anonymized use of imaging data for research purposes, was obtained from all patients before radiologic evaluation as part of routine clinical practice at our institution.

### Study participants

Patients who had an ankle MRI at Koç University Hospital, a tertiary care institution, between January 2022 and November 2025 were reviewed retrospectively. Individuals were included if the clinical referral or preliminary diagnosis indicated lateral ankle sprains or if their MRI reports contained terms such as "lateral ankle sprain" or "lateral ankle injury." A control group was formed from patients examined during the same period who had no history of ankle trauma in the MRI referral indications and whose MRI reports were interpreted as completely normal. Both pediatric and adult patients were eligible. Patients were excluded if they had a history of lower extremity surgery, fractures, tumors, infection, or marked degenerative joint disease. Additional exclusion criteria included injuries resulting from pronation–external rotation or pronation–abduction mechanisms, suboptimal MRI image quality, and cases with unclear or inconsistent trauma mechanisms. Based on these criteria, two analytic datasets were constructed: the first compared patients with ligament tears and sprains with healthy controls, and the second compared healthy controls, patients with sprains without ligament tears, and patients with sprains and ligament tears (Figure 1).

### Magnetic resonance imaging protocol

MRI examinations were performed using a 3.0-T scanner (MAGNETOM Skyra; Siemens Healthineers, Erlangen, Germany) or a 1.5-T scanner (MAGNETOM Aera; Siemens Healthineers), with a dedicated 16-channel

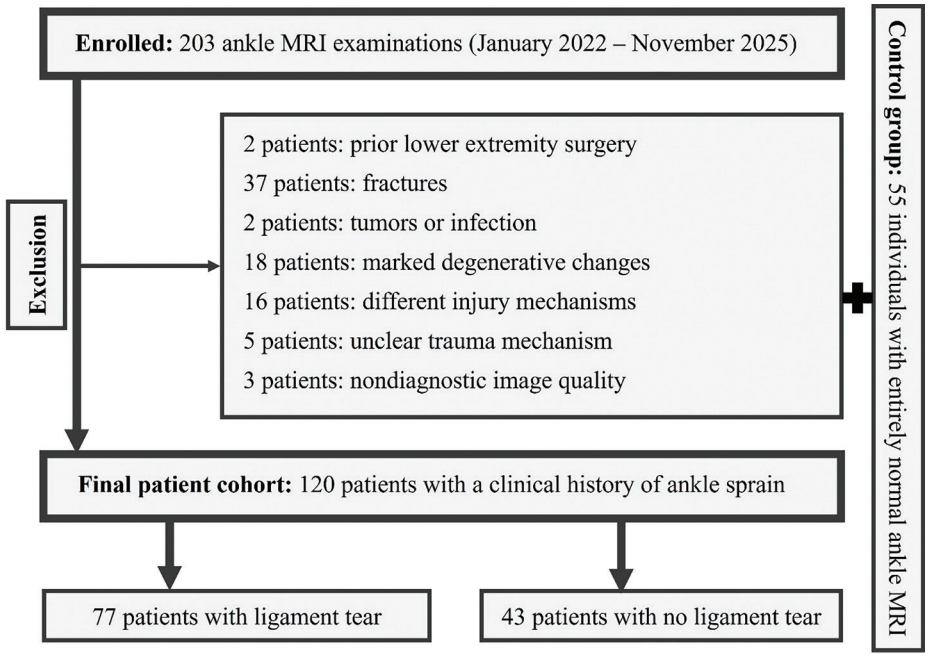



**Figure 1.** Flowchart of the study. MRI, magnetic resonance imaging.

phased-array coil optimized for high-resolution ankle imaging. Patients were positioned supine in a neutral position to avoid plantar flexion or dorsiflexion. The protocol included axial and sagittal T1-weighted spin-echo sequences, as well as axial, coronal, and sagittal fat-suppressed proton density sequences. In all cases, imaging was obtained on the same day or within a few days after the traumatic event and, therefore, generally reflects the acute phase of injuries.

### Image analysis

All MRI examinations were independently reviewed by two radiologists with 10 and 15 years of experience, who were blinded to clinical details and patient groups. Morphometric MRI measurements included the angle between the anterior and the posterior facets, the anterior and posterior facet lengths of the fibular notch, the fibular notch depth, tibial and fibular thickness, medial and lateral malleolar articular surface heights, talar articular surface width, the length of the tibial incisura, the anterior and posterior widths of the tibiofibular syndesmosis, and the tibiofibular clear space (Figures 2 and 3). As described in previous morphometric MRI studies, axial measurements of the distal tibiofibular syndesmosis were obtained from images acquired approximately 10 mm above the tibial subchondral bone and parallel to the tibial plafond, where the anterior tibial tubercle and the fibular incisura are most clearly defined.[17] The line for the length of the tibial incisura was drawn tangentially

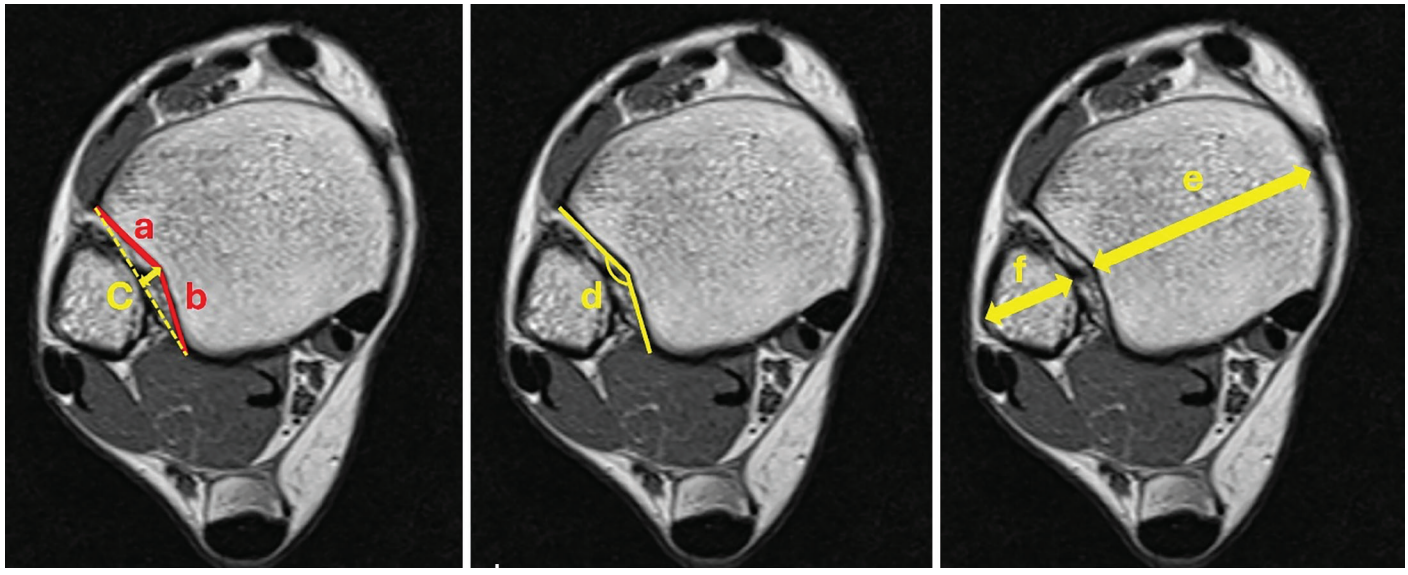



**Figure 2.** Magnetic resonance imaging (MRI)-based morphometric measurements of the distal tibiofibular syndesmosis. Axial MRI images demonstrating morphometric measurements: **(a)** anterior facet length, **(b)** posterior facet length, **(c)** fibular notch depth, **(d)** angle between the anterior and posterior facets, **(e)** tibial thickness, and **(f)** fibular thickness.

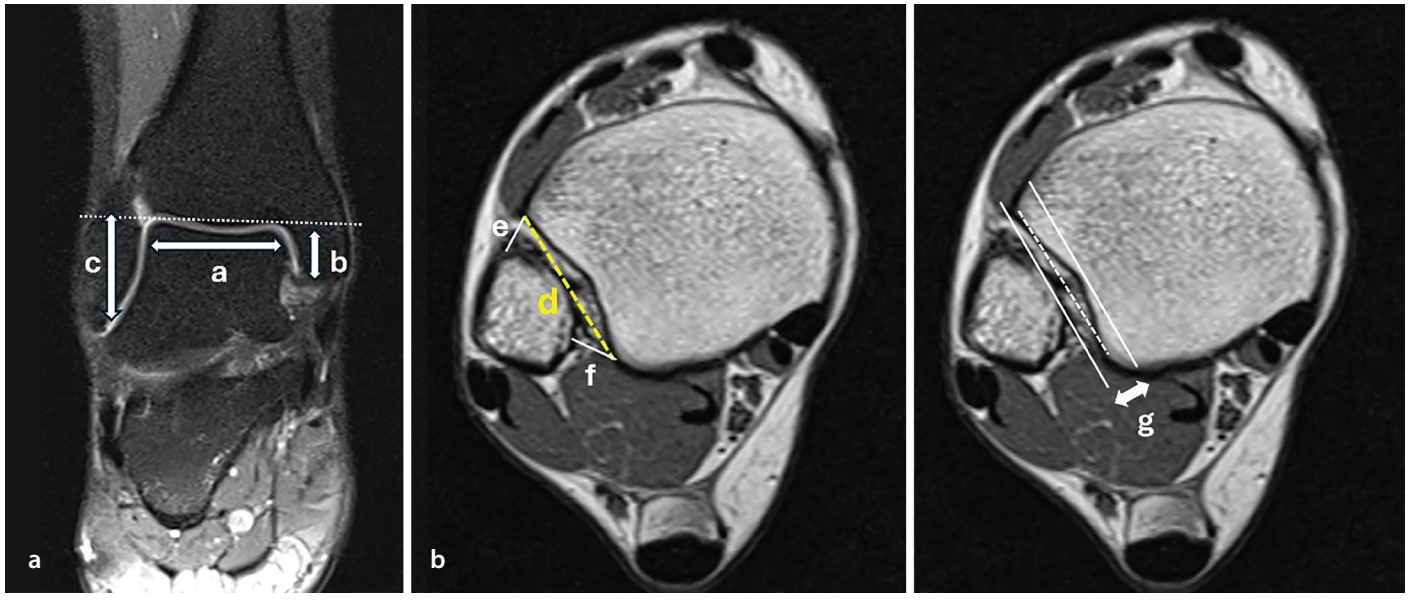



**Figure 3.** Magnetic resonance imaging (MRI)-based morphometric measurements of the ankle. Coronal and axial MRI images demonstrating morphometric measurements: **(a)** talar articular surface width, **(b)** medial malleolar articular surface height, **(c)** lateral malleolar articular surface height, **(d)** length of the tibial incisura, **(e)** anterior width of the tibiofibular syndesmosis, **(f)** posterior width of the tibiofibular syndesmosis, and **(g)** tibiofibular clear space.

between the most prominent portions of the anterior and posterior tibial tubercles. Using this reference, the anterior and posterior widths of the tibiofibular syndesmosis were measured from the intersections of this line with the tibial tubercles to the closest point of the fibula. The tibiofibular clear space was defined as the true distance between the fibula and the deepest point of the tibial incisura. Coronal measurements of the talocrural joint were obtained at the level where the talar articular surface was maximally visualized. A reference line parallel to the superior cortical surface of the talus was first drawn, along which the talar articular surface width was measured. The lateral and medial malleolar articular surface heights were then measured as the perpendicular distances from this reference line to the distal margins of the respective malleoli.

In addition to these absolute measurements, ratio-based parameters were calculated to normalize for inter-individual variations. These included the anterior-to-posterior facet length ratio, the lateral malleolar height–to–talar articular width ratio, the fibular notch depth–to–tibial thickness ratio, the fibular notch depth–to–fibular thickness ratio, the medial malleolar articular surface height–to–talar articular surface width ratio, the lateral malleolar articular surface height–to–medial malleolar articular surface ratio, and the ratio of the anterior width of the tibiofibular syndesmosis to the posterior width. Ligament tears include injuries of the anterior talofibular ligament (ATFL), the posterior talofibular ligament, and the deltoid ligament. Additionally, incidental talar dome osteochondral defects (OCDs) were noted. Ligament injuries were evaluated based on standard MRI criteria: Grade 1 injuries, representing ligament sprains without fiber disruption, were classified as sprains, whereas grade 2 (partial tears) and grade 3 (complete tears) injuries were categorized as ligament tears for the statistical analysis. All measurements were first performed independently by both radiologists, and these independent readings were used to assess inter-reader agreement. Subsequently, the measurements were jointly reviewed, and consensus values were determined for each case, which were then used for the statistical analyses.

### Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics (version 28.0; IBM Corp., Armonk, NY, USA). Normality was assessed using the Shapiro–Wilk test. Group comparisons in Dataset 1 were performed using the Mann–Whitney U test or Student's t-test for continuous variables and chi-square or Fisher's exact tests for categorical variables. Primary and exploratory parameters were evaluated using raw *P* values and false discovery rate (FDR)–adjusted q values to account for multiple testing. A multivariable logistic regression model was used to identify independent predictors of ligament tears, adjusting for age, sex, and laterality, and results were reported as odds ratios with 95% confidence intervals (CIs). Model performance was quantified using the receiver operating characteristic (ROC) the area under curve (AUC). In Dataset 2, differences among controls, patients with sprains only, and patients with sprains and ligament tears were evaluated using the Kruskal–Wallis test, followed by Holm-adjusted pairwise comparisons. A separate binary logistic regression model was applied to identify predictors of ligament tears among patients with sprains. For parameters that remained significant in multivariable analysis, optimal cut-off values were determined using ROC analysis based on the Youden index (sensitivity + specificity − 1). The Youden index was used to identify statistically optimal thresholds; however, clinical application may prioritize sensitivity or specificity depending on the clinical context. For all analyses, a $P$ value < 0.05 was considered statistically significant.

Inter-reader agreement for morphometric measurements was assessed using the intraclass correlation coefficient (ICC) [two-way random model, absolute agreement; ICC (2,1)]. Given the distributional characteristics of ratio-based parameters, CIs were estimated using bootstrap resampling with 2,000 iterations. Bland–Altman analysis was additionally performed to investigate the acceptable interval for the inter-reader differences.

## Results

A total of 203 ankle MRI examinations with a preliminary diagnosis of lateral ankle sprains, or with a report including terms such as "lateral ankle sprain" or "lateral ankle injury," performed between January 2022 and November 2025, were initially reviewed. Of these, 2 patients with prior lower extremity surgery, 37 with fractures, 2 with tumors or infection, 18 with marked degenerative changes, 16 with pronation–external rotation or pronation–abduction injury mechanisms, 5 with unclear trauma mechanisms, and 3 with non-diagnostic trauma were excluded according to the predefined criteria. The final study cohort consisted of 120 patients with a clinical history of ankle sprains. Among these, 77 patients (64.2%) had ligament tears, which were confirmed by MRI, whereas 43 patients (35.8%) had no ligament tears. Additionally, 55 individuals with entirely normal ankle MRIs were included as a control group.

The mean age of the cohort was 37.1 years (median: 38 years). Among all groups, 98 patients (56.0%) were women and 77 (44.0%) were men. The sprained or evaluated side was the left ankle in 70 patients (40.0%) and the right ankle in 105 patients (60.0%). No bilateral examinations were included. The mean age was 38.2 years (median: 41 years) in the group with ligament tears, 37.7 years (median: 38 years) in the group with sprains without tears, and 35.2 years (median: 34 years) in the control group. Sex distribution was 46 women and 31 men in the group with ligament tears, 24 women and 19 men in the group with sprains without tears, and 28 women and 27 men in the control group. Baseline comparisons demonstrated no significant differences between groups in terms of age or sex distribution ($P > 0.05$).

Among the 120 patients with ankle sprains, ATFL tears were the most common injury type and were identified in 73 patients (60.8%). Posterior talofibular ligament tears were noted in 10 patients (8.3%), and deltoid ligament tears were identified in 20 patients (16.7%). All patients with deltoid ligament tears also had concomitant lateral ligament tears, and no isolated deltoid ligament tears were observed. Additionally, four patients had grade 1 OCD, 3 patients had grade 2, and 3 patients had grade 3.

### Patients with lateral ankle sprains and ligament tears vs. healthy controls

Dataset 1 compared patients with ankle sprains and MRI-confirmed ligament tears with healthy individuals who had entirely normal ankle MRIs to identify morphological features associated with ligamentous injuries. In this dataset, patients with ankle sprains and MRI-confirmed ligament tears demonstrated significantly greater tibiofibular clear space than the controls ($P = 0.005$). The fibular notch depth–to–tibial thickness ratio was significantly higher and the lateral malleolar height–to–talar articular width ratio was significantly lower in the group with ligament tears than in the healthy controls ($P = 0.034$ and $P = 0.019$, respectively) (Table 1). Among exploratory variables, several parameters reflecting distal tibiofibular morphology showed FDR-significant differences. The anterior and posterior widths of the ti-

biofibular syndesmosis, as well as the lateral-to-medial malleolar height ratio, were significantly greater in the group with sprains and ligament tears than in the healthy controls. Conversely, the medial malleolar articular surface height and the medial malleolar height–to–talar width ratio were significantly lower in the group with sprains and ligament tears than in the healthy controls (Table 2).

In the multivariable logistic regression model for Dataset 1, tibiofibular clear space remained a strong and statistically significant predictor ($P = 0.004$), with greater values independently associated with lateral ankle sprains and ligament tears. The fibular notch depth–to–tibial thickness ratio also demonstrated an independent association ($P < 0.001$). The lateral malleolar height–to–talar articular width ratio showed a borderline association after covariate adjustment ($P = 0.046$). The overall performance of the final model was good; the AUC was 0.785, indicating good discriminatory ability.

ROC analysis was additionally performed to identify optimal thresholds for the parameters that remained significant in multivariable analysis. For the tibiofibular clear space, a cut-off value of 4.3 mm yielded 66% sensitivity and 67% specificity for discriminating ankles with tears from controls (AUC: 0.644).

**Table 1.** Primary morphometric parameters (Dataset 1): comparison of patients with ligament tear vs. healthy controls

| Parameter | Healthy controls | Ligament tear group | Difference | Test | *P* value |
|---|---|---|---|---|---|
| | Median (IQR) | Median (IQR) | | | |
| Tibiofibular clear space (mm) | 4.1 (3.7–4.6) | 4.6 (4.0–5.3) | +0.50 higher | MWU | **0.005** |
| Fibular notch depth–to–tibial thickness ratio | 0.094 (0.074–0.117) | 0.107 (0.084–0.131) | +0.0116 higher | t-test | **0.034** |
| Lateral malleolar height–to–talar width ratio | 0.866 (0.798–0.969) | 0.833 (0.763–0.917) | −0.0469 lower | t-test | **0.019** |
| Anterior-to-posterior facet length ratio | 0.845 (0.614–1.176) | 0.794 (0.613–1.100) | No significant difference | MWU | 0.773 |
| Angle between anterior and posterior facets (°) | 134.9 (128.8–142.8) | 133.6 (127.2–142.5) | No significant difference | MWU | 0.780 |

IQR, interquartile range; MWU, Mann–Whitney U.

**Table 2.** Exploratory morphometric parameters (Dataset 1): multiple-comparison–adjusted analysis (FDR corrected)

| Parameter | Healthy controls | Ligament tear group | FDR q value | Effect |
|---|---|---|---|---|
| | Median (IQR) | Median (IQR) | | |
| Medial malleolar articular surface height (mm) | 11.2 (9.6–12.9) | 9.3 (7.4–10.9) | **$1.7 \times 10^{-5}$** | Lower |
| Anterior width of tibiofibular syndesmosis (mm) | 2.0 (1.7–2.5) | 2.5 (1.9–3.2) | **0.0148** | Higher |
| Posterior width of tibiofibular syndesmosis (mm) | 4.2 (3.9–4.8) | 5.2 (4.7–5.8) | **$2.0 \times 10^{-6}$** | Higher |
| Medial malleolar height–to–talar width ratio | 0.433 (0.363–0.493) | 0.358 (0.307–0.416) | **$2.0 \times 10^{-6}$** | Lower |
| Lateral-to-medial malleolar height ratio | 1.985 (1.809–2.345) | 2.384 (1.991–3.015) | **$2.74 \times 10^{-4}$** | Higher |
| Anterior facet length of fibular notch (mm) | 10.1 (8.8–11.9) | 10.1 (8.7–12.4) | ns | - |
| Posterior facet length of fibular notch (mm) | 12.0 (10.1–13.5) | 12.5 (10.8–14.3) | ns | - |
| Fibular notch depth (mm) | 3.5 (2.8–4.5) | 3.9 (3.3–4.8) | ns | - |
| Tibial thickness (mm) | 38.0 (35.0–41.0) | 36.8 (35.1–40.1) | ns | - |
| Fibular thickness (mm) | 12.8 (11.9–13.9) | 13.1 (11.9–14.4) | ns | - |
| Lateral malleolar articular surface height (mm) | 22.6 (21.2–23.5) | 21.7 (19.9–23.0) | ns | - |
| Talar articular surface width (mm) | 25.9 (23.2–28.4) | 26.0 (23.9–28.4) | ns | - |
| Length of tibial incisura (mm) | 21.1 (19.6–22.9) | 22.0 (19.8–24.3) | ns | - |

ns, non-significant; FDR, false discovery rate; IQR, interquartile range.

For the fibular notch depth–to–tibial thickness ratio, the optimal threshold was 0.099, yielding 60% sensitivity and 58% specificity, although its discriminatory performance was limited (AUC: 0.609).

### Healthy controls vs. patients with sprains without ligament tears vs. patients with sprains with ligament tears

Dataset 2 was used to evaluate differences between three groups (healthy controls, patients with sprains without ligament tears, and patients with sprains and tears) and to identify parameters linked to both sprain susceptibility and an increased risk of ligament tears after sprains. In this evaluation, significant differences were observed for tibiofibular clear space ($P$ = 0.021), the fibular notch depth–to–tibial thickness ratio ($P$ = 0.044), and the lateral malleolar height–to–talar articular width ratio ($P$ = 0.012) (Table 3). For the tibiofibular clear space, post-hoc comparisons demonstrated a significant difference between controls and the group with ligament tears (Holm-adjusted $P$ = 0.0143), whereas no significant differences were observed between controls and patients with sprains without tears or between those with sprains without tears and those with sprains and tears. Also, for the lateral malleolar height–to–talar articular width ratio, a significant difference was observed between patients with sprains without tears and those with sprains and tears (Holm-adjusted $P$ = 0.0209), whereas comparisons involving the control group were not significant; this is more consistent with an increased risk of tears rather than sprain susceptibility. No significant differences were observed for the remaining primary parameters. Among exploratory variables, several demonstrated robust groupwise differences after FDR correction, many of which exhibited graded changes from controls to patients with sprains without tears and finally to patients with sprains and tears (Table 4).

In the logistic regression model restricted to patients with ankle sprains, the lateral malleolar height–to–talar articular width ratio was the only independent predictor of ligament tears ($P$ = 0.025), whereas other parameters did not reach significance ($P$ > 0.05). The discriminatory performance of the model was moderate, with an AUC of 0.699. ROC analysis further identified an optimal threshold for the lateral malleolar height–to–talar articular width ratio. A cut-off value of ≤ 0.86 was associated with an increased likelihood of ligament tears among patients with sprains, yielding a sensitivity of 62% and a specificity of 65% according to the Youden index–based threshold.

**Table 3.** Primary morphometric parameters (Dataset 2): Comparison among controls, patients with sprains without ligament tears, and patients with sprains with ligament tears

| Parameter | Control | Sprain without tear | Sprain with tear | *P* value | Holm-adjusted pairwise results |
|---|---|---|---|---|---|
| | Median (IQR) | Median (IQR) | Median (IQR) | | |
| Tibiofibular clear space (mm) | 4.1 (3.7–4.6) | 4.3 (3.7–5.2) | 4.6 (4.0–5.3) | **0.0217** | Control vs. tear+: 0.0143<br>Control vs. tear–: ns<br>Tear– vs. tear+: ns |
| Fibular notch depth–to–tibial thickness ratio | 0.094 (0.074–0.117) | 0.110 (0.087–0.136) | 0.107 (0.084–0.131) | 0.0442 | All pairwise ns |
| Lateral malleolar height–to–talar articular width ratio | 0.866 (0.798–0.969) | 0.906 (0.811–0.977) | 0.833 (0.763–0.917) | **0.0126** | Tear– vs. tear+: 0.0209<br>Control vs. tear+: ns<br>Control vs. tear–: ns |
| Anterior-to-posterior facet length ratio | 0.845 (0.614–1.176) | 0.877 (0.691–1.194) | 0.794 (0.613–1.100) | 0.6499 | All pairwise ns |
| Angle between anterior and posterior facets (°) | 134.9 (128.8–142.8) | 134.4 (124.8–143.3) | 133.6 (127.2–142.5) | 0.8043 | All pairwise ns |

ns, non-significant; IQR, interquartile range.

**Table 4.** Exploratory morphometric parameters (Dataset 2): FDR-corrected comparison among the three groups

| Parameter | Control | Sprain without tear | Sprain with tear | FDR q value | Pattern/effect direction |
|---|---|---|---|---|---|
| | Median (IQR) | Median (IQR) | Median (IQR) | | |
| Posterior width of the tibiofibular syndesmosis (mm) | 4.2 (3.9–4.8) | 4.7 (4.3–5.5) | 5.2 (4.7–5.8) | **0.000009** | Monotonic ↑ (Control < tear– < tear+) |
| Medial malleolar height–to–talar width ratio | 0.433 (0.363–0.493) | 0.346 (0.316–0.418) | 0.358 (0.307–0.416) | **0.000012** | Lower in sprain groups |
| Lateral-to-medial malleolar height ratio | 1.985 (1.809–2.345) | 2.417 (2.233–2.880) | 2.384 (1.991–3.015) | **0.000061** | Higher in sprain groups |
| Medial malleolar articular surface height (mm) | 11.2 (9.6–12.9) | 8.9 (7.7–11.4) | 9.3 (7.4–10.9) | **0.000076** | Lower in sprain groups |
| Lateral malleolar articular surface height (mm) | 22.6 (21.2–23.5) | 23.1 (21.2–24.8) | 21.7 (19.9–23.0) | **0.0271** | Higher in sprain groups |
| Anterior width of the tibiofibular syndesmosis (mm) | 2.0 (1.7–2.5) | 2.3 (1.7–2.7) | 2.5 (1.9–3.2) | **0.0348** | Monotonic ↑ (Control < tear– < tear+) |
| Other variables | - | - | - | ns | No pattern |

ns: non-significant; FDR: false discovery rate; IQR, interquartile range.

### Inter-reader reliability

Inter-reader agreement for morphometric measurements ranged from fair to good. The highest agreement was observed for the fibular notch depth–to–tibial thickness ratio, with an ICC of 0.77 (95% CI: 0.70–0.83). Tibiofibular clear space demonstrated moderate agreement (ICC: 0.65, 95% CI: 0.53–0.74), whereas the angle between the anterior and posterior facets showed good agreement (ICC: 0.72, 95% CI: 0.55–0.88). The anterior-to-posterior facet length ratio exhibited moderate agreement (ICC: 0.58, 95% CI: 0.41–0.71). In contrast, the lateral malleolar height–to–talar articular width ratio showed lower inter-reader agreement (ICC: 0.27, 95% CI: 0.18–0.37). Bland–Altman analyses demonstrated good agreement for the anterior-to-posterior facet length ratio and the fibular notch depth–to–tibial thickness ratio, whereas the lateral malleolar height–to–talar articular width ratio showed a small systematic bias with wider inter-reader variability (Figures 4 and 5).

## Discussion

In this study, we systematically evaluated distal tibiofibular and talocrural joint morphology on MRI to determine which anatomical features increase the risk of ankle sprains and which parameters are specifically associated with an increased risk of ligament tears. Unlike previous research that typically examined either sprain-vs.-control differences or post-injury structural variations in isolation, our study incorporated a well-defined three-group design, allowing both sprain-related and tear-related features to be identified within the same analytical framework. Across datasets, tibiofibular clear space and the fibular notch depth–to–tibial thickness ratio emerged as consistently elevated in patients with ligament tears, indicating a broader syndesmotic openness and deepening that characterize ankles with confirmed tears (Figure 6). Most notably, the lateral malleolar height–to–talar width ratio distinguished tear risk, highlighting a lateral talocrural configuration that appears specifically relevant to an increased risk of tears rather than sprain susceptibility alone (Figure 7). In addition, several exploratory parameters related to syndesmotic width and malleolar geometry—including anterior and posterior tibiofibular syndesmotic width, the lateral-to-medial malleolar height ratio, medial malleolar articular surface height, and the medial malleolar height–to–talar width ratio—also demonstrated FDR-significant differences. These findings further support the presence of structural differences in distal tibiofibular configuration among patients with ligament tears.

Several of our findings are supported by prior MRI-based morphometric studies investigating distal tibiofibular and talocrural anatomy in lateral ankle instability. Vieira et al.[11] demonstrated that syndesmotic geometry and malleolar–talar relationships play a critical role in ankle sprain susceptibility, reporting significant alterations in fibular notch–related ratios and malleolar parameters in patients with sprains. Similarly, Ataoğlu et al.[15] showed that variations in anterior and posterior facet lengths, tibial thickness, and fibular notch–based ratios were associated with arthroscopically proven ankle instability. Our findings are in line with these observations, particularly with respect to the fibular notch depth–to–tibial thickness ratio and lateral malleolar height–to–talar width ratios, supporting the concept that reduced osseous containment predisposes the ankle to instability. However, important distinctions differentiate our study from these earlier works. Both Vieira et al.[11] and Ataoğlu et al.[15] primarily focused on comparing patients with unstable or sprain-positive ankles to control groups, without stratifying patients according to ligament integrity. By separately analyzing patients with sprains with and without MRI-confirmed ligament tears, our study extends this framework and demonstrates that although several morphometric parameters are already altered in

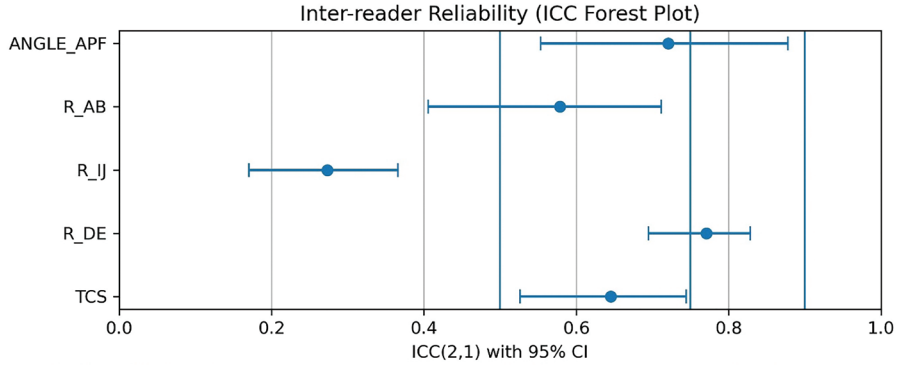



**Figure 4.** Inter-reader reliability analysis. ICC, intraclass correlation coefficient; CI, confidence interval; ANGLE_APF, angle between anterior and posterior facets; R-AB, anterior-to-posterior facet length ratio; R-IJ, lateral malleolar height to talar articular width ratio; R-DE, fibular notch depth to tibial thickness ratio; TCS, tibiofibular clear space.

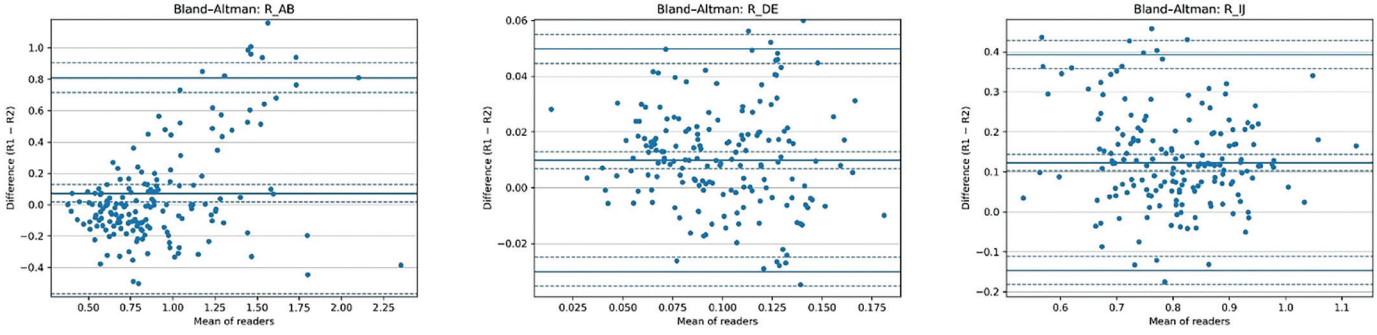



**Figure 5.** Bland–Altman plots for ratio-based measurement. R-AB, anterior-to-posterior facet length ratio; R-DE, fibular notch depth to tibial thickness ratio; R-IJ, lateral malleolar height to talar articular width ratio.

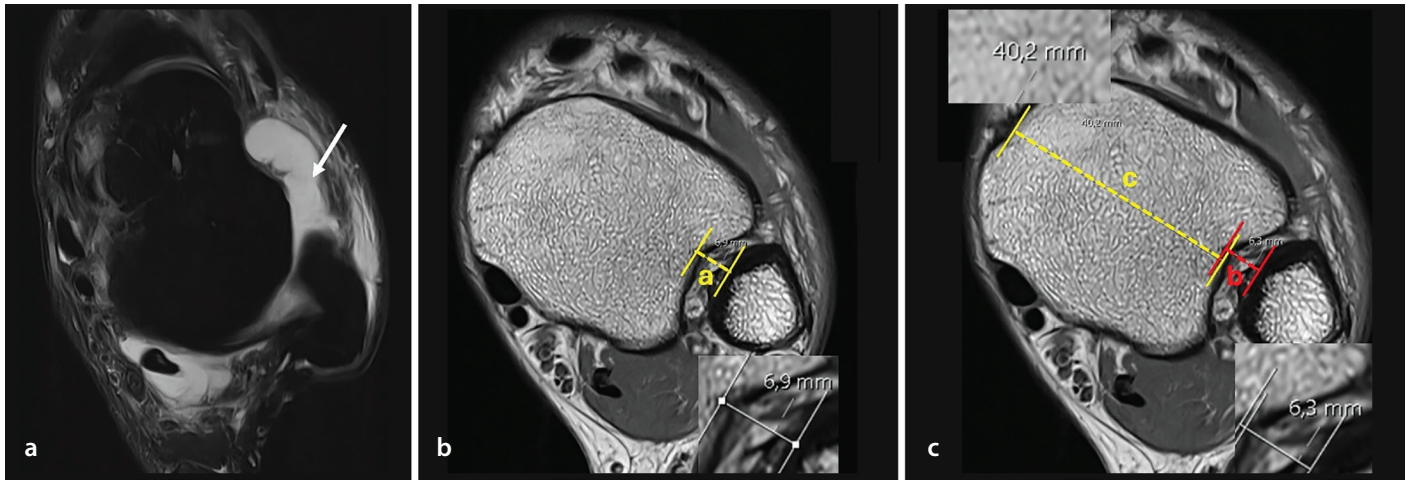



**Figure 6.** Axial ankle magnetic resonance imaging of a 48-year-old patient presenting with an ankle sprain demonstrates a complete tear of the anterior talofibular ligament (arrow in **Figure 6a**). Quantitative morphometric assessment (**Figure 6b** and **6c**) shows an increased tibiofibular clear space (**a**: 6.9 mm), exceeding the identified cut-off value of 4.3 mm. Additionally, the fibular notch depth **(b)**–to–tibial thickness **(c)** ratio is elevated (0.15), surpassing the threshold value of 0.099.

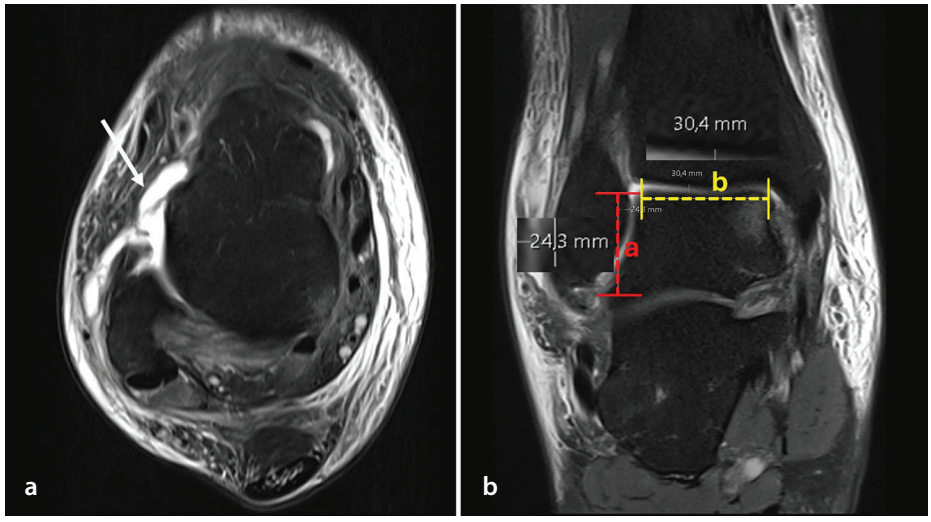



**Figure 7.** Axial and coronal ankle magnetic resonance imaging of a 46-year-old man presenting with an ankle sprain demonstrates a tear of the anterior talofibular ligament (arrow in **Figure 7a**). Morphometric analysis (**Figure 7b**) reveals a decreased lateral malleolar height **(a)**–to–talar articular width **(b)** ratio of 0.79, which falls below the identified cut-off value (≤ 0.86) associated with an increased likelihood of ligament tears.

patients with sprains, only a subset—most notably the lateral malleolar height–to–talar articular width ratio—is independently associated with a ligament tear rather than a sprain alone. Additionally, our inclusion of tibiofibular clear space and anterior–posterior syndesmotic width measurements provided further insight into syndesmotic openness, parameters that were not evaluated in prior studies.

In another study, Lee et al.[8] reported that the medial malleolar length–to–talar width ratio was associated with lateral ankle fractures in the ankle injury group. In our study, however, we found that lateral malleolar–related measurements demonstrated significant groupwise differences and showed changes from controls to injured groups. Unlike Lee et al.,[8] whose analysis primarily focused on the presence of sprains and ligament tears, our study further evaluated progression within injured patients. In this context, we identified the lateral malleolar height–to–talar articular width ratio as a specific discriminator of an increased risk of tears rather than sprain susceptibility alone.

Fibular position and syndesmotic configuration have been increasingly recognized as key contributors to ankle instability. Prior works have demonstrated that altered fibular positioning is associated with lateral ankle instability.[13,16] For example, Berkowitz and Kim[13] reported that a more posteriorly positioned fibula within the ankle mortise was significantly more common in patients undergoing lateral ankle stabilization procedures, suggesting that fibular malposition may be associated with an increased risk of chronic ankle instability. Also, Liu et al.[16] showed that specific shapes of the distal tibiofibular syndesmosis, particularly a shallow incisura, were associated with an increased risk of recurrent lateral ankle sprains. Although these studies primarily relied on qualitative classification or positional assessment, our work extends this concept by quantitatively integrating syndesmotic width, fibular notch depth, and ratio-based morphometric parameters within a unified analytical framework. In our cohort, fibular notch depth–related ratios were significantly increased in patients with ligament tears, supporting the role of syndesmotic openness and reduced fibular containment in ligament failure.

Although most primary morphometric parameters demonstrated good to moderate inter-reader agreement, the lateral malleolar height–to–talar articular width ratio showed comparatively lower ICC values in this study. This finding likely reflects the composite nature of this parameter, which may be influenced by slice selection variability. In contrast, other parameters rely on more sharply defined boundaries, resulting in higher reproducibility. Importantly, despite lower reproducibility, consensus measurements still demonstrated that the lateral malleolar height–to–talar articular width ratio emerges as an independent predictor of ligament tears in multivariable analyses. For improved standardization and to facilitate potential clinical application in future studies, we describe the measurement approach in detail.

On coronal MRIs, a reference line was first drawn parallel to the superior cortical surface of the talus at the level where the talar articular surface was maximally visualized. The talar articular width was measured along this line. On the same slice, the lateral malleolar articular height was measured as the perpendicular distance from this reference line to the distal tip of the lateral malleolus. The inferior boundary of the measurement corresponded to the level where the talofibular ligament originates from the fibula. All measurements were initially performed independently by two radiologists for inter-reader reliability analysis, and the values used for the final statistical analyses were obtained through consensus.

An additional consideration when interpreting our findings is the potential influence of post-traumatic changes on certain morphometric measurements. Parameters reflecting syndesmotic width, especially tibiofibular clear space, may partially represent secondary widening related to ligament disruption rather than purely predisposing anatomical configurations. In contrast, fibular notch depth–related measurements or the lateral malleolar height–to–talar articular width ratio represent geometric relationships between bony structures. Furthermore, the observation that the lateral malleolar height–to–talar articular width ratio differentiates patients with sprains and ligament tears from those without tears supports its potential role as a predisposing anatomical factor rather than solely a consequence of trauma.

This study has several limitations. First, it has a retrospective design and was performed at a single institution. However, the sample size was considered adequate for the regression analyses. Second, MRI examinations were performed on both 1.5-T and 3.0-T scanners. Although the same routine institutional protocol was used, this may still have introduced technical variability. However, as the evaluated parameters primarily represent osseous morphometric measurements, substantial effects of field strength differences on these measurements are unlikely. Third, the healthy control group includes individuals with entirely normal ankle MRIs, but no long-term clinical follow-up was performed. Therefore, we cannot definitively confirm that none of these individuals would later sustain an ankle sprain. Fourth, our study did not include clinical information such as activity levels or dynamic/weight-bearing assessments, which may have provided additional insights. Neurological conditions that could predispose patients to ankle sprains were not systematically evaluated due to the retrospective design. In addition, further analysis according to partial vs. complete tears was not performed because of the limited sample size. Future studies with larger cohorts may help clarify this point. Another limitation is the inclusion of both pediatric and adult patients, which may introduce heterogeneity related to skeletal maturity and epiphyseal closure status. Although age was included as a covariate in the analyses, subgroup analyses according to maturity were not performed. Finally, although the cohort size was relatively large, subgroup analyses may have been underpowered to detect smaller effect sizes.

In conclusion, this study indicates that specific MRI-based morphologic features of the distal tibiofibular syndesmosis and talocrural joint are associated with both susceptibility to ankle sprains and an increased risk of ligament tears. From a clinical perspective, the identified morphometric parameters may represent different stages of structural predisposition and injury. The fibular notch depth–to–tibial thickness ratio appears to reflect an underlying anatomical predisposition related to reduced osseous containment of the distal tibiofibular joint. Tibiofibular clear space may partly represent secondary widening associated with ligament disruption and, therefore, may reflect structural changes in ankles with confirmed tears. Finally, the lateral malleolar height–to–talar articular width ratio, which independently differentiated patients with sprains and ligament tears from those without tears, may represent a transition indicator for increased susceptibility to ligament tears following a sprain. Although these parameters alone cannot determine clinical outcomes, together they provide complementary morphologic information that may help improve structural risk stratification. Reporting markedly abnormal values may help identify patients at higher risk of ligament tears following ankle sprains. However, the discriminatory performance of these morphologic parameters was moderate to good, and the findings should be interpreted cautiously until validated in independent cohorts. Future prospective studies incorporating clinical and functional outcomes are required to validate these findings and to confirm the clinical utility of these morphometric parameters in risk stratification and individualized management.


## Conflict of interest disclosure

Evrim Özmen, MD, serves as Section Editor for Diagnostic and Interventional Radiology. She had no involvement in the peer review of this article and had no access to information regarding its peer review. The other authors declared no conflicts of interest.


## References


1. Gribble PA, Bleakley CM, Caulfield BM, et al. Evidence review for the 2016 International Ankle Consortium consensus statement on the prevalence, impact and long-term consequences of lateral ankle sprains. *Br J Sports Med*. 2016;50(24):1496-1505. [Crossref]
2. Sharma S, Dhillon MS, Kumar P, Rajnish RK. Patterns and trends of foot and ankle injuries in Olympic athletes: a systematic review and meta-analysis. *Indian J Orthop*. 2020;54(3):294-307. [Crossref]
3. Machado M, Amado P, Babulal J. Ankle instability – review and new trends. *J Orthop Trauma Rehabil*. 2021;28. [Crossref]
4. Hertel J. Functional anatomy, pathomechanics, and pathophysiology of lateral ankle instability. *J Athl Train*. 2002;37(4):364-375. [Crossref]
5. Kim H, Son SJ, Seeley MK, Hopkins JT. Altered movement biomechanics in chronic ankle instability, coper, and control groups: energy absorption and distribution implications. *J Athl Train*. 2019;54(6):708-717. Erratum in: *J Athl Train*. 2020;55(1):5. [Crossref]
6. Golanó P, Vega J, de Leeuw PA, et al. Anatomy of the ankle ligaments: a pictorial essay. *Knee Surg Sports Traumatol Arthrosc*. 2016;24(4):944-956. [Crossref]
7. Hermans JJ, Beumer A, de Jong TA, Kleinrensink GJ. Anatomy of the distal tibiofibular syndesmosis in adults: a pictorial essay with a multimodality approach. *J Anat*. 2010;217(6):633-645. [Crossref]
8. Lee KM, Chung CY, Sung KH, et al. Anatomical predisposition of the ankle joint for lateral sprain or lateral malleolar fracture evaluated by radiographic measurements. *Foot Ankle Int*. 2015;36(1):64-69. [Crossref]
9. Magerkurth O, Frigg A, Hintermann B, Dick W, Valderrabano V. Frontal and lateral characteristics of the osseous configuration in chronic ankle instability. *Br J Sports Med*. 2010;44(8):568-572. [Crossref]
10. Verhagen EALM, van der Beek AJ. Optimising ankle sprain prevention: a critical review and practical appraisal of the literature. *Br J Sports Med*. 2010;44(15):1082-1088. [Crossref]
11. Vieira J, Vieira AC, Vieira A. MRI analysis of distal tibiofibular joint and ankle anatomy to assess lateral ankle sprain risk. *Insights Imaging*. 2025;16(1):213. [Crossref]

12. Beynnon BD, Renström PA, Alosa DM, Baumhauer JF, Vacek PM. Ankle ligament injury risk factors: a prospective study of college athletes. *J Orthop Res*. 2001;19(2):213-220. [Crossref]
13. Berkowitz MJ, Kim DH. Fibular position in relation to lateral ankle instability. *Foot Ankle Int*. 2004;25(5):318-321. [Crossref]
14. Mason J, Kniewasser C, Hollander K, Zech A. Intrinsic risk factors for ankle sprain differ between male and female athletes: a systematic review and meta-analysis. *Sports Med Open*. 2022;8(1):139. [Crossref]
15. Ataoğlu MB, Tokgöz MA, Köktürk A, Ergişi Y, Hatipoğlu MY, Kanatlı U. Radiologic evaluation of the effect of distal tibiofibular joint anatomy on arthroscopically proven ankle instability. *Foot Ankle Int*. 2020;41(2):223-228. [Crossref]
16. Liu Q, Lin B, Guo Z, Ding Z, Lian K, Lin D. Shapes of distal tibiofibular syndesmosis are associated with risk of recurrent lateral ankle sprains. *Sci Rep*. 2017;7(1):6244. [Crossref]
17. Lepojärvi S, Niinimäki J, Pakarinen H, Leskelä HV. Rotational dynamics of the normal distal tibiofibular joint with weight-bearing computed tomography. *Foot Ankle Int*. 2016;37(6):627-635. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2025.253651

NEURORADIOLOGY

ORIGINAL ARTICLE



# Prognostic value of nidus sphericity in brain AVMs treated with Gamma Knife Radiosurgery: a preliminary study

Yunus Emre Şentürk[1]
Enes Muhammed Cantürk[1]
Ahmet Peker[1]
Sabahattin Yüzkan[1]
Selçuk Peker[2]

[1]Koç University Hospital, Department of Radiology, İstanbul, Türkiye

[2]Koç University Hospital, Department of Neurosurgery, İstanbul, Türkiye

**PURPOSE**

To evaluate the association of the three-dimensional (3D)-modelled sphericity index of brain arteriovenous malformation (AVM) with nidus obliteration rate and time following Gamma Knife® Radiosurgery (GKRS), and to compare the predictive value of the AVM nidus sphericity index with previously established morphological predictors, such as AVM volume.

**METHODS**

This institutional review board-approved retrospective study included 44 patients with cerebral AVMs who underwent single-session or hypofractionated GKRS between 2020 and 2023. Patients who received multimodal therapy, including prior endovascular embolization or microsurgical resection of the AVM nidus, were excluded. A minimum follow-up of 24 months was required for study inclusion. The primary endpoint was defined as complete angiographic obliteration following the initial GKRS, without any latent intracranial hemorrhage requiring hospitalization or surgical intervention. Pretreatment threshold-based semi-automatic segmentation of the AVM nidus was performed to obtain its volume and surface area, from which the sphericity index (Φ) was calculated.

**RESULTS**

Nineteen (43.2%) AVMs achieved obliteration at a mean of 35.7 ± 11.4 months, whereas 25 (56.8%) had residual nidus at 43.7 ± 13.4 months. Sphericity values were more compact and stable, whereas volume showed wide variability across the groups. The median volume of obliterated AVMs was 1.6 (3.9) $cm^3$, and the median volume for AVMs with residual nidus was 4.9 (13.7) $cm^3$ ($P = 0.002$). Median AVM sphericity was 0.53 (0.26) for obliterated AVMs and 0.32 (0.19) for AVMs with residual nidus ($P = 0.003$). Sphericity demonstrated fair discriminative performance, comparable to AVM volume (Φ cut-off: 0.41; sensitivity 79%, specificity 68%, area under the curve: 0.77). However, optimal cut-off values of 0.30 and 0.66 yielded a sensitivity and specificity of 100% and 96%, respectively. Kaplan–Meier analysis revealed a shorter median obliteration time for high-sphericity AVMs (> 0.41) compared with low-sphericity AVMs (45 vs. 60 months, $P = 0.001$). Among patient-related and morphological parameters, sphericity was associated with earlier AVM obliteration (hazard ratio: 36.29, 95% confidence interval: 2.89–454.37, $P = 0.005$), although it did not remain an independent predictor in multivariate analysis.

**CONCLUSION**

This preliminary study found that higher AVM nidus sphericity was associated with increased obliteration rates and shorter time to obliteration following GKRS. Although not an independent predictor, sphericity exhibited a more stable distribution than volume, suggesting its potential as a complementary 3D biomarker for predicting radiosurgical outcomes of AVMs.

**CLINICAL SIGNIFICANCE**

AVM nidus sphericity may serve as a practical 3D geometric biomarker for predicting long-term outcomes following GKRS.

**KEYWORDS**

Arteriovenous malformation, Gamma Knife Radiosurgery, magnetic resonance imaging, Spetzler–Martin grade, sphericity

**Handling editor:** Gülgün Yılmaz Ovalı

**Corresponding author:** Yunus Emre Şentürk

**E-mail:** ysenturk@kuh.ku.edu.tr

Received 15 September 2025; revision requested 24 October 2025; accepted 27 November 2025.

Epub: 13.01.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253651

You may cite this article as: Şentürk YE, Cantürk EM, Peker A, Yüzkan S, Peker S. Prognostic value of nidus sphericity in brain AVMs treated with Gamma Knife Radiosurgery: a preliminary study. *Diagn Interv Radiol*. 2026;32(4):510-518.

Cerebral arteriovenous malformations (AVM) are rare complex vascular abnormalities with an annual incidence of 1.12 per 100,000 person-years.[1] The annual hemorrhage rate is 2.2% for unruptured AVMs and 3.0% for ruptured AVMs. The risk of hemorrhage is mostly associated with prior hemorrhage, deep location, presence of deep venous drainage, and associated aneurysms.[2] Cerebral AVM management remains under debate, especially for unruptured AVMs. Current strategies include observation, microsurgical resection, endovascular embolization, and stereotactic radiosurgery, or various combinations of these methods. Of these, the radiosurgical approach, particularly Gamma Knife® Radiosurgery (GKRS), is a well-recognized and effective approach for smaller and compact AVMs.[3,4] For larger AVMs [Spetzler–Martin (SM) grades 4 and 5], cerebral AVM treatment often requires a multimodal approach, including GKRS and neoadjuvant endovascular embolization to achieve AVM obliteration.[5] However, some reports have suggested that pre-GKRS neoadjuvant endovascular embolization may hinder accurate AVM delineation in high-grade AVMs during treatment planning and promote collateralization around the nidus, potentially reducing the likelihood of obliteration.[6] Although this remains controversial and neoadjuvant endovascular embolization is the more favored approach, GKRS alone may be considered for high-grade AVMs (SM grades 4 and 5) in some centers, typically requiring staged hypofractionated sessions to reduce nidus volume while preventing latent hemorrhage, which otherwise carries an annual bleeding risk of 2%–4% or more if untreated.[7] Nevertheless, achieving complete obliteration with standalone GKRS remains the primary endpoint for SM grade 4 and 5 AVMs, with reported rates ranging between 33% and 53%.[8]

Morphological determinants of cerebral AVM response to GKRS were comprehensively assessed in the largest pooled meta-analysis, which comprised 12 cohorts.[9] In addition to marginal dose, this study mostly examined the effect of AVM nidus size, volume, venous drainage pattern, anatomical location, and SM grade to GKRS response. The most impactful predictor of obliteration was a smaller AVM volume (< 10 $cm^3$), which consistently emerged as the strongest independent parameter associated with AVM obliteration following GKRS. Similarly, a lower SM grade correlated with higher obliteration rates; however, this effect was largely confounded by nidus size in multivariate modeling. Interestingly, the presence of deep venous drainage was associated with an increased probability of obliteration, although its contribution was far less pronounced than AVM volume. In contrast, deeper location and a history of prior embolization were identified as independent negative predictors, reducing the long-term AVM obliteration rate following GKRS.

**Main points**

- Predicting long-term obliteration of brain arteriovenous malformations (AVMs) following Gamma Knife® Radiosurgery (GKRS) remains challenging.
- AVM nidus sphericity quantifies surface irregularity or elongations as a novel three-dimensional geometric biomarker.
- Higher nidus sphericity is associated with earlier and more frequent AVM obliteration following GKRS.

Besides the classical morphological determinants for GKRS response of AVMs, the current segmentation techniques can allow measurement of three-dimensional (3D) features of AVMs and their potential impact on response status. In this context, a series study reported that a more compact AVM achieves better obliteration, and claimed that, in addition to nidus size, a well-defined AVM margin relative to brain parenchyma is another beneficial feature for achieving better radiosurgery response.[10] In addition to the compactness index—which quantifies the vascular proportion of the AVM nidus relative to the surrounding brain parenchyma—the sphericity index provides a numerical measure of a 3D structure, describing how closely the nidus approximates a perfect sphere and thereby reflecting the degree of AVM nidus dispersion. In contrast to the compactness of AVM, the sphericity of AVM measurement is a ratio of surface area to volume, which is most sensitive to AVM nidus elongation, lobulation, and surface irregularities that were mostly demonstrated in studies with solid tumors.[11,12] Regardless of the maximal diameter of AVMs, lower sphericity represents higher AVM surface area, thereby indicating more contact with surrounding brain parenchyma. This is an underexplored 3D parameter in the context of AVMs, with the first report demonstrating that low-nidal sphericity below 0.44 was strongly correlated with a high-risk of seizure incidence in AVMs.[13]

This preliminary study assesses the association between AVM nidus sphericity and the cerebral AVM obliteration in patients treated exclusively with single-session or hypofractionated GKRS, without adjunctive endovascular embolization or microsurgery. Furthermore, the study evaluates the predictive performance of sphericity and compares it with established parameters, such as AVM volume, in achieving obliteration following GKRS.

## Methods

### Study participants and neuroimaging follow-up

This single-center retrospective study was approved by the Clinical Ethical Committee of the Koç University Ethical Committee on Human Research (protocol number: 2025.386.IRB2.174, approval date: 11.09.2025). Given the retrospective design, informed consent was waived.

The inclusion criterion for this study comprised patients with AVMs who underwent GKRS at Koç University Hospital between January 2020 and December 2023. The following criteria determined exclusion: patients 1) aged < 3 years; 2) with active imaging surveillance of < 24 months following GKRS;[14] and 3) who had undergone prior neoadjuvant endovascular embolization or open surgical resection with residual cerebral AVM nidus. During the study period, a total of 168 patients with AVMs were treated with GKRS. Of these, 27 had undergone prior embolization, and 11 had prior surgical resection with residual nidus. Seventy-six cases did not meet the minimum 24-month follow-up requirement. After applying these exclusion criteria, 44 AVMs were eligible for final analysis. The patient selection flowchart is presented in Figure 1. The AVMs were graded according to the SM classification system.

All patients underwent annual imaging surveillance with time-of-flight (ToF) magnetic resonance angiography (MRA) and contrast-enhanced 3D T1-VIBE magnetic resonance imaging (CE-MRI). After 24 months, if no residual nidus was detected on MRA with CE-MRI, digital subtraction angiography (DSA) was performed to confirm AVM obliteration, defined as the absence of a visible nidus and early cerebral venous filling. If a residual nidus was detected beyond 24 months on MRA with CE-MRI, repeat GKRS was performed, and follow-up imaging was continued with MRA every 6 months until no residual nidus was observed. Final confirmation of AVM obliteration was then obtained via DSA. The primary endpoint was defined as complete AVM obliteration following GKRS without major delayed hemorrhage necessitating hospitalization or surgical in-

tervention. The need for a second single or hypofractionated GKRS during follow-up did not preclude meeting the primary endpoint, as, according to the series conducted by Chung et al.,[15] angiographic obliteration is often expected in an average of 66 months even after repeated GKRS sessions.

### Gamma Knife Radiosurgery procedure and dose planning

On the day of radiosurgery, stereotactic MRI and computed tomography (CT) were obtained without a frameless workflow. The stereotactic MRI protocol included isotropic 3D T2-SPACE (slice thickness: 1.5 mm; TR/TE: 2,500/321 ms; field of view: 230 × 230 mm; flip angle: 150°) and 3D post-contrast T1-VIBE (slice thickness: 1.5 mm; TR/TE: 11/4.7 ms; field of view: 230 × 230 mm; flip angle: 23°), both acquired on the same 1.5 T scanner. A 3D ToF-MRA sequence was also obtained (slice thickness: 1.5 mm; TR/TE: 25/7 ms; field of view: 230 × 230 mm; flip angle: 25°). In addition, a thin-slice stereotactic CT (slice thickness: 1.0 mm) was acquired with the head oriented parallel to the scanner table. The 3D post-contrast T1-VIBE, 3D T2-SPACE, and ToF-MRA datasets were rigidly fused in the GammaPlan planning system (Elekta AB, Stockholm, Sweden) for AVM nidus delineation. Stereotactic CT was used for distortion correction. When the AVM nidus was poorly visualized on both post-contrast T1-VIBE and ToF-MRA sequences, or when a dural/pial arteriovenous fistula or incomplete visualization of feeding arteries was suspected, additional two-dimensional (2D) biplanar DSA (posteroanterior and lateral projections) was performed to improve nidus delineation. In the present series, 2D biplanar DSA runs were co-registered in 11 (25%) cerebral AVM cases as an additional fourth element to the fused MRI stacks within the GammaPlan vascular module to improve AVM nidus delineation.

All patients were treated using the Leksell Gamma Knife® Icon™ (Elekta AB, Stockholm, Sweden) system. Mild sedation was administered, and a stereotactic frame was applied to achieve rigid head fixation. Treatment planning was performed individually using the GammaPlan module planning system with semiautomatic segmentation of the AVM nidus. Here, GKRS plans were generated to achieve the highest possible conformity to the segmented nidus while excluding feeding arteries. Isodose shaping was accomplished using multiple small isocenters to minimize dose to the perinidal brain parenchyma. The coverage index, defined as the percentage of nidus volume encompassed by the prescription (minimum) dose, was maintained at ≥ 95% for all targeted AVM volumes. The selectivity index was maintained above 0.75 for all AVM cases. The conformity index, representing the ratio between the prescription isodose volume and the AVM nidus volume, typically ranges between 1.0 and 1.5. A steep dose fall-off strategy (high-gradient index) was systematically preferred to avoid irradiation of adjacent eloquent parenchyma. Hence, the gradient index was maintained under 3 in all AVMs, and the marginal prescription isodose was set to the 50% isodose line. All single-fraction GKRS procedures were performed using the Leksell G frame.

Patients with an AVM < 3 cm and located in non-eloquent brain regions were typically treated with single-session GKRS, with marginal doses ranging from 15–25 Gy depending on nidus volume and proximity to critical structures. For larger AVMs (> 3 cm), or for cases with high-risk features (e.g., intranidal aneurysm, tortuous high-flow deep venous drainage, history of prior hemorrhage), or AVMs located in eloquent/critical regions (brainstem, optic tracts, corona radiata, perirolandic cortex, etc.), a hypofractionated GKRS scheme was used. This consisted of five consecutive daily fractions, each delivering 3–6 Gy, to reduce the risk of radiation-induced toxicity in adjacent white matter tracts and to minimize the chance of post-treatment hemorrhage. In the 5-day hypofractionated GKRS protocol, the frameless thermoplastic mask system was used to ensure reproducible positioning throughout all treatment fractions. Before each session, cone-beam CT verification was performed to confirm accurate head positioning and alignment with the treatment plan. An infrared laser marker was used to monitor any abnormal immobilization of the mask during the GKRS treatment.

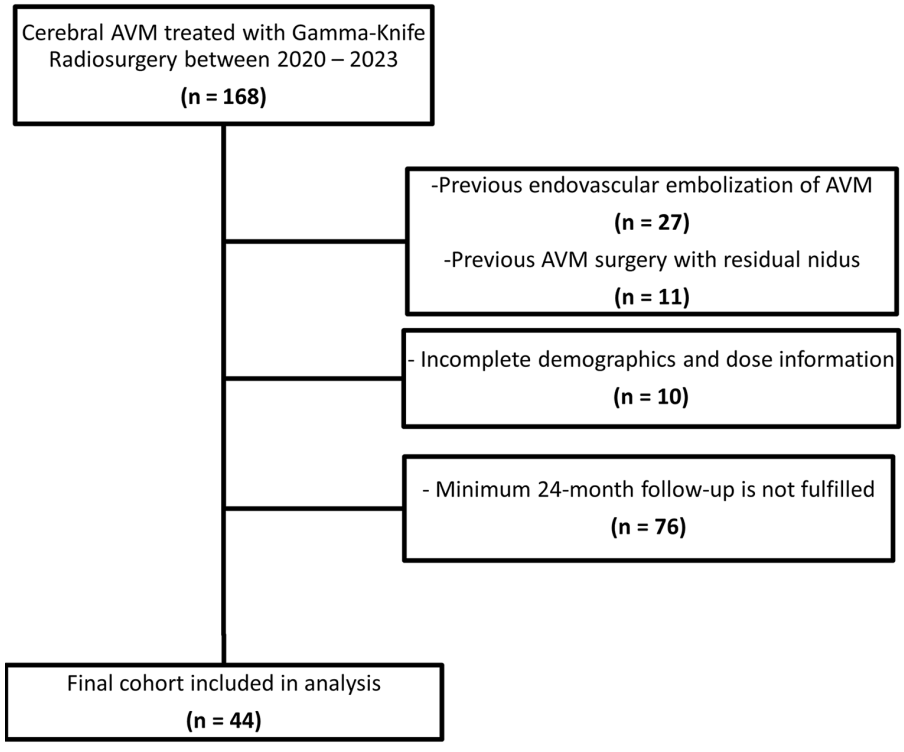



**Figure 1.** Summary flow chart of the study group selection process. AVM, arteriovenous malformation.

### Segmentation and image processing

All anonymized MRI data were analyzed by a neuroradiologist with 6 years of experience in the field of neuroimaging. The radiologist was blinded to all clinical information of the patients during the analysis. Three-dimensional volumetric T1-weighted gradient-echo images (3D T1-VIBE; voxel size: 1.0 × 1.0 × 1.0 mm) were imported into 3D Slicer (version 5.8.1; https://www.slicer.org) for image processing and sphericity analysis. Although both CE-3D T1-VIBE and ToF-MRA were performed, only the CE-T1-VIBE sequence was used as the reference dataset for AVM segmentation; ToF-MRA was not registered with the CE-T1-VIBE images, since flow-related signal attenuation and saturation effects in the former frequently lead to underestimation of the nidus size, particularly in regions of slow or turbulent flow. Flow-related signal attenuation in the AVM

nidus was more evident in the larger aneurysm (nidus size > 3 cm) in ToF-MRA, which potentially biased the segmentation and led to a more artificially contracted nidus relative to an isotropic CE-T1-VIBE series.

The AVM nidus was initially delineated semiautomatically using a threshold-based segmentation approach on CE 3D T1-VIBE images. This process captured the enhancing nidus together with adjacent vascular structures, including feeding arteries, draining veins, and portions of normal vasculature. To refine the segmentation, the scissors tool in 3D Slicer was used to manually remove non-nidal vessels. An approximate 3-mm margin was preserved at the interface with feeding arteries to avoid inadvertent over-cutting of the nidus. The resulting volume was then rendered as a 3D surface. The segmentation method is illustrated in Figure 2. In contrast to prior radiomics studies of solid brain tumors, no Gaussian voxel-smoothing algorithm was applied since surface smoothing substantially inflates surface area measurements and artificially increases sphericity. Following completion of manual segmentation, the segment statistics module in 3D Slicer was used to compute the voxel-based volume and the surface mesh geometry of the AVM nidus. Given isotropic voxel dimensions of 1.0 $mm^3$, volume was calculated directly from voxel counts. Sphericity (Φ) was derived from volume (V) and surface area (A) using the following formula:

$$\Phi = \frac{\pi^{\frac{1}{3}}(6V)^{2/3}}{A}$$

### Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics, version 28.0 (IBM Corp., Armonk, NY, USA). The analyses were primarily conducted by one of the authors (Y.E.S.), and all results were independently reviewed and verified by the co-authors, each of whom has substantial experience in medical statistics and data interpretation. Descriptive statistics were calculated for continuous variables, whereas categorical variables were summarized as frequencies and percentages. Normality of distribution for continuous variables was assessed using the Shapiro–Wilk test. The primary endpoint was defined as long-term GKRS outcome (> 24 months), dichotomized as complete obliteration vs. residual nidus. Group comparisons were performed according to the status of the primary endpoint. Normally distributed continuous variables were compared using the independent-samples t-test, whereas non-normally distributed variables were analyzed using the Mann–Whitney U test. Categorical variables were compared using Pearson's chi-square test. To assess segmentation reproducibility, one co-rater (A.P.) independently segmented the AVM nidus across the entire cohort. Inter-rater reliability for AVM volume and sphericity was evaluated using the intraclass correlation coefficient (ICC) derived from a two-way mixed-effects model. The diagnostic performance of AVM nidus sphericity and volume was evaluated using receiver operating characteristic (ROC) analysis, and optimal cut-off values were determined using the maximum Youden index in addition to maximum sensitivity and specificity strategies. Based on this threshold, AVMs were dichotomized into high-sphericity vs. low-sphericity groups. Kaplan–Meier survival analysis was then conducted to evaluate the association between sphericity status and time to obliteration following GKRS. Patients with residual AVM nidus who remained under imaging surveillance were censored at their last follow-up. Finally, univariate and multivariate Cox proportional hazards regression models were constructed to assess the relationship between time to obliteration and AVM morphologic parameters, age, and sex. A $P$ value < 0.05 was considered statistically significant in all analyses.

## Results

### Patient demographics and Gamma Knife Radiosurgery outcomes

A total of 44 patients (22 women/22 men) were deemed eligible for the study. The average follow-up time after the GKRS was 44.2 ± 15.7 months. Nineteen (43.2%) patients with AVM achieved the primary endpoint by displaying obliteration of the AVM nidus with a mean obliteration time of 35.7 ± 11.4 months. Out of 19 obliterated AVMs, 4 (13.8%) required a second GKRS with a mean marginal dose of 15.8 ± 1.5 Gy. Twenty-five (56.8%) patients with AVM did not achieve the primary endpoint and showed residual nidus despite marked volume reduction, with a mean last follow-up time after GKRS of 43.7 ± 13.4 months. Each of these cases required a second GKRS retreatment with a mean marginal dose of 20.2 ± 6.3 Gy. The patient demographics, GKRS parameters, and basic AVM features are listed in Table 1.

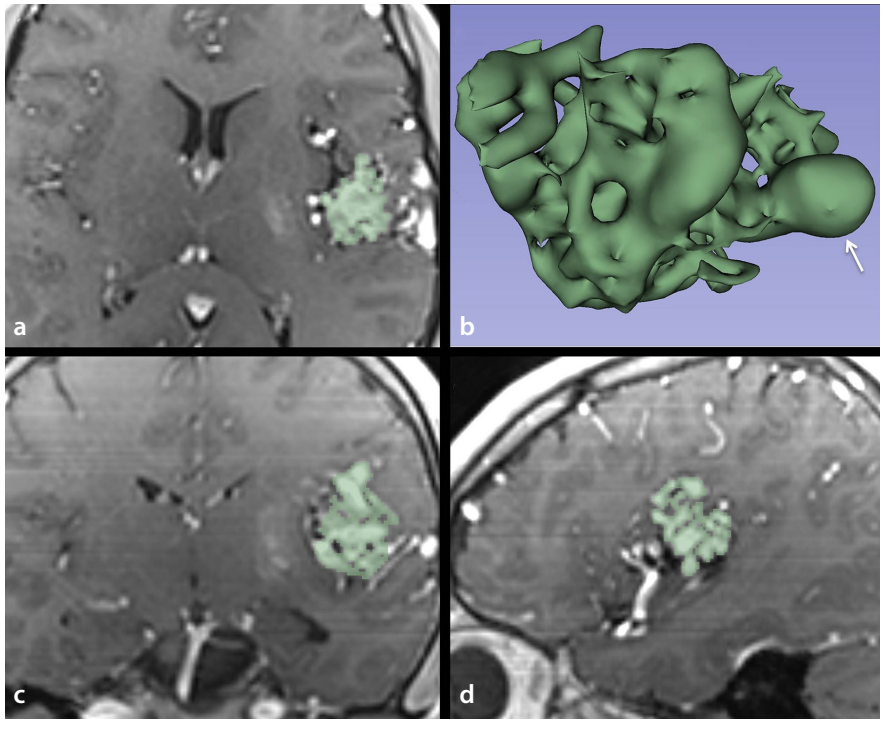



**Figure 2.** Illustration of brain arteriovenous malformation (AVM) segmentation. **(a, c, d)** Segmentation was performed on isotropic T1-weighted gradient-echo images (voxel size: 1.0 mm) for sphericity analysis. The AVM nidus was initially delineated semi-automatically using a threshold-based segmentation approach to distinguish the nidus from the surrounding brain parenchyma. Subsequently, the scissor tool was used to manually remove non-nidal vessels, such as feeding arteries beyond a 3-mm safety margin from the central nidus and large draining veins exiting the nidus. **(b)** Volume rendered three-dimensional view of segmented left insular AVM nidus. The Gaussian smoothing algorithm was exclusively avoided to prevent misrepresentation of the surface area. The measured AVM nidus surface area (A) was 21.16 $cm^2$, the nidus volume (V) was 2.68 $cm^3$, and the calculated sphericity index (Φ) was 0.44 for this patient. The arrow shows an intranidal aneurysm.

Among the 19 patients who achieved AVM obliteration, none experienced major intracranial hemorrhage requiring surgical intervention or hospitalization. One patient developed a minor localized hemorrhage at the site of the treated nidus following a single session of GKRS. In contrast, among the 25 patients with residual AVM nidus, 2 presented with self-limited subarachnoid hemorrhage within 1 year after GKRS, both requiring hospitalization without surgical or endovascular intervention.

### Sphericity and volume analysis of arteriovenous malformation in response to Gamma Knife Radiosurgery

Table 2 summarizes the distribution of patients, treatment characteristics, and AVM morphology according to GKRS response. Hypofractionated GKRS was 8.8-times more frequent in AVMs with residual nidus compared with those achieving obliteration ($P$ = 0.005). There was a trend toward significance in the marginal dose, with residual AVMs receiving a median of 25 (12) Gy vs. 20 (6) Gy in obliterated AVMs ($P$ = 0.060). Retreatment was required in four (13.8%) obliterated AVMs and in each of the residual AVMs, corresponding to a 7.3-fold-higher likelihood of retreatment in the residual group ($P$ = 0.001). The median AVM volume was 4.9 $cm^3$ in the residual group and 1.6 $cm^3$ in the obliterated group ($P$ = 0.002). Likewise, the median AVM sphericity was 0.53 in obliterated AVMs and 0.32 in those with residual nidus ($P$ = 0.001). Among patients achieving the primary endpoint, AVM size of 3–6 cm was the most frequent category (11 cases, 57.9%), whereas in the residual group, AVM size > 6 cm predominated (13 cases, 52%) ($P$ = 0.046). In line with this, the distribution of SM grade differed between groups, with SM grade 3 most common in obliterated AVMs and SM grade 4 most prevalent in residual AVMs ($P$ = 0.021).

**Table 1.** Patient demographics, AVM features, and GKRS outcomes

| Characteristic | Total (n = 44) |
|---|---|
| **Primary endpoint, n (%)** | |
| Obliteration | 19 (43.2%) |
| Residual nidus | 25 (56.8%) |
| **Age,** (mean ± SD) | 25.3 ± 13.4 |
| **Sex,** n (%) | |
| Male | 22 (50%) |
| Female | 22 (50%) |
| **Follow-up,** mo (mean ± SD) | 44.2 ± 15.7 |
| **Time to obliteration,** mo (mean ± SD) | 35.7 ± 11.4 |
| **Marginal dose,** (50 % isodose line), Gy, (mean ± SD) | 22.8 ± 7.1 |
| **Retreatment rate,** n (%) | 29 (65.9%) |
| **Latent bleeding following GKRS,** n (%) | 3 (6.8%) |
| **Location,** n (%) | |
| Lobar | 32 (72.8%) |
| Cerebellum | 4 (9.1%) |
| Thalamus | 3 (6.8%) |
| Corpus callosum | 2 (4.5%) |
| Basal ganglia | 1 (2.3%) |
| Brain stem | 2 (4.5%) |
| **Nidus size, cm** | |
| < 3 cm | 20 (45.5%) |
| 3–6 cm | 17 (38.6%) |
| > 6 cm | 7 (15.9%) |
| **Intranidal aneurysm**, n (%) | 11 (25%) |
| **Deep venous drainage,** n (%) | 27 (61.4%) |
| **Spetzler–Martin grade,** n (%) | |
| 2 | 8 (18.2%) |
| 3 | 18 (40.9%) |
| 4 | 14 (31.8%) |
| 5 | 4 (9.1%) |
| **Clinical presentation,** n (%) | |
| Asymptomatic | 20 (45.5%) |
| Seizure | 15 (34.1%) |
| Bleeding | 4 (9.1%) |
| Sensorimotor symptoms | 4 (9.1%) |
| Visual symptoms | 1 (2.2%) |

AVM, arteriovenous malformation; GKRS, Gamma Knife Radiosurgery; SD, standard deviation

### Inter-rater reliability of segmented arteriovenous malformation volume and sphericity

The ICC was 0.94 [95% confidence interval (CI): 0.55–0.99] for AVM volume and 0.97 (95% CI: 0.86–0.99) for sphericity, indicating excellent inter-rater reliability in both parameters. However, the lower bound of the 95% CI for AVM volume was less than the ICC threshold of 0.70, commonly regarded as the cut-off for good agreement, suggesting that AVM volume measurements may be more susceptible to segmentation-related variability.

### Receiver operating characteristic curve analysis of arteriovenous malformation volume and sphericity for predicting long-term radiosurgical response

The ROC analysis results for AVM sphericity and volume are presented in Table 3. Here, AVM nidus sphericity demonstrated fairly good discriminative performance and a moderate positive correlation with long-term obliteration following GKRS (Spearman's ρ: 0.45, $P$ = 0.002). This finding indicates that as nidus sphericity increases, the likelihood of AVM obliteration following GKRS rises. The optimal sphericity cut-off of 0.41, derived from the best Youden index (0.469), yielded a sensitivity of 79%, specificity of 68%, and an area under the curve (AUC) of 0.77. The highest sensitivity (100%) was achieved at a cut-off of 0.30, whereas the highest specificity (96%) was achieved at a cut-off of 0.66. In contrast, AVM volume showed a moderate negative correlation with GKRS response (Spearman's ρ: –0.42, $P$ = 0.004). The distribution of AVM volume was highly variable, with extreme outliers particularly evident in large nidus sizes, ranging from 0.06 to 60.7

$cm^3$. This wide variability negatively impacted the robustness of volume-based cut-offs for predicting long-term GKRS outcomes. For the AVM nidus volume, the best cut-off value was 1.61 $cm^3$, which achieved a sensitivity of 53%, specificity of 92%, and AUC of 0.78, with the highest Youden index of 0.450. Maximum sensitivity (100%) was achieved at a 0.14-$cm^3$ cut-off, whereas maximum specificity (100%) was observed at a 15.6-$cm^3$ cut-off. Figure 3 illustrates a more stable and compact distribution of AVM nidus sphericity values compared with AVM volume, which demonstrated pronounced variability across the group.

### Effect of arteriovenous malformation morphologic features on obliteration time following Gamma Knife Radiosurgery

When patients were stratified by AVM sphericity according to the optimal cut-off (≤ 0.41: low sphericity, > 0.41: high sphericity), the median time to obliteration following GKRS was 45 months in the high-sphericity group and 60 months in the low-sphericity group ($P$ = 0.001), indicating a faster obliteration rate in AVMs with higher sphericity. Twenty-five patients with residual AVM nidus who did not reach the primary endpoint were censored in the survival analysis, as illustrated in Figure 4.

**Table 2.** Comparison of demographic data, AVM morphological features, and GKRS parameters based on treatment outcomes

| Variable | Obliterated AVM nidus (n = 19) | Residual AVM nidus (n = 25) | *P* value |
|---|---|---|---|
| **Age,** (mean ± SD) | 27.4 ± 15.8 | 23.7 ± 11.4 | 0.371 |
| **Sex** (F/M), | 10/9 | 12/13 | 0.763 |
| **Fractionation scheme (single/hypofractionated),** n | 15/4 | 11/14 | **0.005** |
| **Marginal dose,** median [IQR] | 20 [6] | 25 [12] | 0.060 |
| **Multiple-isocenter GKRS plan,** n (%) | 16 (84.2%) | 24 (96%) | 0.410 |
| **Number of isocenters per nidus,** median [IQR] | 9 [12] | 13 [8.0] | 0.204 |
| **Retreatment rate,** n (%) | 4 (13.8%) | 25 (100.0%) | **0.001** |
| **Retreatment dose,** median [IQR] | 15 [9] | 16 [10] | 0.090 |
| **AVM volume $cm^3$,** median [IQR] | 1.6 [3.9] | 4.9 [13.7] | **0.002** |
| **AVM surface area, $cm^2$** median [IQR] | 13.4 [26.3] | 44.2 [84.9] | **0.001** |
| **AVM sphericity, Φ,** median [IQR] | 0.53 [0.26] | 0.32 [0.19] | **0.003** |
| **Nidus size** | | | **0.046** |
| < 3 cm | 1 (5.3%) | 6 (24%) | |
| 3–6 cm | 11 (57.9%) | 6 (24%) | |
| > 6 cm | 7 (36.8%) | 13 (52%) | |
| **Intra-nidal aneurysm,** n (%) | 5 (26.3%) | 6 (24%) | 0.861 |
| **Deep venous drainage,** n (%) | 10 (52.6%) | 7 (28.0%) | 0.103 |
| **Spetzler–Martin grade,** n (%) | | | **0.021** |
| 2 | 6 (31.6%) | 2 (8%) | |
| 3 | 10 (52.6%) | 8 (32%) | |
| 4 | 2 (10.5%) | 12 (48%) | |
| 5 | 1 (5.3%) | 3 (12%) | |

AVM, arteriovenous malformation; GKRS, Gamma Knife Radiosurgery; IQR, interquartile range; SD, standard deviation; F, female; M, male

The impact of demographic and morphological factors on long-term AVM obliteration was first evaluated using a univariate Cox proportional hazards model. Variables that reached statistical significance were subsequently entered into a multivariate model to explore interdependence. The results are summarized in Table 4. Among the dose- and treatment-related variables, both the hypofractionation scheme and the use of multiple-isocenter treatment plans demonstrated only a trend toward a reduced hazard of earlier AVM obliteration without reaching statistical significance ($P$ = 0.060 for both). Among the morphological predictors, increased AVM sphericity had the highest positive hazard with faster obliteration [hazard ratio (HR): 36.29, 95% CI: 2.89–454.37, $P$ = 0.005], in contrast, larger AVM volume (HR: 0.88, 95% CI: 0.78–0.98, $P$ = 0.026) and higher SM grade (HR: 0.39, 95% CI: 0.21–0.71, $P$ = 0.002) were inversely associated with the early AVM obliteration, although their absolute hazard were notably smaller than that of sphericity. When these three morphological variables were analyzed together in the multivariate model, none retained independent predictive significance, suggesting that the predictive effect of sphericity is attenuated when adjusted for AVM volume and SM grade. Further variance inflation factor analysis yielded values of 2.82 for sphericity, 1.72 for volume, and 2.37 for SM grade, all below the conventional threshold for severe multicollinearity. Nevertheless, the limited sample size likely reduced the stability and independent predictive power of these parameters in the multivariate setting.

## Discussion

The present preliminary study highlights the prognostic value of the surface area-based sphericity index in predicting cerebral AVM obliteration following GKRS. In this preliminary cohort, higher sphericity was associated with both improved long-term AVM nidus obliteration outcomes and earlier achievement of obliteration following GKRS treatment. Unlike solid brain tumors

**Table 3.** Performance of AVM sphericity and volume in achieving the obliteration of AVM after the GKRS

| Variable | Cut-off | Youden index | AUC | Accuracy | Sensitivity,% | Specificity, % |
|---|---|---|---|---|---|---|
| **Sphericity** | 0.30 | 0.440 | 0.77 | 0.64 | 100 | 44 |
| **Sphericity** | **0.41** | **0.469*** | **0.77** | **0.73** | **79** | **68** |
| **Sphericity** | 0.66 | 0.223 | 0.77 | 0.64 | 26 | 96 |
| **Volume, $cm^3$** | 0.14 | 0.118 | 0.78 | 0.43 | 100 | 16 |
| **Volume, $cm^3$** | **1.61** | **0.450*** | **0.78** | **0.75** | **53** | **92** |
| **Volume, $cm^3$** | 15.6 | 0.280 | 0.78 | 0.57 | 28 | 100 |

*Highest Youden index value. AVM, arteriovenous malformation; GKRS, Gamma Knife Radiosurgery; AUC, area under the curve

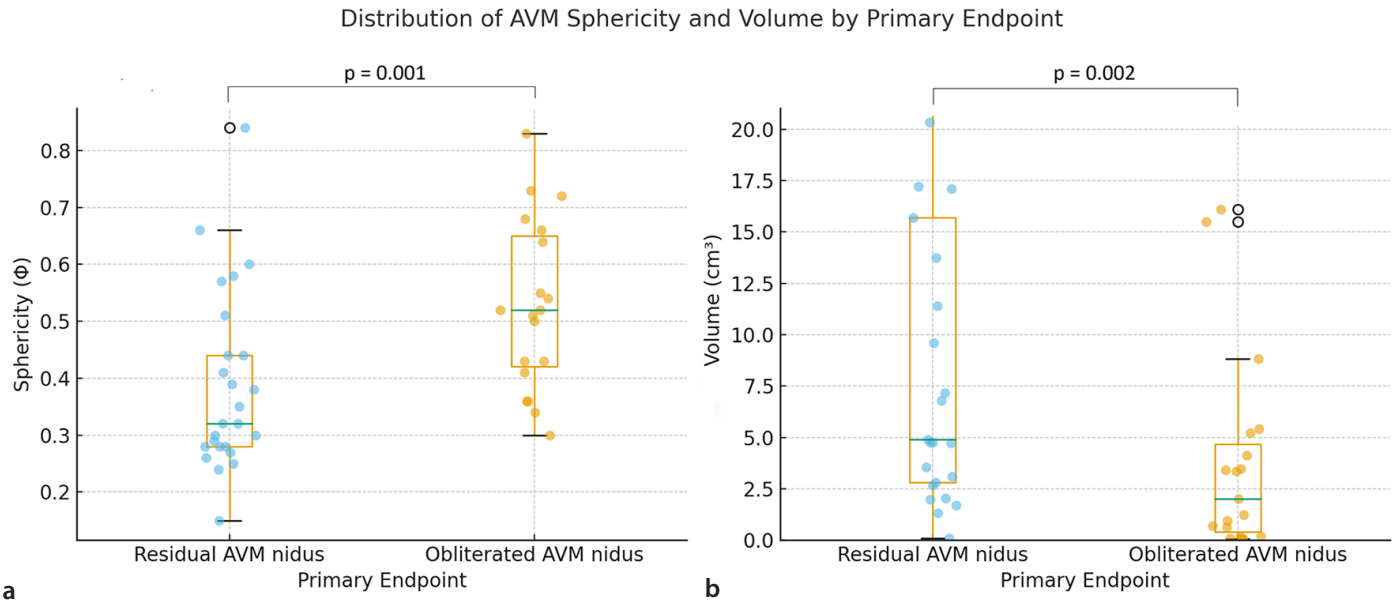



**Figure 3.** Boxplots showing the distribution of arteriovenous malformation (AVM) sphericity (Φ) and volume (V) according to treatment outcome, which is defined as residual vs. obliterated AVM nidus following Gamma Knife Radiosurgery. Sphericity index **(a)** showed a more compact and stable distribution compared with AVM volume **(b)**, which demonstrated wide variability and substantial overlap between groups.

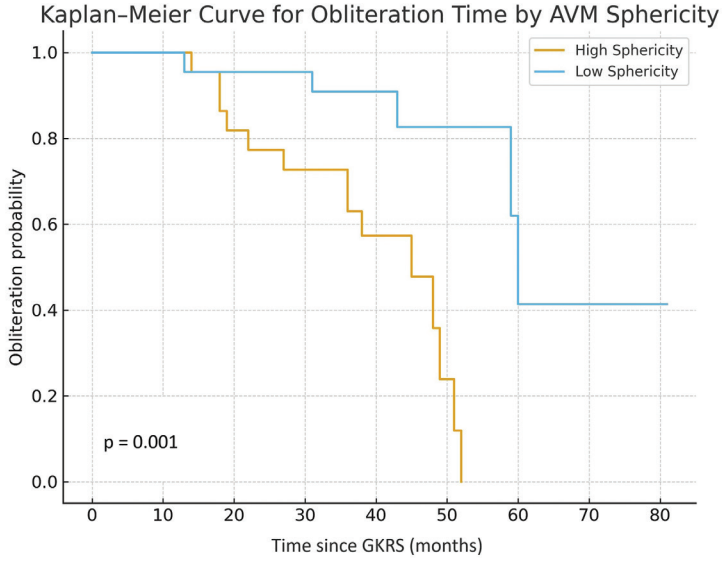



**Figure 4.** Kaplan–Meier analysis of arteriovenous malformation (AVM) obliteration time based on nidus sphericity. The AVMs were dichotomized according to the sphericity threshold of 0.41 (low sphericity: ≤ 0.41, high sphericity: > 0.41). Obliteration is defined as the primary endpoint. The median time to obliteration was 45 months for AVMs with a high-sphericity nidus and 60 months for AVMs with a low-sphericity nidus ($P$ = 0.001). GKRS; Gamma Knife® Radiosurgery.

or metastases, AVMs are inherently irregular and non-spherical structures; therefore, the sphericity values in the current study were lower than those typically reported in studies with solid tumors.[11,12] Nevertheless, obliterated AVMs demonstrated a median sphericity of 0.53 (0.26), compared with 0.32 (0.19) in AVMs with residual nidus. Using an optimal cut-off of 0.41, high-sphericity AVMs achieved obliteration at a median of 45 months, vs. 60 months in the low-sphericity group. These findings suggest that an AVM nidus with high sphericity correlates with shorter AVM obliteration time, whereas AVM volume and SM grade exert relatively smaller effects. However, sphericity did not retain independent predictive value in multivariate analysis for the AVM obliteration time, likely attributable to the limited sample size of the study.

Previous reports have consistently emphasized the nidus volume as the most robust structural feature for AVM obliteration.[9,16,17] Although our findings support this, the volume-based distribution of cerebral AVMs is extremely heterogeneous in the previous studies, leading to difficulties in determining the useful volume thresholds for achieving AVM obliteration. Marked differences in cerebral AVM volume make the pre-radiosurgical AVM volume clinically impractical to use as a prognostic parameter. Another caveat concerning AVM volume lies in the lack of knowledge on the extent of the nidal vascular distribution, such as a more compact AVM or dispersed AVM nidus may have a closer absolute AVM volume, although marginal radiation dose and time of obliteration, as well as the latent time to hemorrhage may differ from one another.[18] Therefore, the 3D features of AVM can be further elucidated to build a prognostic model for GKRS response. In the present cohort, the distribution of absolute nidal sphericity values effectively reflected the degree of surface irregularity within the nidus. The therapeutic principle of GKRS is based on endothelial injury and subsequent progressive fibrointimal hyperplasia leading to occlusion of the AVM nidus.[19] According to the radiomics-based tumor-shape study by Limkin et al.,[20] as sphericity decreases, intra-nidal irregularities and dispersion are expected to increase. Accounting for this consideration, a highly spherical AVM nidus can be expected to receive a more uniform radiation dose; however, when multiple isocenters and complex plugging techniques are applied, uniform dose distribution can be achievable across variable 3D nidus conformations. In our cohort, 40 (89.9%) AVMs were treated using a multiple-isocenter, complex plugging strategy, whereas only 4 cases were

**Table 4.** Cox proportional hazards model of AVM sphericity, morphological features, and demographics in relation to time to obliteration following GKRS

| | Univariate ( n = 44) | | Multivariate ( n = 44) | |
|---|---|---|---|---|
| Predictor | HR (95% CI) | *P* | HR (95% CI) | *P* |
| Age (year) | 1.02 (0.99–1.07) | 0.191 | | |
| Sex (female) | 0.91 (0.38–2.21) | 0.712 | | |
| Total marginal dose | 0.95 (0.88–1.03) | 0.213 | | |
| Fractionation scheme (hypofractionated) | 0.24 (0.06–1.04) | 0.060 | | |
| Multiple-isocenter GKRS plan | 0.29 (0.12–1.04) | 0.060 | | |
| Number of isocenters per nidus | 1.03 (0.96–1.10) | 0.394 | | |
| Spetzler–Martin grade | 0.39 (0.21–0.71) | **0.002** | 0.45 (0.160–1.29) | 0.142 |
| Intra-nidal aneurysm | 1.04 (0.36–2.99) | 0.936 | | |
| AVM volume ($cm^3$) | 0.88 (0.78–0.98) | **0.026** | 0.94 (0.82–1.08) | 0.383 |
| AVM sphericity | 36.29 (2.89–454.37) | **0.005** | 0.430 (0.01–59.74) | 0.734 |

AVM, arteriovenous malformation; HR, hazard ratio; GKRS, Gamma Knife Radiosurgery, CI, confidence interval

treated with a single, unplugged isocenter. The median number of isocenters did not differ significantly between the obliterated and residual AVM groups, suggesting that the use of multiple-isocenter planning is a consistent requirement across diverse nidus geometries to ensure therapeutic coverage. Therefore, the more favorable long-term obliteration outcomes observed in highly spherical AVMs may not be attributed only to distributed dose heterogeneity in the nidus but also to intrinsic anatomical factors, such as increased nidal vessel volume relative to surface area, which may enhance the effective bioavailability of therapeutic radiation even in complex AVM conformations.[21]

Earlier AVM obliteration has traditionally been associated with low nidus volume (< 10 $cm^3$) and lower SM grades. More recently, Pacini et al.[22] suggested that a high compactness index of the AVM nidus can be noted as a single independent quantitative predictor of AVM obliteration, demonstrating strong predictive capacity (AUC: 0.82; sensitivity: 97%; specificity: 52%). Our preliminary findings in this smaller cohort suggested that higher nidal sphericity is also associated with earlier AVM obliteration. However, the predictive performance of the sphericity index was not as robust as that of the compactness index in that series. Nonetheless, the direct comparison is not feasible as our cohort uniquely included AVMs treated exclusively with GKRS, whereas the Pacini et al. study comprised a larger group treated by multimodal strategies including endovascular embolization and open microsurgery.[22] Similar to the compactness index, our study demonstrated that a sphericity value > 0.41 may have predictive utility in achieving earlier AVM obliteration. Conceptually, compactness is defined as the vascular density within the parenchyma encompassing the AVM nidus,[23] whereas sphericity reflects surface irregularities and is presumed to be more sensitive to lobulations within an AVM nidus.[24] Notably, two AVM niduses with similar compactness values may differ substantially in terms of lobulation or sphericity. Therefore, integrating both compactness and sphericity indices as 3D parameters of AVM nidus, or developing automated software to calculate them, represents a promising research area that may offer new insights and facilitate early prediction of AVM obliteration following GKRS.

This study has several limitations, with the first being the retrospective single-center design and relatively small cohort size. The small cohort size may have reduced the statistical power for multivariate analysis, where the sphericity alone did not retain independent predictive capacity for AVM obliteration time. Nevertheless, the cohort was homogeneous, comprising only brain AVMs treated exclusively with GKRS. Second, AVM nidus segmentation was manually performed with semi-automated refinement, which is susceptible to inter-observer variability. Although the contouring was performed carefully and surface smoothing was avoided to preserve true geometric complexity, a certain extent of segmentation bias remains possible. Third, heterogeneity of treatment parameters, including marginal dose selection and use of hypofractionated vs. single-session GKRS, could have confounded AVM obliteration outcomes. Moreover, the compactness index was not specifically evaluated in this cohort, as the segmented 3D AVM nidus in the 3D Slicer environment did not precisely correspond to the planned treatment volume delineated by the prescription isodose boundaries within the radiosurgical treatment planning system. This discrepancy between anatomical and dosimetric segmentation volumes precluded reliable computation of the compactness index within the same workflow. Finally, although this study highlights the potential of sphericity as a novel geometric parameter, the absence of external validation of AVM nidus sphericity measurements represents another limitation, and the methodology should be further verified in larger independent cohorts. In the future, we plan to continue the analysis with a larger cohort.

In conclusion, the sphericity of the AVM nidus represents a measure of intra-nidal irregularity and elongation, and, therefore, the extent of nidal vascular surface area contacting the surrounding brain parenchyma. Although most AVM niduses do not approximate a true spherical shape, relatively higher sphericity values ($\Phi$ > 0.41) were moderately associated with improved radiosurgical response and higher obliteration rates. Compared with the absolute AVM nidus volume, the sphericity exhibited far more stable distribution, highlighting its potential role in refining prognostic models following GKRS. However, sphericity alone did not retain independent predictive value for AVM obliteration time, likely due to the limited sample size, which highlights the need for external validation. Despite these limitations, the present findings suggest that AVM nidus sphericity represents a practical and promising 3D geometric marker that warrants further validation in larger multicenter series to establish its predictive role for AVM obliteration.

## Footnotes

### Conflict of interest disclosure

The authors declared that there is no conflict of interest.

## References


1. Al-Shahi R, Bhattacharya JJ, Currie DG, et al. Prospective, population-based detection of intracranial vascular malformations in adults: the Scottish Intracranial Vascular Malformation Study (SIVMS). *Stroke*. 2003;34(5):1163-1169. [Crossref]
2. Gross BA, Du R. Natural history of cerebral arteriovenous malformations: a meta-analysis. *J Neurosurg*. 2013;118(2):437-443. [Crossref]
3. Kano H, Kondziolka D, Flickinger JC, et al. Stereotactic radiosurgery for arteriovenous malformations after embolization: a case-control study. *J Neurosurg*. 2012;117(2):265-275. [Crossref]
4. Yamamoto M, Jimbo M, Hara M, Saito I, Mori K. Gamma Knife Radiosurgery for arteriovenous malformations: long-term follow-up results focusing on complications occurring more than 5 years after irradiation. *Neurosurgery*. 1996;38(5):906-914. [Crossref]
5. Xiaochuan H, Yuhua J, Xianli L, Hongchao Y, Yang Z, Youxiang L. Targeted embolization reduces hemorrhage complications in partially embolized cerebral AVM combined with Gamma Knife Surgery. *Interv Neuroradiol*. 2015;21(1):80-87. [Crossref]
6. Kwon Y, Jeon SR, Kim JH, et al. Analysis of the causes of treatment failure in Gamma Knife Radiosurgery for intracranial arteriovenous malformations. *J Neurosurg*. 2000;93 Suppl 3:104-106. [Crossref]
7. Lauren C, Niryana IW, Mahadewa TGB. Impact of embolization on stereotactic radiosurgery outcomes for intracranial arteriovenous malformations Spetzler-Martin grades III-V: a systematic review and meta-analysis. *Front Surg*. 2025;12:1563256. [Crossref]
8. Byun J, Kwon DH, Lee DH, Park W, Park JC, Ahn JS. Radiosurgery for cerebral arteriovenous malformation (AVM): current treatment strategy and radiosurgical technique for large cerebral AVM. *J Korean Neurosurg Soc*. 2020;63(4):415-426. [Crossref]
9. Li W, Wang Y, Lu L, Zhang Y. The factors associated with obliteration following stereotactic radiosurgery in patients with brain arteriovenous malformations: a meta-analysis. *ANZ J Surg*. 2022;92(5):970-979. [Crossref]
10. Huang PW, Peng SJ, Pan DH, et al. Compactness index: a radiosurgery outcome predictor for patients with unruptured brain arteriovenous malformations. *J Neurosurg*. 2022;138(1):241-250. [Crossref]
11. Tarsitano A, Ricotta F, Cercenelli L, et al. Pretreatment tumor volume and tumor sphericity as prognostic factors in patients with oral cavity squamous cell carcinoma. *J Craniomaxillofac Surg*. 2019;47(3):510-515. [Crossref]
12. Ko PH, Kim HJ, Lee JS, Kim WC. Tumor volume and sphericity as predictors of local control after stereotactic radiosurgery for limited number (1-4) brain metastases from nonsmall cell lung cancer. *Asia Pac J Clin Oncol*. 2020;16(3):165-171. [Crossref]
13. Lin JY, Lu CF, Hu YS, et al. Magnetic resonance radiomics-derived sphericity correlates with seizure in brain arteriovenous malformations. *Eur Radiol*. 2024;34(1):588-599. [Crossref]
14. Graffeo CS, Sahgal A, De Salles A, et al. Stereotactic radiosurgery for Spetzler-Martin grade I and II arteriovenous malformations: International Society of Stereotactic Radiosurgery (ISRS) Practice Guideline. *Neurosurgery*. 2020;87(3):442-452. [Crossref]
15. Chung Y, Park CK, Choi SK, Lim YJ. Long-term follow-up study of Gamma Knife Radiosurgery for arteriovenous malformations with diffuse-type nidi. *J Korean Soc Ster Func Neurosurg*. 2021;17(2):92-96. [Crossref]
16. Karlsson B, Lindquist C, Steiner L. Prediction of obliteration after Gamma Knife Surgery for cerebral arteriovenous malformations. *Neurosurgery*. 1997;40(3):425-430. [Crossref]
17. Kim MJ, Jung HH, Kim YB, et al. Comparison of single-session, neoadjuvant, and adjuvant embolization Gamma Knife Radiosurgery for arteriovenous malformation. *Neurosurgery*. 2023;92(5):986-997. [Crossref]
18. Huang PW, Peng SJ, Pan DH, et al. Vascular compactness of unruptured brain arteriovenous malformation predicts risk of hemorrhage after stereotactic radiosurgery. *Sci Rep*. 2024;14(1):4011. [Crossref]
19. Schneider BF, Eberhard DA, Steiner LE. Histopathology of arteriovenous malformations after gamma knife radiosurgery. *J Neurosurg*. 1997;87(3):352-357. [Crossref]
20. Limkin EJ, Reuzé S, Carré A, et al. The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features. *Sci Rep*. 2019;9(1):4329. [Crossref]
21. Sun MH, Tsuei YS, Wang TW, et al. Outcomes of gamma knife treated large symptomatic arteriovenous malformations according to guidelines of Taiwan neurosurgical consensus. Radiat Oncol. 2025;20(1):89. [Crossref]
22. Pacini A, Shotar E, Granger B, et al. Niduscompacity determined by semi-automated segmentation is a strong quantitative predictor of brain arterio-venous malformation cure. *Clin Neuroradiol*. 2023;33(4):1095-1104. [Crossref]
23. Frisoli FA, Lang SS, Vossough A, et al. Intrarater and interrater reliability of the pediatric arteriovenous malformation compactness score in children. *J Neurosurg Pediatr*. 2013;11(5):547-551. [Crossref]
24. Cruz-Matías I, Ayala D, Hiller D, et al. Sphericity and roundness computation for particles using the extreme vertices model. *Journal of Computational Science*. 2019;30:28-40. [Crossref]

Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263854

NEURORADIOLOGY

LETTER TO THE EDITOR



# Letter to the Editor: Comment on the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans

Yiğit Can Kartal

University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Department of Radiology, İstanbul, Türkiye

**Dear Editor,**

I read with great interest the article by Bayar-Kapıcı et al.[1] evaluating the diagnostic sensitivity of a multimodal large language model (MLLM) in detecting intracranial hemorrhage in non-contrast cranial computed tomography. The authors are to be commended for addressing a timely and clinically relevant topic and for providing a systematic evaluation of an MLLM in the context of acute neuroimaging.

The reported findings offer valuable insight into the current capabilities and limitations of MLLMs in image-based diagnostic tasks. In particular, the observed performance of MLLMs in subtle and borderline hemorrhagic findings may be interpreted in light of the inherent diagnostic complexity of these cases. In routine clinical practice, such findings often fall into a diagnostic gray zone, in which a degree of interpretive variability is unavoidable. From this perspective, discordance between an MLLM's output and the reference interpretation may, in part, reflect intrinsic diagnostic uncertainty rather than the true inadequacy of the model.

The image-only evaluation framework employed in the study represents a deliberately controlled and methodologically sound approach. However, given the reasoning-based architecture of MLLMs, diagnostic behavior may reasonably be influenced by the availability of a minimal, clinically relevant context. The inclusion of limited clinical cues—such as trauma history, anticoagulant use, or the patient's overall clinical condition—may provide a more representative assessment of how these models could function in real-world decision-support scenarios. In this regard, recent evidence suggests that prompt engineering and input conditions play a decisive role in how MLLMs integrate the clinical context with imaging data, with measurable effects on diagnostic performance.[2] Collectively, these observations underscore the synergistic role of clinical contexts in diagnostic reasoning.

Another aspect worth considering is the reliance on one or two preselected image slices for evaluation. Radiologic interpretation in clinical practice often benefits from reviewing adjacent slices across an image series together with dynamic window and level adjustments, particularly for subtle hemorrhagic findings or for distinguishing a true pathology from artifacts. Evaluation based on isolated images without the ability to adjust window settings, although practical for experimental design, may therefore differ from routine diagnostic workflows and influence measured performance metrics.

Finally, although sensitivity remains a critical metric, the potential for incorrect affirmative or negative outputs in the absence of sufficient contextual and visual grounding represents an important safety consideration, particularly in high-stakes neuroimaging scenarios, in which confidently presented but incorrect model outputs may have clinical consequences. Future investigations incorporating clinically realistic prompts, slice-to-slice correlation with adjustable windowing, and human-in-the-loop frameworks may help clarify how MLLMs—designed to jointly process visual and textual information—can best be integrated into clinical practice as assistive tools rather than standalone diagnostic systems. Within such frameworks, human–artificial intelligence collaboration, in which MLLMs support radiologists by highlighting potential abnormalities while final interpretation and decision-making remain

**Handling editor:** Mehmet Ruhi Onur

**Corresponding author:** Yiğit Can Kartal

**E-mail:** yckartal@hotmail.com

Received 05 January 2026; accepted 16 January 2026.

Epub: 16.02.2026

Publication date: 01.07.2026

DOI: 10.4274/dir.2026.263854

You may cite this article as: Kartal YC. Letter to the Editor: comment on the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol*. 2026;32(4):519-520.

clinician driven, may represent the most appropriate and safest paradigm for clinical deployment.

I congratulate the authors on this valuable contribution and hope that these considerations may support continued efforts to develop clinically meaningful and safe applications for MLLMs in radiology.

## Conflict of interest disclosure

The author declared no conflicts of interest.

## References


1. Bayar-Kapıcı O, Altunışık E, Musabeyoğlu F, Dev Ş, Kaya Ö. Artificial intelligence in radiology: diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol*. 2026;32(1):27-32. [Crossref]
2. Han T, Jeong WK, Shin J. Diagnostic performance of multimodal large language models in radiological quiz cases: the effects of prompt engineering and input conditions. *Ultrasonography*. 2025;44(3):220-231. [Crossref]