

EISSN 1305-3612

DIR

Diagnostic and Interventional
Radiology

TSR
1924
TURKISH SOCIETY
OF RADIOLOGY

dirjournal.org

VOLUME 31
ISSUE 4
July 2025

Editor in Chief

Mehmet Ruhi Onur, MD

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

ORCID ID: 0000-0003-1732-7862

Section Editors and Scientific Editorial Board

Abdominal Imaging

İlkay S. İdilman, MD 

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

ORCID ID: 0000-0002-1913-2404

Sonay Aydın, MD 

Department of Radiology, Erzincan Binali Yıldırım University Faculty of Medicine, Erzincan, Türkiye

ORCID ID: 0000-0002-3812-6333

Artificial Intelligence and Informatics

Burak Koçak, MD 

Department of Radiology, University of Health Sciences, Başakşehir Çam and Sakura City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0002-7307-396X

Tuğba Akıncı D'Antonoli, MD 

Institute of Radiology and Nuclear Medicine, Cantonal Hospital Baselland, Liestal, Switzerland

ORCID ID: 0000-0002-7237-711X

Breast Imaging

Füsun Taşkın, MD 

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

ORCID ID: 0000-0001-7985-3660

Chest and Cardiovascular Imaging

Furkan Ufuk, MD 

Department of Radiology, The University of Chicago, Chicago, USA

ORCID ID: 0000-0002-8614-5387

Hybrid Imaging and Nuclear Medicine

Evrım Bengi Türkbey, MD 

Radiology and Imaging Sciences, Clinical Center, National Institutes of Health Bethesda, Maryland, United States

ORCID ID: 0000-0002-5216-3528

Interventional Radiology

Barbaros Çil, MD, FCIIRSE 

Department of Radiology, Koç University School of Medicine, İstanbul, Türkiye

ORCID ID: 0000-0003-1079-0088

Bahri Üstünsöz, MD 

Department of Radiology, LSUHSC (Louisiana State University Health Science Center) School of Medicine, New Orleans, United States

ORCID ID: 0000-0003-4308-6708

James Milburn, MD 

Department of Radiology, Ochsner Medical System, New Orleans, Louisiana, USA

ORCID ID: 0000-0003-3403-2628

Musculoskeletal Imaging

Zeynep Maraş Özdemir, MD 

Department of Radiology, İnönü University Faculty of Medicine, Malatya, Türkiye

ORCID ID: 0000-0003-1085-8978

Neuroradiology

Gülgün Yılmaz Ovalı, MD 

Department of Radiology, Celal Bayar University Faculty of Medicine, Manisa, Türkiye

ORCID ID: 0000-0001-8433-5622

Erkan Gökçe, MD 

Department of Radiology, Tokat Gaziosmanpaşa University Faculty of Medicine, Tokat, Türkiye

ORCID ID: 0000-0003-3947-2972

Pediatric Radiology

Meltem Ceyhan Bilgici, MD 

Department of Radiology, 19 Mayıs University Faculty of Medicine, Samsun, Türkiye

ORCID ID: 0000-0002-0133-0234

Evrım Özmen, MD 

Department of Radiology, Koç University Hospital, İstanbul, Türkiye

ORCID ID: 0000-0003-3100-4197

Publication Coordinator

Nermin Tunçbilek, MD 

Department of Radiology, Trakya University Faculty of Medicine, Edirne, Türkiye

ORCID ID: 0000-0002-8734-1849

Biostatistical Consultant

İlker Ercan, PhD 

Department of Biostatistics, Uludağ University School of Medicine, Bursa, Türkiye

ORCID ID: 0000-0002-2382-290X

Publication Services

Galenos Publishing, İstanbul, Türkiye

Past Editors

Editors in Chief

Mustafa Seçil, MD (2016-2023)

Nevzat Karabulut, MD (2011-2016)

Üstün Aydingöz, MD (2010-2011)

Okan Akhan, MD (2001-2010)

Ferhun Balkancı, MD (1999-2001)

Aytekin Besim, MD (1994-1999)*

* Dr. Aytekin Besim actually served as the General Coordinator. His work in this capacity, however, was in effect that of an Editor in Chief.

Editors

Ayşenur Cila, MD (2001-2002)

Suat Kemal Aytaç, MD (1997-2001)

Erhan Ilgıt, MD (1994-2001)

Okan Akhan, MD (1994-2001)

Ferhun Balkancı, MD (1994-2000)

Serdar Akyar, MD (1994-1997)

Section Editors

Section Editorship was established in 2002 at the tenure of Dr Okan Akhan, Editor in Chief.

Abdominal Imaging

Bengi Gürses, MD (2020-2023)

Mehmet Ruhi Onur, MD (2016-2023)

Barış Türkbey, MD (2014-2020)

Mustafa N. Özmen, MD (2012-2018)

Murat Acar, MD (2015-2016)

Mustafa Seçil, MD (2011-2016)

Ahmet Tuncay Turgut, MD (2011)

Deniz Akata, MD (2007-2011)

Ayşe Erden, MD (2002-2011)

Okan Akhan, MD (2002-2010)

Hakan Özdemir, MD (2002-2010)

Artificial Intelligence and Informatics

Barış Türkbey, MD (2020-2023)

Breast Imaging

Mustafa Erkin Arıbal, MD (2016-2023)

Sibel Kul (2015-2018)

Ayşenur Oktay, MD (2009-2014)

Ayşegül Özdemir, MD (2004-2009)

Cardiovascular Imaging

Uğur Bozlar, MD (2016-2023)

Muşturay Karçaaltıncaba, MD (2007-2010)

Mecit Kantarcı (2010-2016)

Chest Imaging

Nevzat Karabulut, MD (2010-2014)

Çetin Atasoy, MD (2007-2010)

Macit Arıyürek, MD (2002-2007)

Figen Demirkazık, MD, (2014-2018)

General Radiology

Ersin Öztürk, MD (2014-2017)

Utku Şenol, MD (2010-2013)

Oğuz Dicle, MD (2007-2010)

Interventional Radiology

Cüneyt Aytekin, MD (2016-2023)

Bora Peynircioğlu, MD (2012-2015)

Levent Oğuzkurt, MD (2011-2014)

Fatih Boyvat, MD (2007-2010)

İsmail Oran, MD (2015-2019)

Musculoskeletal Imaging

Hatice Tuba Sanal, MD (2016-2023)

Fatih Kantarcı, MD (2014-2016)

Ayşenur Oktay, MD (2011-2013)

Üstün Aydınöz, MD (2002-2011)

Berna Dirim Mete (2016-2017)

Neuroradiology and Head & Neck Imaging

Kubilay Aydın, MD (2016-2023)

Nafi Aygün, MD (2016-2023)

Kader Karlı Oğuz, MD (2011-2015)

Süleyman Men, MD (2007-2013)

Muhteşem Ağıldere, MD (2002-2011)

Nuclear Medicine

A. Cahid Civelek, MD (2016-2023)

Oktay Sarı, MD (2015)

Akın Yıldız, MD (2011-2014)

Pediatric Radiology

Korgün Koral, MD (2016-2023)

Murat Kocaoğlu, MD (2016-2023)

Ensar Yekeler, MD (2014-2016)

Suat Fitöz, MD (2007-2013)

Diagnostic and Interventional Radiology (Diagn Interv Radiol) is a bimonthly periodical of the Turkish Society of Radiology and the content of the journal is available at <https://www.dirjournal.org/>. It is peer-reviewed and adheres to the highest ethical and editorial standards. The editors of the journal endorse the Editorial Policy Statements Approved by the Council of Science Editors Board of Directors (<https://cse.memberclicks.net/>). The journal is in compliance with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals published by the International Committee of Medical Journal Editors (updated May 2022, www.icmje.org).

First ten volumes of Diagnostic and Interventional Radiology have been published in Turkish under the name of Tanısal ve Girişimsel Radyoloji (Index Medicus® abbreviation: Tani Girişim Radyol), the current title's exact Turkish translation.

Diagnostic and Interventional Radiology is an open access publication, and the journal's publication model is based on Budapest Open Access Initiative (BOAI) declaration. All published content is available online, free of charge at <https://www.dirjournal.org/>. Authors retain the copyright of their published work in Diagnostic and Interventional Radiology. The journal's content is licensed under a Creative Commons Attribution-NonCommercial (CC BY-NC) 4.0 International License which permits third parties to share and adapt the content for non-commercial purposes by giving the appropriate credit to the original work.

Please refer to the journal's webpage (<https://dirjournal.org/>) for "Aims and Scope", "Instructions to Authors" and "Instructions to Reviewers".

The editorial and publication processes of the journal are shaped in accordance with the guidelines of the ICMJE, WAME, CSE, COPE, EASE, and NISO.

Diagnostic and Interventional Radiology is indexed in **SCI-Expanded, Pubmed/Medline, Pubmed Central, TUBITAK ULAKBIM TR Index, DOAJ, HINARI, EMBASE, CINAHL, Scopus, Gale and CNKI.**

The journal is published online.

Owner: Can Çevikol on behalf of Turkish Society of Radiology

Responsible Manager: Mehmet Ruhi Onur

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 2025

EISSN 1305-3612

International scientific journal published bimonthly.



Contents

ABDOMINAL IMAGING

285 Invited Review Tumor-like conditions that mimic liver tumors. *Nir Stanietzky, Ahmed Ebada Salem, Khaled M. Elsayes, Maryam Rezvani, Kurt Fraivillig, Usama Salem, Sergio Klimkowski, Mahmoud Diab, Sagar Naik, Ahmed Sobieh, Christine O. Menias, Akram M. Shaaban*

295 Original Article The role of multiparametric magnetic resonance imaging in the differentiation of low- and high-grade non-muscle invasive bladder cancer. *Merve Nur Taşdemir, Uluhan Eryürük, Ural Oğuz, Birgül Tok, Serdar Aslan*

ARTIFICIAL INTELLIGENCE AND INFORMATICS

303 Invited Review The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. *Benjamin D. Simon, Kutsev Bengisu Ozyoruk, David G. Gelikman, Stephanie A. Harmon, Barış Türkbey*

313 Original Article Machine learning models for discriminating clinically significant from clinically insignificant prostate cancer using bi-parametric magnetic resonance imaging. *Hakan Ayyıldız, Okan İnce, Esin Korkut, Merve Gülbiz Dağoğlu Kartal, Atadan Tunacı, Şükrü Mehmet Ertürk*

321 Invited Review Reproducibility and interpretability in radiomics: a critical assessment. *Aydın Demircioğlu*

329 Letter To The Editor Letter to the editor re: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging. *Hinpetch Daungsupawong, Viroj Wiwanitkit*

331 Letter To The Editor Reply: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Alperen Elek, Duygu Doğa Ekizalioglu, Ezgi Güler*

BREAST IMAGING

333 Original Article Diagnostic value of the flare sign in predicting extracapsular extension in metastatic axillary lymph nodes and nodal status on breast magnetic resonance imaging. *Cihan Özgür, Baran Serdar Sunal, Savaş Hereklioğlu, Meltem Öznur, Sibel Özkan Gürdal*

CARDIOVASCULAR IMAGING

340 Commentary Revolutionizing cardiac imaging: how photon-counting computed tomography is redefining coronary artery stent assessment. *Çağdaş Topel, Furkan Ufuk*

GENERAL RADIOLOGY

342 Invited Review Multidisciplinary approach to diagnostic radiology education: a novel educational intervention for Turkish medical students. *Parth Patel, Emre Altınmakas, Görkem Ayas, Rachel Stanietzky, Madeline L. Stewart, Abdelrahman Elshikh, Disha Ram, Hrishika Bhosale, Mohamed Eltaher, Serageldin Kamel, Munevver N. Duran, Umut Yücel, Mohamed Badawy, Scott Rohren, Khaled M. Elsayes*

HEAD AND NECK IMAGING

347 Original Article A diagnostic model based on magnetic resonance imaging for Menière's disease: a multicentre study. *Xinyi Chen, Yanfeng Zhao, Yunchong Han, Kai Wei, Shufang Cheng, Yongjun Ye, Jie Feng, Xinchun Huang, Jingjing Xu*

INTERVENTIONAL RADIOLOGY

359 Original Article Splenic artery embolization in the treatment of blunt splenic injury: single level 1 trauma center experience. *Katelyn Gill, Sarah Aleman, Alexandra H. Fairchild, Bahri Üstünsöz, Dan Laney, Alison A. Smith, Hector Ferral*

366 Original Article Safety and efficacy of flow diverter stents in the treatment of bifurcation cerebral aneurysms: single-center experience. *Yerbol Makhambetov, Aiman Maida, Chingiz Nurimanov, Assylbek Kaliyev, Baurzhan Kunakbayev, Nurtay Nurakay, Serik Dyussebayev, Nursultan Makhambetov*

372 Original Article Computed tomography-guided cryoablation in treating adrenal metastases: a retrospective single-center study. *Claudio Pusceddu, Eliodoro Faiella, Claudio Cau, Pierluigi Rinaldi, Luca Melis, Salvatore Marsico*

377 Original Article Evaluating the prognostic impact of inflammatory markers on treatment outcomes in patients with intrahepatic cholangiocarcinoma undergoing radioembolization. *Sinan Sözütok, Ferhat Can Pişkin, Hüseyin Tuğsan Ballı, Berkay Dik*

Contents

384 Original Article Efficacy of endovascular circulating false lumen occlusion in chronic aneurysmal descending aortic dissections. *Emeric Gremen, Mathieu Finas, Elliott Mathieu, Frédéric Thony, Mathieu Rodiere, Julien Ghelfi*

MOLECULAR IMAGING

392 Original Article Amide proton transfer-weighted magnetic resonance imaging for the evaluation of testicular spermatogenic function: a preliminary study. *Guanglei Tang, Shulin Ma, Wenhao Fu, Weijian Yun, Yang Peng, Jian Guan*

PEDIATRIC RADIOLOGY

399 Original Article Validation of R2* magnetic resonance imaging for quantifying secondary iron overload in pediatric patients. *Tahani M. Ahmad, Fareed M. Khedair Ahmad, Mohamed Abdoell, Olfat M. Ahmad, Matthew T. Rogers*



Tumor-like conditions that mimic liver tumors

- Nir Stanietzky¹
- Ahmed Ebada Salem^{2,3}
- Khaled M. Elsayes¹
- Maryam Rezvani²
- Kurt Fraivillig¹
- Usama Salem¹
- Sergio Klimkowski¹
- Mahmoud Diab⁴
- Sagar Naik¹
- Ahmed Sobieh⁵
- Christine O. Menias⁶
- Akram M. Shaaban²

¹The University of Texas MD Anderson Cancer Center, Department of Abdominal Radiology, Texas, USA

²The University of Utah, Department of Radiology, Utah, USA

³Alexandria University Faculty of Medicine, Department of Radiology, Alexandria, Egypt

⁴Suez Canal University Faculty of Medicine, Department of Radiology, Ismailia, Egypt

⁵University of Kentucky, Department of Radiology, Kentucky, USA

⁶Mayo Clinic, Department of Radiology, Arizona, USA

ABSTRACT

Non-neoplastic tumor-like conditions of the liver can appear similar to hepatic neoplasms. In many cases, a biopsy is required to confirm the pathology. However, several tumor-like conditions can be correctly diagnosed or suggested prospectively, thus saving patients from unnecessary anxiety and expense. In this image-focused review, we present the ultrasound, computed tomography, magnetic resonance imaging, and positron emission tomography scan features of eight such entities. Clues that indicate the correct pathology are discussed, and the usual clinical setting is described. Many of these lesions are treated differently from true neoplasms, and the current treatment plan is discussed in many of the cases presented. After reviewing this article, the reader will have a better understanding of these lesions and the situations in which they should be included in the differential diagnosis.

KEYWORDS

Benign hepatic lesion, hepatic amyloidosis, hepatic extramedullary hematopoiesis, hepatic pseudotumor, hepatic sarcoidosis, hepatic tumor mimics, hepatic tumor-like conditions, hepatobiliary tuberculomas, liver imaging, mesenchymal hamartoma, myofibroblastoma, peliosis hepatis

Differentiating hepatic tumors from tumor-like conditions can be challenging because the imaging features may overlap. An understanding of several of the more common tumor-like lesions that can be misdiagnosed as tumors is vital so that the correct pathology can be included in the differential diagnosis. In this article, we describe the clinical and imaging features of some of the more common tumor-like conditions to assist radiologists in determining the correct diagnosis.

Mesenchymal hamartoma

Mesenchymal hamartomas of the liver (MHL) represent only 5% of pediatric tumors. However, after infantile hemangiomas, they are the second most common benign liver tumors in children younger than 5 years. The lesions normally present as painless abdominal distention in a patient with a normal or slightly elevated alpha fetoprotein level.¹ In some cases, the alpha fetoprotein level is substantially elevated, leading to diagnostic confusion. The tumors can be very large (up to 20–30 cm).¹ Mesenchymal hamartomas are associated with Beckwith–Wiedemann syndrome and multiple congenital fetal anomalies.² Most cases (75%) arise from the right hepatic lobe, and up to 20% of tumors are pedunculated.¹

On imaging, an MHL appears as a large, well-demarcated, multilocular cystic mass with septal and/or stromal enhancement (Figure 1). On magnetic resonance imaging (MRI), the cystic components demonstrate high signal intensity on T2-weighted images and variable signal intensity on T1-weighted images as a result of hemorrhagic or proteinaceous debris (Figure 2). The solid and stromal components are mildly enhanced by the administration of intravenous gadolinium.³

Histologically, MHLs are composed of hepatocytes, haphazardly arranged bland spindle cells, and benign bile ducts in a collagenous stroma.^{1,4} They have no malignant features and are usually cured through surgical resection. On rare occasions, MHLs have been theorized to progress to embryonal sarcoma, as the two entities have similar molecular alterations in chro-

Corresponding author: Khaled M. Elsayes

E-mail: kmelsayes@mdanderson.org

Received 25 April 2024; revision requested 12 May 2024; accepted 05 June 2024.



Epub: 14.06.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242826

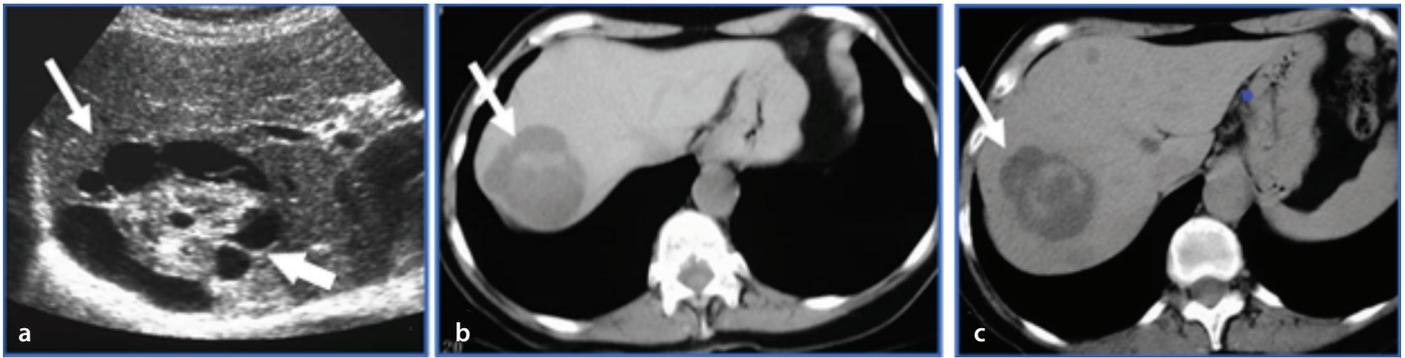


Figure 1. Grayscale ultrasound scan (a) showing a complex cystic mass with a solid component (white arrow). Non-enhanced computed tomography images (b, c) showing a mass in the right hepatic lobe (white arrows). The mass was predominantly cystic and had a central solid component. A mesenchymal hamartoma was identified through pathology.

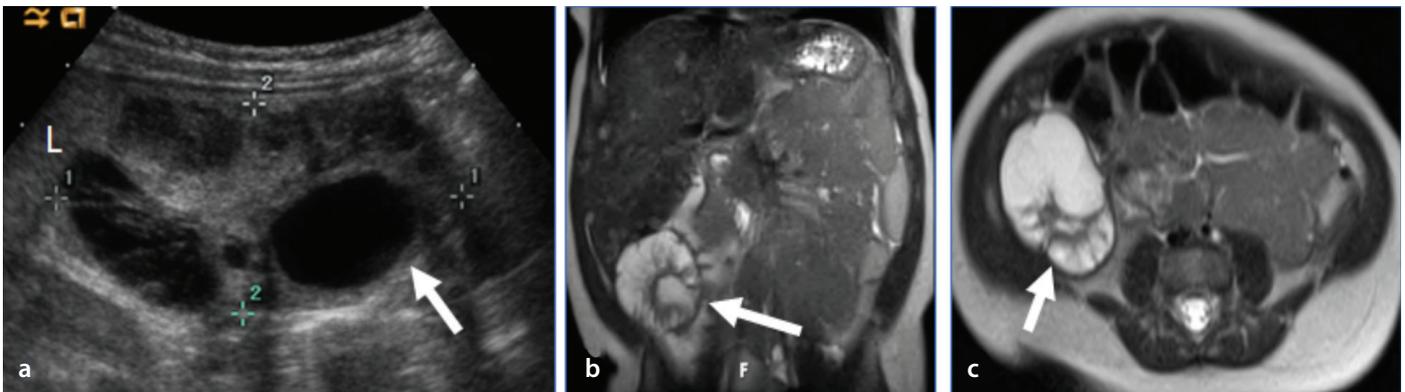


Figure 2. Grayscale ultrasound scan (a) showing a mixed solid and cystic mesenchymal hamartoma (white arrow) arising from the inferior aspect of the liver (L). Coronal (b) and axial (c) T2-weighted images showing a complex cystic mass with internal septations (white arrows).

mosome 19.¹ Additionally, there have been case reports of benign bile duct-like structures in the periphery of embryonal sarcomas with a similar histologic configuration to MHL, raising the possibility that the MHL was the precursor lesion.¹ Therefore, complete surgical excision is crucial.² Rarely, symptomatic unresectable disease has necessitated liver transplantation.²

Myofibroblastoma

Myofibroblastomas (inflammatory pseudotumors), which represent less than 1% of liver tumors, are a heterogeneous group of masses that form benign tumors composed of inflam-

matory cells and fibrous stroma. These tumors may be very slightly more prevalent in men. Patients are diagnosed at an average age of 50 years, and there is no association with cirrhosis. The etiology of myofibroblastomas is unknown; however, it has been hypothesized that they result from hepatic infection, cholangitis, or a vascular injury.^{4,5} In most cases, the infectious agent is not identified. Some of these tumors occur in patients with immunoglobulin G4 sclerosing disease, and these patients may have a history of autoimmune pancreatitis or sclerosing cholangitis.^{2,5} Patients may present with fever, abdominal pain, and weight loss. Most tumors are 2–5 cm in diameter when detected, but some patients have presented with tumors greater than 20 cm.⁵ Patients may have elevated erythrocyte sedimentation rates, neutrophil counts, and C-reactive protein levels.⁵ The pathologic features can overlap with other disease entities, and a subset requires next generation sequencing to aid in diagnosis.²

On ultrasound, myofibroblastomas present as well-circumscribed heterogeneous masses with solid and cystic components. On computed tomography (CT), they are commonly hypoaattenuating compared with the liver on unenhanced images, and they usually demonstrate peripheral or septal en-

hancement on delayed images. At least part of the lesion may demonstrate delayed enhancement, presumed to be related to abundant fibrous tissue (Figure 3). On MRI, they are hypointense in T1-weighted images and isointense or hyperintense on T2-weighted images, with a similar postcontrast enhancement pattern to that seen on CT (Figure 4). Typically, no associated desmoplastic reaction occurs. The differential diagnosis includes cholangiocarcinoma, which can also demonstrate delayed enhancement, abscess, metastasis, and hepatocellular carcinoma.⁶

Myofibroblastomas are often treated conservatively because they may resolve spontaneously or respond to antibiotic therapy or steroid therapy. Although surgical resection is curative, it is generally reserved for cases in which no biopsy was performed or in which the diagnosis remained unclear after biopsy. Surgery may also be performed in cases in which the lesion does not respond to conventional therapy.

Extramedullary hematopoiesis

Extramedullary hematopoiesis (EMH) refers to the production of blood cells outside the bone marrow, a phenomenon observed

Main points

- Non-neoplastic lesions of the liver can appear similar to hepatic tumors. Although pathologic evidence is often necessary for a definitive diagnosis, including these entities in the differential diagnosis can aid the clinician with the diagnostic workup.
- Multiple clinical clues might suggest that a liver lesion is benign. Although these may not be definitive, key factors include normal tumor markers, a history of infection, and a history of a systemic disease known to involve the liver.

in patients with defective marrow synthesis (e.g., chronic myelofibrosis) or with conditions associated with peripheral red blood cell destruction (such as hemoglobinopathies and hemolytic anemias).⁷ The liver and spleen are the most common sites for EMH. Organ enlargement, particularly hepatomegaly, is more frequently observed than discrete lesions. When discrete lesions do occur, they may present as single or multiple focal masses or as infiltrative periportal or peribiliary soft tissue lesions.^{8,9}

On ultrasound, lesions associated with EMH can appear hypoechoic or hyperechoic and typically present as heterogeneously

attenuated nodules.^{10,11} CT generally reveals hypodense lesions with variable attenuation and may show no, mild, or heterogeneous enhancement (Figure 5).^{10,12,13} In rare cases, lesions may exhibit fat density without enhancement.¹⁴

On MRI, EMH lesions have variable signal characteristics, depending on whether the lesion is functional (actively hematopoietic) or inactive. Active lesions typically exhibit intermediate T1 and high T2 signal intensities with mild to moderate enhancement (Figure 6). By contrast, inactive lesions, which are predominantly composed of fibrous tissue, indicate low T1 and T2 signal intensity

with minimal or no enhancement.^{12,15} Macroscopic fat, which can occur in inactive lesions, produces a signal intensity that detects fat in all sequences. In patients with transfusion-related hemochromatosis, decreased background hepatic signal intensity is seen on sequences most sensitive to magnetic susceptibility.¹⁵ Despite these imaging findings, a definitive diagnosis usually requires pathologic evaluation, which often reveals erythroid or myeloid precursors along with fatty or fibrous tissue.¹⁶

Epstein-Barr virus-associated smooth muscle tumor

Epstein-Barr virus (EBV)-associated smooth muscle tumors are rare slow-growing mesenchymal tumors that primarily affect immunocompromised patients. The incidence is highest among individuals who are HIV positive, followed by patients with drug-induced immunosuppression after organ transplantation, and, less commonly, patients with congenital immunodeficiency disorders.¹⁷⁻²⁰

These tumors impact both adult and pediatric patients and are slightly more prevalent among women.²¹ Abdominal pain is frequently reported as the main presenting symptom.^{17,21}

In patients with AIDS, multicentric involvement is common, either occurring simultaneously or sequentially.²² Molecular analyses of such multicentric presentations have revealed distinct clonal origins for the lesions at each site, suggesting that these lesions represent multiple primary tumors rather than metastases.^{20,23-25} The tumors can affect various organs, including the central nervous system, spleen, lungs, gastrointestinal tract, larynx, pharynx, skin, adrenal glands, and eyes. The liver is the most commonly involved organ in immunosuppressed organ transplant recipients, the second most

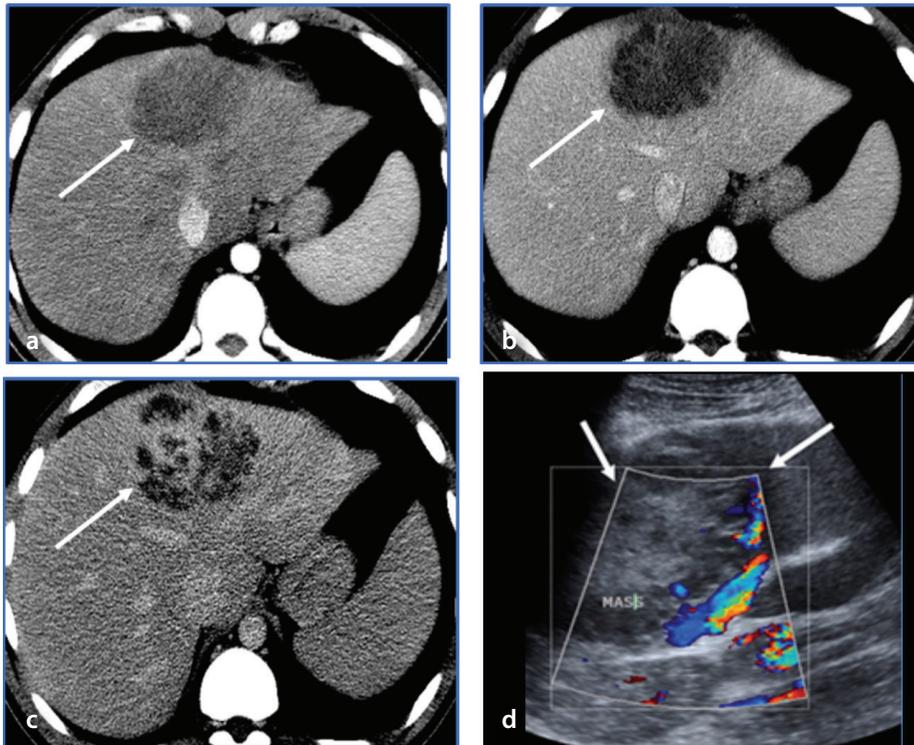


Figure 3. Axial contrast-enhanced computed tomography scans of a myofibroblastoma taken during the arterial (a), portal venous (b), and delayed (c) phases with a mass in the left hepatic lobe (white arrows), demonstrating poor enhancement during the arterial and portal phases and heterogeneous internal delayed enhancement during the delayed phase. Color Doppler ultrasound images (d) showing a multilobulated mass with an echogenic center, hypoechoic periphery, and minimal internal blood flow.

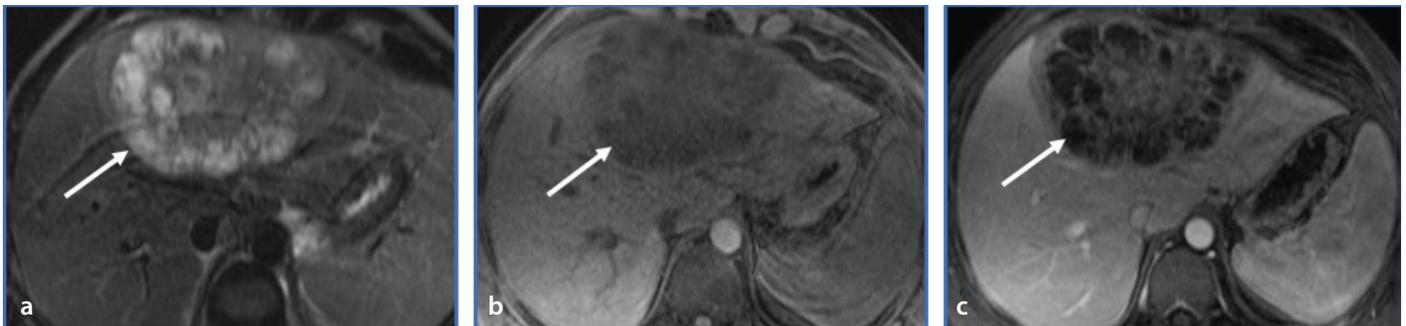


Figure 4. Additional case of a biopsy-verified myofibroblastoma. Axial T2-weighted image (a), pre-contrast T1-weighted image (b), and post-gadolinium T1-weighted image with fat suppression (c) revealing a multilobulated solid and cystic mass (white arrows) with numerous small T2-bright peripheral components and a more solid central portion that is mildly hyperintense on the T2-weighted image and mildly hypointense on the T1-weighted images. With contrast administration, septal enhancement and progressive delayed enhancement of the central component of the mass were identified.

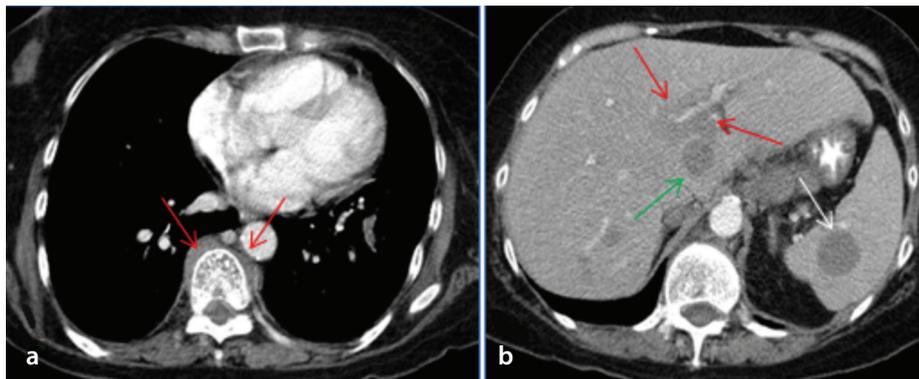


Figure 5. Contrast-enhanced computed tomography (CECT) images revealing extramedullary hematopoiesis in a 61-year-old woman with myelodysplastic syndrome. Axial CECT image (a) at the level of the lower chest showing paravertebral masses (red arrows). Axial CECT image (b) through the upper abdomen showing hepatic (green arrow) and splenic (white arrow) rounded hypoattenuating masses and periportal poorly enhancing tissue (red arrows).

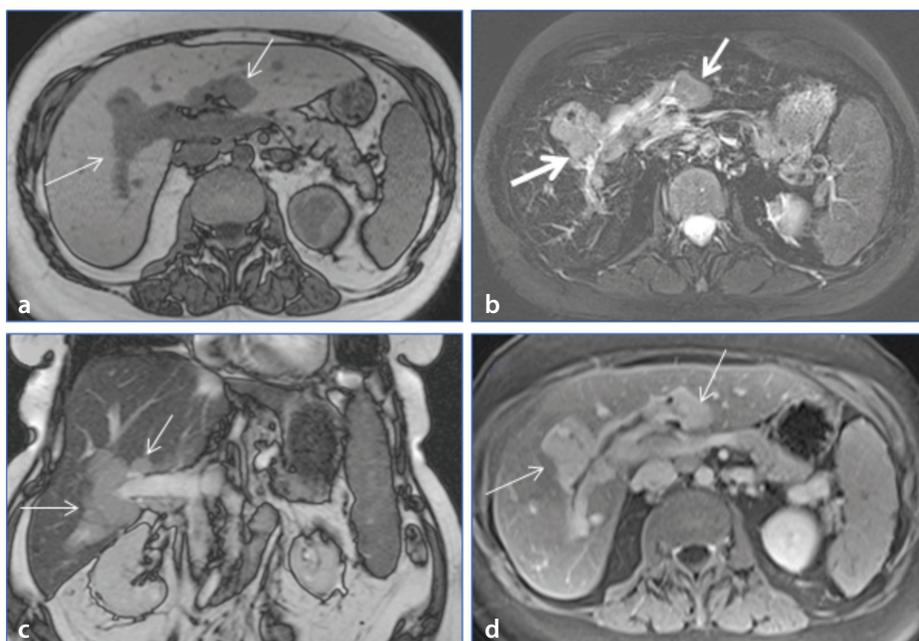


Figure 6. Axial T1-weighted image (a), axial T2-weighted image with fat suppression (b), coronal T2-weighted image (c), and axial T1-weighted image with fat suppression in the late arterial phase (d) in a 56-year-old woman with extramedullary hematopoiesis and myelofibrosis showing periportal masses (white arrows), demonstrating low signal intensity on the T1-weighted images, high signal intensity in the T2-weighted images, and progressive enhancement after contrast administration.

common site in patients with AIDS, and the least common site in patients with congenital immunodeficiencies.^{21,25-31}

The imaging characteristics of EBV-associated smooth muscle tumors are generally non-specific. On CT, the tumors present as hypodense lesions, often with rim enhancement^{22,32} and sometimes with central necrosis.^{21,27} Findings on MRI typically include isointensity on T1-weighted images and isointensity to mild hyperintensity on T2-weighted images, with substantial contrast enhancement (Figure 7).³³ Diagnosis is confirmed through immunohistochemistry, which shows positivity for smooth muscle actin and EBV-encoded RNA.²²

Differentiating EBV-associated smooth muscle tumors from non-EBV-related primary or metastatic smooth muscle tumors of the liver, such as leiomyosarcomas, is crucial because of the latter's less favorable prognosis. Even without intervention, EBV-associated tumors tend to progress slowly.^{17,22}

Hepatobiliary tuberculosis and tuberculomas

Hepatobiliary tuberculosis and tuberculomas of the liver most commonly occur in patients with disseminated disease. In an autopsy series, a hepatosplenic prevalence of 80% to 100% was seen in the setting of pulmonary miliary disease.^{34,35} There are

three types of hepatobiliary tuberculosis: parenchymal (including the miliary, nodular, and mixed subtypes), biliary, and serohepatic.³⁶ Imaging plays a key role in the diagnosis and management of these patients because the symptoms of patients with these lesions are often vague and non-specific, potentially delaying appropriate treatment. Most cases ultimately require tissue sampling to confirm the diagnosis.

In parenchymal hepatobiliary tuberculosis, imaging findings typically identify multiple (<2 cm) well-defined nodules throughout the liver that may contain calcifications (Figure 8). Larger lesions (>2 cm) in the less commonly seen macronodular type of hepatobiliary tuberculosis tend to demonstrate peripheral rim enhancement and central necrosis on CT, and they may contain chunky peripheral or central calcifications (Figure 9).³⁷ However, their appearance is dependent on the degree of caseous necrosis and liquefaction. On CT imaging, non-caseating granulomas appear hypoattenuating with no or mild peripheral rim enhancement.³⁸ However, lesions with caseous necrosis and liquefaction can resemble pyogenic abscesses with a honeycomb appearance, multiple enhancing septations, and regions of scattered necrosis.³⁸ On MRI, macronodular lesions typically appear as T1 hypointense. On T2-weighted images, they have varied intensity (hypointense to hyperintense) with peripheral T2-hypointense rims, and on postcontrast sequences, they have heterogeneous enhancement.^{36,39} Larger macronodular nodules may represent the fusion of miliary and/or micronodular nodules. Mixed parenchymal lesions have imaging findings of both the miliary and nodular types. Sonographically, these lesions usually appear uniformly hypoechoic or heterogeneous, with intermixed echogenic and hypoechoic regions.⁴⁰ However, lesions with other patterns have been observed, including -rarely- hypoechoic to echogenic lesions with echogenic centers.⁴⁰ The lesion borders tend to be poorly defined, but this too varies. The imaging characteristics of these lesions can be non-specific and can vary depending on the stage of the hepatic granuloma. Therefore, tissue sampling is often required to confirm the diagnosis.

Tuberculous cholangitis is a rare presentation of hepatobiliary tuberculosis and primarily occurs in children.³⁸ Imaging findings consist of dilated irregular intrahepatic

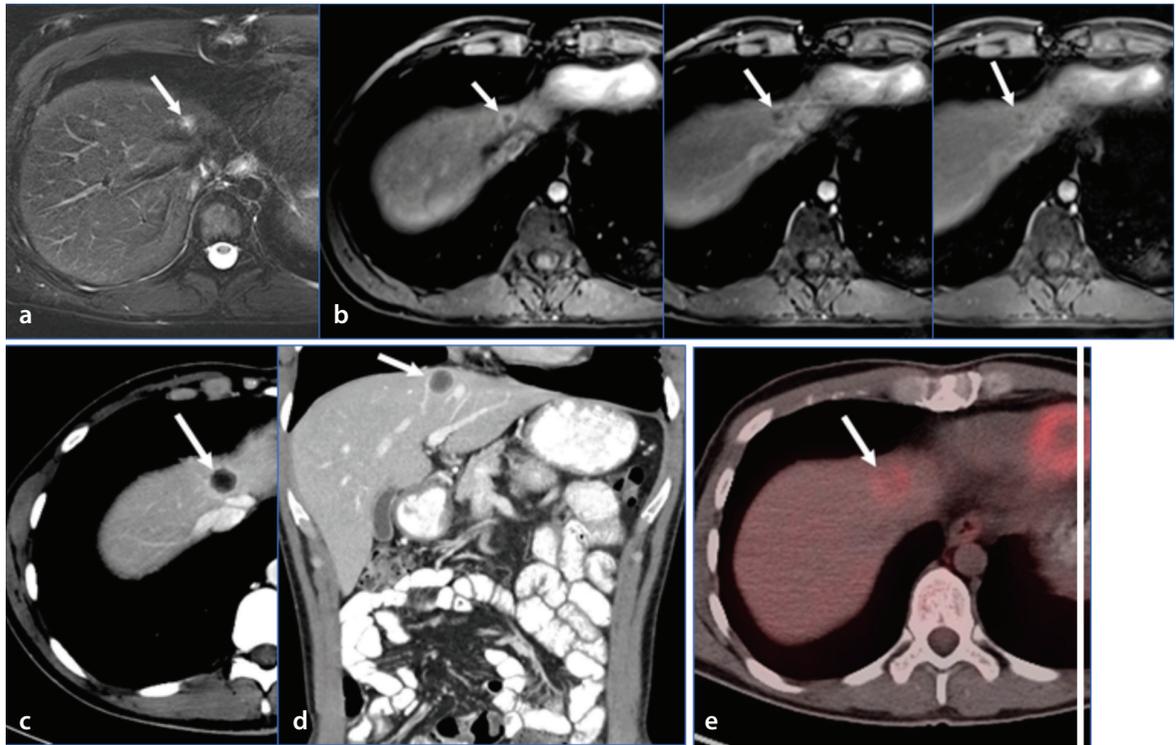


Figure 7. Images from a 26-year-old woman with stage IVb monomorphic posttransplant lymphoproliferative disorder involving the bones and lymph nodes. The patient presented with a slowly growing liver mass identified as an Epstein–Barr virus-associated smooth muscle tumor. Axial T2-weighted images (a) and three dynamic gadolinium-enhanced T1-weighted images with fat suppression (b) showing a small T2-hyperintense lesion in the left hepatic lobe (arrows). The lesion exhibits early peripheral enhancement and some delayed fill-in of contrast material with persistent peripheral enhancement. Axial (c) and coronal (d) contrast-enhanced computed tomography images obtained 15 months later showing a mild increase in lesion size (arrows). Peripheral enhancement was observed, and the center of the lesion had an attenuation of approximately 15 HU. An axial positron emission tomography/computed tomography scan (e) showing the lesion's increased metabolic activity (standardized uptake value: 3.87). HU, Hounsfield units

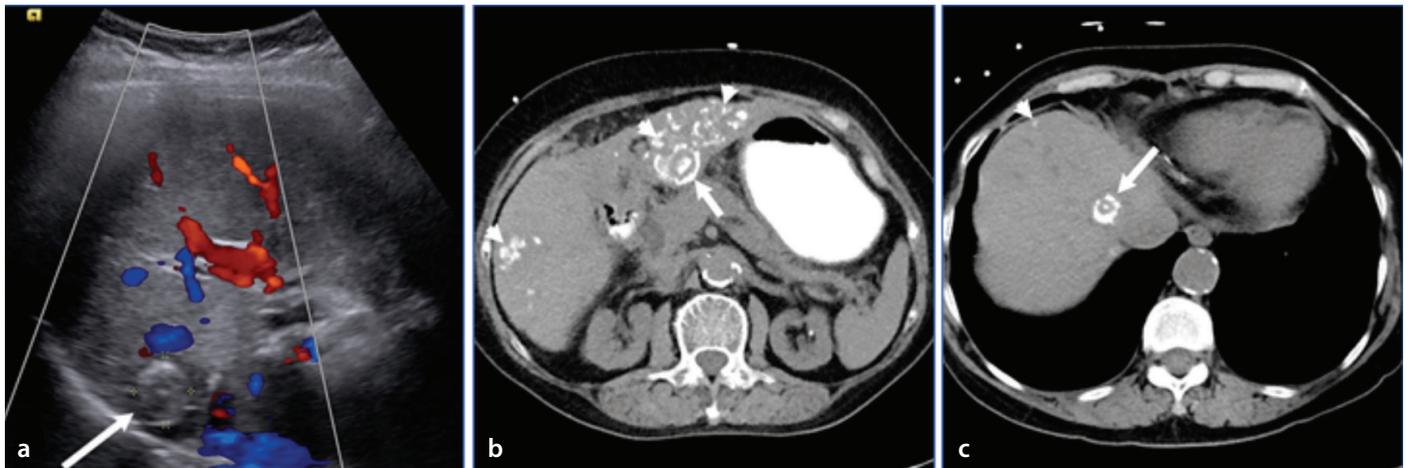


Figure 8. Transabdominal ultrasound scan (a) showing a heterogeneously echogenic lesion (arrow) with a peripheral rim of increased echogenicity. Axial unenhanced computed tomography images (b, c) showing multiple calcified lesions (arrows). Some of the lesions have a target-like appearance, with central and peripheral calcifications separated by a zone of soft-tissue attenuation. This was a biopsy-verified tuberculoma.

ducts or diffuse miliary calcifications along the affected bile ducts.⁴¹ Involvement of the biliary tree may be caused by biliary tract contamination from enteric mycobacterium infection, hematogenous spread, or direct extension from adjacent infected structures (e.g., caseating granuloma or hilar lymphadenopathy).⁴⁰

The serohepatic type of hepatobiliary tuberculosis is the rarest form. Imaging findings consist of multiple hypoattenuating nodules in a subcapsular distribution with a thickened hepatic capsule.³⁶ These findings give the liver a characteristic appearance of subcapsular nodularity, which has been characterized as having “sugar coating” or a “frosted liver” appearance.⁴¹

The diagnosis of serohepatic hepatobiliary tuberculosis should be considered among high-risk patients or in those who have suspected tuberculous involvement of the lungs, spleen, or lymph nodes. A definitive diagnosis is made with acid-fast staining and with histology revealing caseating granulomas.⁴²

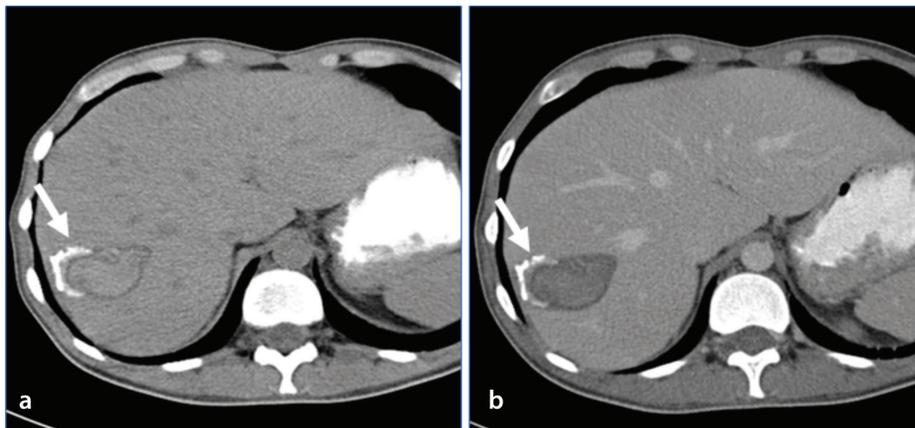


Figure 9. Biopsy-verified tuberculoma with a history of AIDS and disseminated tuberculosis. Axial unenhanced (a) and contrast-enhanced (b) computed tomography images demonstrating a lesion in the right hepatic lobe (arrows). The unenhanced image shows central high attenuation, a peripheral rim of low attenuation, and an incomplete ring of calcification. No enhancement was identified after contrast administration.

Peliosis hepatis

Peliosis hepatitis is an uncommon lesion characterized by abnormal sinusoidal dilatation and multiple blood-filled lacunar spaces.⁴³ It is thought to be caused by hepatic flow obstruction at the sinusoidal level and is potentially caused by steroids, oral contraceptives, cytotoxic drugs, chronic lung disease, infections (e.g., HIV and tuberculosis), and various malignancies (especially hepatomas).⁴⁴ In HIV-related peliosis hepatitis, *Bartonella henselae* has been identified as a causative agent, with regression after appropriate antibiotic therapy.⁴⁵ Peliosis hepatitis can also develop in patients with a renal or cardiac transplant. Hematologic disorders, diabetes, and necrotizing vasculitis also seem to be associated with peliosis hepatitis. However, up to 50% of peliosis hepatitis cases are idiopathic.⁴⁴ These lesions are normally asymptomatic, but they may cause symptoms if they rupture and hemorrhage. Additional complications include portal hypertension, cholestasis, liver enlargement, ascites, and even liver failure.

Two different pathologic forms of peliosis hepatitis have been described: the phlebotatic and parenchymal types. The phlebotatic type is characterized by blood-filled cavities lined with endothelial cells and an aneurysmal dilatation of the central vein. By contrast, in the parenchymal type, the blood-filled cavities are not lined by endothelial cells, and it occurs in the setting of hemorrhagic parenchymal necrosis.^{46,47}

Lesions typically measure up to a few centimeters and usually demonstrate no mass effect on transiting vessels. Imaging features vary according to the age of the blood

components and the presence or absence of hepatic steatosis. On ultrasound, these lesions tend to be hypoechoic in the setting of background steatosis and hyperechoic in the setting of normal liver tissue with increased perinodular or intranodular vascular flow on Doppler imaging.

On multiphase CT and MRI, peliosis hepatitis lesions exhibit intralesional hemorrhage and predominantly progressive centripetal enhancement, although centripetal enhancement is also possible (Figure 10). The appearance of these lesions can vary based on the age of the hemorrhagic components.⁴⁸ Cavities within these lesions that communicate with sinusoids demonstrate enhancement in line with that of blood vessels, whereas thrombosed portions are not enhanced. These lesions usually demonstrate early discontinuous (globular) arterial enhancement. In atypical cases where the enhancement is centripetal rather than classic centrifugal, the lesions may be mistaken for hemangiomas.⁴⁸

Sarcoidosis

Sarcoidosis is a multisystemic inflammatory disease that is histologically defined by the formation of non-necrotic granulomas. The cause of sarcoidosis remains unclear, but currently accepted theories suggest that a genetic predisposition and exposure to environmental or occupational antigens lead to a dysfunctional immune response.⁴⁹ The disease mainly affects young and middle-aged adults, and there is a slight predominance in women compared with men.⁴⁹ Although all populations worldwide are affected, some studies report increased rates among African Americans and people from the Nordic

countries.⁵⁰

The lungs are the most common site of involvement; however, every organ system can be affected. Hepatic disease is reported in up to 80% of cases.^{50,51} The formation of hepatic granulomas can incite an inflammatory response leading to fibrosis and, eventually, cirrhosis.⁵⁰ Most patients with hepatic sarcoidosis (50%–80%) are asymptomatic. Symptoms include fatigue, fever, weight loss, pruritus, jaundice, and abdominal pain.⁵⁰

Imaging findings in sarcoid-related liver disease are often non-specific and vary substantially. In most patients with sarcoidosis, the liver appears normal on imaging. The most common finding, reported in up to 29% of cases, is hepatomegaly.^{51,52} Other abnormalities include portal hypertension, portal vein or hepatic vein thrombosis, and cirrhosis.^{51,53}

Focal hepatic lesions are also a feature in a small percentage of patients with sarcoidosis. They often appear as multiple small lesions, which can coalesce and form larger lesions.^{51,52} On non-contrast CT, these lesions are usually hypodense (Figure 11), whereas on MRI, the lesions are typically T1 isointense or slightly hypointense and T2 hypointense. However, increases in lesion size, the coalescence of larger granulomas, or active inflammation can result in hyperintense T2-weighted signal intensity.⁵⁴ On both imaging modalities, lesions are hypoenhancing relative to the surrounding parenchyma (Figure 12). Concomitant lesions can also be seen within the spleen, and portal lymphadenopathy may also be present. Both of these findings can aid in diagnosis.⁵⁵

Cholestasis is another feature of hepatic sarcoidosis, and it can result in intrahepatic or extrahepatic biliary ductal dilatation.^{52,53} The latter could also be secondary to disease involving the common hepatic duct or extrinsic compression from enlarged lymph nodes. The appearance may mimic that of a primary biliary tumor, necessitating caution during imaging interpretation.⁵²

Amyloidosis

Amyloidosis is a general term referring to a group of disorders characterized by the deposition of misfolded serum proteins—amyloid fibrils—in extracellular spaces.⁵⁶ Amyloid fibrils result when precursor proteins that are usually soluble aggregate, forming insoluble fibers that are resistant to degradation.^{56–58} In localized amyloidosis, the formation and

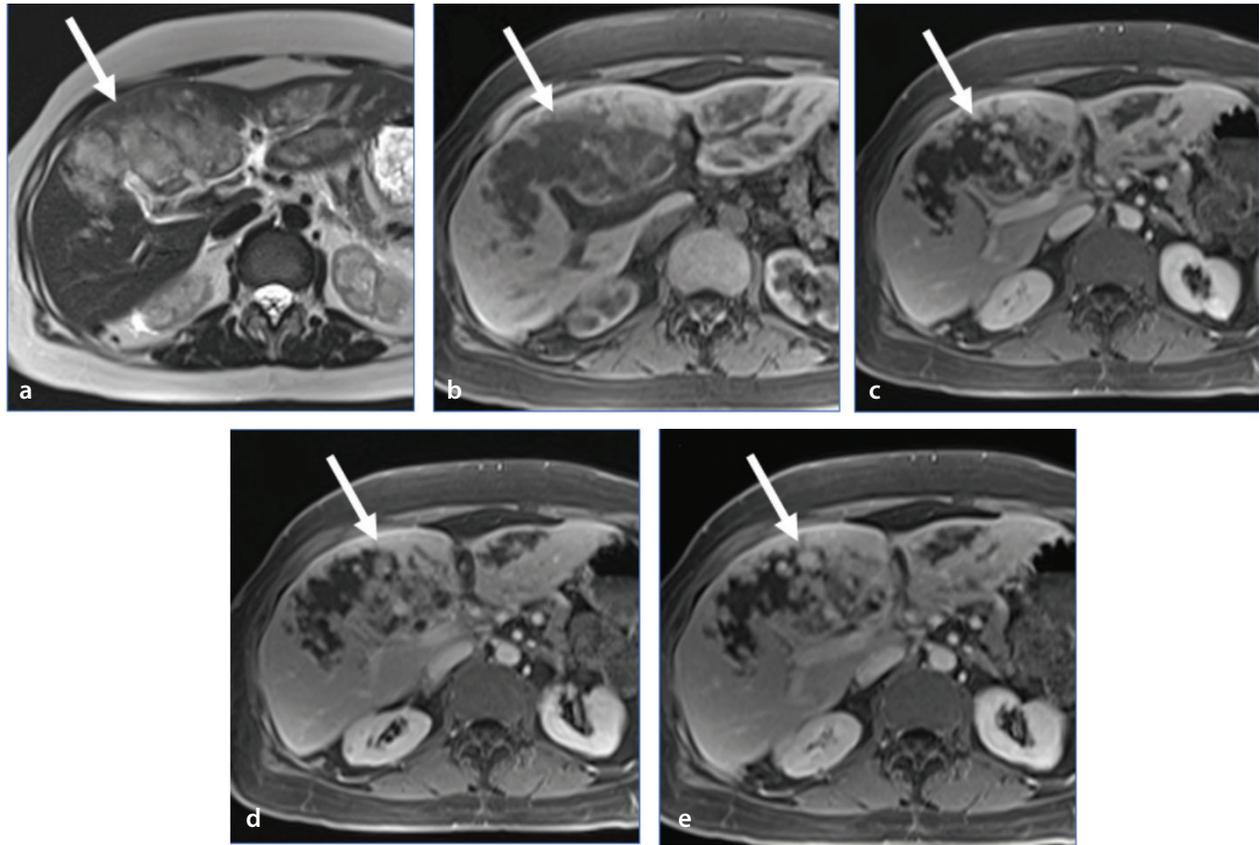


Figure 10. Axial T2-weighted images (a) and dynamic post-gadolinium T1-weighted images with fat suppression (b-e) showing a large hepatic mass (long arrows) spanning both lobes of the liver and demonstrating high T2 signal intensity, low T1 signal intensity, and progressive centripetal enhancement following contrast administration. The lesion represents a biopsy-verified peliosis hepatis that was not present 3 years earlier. Despite its large size, no appreciable mass effect was identified. Vessels are observed coursing through the lesion without significant attenuation.

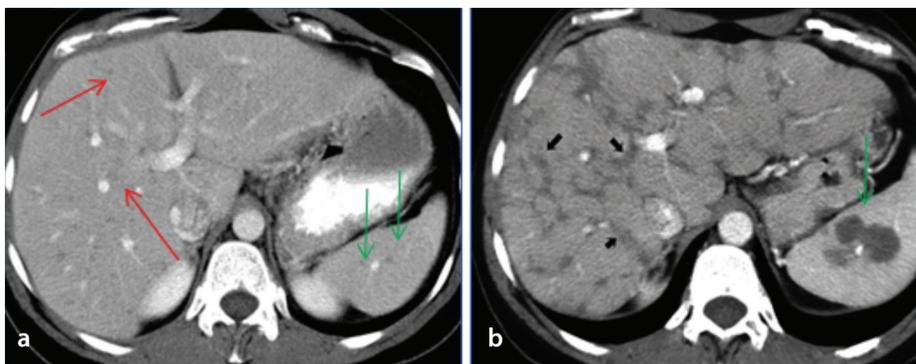


Figure 11. Contrast-enhanced computed tomography (CECT) images from a patient with a long history of pulmonary sarcoidosis. (a) Initial CECT image showing multiple subtle hypodense hepatic (red arrows) and splenic (green arrows) lesions. CECT image from the same patient obtained 5 years after the initial scan (b) showing extensive periportal and bridging non-enhancing fibrous septa throughout the liver (black arrows) with periportal predominance. Note the interval increase in the multiple hypoattenuating splenic lesions (green arrow).

deposition of amyloid fibrils occurs in the same organ. In systemic amyloidosis, the formation of the amyloid fibrils occurs in one organ, and the deposition of the fibrils occurs in a distant location.⁵⁸ Systemic amyloidosis can be classified into multiple types, depending on the precursor protein; over 15 types of precursor proteins have been described.⁵⁸ Common conditions associated with the sys-

temic form of the disease include plasma cell disorders, malignancy, and chronic infection or inflammation.⁵⁶⁻⁵⁸

Amyloidosis can affect all organ systems. In the abdomen, amyloid deposition occurs in the genitourinary and gastrointestinal tracts as well as in the liver, spleen, peritoneum, and retroperitoneum.⁵⁶

Within the liver, amyloid fibril deposition occurs in the perisinusoidal spaces and along the blood vessel walls.^{59,60} As with sarcoidosis, imaging findings in the liver can be non-specific, and amyloidosis most commonly manifests as hepatomegaly.⁶¹ Some reported patterns that can mimic other infiltrative liver diseases, such as steatosis and hepatic venous congestion, include diffusely decreased attenuation on CT, heterogeneity of liver parenchyma, and heterogeneous contrast enhancement.^{59,61} More well-defined masses, along with focal and diffuse calcifications, can also occur (Figure 13).⁶¹

On MRI, diffusely decreased signal intensity on T2-weighted images has been observed, possibly from amyloid deposition and/or increased iron deposition in the liver occurring secondary to amyloid-associated chronic renal disease.^{60,61} Increased liver stiffness has also been observed and can be evaluated using magnetic resonance elastography.^{59,62}

In conclusion, a large variety of pathologies can present within the liver, and this can make diagnosing liver lesions challenging. Although most lesions are common tumors,

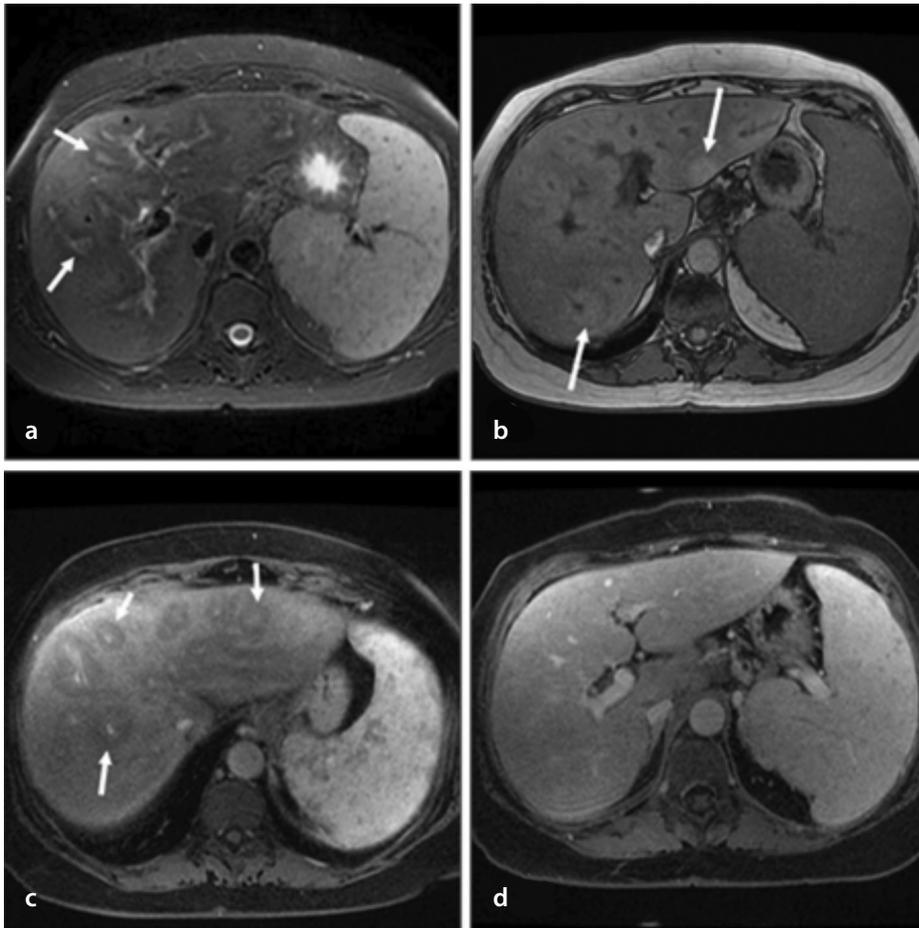


Figure 12. Axial T2-weighted image with fat suppression (a), axial T1-weighted opposed-phase image (b), axial T1-weighted late-arterial-phase image with fat suppression (c), and axial T1-weighted portal-venous-phase image with fat suppression (d) in a patient with sarcoidosis showing periportal masses (white arrows), demonstrating slightly increased signal intensity and poor enhancement during the late arterial phase. The areas of abnormal signaling were not identified during the portal venous phase. There is also evidence of splenic involvement with splenomegaly and multiple small granulomas.

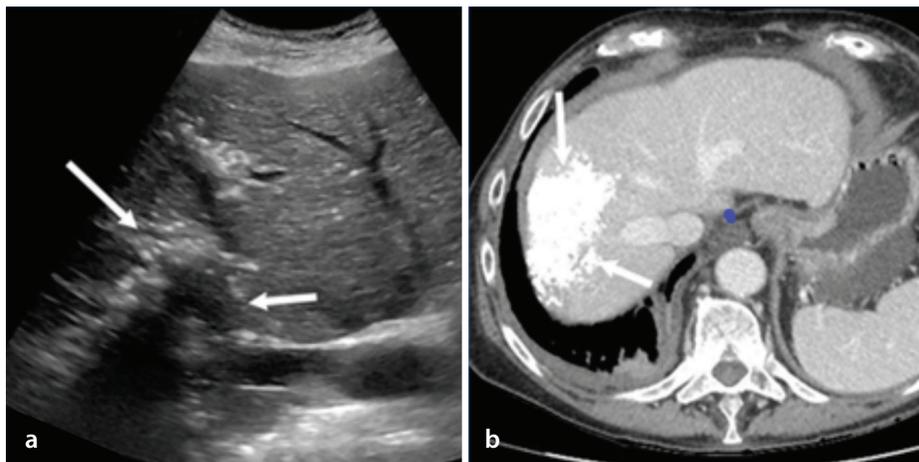


Figure 13. Transabdominal ultrasound scan (a) showing an echogenic mass-like structure (white arrows) with posterior shadowing caused by calcifications. Contrast-enhanced computed tomography scan (b) showing a mass-like structure (white arrows) within the right hepatic lobe related to extensive calcifications. The patient was diagnosed with hepatic amyloidosis.

on rare occasions, radiologists may encounter one of the aforementioned rare lesions. By being familiar with these lesions and including them in the differential diagnosis, radiologists and clinicians can ensure that patients receive the correct diagnosis and treatment.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Stringer MD, Alizai NK. Mesenchymal hamartoma of the liver: a systematic review. *J Pediatr Surg.* 2005;40(11):1681-1690. [\[CrossRef\]](#)
2. Assarzagdegan N, Montgomery E. Uncommon benign neoplasms and pseudotumors of the liver. *Arch Pathol Lab Med.* 2023;147(4):390-402. [\[CrossRef\]](#)
3. Kim SH, Kim WS, Cheon JE, et al. Radiological spectrum of hepatic mesenchymal hamartoma in children. *Korean J Radiol.* 2007;8(6):498-505. [\[CrossRef\]](#)
4. Papke DJ Jr. Mesenchymal neoplasms of the liver. *Surg Pathol Clin.* 2023;16(3):609-634. [\[CrossRef\]](#)
5. Wang D, Misdraji J. Inflammatory pseudotumor of the liver. *Surg Pathol Clin.* 2023;16(3):565-580. [\[CrossRef\]](#)
6. Kim KA, Kim KW, Park SH, et al. Unusual mesenchymal liver tumors in adults: radiologic-pathologic correlation. *AJR Am J Roentgenol.* 2006;187(5):481-489. [\[CrossRef\]](#)
7. Roberts AS, Shetty AS, Mellnick VM, Pickhardt PJ, Bhalla S, Menias CO. Extramedullary haematopoiesis: radiological imaging features. *Clin Radiol.* 2016;71(9):807-814. [\[CrossRef\]](#)
8. Sohawon D, Lau KK, Lau T, Bowden DK. Extramedullary haematopoiesis: a pictorial review of its typical and atypical locations. *J Med Imaging Radiat Oncol.* 2012;56(5):538-544. [\[CrossRef\]](#)
9. La Fianza A, van der Byl G, Maccabelli G, Torretta L, Calliada F. CT and MR findings in extramedullary haematopoiesis with biliary system encasement: a case report. *J Radiol Case Rep.* 2010;4(11):1-8. [\[CrossRef\]](#)
10. Jelali MA, Luciani A, Kobeiter H, et al. MRI features of intrahepatic extramedullary haematopoiesis in sickle cell anaemia. *Cancer Imaging.* 2006;6(1):182-185. [\[CrossRef\]](#)
11. Navarro M, Crespo C, Pérez L, Martínez C, Galant J, González I. Massive intrahepatic extramedullary hematopoiesis in myelofibrosis. *Abdom Imaging.* 2000;25(2):184-186. [\[CrossRef\]](#)
12. Malla S, Razik A, Das CJ, Naranje P, Kandasamy D, Kumar R. Marrow outside marrow: imaging of extramedullary haematopoiesis. *Clin Radiol.* 2020;75(8):565-578. [\[CrossRef\]](#)

13. Kwak HS, Lee JM. CT findings of extramedullary hematopoiesis in the thorax, liver and kidneys, in a patient with idiopathic myelofibrosis. *J Korean Med Sci.* 2000;15(4):460-462. [\[CrossRef\]](#)
14. Gupta P, Naran A, Auh YH, Chung JS. Focal intrahepatic extramedullary hematopoiesis presenting as fatty lesions. *AJR Am J Roentgenol.* 2004;182:1031-1032. [\[CrossRef\]](#)
15. Wong Y, Chen F, Tai KS, et al. Imaging features of focal intrahepatic extramedullary haematopoiesis. *Br J Radiol.* 1999;72(861):906-910. [\[CrossRef\]](#)
16. Kapatia G, Kaur A, Rastogi P, et al. Extramedullary hematopoiesis: clinical and cytological features. *Diagn Cytopathol.* 2020;48(3):191-196. [\[CrossRef\]](#)
17. Dekate J, Chetty R. Epstein-Barr virus-associated smooth muscle tumor. *Arch Pathol Lab Med.* 2016;140(7):718-722. [\[CrossRef\]](#)
18. Pritzker KP, Huang SN, Marshall KG. Malignant tumours following immunosuppressive therapy. *Can Med Assoc J.* 1970;103(13):1362-1365. [\[CrossRef\]](#)
19. Chadwick EG, Connor EJ, Hanson IC, et al. Tumors of smooth-muscle origin in HIV-infected children. *JAMA.* 1990;263(23):3182-3184. [\[CrossRef\]](#)
20. Lee ES, Locker J, Nalesnik M, et al. The association of Epstein-Barr virus with smooth-muscle tumors occurring after organ transplantation. *N Engl J Med.* 1995;332(1):19-25. [\[CrossRef\]](#)
21. Hussein K, Rath B, Ludewig B, Kreipe H, Jonigk D. Clinico-pathological characteristics of different types of immunodeficiency-associated smooth muscle tumours. *Eur J Cancer.* 2014;50(14):2417-2424. [\[CrossRef\]](#)
22. Gallien S, Zuber B, Polivka M, et al. Multifocal Epstein-Barr virus-associated smooth muscle tumor in adults with AIDS: case report and review of the literature. *Oncology.* 2008;74(3-4):167-176. [\[CrossRef\]](#)
23. Deyrup AT, Lee VK, Hill CE, et al. Epstein-Barr virus-associated smooth muscle tumors are distinctive mesenchymal tumors reflecting multiple infection events: a clinicopathologic and molecular analysis of 29 tumors from 19 patients. *Am J Surg Pathol.* 2006;30(1):75-82. [\[CrossRef\]](#)
24. McClain KL, Leach CT, Jenson HB, et al. Association of Epstein-Barr virus with leiomyosarcomas in children with AIDS. *N Engl J Med.* 1995;332(1):12-18. [\[CrossRef\]](#)
25. Jenson HB, Leach CT, McClain KL, et al. Benign and malignant smooth muscle tumors containing Epstein-Barr virus in children with AIDS. *Leuk Lymphoma.* 1997;27:303-314. [\[CrossRef\]](#)
26. Suankratay C, Shuangshoti S, Mutirangura A, et al. Epstein-Barr virus infection-associated smooth-muscle tumors in patients with AIDS. *Clin Infect Dis.* 2005;40(10):1521-1528. [\[CrossRef\]](#)
27. Jossen J, Chu J, Hotchkiss H, et al. Epstein-Barr virus-associated smooth muscle tumors in children following solid organ transplantation: a review. *Pediatr Transplant.* 2015;19(2):235-243. [\[CrossRef\]](#)
28. Yu L, Aldave AJ, Glasgow BJ. Epstein-Barr virus-associated smooth muscle tumor of the iris in a patient with transplant: a case report and review of the literature. *Arch Pathol Lab Med.* 2009;133(8):1238-1241. [\[CrossRef\]](#)
29. Reyes C, Abuzaitoun O, De Jong A, Hanson C, Langston C. Epstein-Barr virus-associated smooth muscle tumors in ataxia-telangiectasia: a case report and review. *Hum Pathol.* 2002;33(1):133-136. [\[CrossRef\]](#)
30. Shaw RK, Issekutz AC, Fraser R, et al. Bilateral adrenal EBV-associated smooth muscle tumors in a child with a natural killer cell deficiency. *Blood.* 2012;119(17):4009-4012. [\[CrossRef\]](#)
31. Tulbah A, Al-Dayel F, Fawaz I, Rosai J. Epstein-Barr virus-associated leiomyosarcoma of the thyroid in a child with congenital immunodeficiency: a case report. *Am J Surg Pathol.* 1999;23(4):473-476. [\[CrossRef\]](#)
32. Conrad A, Brunet AS, Hervieu V, et al. Epstein-Barr virus-associated smooth muscle tumors in a composite tissue allograft and a pediatric liver transplant recipient. *Transpl Infect Dis.* 2013;15(5):182-186. [\[CrossRef\]](#)
33. Lohan R, Bathla G, Gupta S, Hegde AN. Epstein-Barr virus (EBV)-related smooth muscle tumors of central nervous system—a report of two cases and review of literature. *Clin Imaging.* 2013;37(3):564-568. [\[CrossRef\]](#)
34. Kumar V, Pandey D. Isolated hepatosplenic tuberculosis. *Hepatobiliary Pancreat Dis Int.* 2008;7(3):328-330. [\[CrossRef\]](#)
35. Thoeni RF, Margulis AR. Gastrointestinal tuberculosis. *Semin Roentgenol.* 1979;14(4):283-294. [\[CrossRef\]](#)
36. Yu RS, Zhang SZ, Wu JJ, Li RF. Imaging diagnosis of 12 patients with hepatic tuberculosis. *World J Gastroenterol.* 2004;10(11):1639-1642. [\[CrossRef\]](#)
37. Leder RA, Low VH. Tuberculosis of the abdomen. *Radiol Clin North Am.* 1995;33(4):691-705. [\[CrossRef\]](#)
38. Joshi AR, Basantani AS, Patel TC. Role of CT and MRI in abdominal tuberculosis. *Curr Radiol Rep.* 2014;2:66. [\[CrossRef\]](#)
39. Levine C. Primary macronodular hepatic tuberculosis: US and CT appearances. *Gastrointest Radiol.* 1990;15:307-309. [\[CrossRef\]](#)
40. Chong VH, Lim KS. Hepatobiliary tuberculosis. *Singapore Med J.* 2010;51(9):744-751. [\[CrossRef\]](#)
41. Kakkar C, Polnaya AM, Koteswara P, Smiti S, Rajagopal KV, Arora A. Hepatic tuberculosis: a multimodality imaging review. *Insights Imaging.* 2015;6(6):647-658. [\[CrossRef\]](#)
42. Karaosmanoglu AD, Onur MR, Sahani DV, Tabari A, Karcaaltincaba M. Hepatobiliary tuberculosis: imaging findings. *AJR Am J Roentgenol.* 2016;207(4):694-704. [\[CrossRef\]](#)
43. Zak FG. Peliosis hepatis. *Am J Pathol.* 1950;26(1):1-15. [\[CrossRef\]](#)
44. Wanless IR. Vascular disorders. In: MacSweenRNM, Burt AD, Portmann BC, Ishak KG, Scheuer PJ, Anthony PP, eds. *Pathology of the Liver*, 4th ed. Glasgow, UK: Churchill Living-stone; 2002:553-555. [\[CrossRef\]](#)
45. Koehler JE. Bartonella-associated infections in HIV-infected patients. *AIDS Clin Care.* 1995;(12):97-102. PMID: 11362939. [\[CrossRef\]](#)
46. Vignaux O, Legmann P, de Pinieux G, et al. Hemorrhagic necrosis due to peliosis hepatis: imaging findings and pathological correlation. *Eur Radiol.* 1999;9(3):454-456. [\[CrossRef\]](#)
47. Yanoff M, Rawson AJ. Peliosis hepatis. An anatomic study with demonstration of two varieties. *Arch Pathol.* 1964;77:159-165. [\[CrossRef\]](#)
48. Iannaccone R, Federle MP, Brancatelli G, et al. Peliosis hepatis: spectrum of imaging findings. *AJR Am J Roentgenol.* 2006;187(1):43-52. [\[CrossRef\]](#)
49. Drent M, Crouser ED, Grunewald J. Challenges of sarcoidosis and its management. *N Engl J Med.* 2021;385(11):1018-1032. [\[CrossRef\]](#)
50. Moser A, Cheung A. Hepatic sarcoidosis: a review of the diagnosis and management. *Curr Hepatol Rep.* 2024;23(9):137-144. [\[CrossRef\]](#)
51. Karaosmanoğlu AD, Onur MR, Saini S, Taberi A, Karcaaltincaba M. Imaging of hepatobiliary involvement in sarcoidosis. *Abdom Imaging.* 2015;40(8):3330-3337. [\[CrossRef\]](#)
52. Warshauer DM, Lee JK. Imaging manifestations of abdominal sarcoidosis. *AJR Am J Roentgenol.* 2004;182(1):15-28. [\[CrossRef\]](#)
53. Ferreira A, Ramalho M, de Campos RO, et al. Hepatic sarcoidosis: MR appearances in patients with chronic liver disease. *Magn Reson Imaging.* 2013;31(3):432-438. [\[CrossRef\]](#)
54. Jung G, Brill N, Poll LW, Koch JA, Wettstein M. MRI of hepatic sarcoidosis: large confluent lesions mimicking malignancy. *AJR Am J Roentgenol.* 2004;183(1):171-173. [\[CrossRef\]](#)
55. Koyama T, Ueda H, Togashi K, Umeoka S, Kataoka M, Nagai S. Radiologic manifestations of sarcoidosis in various organs. *Radiographics.* 2004;24(1):87-104. [\[CrossRef\]](#)
56. Krauß LU, Schmid S, Mester P, et al. Clinical, Endoscopic, and histopathologic observations in gastrointestinal amyloidosis. *J Gastrointest Liver Dis.* 2023;32(4):497-506. [\[CrossRef\]](#)
57. Rambaran RN, Serpell LC. Amyloid fibrils: abnormal protein assembly. *Prion.* 2008;2(3):112-117.
58. Muchtar E, Dispenzieri A, Magen H, et al. Systemic amyloidosis from A (AA) to T (ATTR): a

- review (Review). *J Intern Med.* 2021;289(3):268-292. [\[CrossRef\]](#)
59. Özcan HN, Haliloğlu M, Sökmensüer C, Akata D, Özmen M, Karçaaltınçaba M. Imaging for abdominal involvement in amyloidosis. *Diagn Interv Radiol.* 2017;23(4):282-285. [\[CrossRef\]](#)
60. Venkatesh SK, Hoodeshenas S, Venkatesh SH, et al. Magnetic resonance elastography of liver in light chain amyloidosis. *J Clin Med.* 2019;8(5):739. [\[CrossRef\]](#)
61. Kim SH, Han JK, Lee KH, et al. Abdominal amyloidosis: spectrum of radiological findings. *Clin Radiol.* 2003;58(8):610-620. [\[CrossRef\]](#)
62. Sugi MD, Kawashima A, Salomao MA, Bhalla S, Venkatesh SK, Pickhardt PJ. Amyloidosis: multisystem spectrum of disease with pathologic correlation. *Radiographics.* 2021;41(5):1454-1474. [\[CrossRef\]](#)



The role of multiparametric magnetic resonance imaging in the differentiation of low- and high-grade non-muscle invasive bladder cancer

Merve Nur Taşdemir¹
 Uluhan Eryürük¹
 Ural Oğuz²
 Birgül Tok³
 Serdar Aslan¹

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

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

³Giresun University Faculty of Medicine, Department of Pathology, Giresun, Türkiye

PURPOSE

To evaluate the diagnostic efficacy of apparent diffusion coefficient (ADC) measurements and semi-quantitative dynamic contrast enhancement (DCE) parameters in predicting the differentiation between low- and high-grade tumors in non-muscle invasive bladder cancers (NMIBC).

METHODS

Patients with NMIBC, who were histopathologically confirmed between August 2020 and July 2023, were analyzed by 2 radiologists with different levels of experience. DCE semi-quantitative parameters such as wash-in rate (WiR), wash-out ratio (WoR), time to peak (TTP), and peak enhancement (PE) were calculated. ADC measurements were performed using the three-region-of-interest (ADCT) and whole volume (ADCw) methods; ADCT ratio (ADCTr) and ADCw ratio (ADCwR) were also calculated. Receiver operating characteristic curve analysis was performed to demonstrate the cut-off values of ADCT, ADCw, ADCTr, and ADCwR to differentiate low- and high-grade tumors. The intra-class correlation coefficient was used to evaluate inter-reader agreement.

RESULTS

A total of 89 patients were included in this study. Of these patients, 48 had low-grade NMIBC, and 41 had high-grade NMIBC. There was no significant difference in mean WoR, WiR, TTP, and PE values between low- and high-grade NMIBC ($P > 0.05$). The ADCT, ADCw, ADCTr, and ADCwR values of high-grade NMIBC were significantly lower than those of low-grade NMIBC ($P < 0.001$). With cut-off values of 0.449 and 0.435, ADCTr had the best diagnostic value for both readers, showing better accuracy, sensitivity, specificity, and area under the curve (85.4%–83.1%, 87.5%–85.4%, 82.9%–80.4%, and 0.879–0.857, respectively, with confidence intervals). Additionally, ADCTr and ADCT showed acceptable diagnostic performance for both readers, with cut-off values of 0.439 and 0.431, respectively, for differentiating Ta- and T1-stages. The inter-reader agreement was almost perfect for ADC measurements.

CONCLUSION

While DCE semiquantitative parameters did not yield significant outcomes in distinguishing between low and high grades, ADCTr holds promise for enhancing patient management in NMIBC cases and stands as a potential preoperative radiological asset.

CLINICAL SIGNIFICANCE

Individuals diagnosed with NMIBC may require different treatment approaches; therefore, it is very important to distinguish between low- and high-grade cases preoperatively. The differentiation between the Ta- and T1-stages is recognized as crucial in patient treatment strategies. Furthermore, ADCTr shows promise for improving patient management in NMIBC cases.

KEYWORDS

Apparent diffusion coefficient reference, apparent diffusion coefficient ratio, non-muscle invasive bladder cancer, magnetic resonance imaging, diffusion-weighted imaging, apparent diffusion coefficient

Corresponding author: Merve Nur Taşdemir

E-mail: mervetsdmr@gmail.com

Received 29 August 2024; revision requested 02 October 2024; last revision received 10 October 2024; accepted 20 October 2024.



Epub: 25.11.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.243004

Bladder cancer ranks as the second most prevalent genitourinary malignancy, following prostate cancer, and accounts for over 500,000 new cases and 200,000 fatalities each year.¹ The majority of bladder cancers are urothelial cell carcinomas and are tissue-based, categorized into low- or high-grade tumors.² While ultrasound and computed tomography are commonly employed in the diagnosis of bladder cancer,³ magnetic resonance imaging (MRI) is regularly utilized in the local staging of bladder cancer due to its capability to evaluate muscle invasion.⁴

The most critical factor that affects the prognosis of bladder cancer is muscle invasion.⁵ Muscle-invasive bladder cancer (T2–T4) has a poor outcome and typically requires aggressive interventions such as cystectomy, systemic treatment, or a mix of both.² Non-muscle invasive bladder cancer (NMIBC) (Ta–T1) typically exhibits low-grade characteristics and demonstrates a non-aggressive demeanor.² Roughly 70% of tumors constitute NMIBC, with over 50% being Ta-stage tumors. Despite the majority of cases being identified at a non-muscle invasive stage, there is a substantial risk of disease progression and recurrence.^{6,7} Treatment approaches primarily concentrate on reducing local recurrence and impeding stage advancement, with the overarching objective of preserving and improving the patient's quality of life.²

According to the treatment guidelines of the American Cancer Society, intravesical chemotherapy is recommended for Ta-stage tumors in the presence of low-grade tumors, whereas intravesical bacillus calmette-guerin is recommended in the presence of high-grade tumors. For T1-stage tumors, cystectomy may be recommended in the presence of high-grade tumors if there are multiple tumors or if the tumor is large when first detected.⁸ Stöckle et al.⁹ highlighted the sig-

nificance of differentiating between Ta- and T1-stages in the distinction of treatment. According to this study, the prognosis of patients with T1-stage tumors who underwent late cystectomy is worse than that of patients with T2 tumors.⁹ Transurethral resection (TUR) is ineffective in managing lymphogenic micrometastases that initiate during the pT1-stage. According to Jakse et al.¹⁰, 50% of all patients with T1 carcinomas developed a muscle-infiltrating recurrence within 40 months after TUR. The crucial point for therapeutic outcomes appears to be the onset of invasive growth (i.e., lamina propria invasion). This means that even tumor stage T1 is too advanced to consider TUR as a reliable curative treatment.⁹ Individuals diagnosed with NMIBC may require different treatment approaches; therefore, it is crucial to preoperatively differentiate between low- and high-grade cases.⁴

Dynamic contrast-enhanced (DCE) MRI, also known as functional MRI, has been shown to offer insights into the characterization of tissue microvasculature and distinguish the tumor from adjacent tissues.¹¹ The efficacy of diffusion-weighted imaging (DWI) in predicting the histologic grade of bladder cancer has also been discussed in the literature. In these studies, apparent diffusion coefficient (ADC) values acquired from DWI have been proposed as being potentially valuable in facilitating differentiation.¹² However, the related studies were not focused exclusively on NMIBC and relied on very small sample sizes.

This study aims to examine the effectiveness of ADC values from DWI and the semiquantitative parameters obtained from DCE-MRI in distinguishing between low- and high-grade tumors in patients with NMIBC, as well as to assess the consistency among readers with varying levels of experience.

Methods

This retrospective investigation was approved by the institutional ethics board, and informed consent was relinquished (Giresun Training and Research Hospital/KAEK-217/23.10.2023/25). The study protocol aligned with the ethical standards of the 1975 Declaration of Helsinki.

Study group

Patients with NMIBC who were histopathologically confirmed between August 2020 and July 2023 were analyzed retrospectively.

Patients were incorporated into the study based on the following criteria:

1. MRI evaluation contained the required sequences.
2. MRI assessment was conducted within 2 weeks before TUR bladder or cystectomy.
3. Low- or high-grade urothelial carcinoma of the bladder was pathologically confirmed.

Patients were removed from the study based on the following criteria:

1. Patients with ADC images of low or invisible quality.
2. Patients with tumors measuring less than 1 cm.
3. Patients with other histopathologically confirmed types of bladder cancer.
4. Patients with hyperintense urine in the bladder lumen on the T1 sequence.¹³

Figure 1 shows the patient selection.

Image acquisition

A 1.5-T MRI system (Magnetom Aera, Siemens Medical Solutions, Erlangen, Germany) was used for MRI examinations. Ultrasonography was performed before the procedure to ensure that the patients had adequate bladder distension. Images were acquired in a supine position with a pelvic phased-array coil. T1-weighted images (T1-WI), axial, coronal, and sagittal fast spin-echo T2-WI, DCE images with three-dimensional high temporal resolution, and DWI with *b*-values of 0, 800, and 1200 s/mm² were acquired. An ADC map was generated using a *b*-value of 1200 s/mm². Gadopentetate dimeglumine (Gadovist, 0.2 mL per kilogram of body weight; Bayer Healthcare, Berlin, Germany) was delivered via a power injector at a rate of 2 mL per second, followed by a further infusion of 20 mL of normal saline. Following the injection of the intravenous contrast agent, axial DCE images were captured in post-contrast phases with no gap between them.

Image analysis

Each MRI scan was uploaded to the picture archiving communication system. Two radiologists with varying levels of expertise (reader 1: a board-certified radiologist with 11 years of urogenital radiology experience; reader 2: a radiology resident with 3 years of training) assessed the images separately from histopathology. The readers maintained a blinded approach and had no access to the patients' demographic or surgical data.

Main points

- It is crucial to preoperatively differentiate between low- and high-grade cases in non-muscle invasive bladder cancer (NMIBC).
- Apparent diffusion coefficient (ADC) measurements of high-grade NMIBC were significantly lower than those of low-grade NMIBC.
- The ADC three region-of-interest ratio is a promising avenue for optimizing NMIBC treatment and a potential preoperative radiological aid.

Measurements were performed on the slice showing the largest diameter of the lesions and the most contrast enhancement, with minimal artifacts. In patients with multiple tumors, measurements were made for the tumor with the maximum diameter. To ensure the accuracy of the ADC values, lesions with a diameter of less than 1 cm and

areas containing artifacts were excluded. The 3 regions of interest (ROIs), each 20 mm², were drawn in distinct regions of the tumor. Next, the average ADC was calculated for the three-ROIs method (ADC_t). Freehand ROIs along the low signal of the tumor's border on ADC maps were applied in the whole-volume ROIs technique (ADC_w).¹⁴ The ROIs were

placed while avoiding blood vessels, necrosis, and tumor stalk. ADC measurements with different ROI methods are shown in Figure 2. The most appropriate ADC reference value for calculating the ADC ratio was obtained from the bladder lumen. The ROI was placed in the center of the bladder lumen urine while avoiding artifacts.¹³ Patients with hyperintense bladder contents on the T1 sequence were excluded. Three 20 mm² ROIs were placed in the center of the bladder, and the average ADC was calculated for reference ADC. Additionally, ADC_t ratio (ADC_tR) was calculated as the ADC (three-ROIs method)/ADC reference, and ADC_w ratio (ADC_wR) was calculated as the ADC (whole-volume ROIs method)/ADC reference.

The ROIs were positioned in regions of tumors displaying maximum enhancement within a homogeneous area. The time signal intensity (SI) curves of all tumors were documented. Furthermore, SI measurements from tumors were normalized using the formula $(SI - S_0) / S_0$ with reference to the pre-contrast SI (S_0). Subsequently, the following parameters, which were initially outlined by Tsili et al.,¹⁵ were computed based on the normalized values. Peak enhancement (PE) was described as the maximum S_i of the tumor. Time to peak (TTP) was described as the duration required to reach the maximum S_i of the tumor. The wash-in rate (WiR) was defined as the greatest slope of tumor enhancement and computed using the following formula: $WiR = \max S_i(PE) - S_{i-1} / \max t_i - t_{i-1}$. Conversely, the wash-out rate (WoR) was described as $\max S_i(PE) - S_7$, indicating the difference between the peak signal and the signal at the last time point.

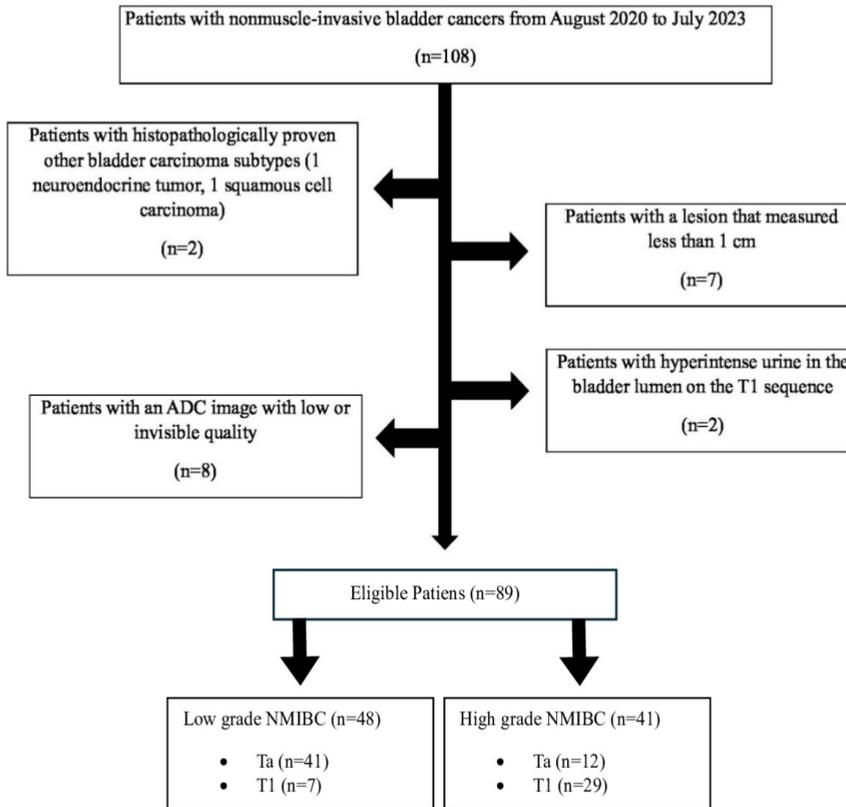


Figure 1. Flowchart of patient selection. ADC, apparent diffusion coefficient; NMIBC, non-muscle invasive bladder cancer.

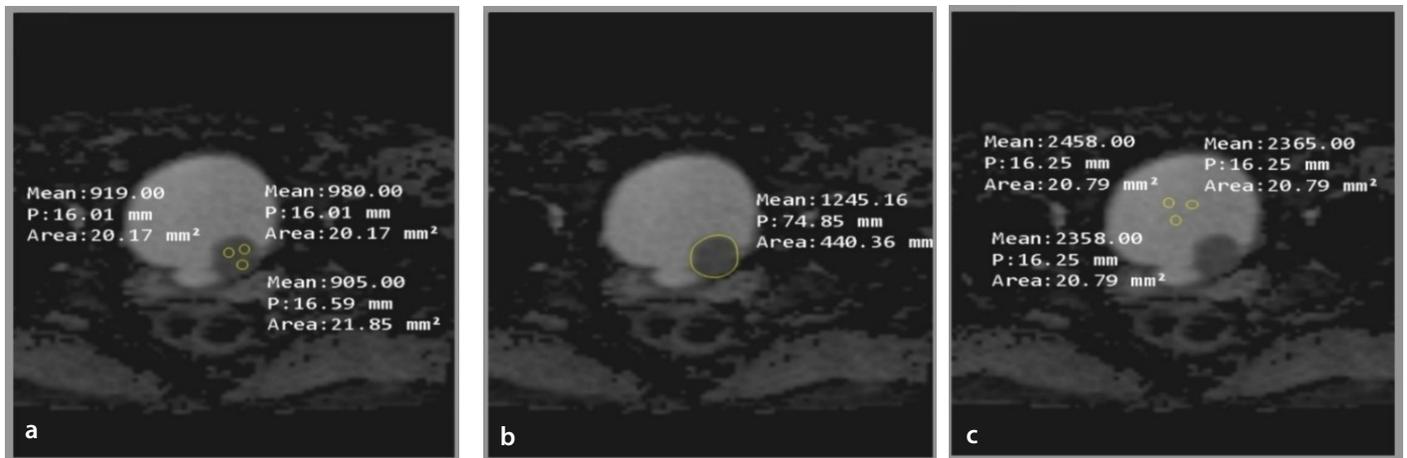


Figure 2. A 54-year-old patient with non-muscle-invasive bladder cancer performed different apparent diffusion coefficient (ADC) values and ADC ratios. (a) Three regions of interest (ROI) were drawn, and the average ADC three ROI (ADC_t) was calculated as 0.934. (b) Freehand ROI along the low signal of the tumor's border on ADC maps. The whole ADC (ADC_w) was calculated as 1.245. (c) Three ROI were drawn in the center of the bladder, and the average ADC was calculated. The reference ADC was 2.393. Based on these findings, ADC_w and ADC_wR are consistent with low-grade tumors, whereas ADC_tR and ADC_t are compatible with high-grade tumors. Histopathological examination revealed high-grade, non-muscle-invasive papillary urothelial carcinoma after transurethral resection of the bladder. ADC_tR, three-ROIs method ADC ratio; ADC_wR, whole-ROIs method ADC ratio.

Statistical analysis

The data analysis was conducted using IBM SPSS Statistics software version 25.0 (IBM SPSS Corp.; Armonk, NY, USA). The normality of the data was assessed using the Kolmogorov–Smirnov test. For normally distributed data, mean values were presented with standard deviations (SD). The independent t-test was utilized to compare the mean values of ADC, PE, TTP, WiR, and WoR between low- and high-grade NMIBC. A receiver operating characteristic (ROC) curve analysis was performed to determine the cut-off values for ADCt, ADCw, ADCtR, and ADCwR in distinguishing between low- and high-grade cases. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated. The intraclass correlation coefficient (ICC) was used to assess inter-reader reliability for ADC measurements. Data were reported as mean ± SD and n (%), with *P* values below 0.05 considered statistically significant.

Results

We identified 108 patients with pathologically confirmed NMIBC. Two patients had other histopathologically confirmed bladder carcinoma subtypes (one neuroendocrine tumor and one squamous cell carcinoma). Seven patients had a lesion that measured less than 1 cm. Eight patients had an ADC image with low or invisible quality, and two patients had hyperintense urine in the bladder lumen on the T1 sequence. These individuals were removed from the study. Consequently, 89 patients (85 men with a median age of 68 years) were enrolled in this study. Within our study population, 48 had low-grade tumors, and 41 had high-grade tumors. The average maximum diameter of low-grade NMIBCs was 21.7 mm (range: 12–89 mm), whereas that of high-grade NMIBCs was 27.6 mm (range: 11–63 mm). A total of 31 tumors were classified as T1-stages, and 58 tumors were classified as Ta-stages based on histopathology.

There was no significant difference in the mean WoR, WiR, TTP, and PE values between low- and high-grade NMIBC for both readers (*P* > 0.05). The ADCt, ADCw, ADCtR, and ADCwR values of high-grade NMIBC were significantly lower than those of low-grade NMIBC for both readers (*P* < 0.001). The mean values of the ADC measurements and semi-quantitative DCE parameters for both readers are shown in Table 1.

Moreover, there was no significant difference in the mean values of WoR, WiR, TTP, and PE between Ta- and T1-stages NMIBC for both readers (*P* > 0.05). The ADCt, ADCw, ADCtR, and ADCwR values of T1-stage NMIBC were significantly lower than those of Ta-stage NMIBC for both readers (*P* < 0.001). Table 2 shows the mean values of ADC measurements and semi-quantitative DCE parameters for both readers.

Receiver operating characteristic analysis of apparent diffusion coefficient measurements for the differentiation of low- and high-grade non-muscle invasive bladder cancer

ROC curve analysis showed that ADCtR had the highest area under the curve (AUC) values for both readers (0.879–0.857) (Figure 3). With cut-off values of 0.449 and 0.435, ADCtR had the best diagnostic performance for both readers, with 85.4%–83.1% accuracy, 87.5%–85.4% sensitivity, and 82.9%–80.4% specificity. Table 3 shows the diagnostic performance of ADC values and ADC ratios for each reader.

Additionally, ADCtR had valuable AUC values for both readers (0.827–0.806) for differentiating the Ta- and T1-stages (Figure 4). With cut-off values of 0.439 and 0.431, ADCtR had acceptable diagnostic performance for both readers, with 76.4%–74.1% accuracy, 82%–80% sensitivity, and 69.2%–66.7% specificity. Table 4 shows the diagnostic performance of ADC values and ADC ratios for each reader for differentiating the Ta- and T1-stages.

Inter-reader agreement was almost perfect for ADC measurements (*P* < 0.001). Inter-reader ICCs between reader 1 and reader 2 were as follows: ADCt = 0.939 [95% confidence interval (CI): 0.908–0.959]; ADCw = 0.968 (95% CI: 0.952–0.979); ADCtR = 0.958 (95% CI: 0.936–0.972); ADCwR = 0.969 (95% CI: 0.953–0.979).

Table 1. Mean values of semiquantitative dynamic contrast-enhanced parameters and apparent diffusion coefficient measurements for low- and high-grade non-muscle invasive bladder cancers

	Low-grade NMIBC	High-grade NMIBC	<i>P</i> value
ADCtR			
Reader 1	0.54 ± 0.1	0.39 ± 0.1	<i>P</i> < 0.001
Reader 2	0.50 ± 0.16	0.42 ± 0.13	<i>P</i> < 0.001
ADCwR			
Reader 1	0.58 ± 0.10	0.44 ± 0.08	<i>P</i> < 0.001
Reader 2	0.56 ± 0.11	0.43 ± 0.09	<i>P</i> < 0.001
ADCt			
Reader 1	1.19 ± 0.24	0.93 ± 0.17	<i>P</i> < 0.001
Reader 2	1.17 ± 0.23	0.94 ± 0.16	<i>P</i> < 0.001
ADCw			
Reader 1	1.27 ± 0.24	1.05 ± 0.18	<i>P</i> < 0.001
Reader 2	1.22 ± 0.23	1.12 ± 0.22	<i>P</i> < 0.001
TTP			
Reader 1	110 ± 7	119 ± 10	<i>P</i> = 0.55
Reader 2	113 ± 31	125 ± 34	<i>P</i> = 0.61
PE			
Reader 1	3.1 ± 0.14	3.6 ± 0.22	<i>P</i> = 0.06
Reader 2	3.4 ± 0.21	3.7 ± 0.26	<i>P</i> = 0.07
WoR			
Reader 1	0.04 ± 0.01	0.07 ± 0.1	<i>P</i> = 0.07
Reader 2	0.03 ± 0.1	0.07 ± 0.2	<i>P</i> = 0.09
WiR			
Reader 1	1.44 ± 0.35	2.03 ± 0.53	<i>P</i> = 0.34
Reader 2	1.28 ± 0.31	1.95 ± 0.67	<i>P</i> = 0.36

NMIBC, non-muscle invasive bladder cancer; ADCtR, three-ROIs method ADC ratio; ADCwR, whole-ROIs method ADC ratio; ADCt, three-ROIs method ADC; ADCw, whole-ROIs method ADC; TTP, time to peak; PE, peak enhancement; WoR, wash-out rate; WiR, wash-in rate; ADC, apparent diffusion coefficient.

Table 2. Mean values of semiquantitative dynamic contrast-enhanced parameters and apparent diffusion coefficient measurements for Ta- and T1-stage non-muscle invasive bladder cancers

	Ta NMIBC	T1 NMIBC	P value
ADCtR			
Reader 1	0.52 ± 0.1	0.39 ± 0.09	<i>P</i> < 0.001
Reader 2	0.51 ± 0.1	0.40 ± 0.08	<i>P</i> < 0.001
ADCwR			
Reader 1	0.56 ± 0.11	0.45 ± 0.08	<i>P</i> < 0.001
Reader 2	0.55 ± 0.12	0.45 ± 0.08	<i>P</i> < 0.001
ADCt			
Reader 1	1.17 ± 0.24	0.93 ± 0.17	<i>P</i> < 0.001
Reader 2	1.17 ± 0.23	0.92 ± 0.18	<i>P</i> < 0.001
ADCw			
Reader 1	1.25 ± 0.25	1.05 ± 0.17	<i>P</i> < 0.001
Reader 2	1.25 ± 0.24	1.06 ± 0.17	<i>P</i> < 0.001
TTP			
Reader 1	115 ± 7	113 ± 11	<i>P</i> = 0.92
Reader 2	91 ± 5	80 ± 6	<i>P</i> = 0.15
PE			
Reader 1	3.1 ± 0.13	3.7 ± 0.23	<i>P</i> = 0.16
Reader 2	3.1 ± 0.12	3.7 ± 0.21	<i>P</i> = 0.17
WoR			
Reader 1	0.04 ± 0.01	0.07 ± 0.1	<i>P</i> = 0.35
Reader 2	0.07 ± 0.1	0.08 ± 0.2	<i>P</i> = 0.26
WiR			
Reader 1	1.43 ± 0.32	2.09 ± 0.58	<i>P</i> = 0.29
Reader 2	2.26 ± 0.42	2.93 ± 0.64	<i>P</i> = 0.38

NMIBC, non-muscle invasive bladder cancer; ADCtR, three-ROIs method ADC ratio; ADCwR, whole-ROIs method ADC ratio; ADCt, three-ROIs method ADC; ADCw, whole-ROIs method ADC; TTP, time to peak; PE, peak enhancement; WoR, wash-out rate; WiR, wash-in rate; ADC, apparent diffusion coefficient.

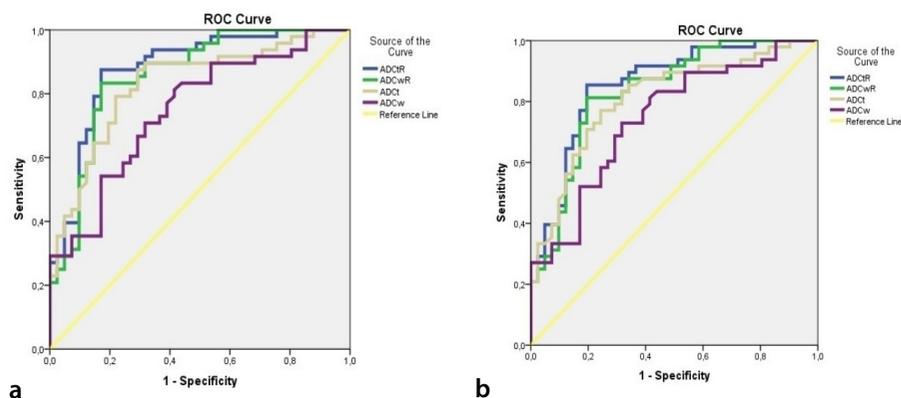


Figure 3. Receiver operating characteristic curve analysis for the differentiation of high- and low-grade non-muscle invasive bladder cancer for reader 1 (a) and reader 2 (b). ADC, apparent diffusion coefficient; ADCtR, three-ROIs method ADC ratio; ADCwR, whole-ROIs method ADC ratio.

Discussion

Of the 89 patients with NMIBC in our study, 48 had low-grade bladder cancer. There was no significant difference in DCE semi-quantitative parameters-WoR, WiR, TTP, and PE-in the differentiation of low- and high-grade NMIBCs. Four methods-ADCt, ADCw, ADCtR, and ADCwR-were compared to distinguish between low- and high-grade NMIBC. The AUC of the ROC for the ADCtR (0.879 ± 0.074) was significantly larger (*P* < 0.001) than that of the other methods for separating low- and high-grade NMIBC. With a cut-off ADCtR value of 0.449, the sensitivity and specificity were 87.5% and 82.9% for reader 1. With a cut-off ADCtR value of 0.435, the sensitivity and specificity were 85.4% and 80.4% for reader 2. Additionally, ADCtR demonstrated the best diagnostic performance in distinguishing between Ta- and T1-stages for both readers, with respective cut-off values of 0.439 and 0.431. In our study, the majority of low-grade tumors were in the Ta-stage, whereas the majority of high-grade tumors were in the T1-stage. This might explain the similar cut-off values and statistical performance in distinguishing between Ta- and T1-stages, as seen in the discrimination between high- and low-grade cases. Thus, it is required to conduct extensive studies that have a more homogeneous distribution.

DWI, in combination with ADC measurement, provides valuable information for quantifying structural tissue changes at a cellular level and aiding in tissue characterization.^{16,17} Low ADC values signify high cellularity, whereas high ADC values signify low cellularity.¹⁶ The intralesional voxels with the lowest ADC values are likely to represent the most aggressive tumors, as they include the highest levels of cellularity.^{18,19} ADC values can be used in multiple myeloma, lymphoma, breast, lung, and testis malignancies and the treatment response of malignancies.²⁰ In previous studies, reference ADC has been useful in brain, liver, pancreas, prostate, and bone lesions, as well as lymph node evaluation.^{21,22} In bladder cancers, ADC values for high- and low-grade tumors were highly variable among the four studies using 1.5T scanners.²³⁻²⁵ Due to variable ADC values, Wang et al.¹³ studied three reference ADC values and obtained the highest accuracy within the bladder lumen. To reduce variability, we also used the bladder lumen as the reference ADC in this study.

Table 3. Diagnostic performance of apparent diffusion coefficient values and ratios for each reader in differentiating patients with low- and high-grade non-muscle invasive bladder cancer

	Reader 1	Reader 2
ADCtR		
Cut-off	0.449	0.435
AUC	0.879 (0.805–0.952)	0.857 (0.778–0.936)
P	<0.001	<0.001
Sensitivity (95% CI)	87.5 (74.7–95.2)	85.4 (72.2–93.9)
Specificity (95% CI)	82.9 (67.9–92.8)	80.4 (65.1–91.1)
PPV (95% CI)	85.7 (75.1–92.2)	83.6 (73.1–90.6)
NPV (95% CI)	85 (72.5–92.3)	82.5 (70.1–90.4)
Accuracy (95% CI)	85.4 (76.3–92)	83.1 (73.7–90.2)
ADCwR		
Cut-off	0.494	0.490
AUC	0.857 (0.776–0.937)	0.833 (0.747–0.918)
P	<0.001	<0.001
Sensitivity (95% CI)	83.3 (69.7–92.5)	81.2 (67.3–91)
Specificity (95% CI)	82.9 (67.9–92.8)	80.4 (65.13–91.1)
PPV (95% CI)	85.1 (74.2–91.9)	82.9 (72–90.2)
NPV (95% CI)	80.9 (68.9–89)	78.5 (66.6–87)
Accuracy (95% CI)	83.1 (73.7–90.2)	80.9 (71.1–88.4)
ADCt		
Cut-off	1.030	0.998
AUC	0.829 (0.743–0.915)	0.811 (0.721–0.901)
P	<0.001	<0.001
Sensitivity (95% CI)	79.1 (65–89.5)	77 (62.6–87.9)
Specificity (95% CI)	78 (62.3–89.4)	75.6 (59.7–87.64)
PPV (95% CI)	80.8 (69.9–88.4)	78.7 (67.8–86.6)
NPV (95% CI)	76.1 (64.3–85)	73.8 (61.9–82.9)
Accuracy (95% CI)	78.6 (68.6–86.6)	76.4 (66.2–84.7)
ADCw		
Cut-off	1.101	1.093
AUC	0.748 (0.648–0.849)	0.745 (0.643–0.847)
P	<0.001	<0.001
Sensitivity (95% CI)	70.8 (55.9–83)	72.9 (58.1–84.7)
Specificity (95% CI)	68.2 (51.9–81.9)	68.2 (51.9–81.9)
PPV (95% CI)	72.3 (61.7–80.9)	72.9 (62.4–81.3)
NPV (95% CI)	66.6 (55.1–76.5)	68.2 (56.4–78.1)
Accuracy (95% CI)	69.6 (59–78.9)	70.7 (60.1–79.9)

ADCtR, three-ROIs method ADC ratio; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; ADCt, three-ROIs method ADC; ADCw, whole-ROIs method ADC; ADCwR, whole-ROIs method ADC ratio; ADC, apparent diffusion coefficient.

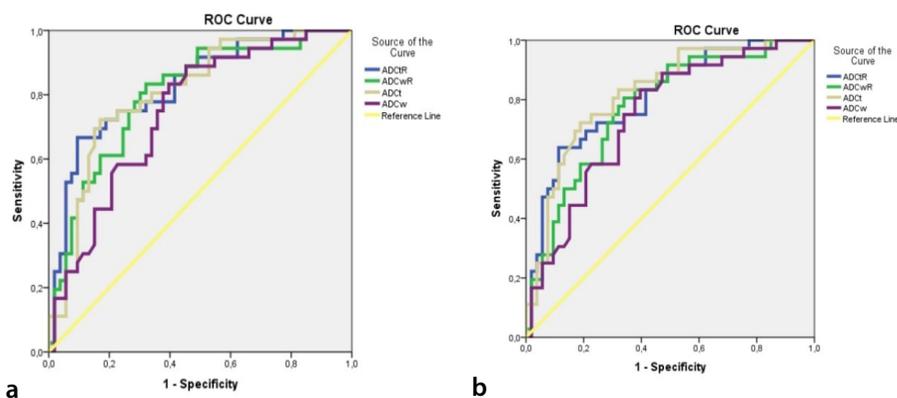


Figure 4. Receiver operating characteristic curve analysis for the differentiation of pTa- and pT1-stage non-muscle invasive bladder cancer for reader 1 (a) and reader 2 (b). ADC, apparent diffusion coefficient; ADCtR, three-ROIs method ADC ratio; ADCwR, whole-ROIs method ADC ratio.

In our study, as each method was evaluated separately to differentiate between low- and high-grade NMIBC, the ADCtR with 87.5% sensitivity and 82.9% specificity was the best method. In comparing the diagnostic performance of ADC values for the differentiation of low- and high-grade bladder cancer in the literature, Wang et al.¹² reported higher sensitivity and specificity values (100% and 95%), with a cut-off of 0.899 mm²/s. The lower specificity and sensitivity in our study can be related to the differences in the research population. The study conducted by Wang et al.¹² included both T1 and T2 bladder cancers. However, we specifically focused on bladder tumors that were non-muscle invasive and utilized a larger sample of patients.

Li et al.¹⁴ reported interobserver agreement for three different methods of measuring ADC values in bladder cancer: single section ROI, three ROI, and whole volume ROI. The average ADC value did not vary significantly in terms of inter-observer consistency across any of the ROI positioning methods in the assessment of tumor grade.¹⁴ Our results were similar to those of Li et al.¹⁴ in that there was excellent consistency between readers with varying levels of experience across all four methods. The agreement between different readers may be attributed to the decreased fibrosis and necrosis in bladder cancer, which suggests that the degree of diffusion is almost homogeneous in bladder cancer.¹⁴ This nature of bladder tumors also aids in the high diagnostic performance of the ADC value in low- and high-grade differentiation for all readers, regardless of experience.

In our study, we found that DCE semi-quantitative parameters-WiR, TTP, and PE-were not efficient in differentiating low- and high-grade NMIBCs. Zhou et al.¹¹ classified bladder tumors into three groups based on their pathological phenotype: low aggressiveness, intermediate aggressiveness, and high aggressiveness, and they examined the effectiveness of semiquantitative parameters derived from DCE imaging in distinguishing between each of these groups. In contrast to our study, Zhou et al.¹¹ obtained a high diagnostic performance in determining the aggressiveness of bladder cancer with a WoR. The difference in our results could be due to our exclusive focus on patients with non-muscle-invasive bladder cancer patients.

Our study has some limitations. First, we did not include lesions smaller than 1 cm.

Table 4. Diagnostic performance of apparent diffusion coefficient values and ratios for each reader in differentiating Ta- and T1-stages in patients with low- and high-grade non-muscle invasive bladder cancer

	Reader 1	Reader 2
ADcTR		
Cut-off	0.439	0.431
AUC	0.827 (0.739–0.914)	0.806 (0.714–0.897)
<i>P</i>	<0.001	<0.001
Sensitivity (95% CI)	82 (68.5–91.4)	80 (66.3–90)
Specificity (95% CI)	69.2 (52.4–83)	66.7 (49.8–81)
PPV (95% CI)	77.3 (67.7–84.8)	75.4 (65.9–83.1)
NPV (95% CI)	75 (61.6–84.9)	72.2 (58.9–82.5)
Accuracy (95% CI)	76.4 (66.2–84.8)	74.16 (63.8–82.9)
ADcWR		
Cut-off	0.486	0.494
AUC	0.800 (0.707–0.894)	0.776 (0.678–0.873)
<i>P</i>	<0.001	<0.001
Sensitivity (95% CI)	81.3 (67.4–91.1)	79.2 (46.9–77.9)
Specificity (95% CI)	65.9 (49.4–79.9)	63.4 (46.9–77.9)
PPV (95% CI)	73.6 (64.1–81.3)	71.7 (62.3–79.5)
NPV (95% CI)	75 (61.5–84.9)	72.2 (58.8–82.5)
Accuracy (95% CI)	74.2 (63.8–82.9)	71.9 (61.4–80.9)
ADcT		
Cut-off	1.030	0.993
AUC	0.806 (0.715–0.898)	0.822 (0.734–0.910)
<i>P</i>	<0.001	<0.001
Sensitivity (95% CI)	77.3 (63.8–87.7)	81.1 (68–90.5)
Specificity (95% CI)	75 (57.8–87.9)	72.2 (54.8–85.8)
PPV (95% CI)	82 (71.8–89.1)	81.1 (71.4–88)
NPV (95% CI)	69.2 (56.9–79)	72.2 (58.9–82.4)
Accuracy (95% CI)	76.4 (66.2–84.7)	77.5 (67.4–85.7)
ADcW		
Cut-off	1.136	1.115
AUC	0.745 (0.644–0.847)	0.755 (0.653–0.856)
<i>P</i>	<0.001	<0.001
Sensitivity (95% CI)	77.2 (62.2–88.5)	76.6 (62–87.7)
Specificity (95% CI)	57.8 (42.2–72.3)	59.5 (43.2–74.4)
PPV (95% CI)	64.2 (55.1–72.3)	67.9 (58.7–75.9)
NPV (95% CI)	72.2 (58.8–82.6)	69.4 (56.1–80.1)
Accuracy (95% CI)	67.4 (56.7–77)	68.5 (57.8–78)

ADcTR, three-ROIs method ADC ratio; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; ADcT, three-ROIs method ADC; ADcW, whole-ROIs method ADC; ADcWR, whole-ROIs method ADC ratio; ADC, apparent diffusion coefficient.

In conclusion, while DCE semiquantitative parameters did not yield significant outcomes in distinguishing between low- and high-grade tumors, ADcTR holds promise for enhancing patient management in NMIBC cases and stands as a potential preoperative radiological asset for NMIBC. The results of our study demonstrated consistency even between readers with different experience levels.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Eryuruk U, Tasdemir MN, Aslan S. Comparison of the diagnostic performance of biparametric and multiparametric MRI in detecting muscle invasion of bladder cancer located at the ureteral orifice. *Abdom Radiol (NY)*. 2023;48(10):3174–3182. [\[CrossRef\]](#)
2. Josephson D, Pasin E, Stein JP. Superficial bladder cancer: part 2. management. *Expert Rev Anticancer Ther*. 2007;7(4):567–581. [\[CrossRef\]](#)
3. Fujimura M, Sakamoto S, Sekita N, et al. Apparent diffusion coefficient value for estimating clinicohistological factors in bladder cancer including infiltration style and lymphatic invasion. *Springerplus*. 2016;5(1):848. [\[CrossRef\]](#)
4. Soukup V, Čapoun O, Cohen D, et al. Prognostic performance and reproducibility of the 1973 and 2004/2016 World Health Organization Grading Classification Systems in non-muscle-invasive bladder cancer: A European Association of Urology Non-muscle Invasive Bladder Cancer Guidelines Panel Systematic Review. *Eur Urol*. 2017;72(5):801–813. [\[CrossRef\]](#)
5. Akcay A, Yagci AB, Celen S, Ozlulderen Y, Turk NS, Ufuk F. VI-RADS score and tumor contact length in MRI: a potential method for the detection of muscle invasion in bladder cancer. *Clin Imaging*. 2021;77:25–36. [\[CrossRef\]](#)
6. Delli Pizzi A, Mastrodicasa D, Marchioni M, et al. Bladder cancer: do we need contrast injection for MRI assessment of muscle invasion? A prospective multi-reader VI-RADS approach. *Eur Radiol*. 2021;31(6):3874–3883. [\[CrossRef\]](#)
7. Panebianco V, Pecoraro M, Del Giudice F, et al. VI-RADS for bladder cancer: current applications and future developments. *J Magn Reson Imaging*. 2022;55(1):23–36. [\[CrossRef\]](#)
8. American Cancer Society. Cancer treating bladder cancer. Accessed 19 December 2022. [\[CrossRef\]](#)

However, this was effective in preventing the partial volume effect. Second, ADC measurements are prone to errors. However, we aimed to minimize this potential by utilizing four different methods and two different readers. Third, this study was a single-institution retrospective study, but it had the larg-

est sample reported in the literature. Larger multicenter studies are required to validate our findings. Finally, the ADC maps were obtained using a monoexponential algorithm; using a multiple exponential fit with additional *b*-values could potentially enhance accuracy and be a more favorable approach.

9. Stöckle M, Alken P, Engelmann U, Jacobi GH, Riedmiller H, Hohenfellner R. Radical cystectomy--often too late? *Eur Urol.* 1987;13(6):361-367. [\[CrossRef\]](#)
10. Jakse G, Loidl W, Seeber G, Hofstädter F. Stage T1, grade 3 transitional cell carcinoma of the bladder: an unfavorable tumor? *J Urol.* 1987;137(1):39-43. [\[CrossRef\]](#)
11. Zhou G, Chen X, Zhang J, Zhu J, Zong G, Wang Z. Contrast-enhanced dynamic and diffusion-weighted MR imaging at 3.0T to assess aggressiveness of bladder cancer. *Eur J Radiol.* 2014;83(11):2013-2018. [\[CrossRef\]](#)
12. Wang HJ, Pui MH, Guo Y, et al. Multiparametric 3-T MRI for differentiating low-versus high-grade and category T1 versus T2 bladder urothelial carcinoma. *AJR Am J Roentgenol.* 2015;204(2):330-334. [\[CrossRef\]](#)
13. Wang HJ, Pui MH, Guo Y, et al. Value of normalized apparent diffusion coefficient for estimating histological grade of vesical urothelial carcinoma. *Clin Radiol.* 2014;69(7):727-731. [\[CrossRef\]](#)
14. Li H, Liu L, Ding L, Zhang Z, Zhang M. Quantitative assessment of bladder cancer reflects grade and recurrence: comparing of three methods of positioning region of interest for adc measurements at diffusion-weighted MR imaging. *Acad Radiol.* 2019;26(9):1148-1153. [\[CrossRef\]](#)
15. Tsili AC, Sylakos A, Ntorkou A, et al. Apparent diffusion coefficient values and dynamic contrast enhancement patterns in differentiating seminomas from nonseminomatous testicular neoplasms. *Eur J Radiol.* 2015;84(7):1219-1226. [\[CrossRef\]](#)
16. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am J Roentgenol.* 2007;188(6):1622-1635. [\[CrossRef\]](#)
17. Tsili AC, Argyropoulou MI, Giannakis D, Tsampalas S, Sofikitis N, Tsampoulas K. Diffusion-weighted MR imaging of normal and abnormal scrotum: preliminary results. *Asian J Androl.* 2012;14(4):649-654. [\[CrossRef\]](#)
18. Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. *Radiology.* 2013;268(2):318-322. [\[CrossRef\]](#)
19. Schnapauff D, Zeile M, Niederhagen MB, et al. Diffusion-weighted echo-planar magnetic resonance imaging for the assessment of tumor cellularity in patients with soft-tissue sarcomas. *J Magn Reson Imaging.* 2009;29(6):1355-1359. [\[CrossRef\]](#)
20. Donners R, Blackledge M, Tunariu N, Messiou C, Merkle EM, Koh DM. Quantitative whole-body diffusion-weighted MR imaging. *Magn Reson Imaging Clin N Am.* 2018;26(4):479-494. [\[CrossRef\]](#)
21. Oh J, Henry RG, Pirzkall A, et al. Survival analysis in patients with glioblastoma multiforme: predictive value of choline-to-N-acetylaspartate index, apparent diffusion coefficient, and relative cerebral blood volume. *J Magn Reson Imaging.* 2004;19(5):546-554. [\[CrossRef\]](#)
22. Park SO, Kim JK, Kim KA, et al. Relative apparent diffusion coefficient: determination of reference site and validation of benefit for detecting metastatic lymph nodes in uterine cervical cancer. *J Magn Reson Imaging.* 2009;29(2):383-390. [\[CrossRef\]](#)
23. Kobayashi S, Koga F, Yoshida S, et al. Diagnostic performance of diffusion-weighted magnetic resonance imaging in bladder cancer: potential utility of apparent diffusion coefficient values as a biomarker to predict clinical aggressiveness. *Eur Radiol.* 2011;21(10):2178-2186. [\[CrossRef\]](#)
24. Kobayashi S, Koga F, Kajino K, et al. Apparent diffusion coefficient value reflects invasive and proliferative potential of bladder cancer. *J Magn Reson Imaging.* 2014;39(1):172-178. [\[CrossRef\]](#)
25. Avcu S, Koseoglu MN, Ceylan K, Bulut MD, Unal O. The value of diffusion-weighted MRI in the diagnosis of malignant and benign urinary bladder lesions. *Br J Radiol.* 2011;84(1006):875-882. Erratum in: *Br J Radiol.* 2011;84(1007):1059. [\[CrossRef\]](#)



The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review

Benjamin D. Simon^{1,2}
 Kutsev Bengisu Ozyoruk¹
 David G. Gelikman¹
 Stephanie A. Harmon¹
 Barış Türkbey¹

¹Molecular Imaging Branch, National Cancer Institute, National Institutes of Health, Bethesda, USA

²University of Oxford, Institute of Biomedical Engineering, Department Engineering Science, Oxford, UK

ABSTRACT

With the ongoing revolution of artificial intelligence (AI) in medicine, the impact of AI in radiology is more pronounced than ever. An increasing number of technical and clinical AI-focused studies are published each day. As these tools inevitably affect patient care and physician practices, it is crucial that radiologists become more familiar with the leading strategies and underlying principles of AI. Multimodal AI models can combine both imaging and clinical metadata and are quickly becoming a popular approach that is being integrated into the medical ecosystem. This narrative review covers major concepts of multimodal AI through the lens of recent literature. We discuss emerging frameworks, including graph neural networks, which allow for explicit learning from non-Euclidean relationships, and transformers, which allow for parallel computation that scales, highlighting existing literature and advocating for a focus on emerging architectures. We also identify key pitfalls in current studies, including issues with taxonomy, data scarcity, and bias. By informing radiologists and biomedical AI experts about existing practices and challenges, we hope to guide the next wave of imaging-based multimodal AI research.

KEYWORDS

Artificial intelligence, cancer research, multimodal, pathology, radiology

Artificial Intelligence (AI) is revolutionizing everyday life with its advanced capabilities in image processing, textual analysis, and more. Though this technology has only recently gained widespread public attention, its origins are not new. Research into neural networks began in the early to mid-20th century,¹ making it surprising that mainstream models, such as ChatGPT, which are now frequently cited in scientific literature, have only recently captured public interest.² Comparable to the emergence of computers in the 1940s, modern AI possesses a long-standing mathematical foundation but is still in its infancy.

The field of radiology is data-heavy, signal-rich, and technology-focused, making it a prime target for building AI applications. Thus, it is crucial that radiologists stay informed about methodological and clinical trends in AI. Radiologists routinely review large amounts of signal-rich data in a multimodal manner, making them well-suited to leverage AI and medical data to enhance diagnostic accuracy. At its core, AI is an extremely thorough pattern-detection system, capable of recognizing patterns beyond human capability for certain tasks. In medical imaging, which is nowadays very commonly used and results in work overload for practicing radiologists, AI has the potential to be a robust support tool within the radiology medical ecosystem. However, the introduction of AI raises ethical dilemmas³ and security concerns,⁴ including data leakage, automated medical decisions, biased data, and clinical impact.

While there is a growing body of literature on biomedical AI, much remains unexplored, particularly in the translation to medical applications. There has been a noticeable shift towards multimodal algorithms that incorporate imaging data with at least one other modality. Nevertheless, literature leveraging multimodal imaging data and clinical co-variables remains relatively sparse. For this reason, existing reviews on the topic have generally focused on 1) unimodal AI for imaging alone⁵⁻⁷ or 2) general multimodal deep learning, which is becoming an increasingly heterogeneous field.⁸⁻¹⁰ This review aims to explore multimodal AI in radiolo-

Corresponding author: Barış Türkbey

E-mail: turkbeyi@mail.nih.gov

Received 17 June 2024; revision requested 17 July 2024; accepted 04 September 2024.



Epub: 01.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242631

gy comprehensively by examining both imaging and clinical variables. Throughout, we assess the methodology and clinical translation to inform future directions and organize approaches within the field.

Modern frameworks and multi-modality fusion techniques

The first focus of this study is the cutting-edge methodologies for multimodal AI. These frameworks are increasingly recognized as impactful approaches in advancing healthcare analytics due to their ability to interpret and integrate disparate forms of medical data, similar to the daily tasks of physicians. For detailed definitions and explanations of key terminology, a glossary of key terms with definitions is provided (Table 1). Central frameworks aim to model the relationship between data and corresponding clinical outcomes. Transformer-based models and graph neural networks (GNNs) have demonstrated remarkable promise in combining clinical notes,¹¹⁻¹³ imaging data,¹⁴⁻¹⁶ and genomic information,¹⁷⁻²⁰ enhancing patient care through personalized and precise predictions and recommendations (Figure 1).

Transformers

Initially conceived for natural language processing, transformers have been adapted for other unimodal input data, such as imaging and genomics, and now, for multimodal tasks in healthcare. These models uniquely focus on different data components as needed and are adept at handling sequential data.²¹ They also employ self-attention

mechanisms, allowing for the assignment of weighted importance to different parts of input data, regardless of order. This implementation is especially beneficial for free text or genomic sequencing data, where the significance of a feature greatly depends on its context. These mechanisms have been extended to consider temporal dependencies in electronic health records (EHRs), enabling the model to discern which historical medi-

cal events are most predictive of future outcomes.²²

Transformers are particularly revolutionary, unlike typical recurrent neural networks, in that they employ a parallelized approach, which allows for scalable computation.²³ Recurrent neural networks are a popular type of model that handle information sequentially and cannot do so in parallel.²⁴ This founda-

Table 1. Glossary of key terminology

Term	Our definition
Multimodal AI	AI models that integrate multiple types of data (e.g., imaging, clinical notes, genomic data) to improve diagnostic accuracy and patient outcomes.
Multichannel AI	AI models that integrate multiple inputs of the same type of data (e.g., multiple pathology images, multiple radiology images, multiple genomic sequences).
GNN	A type of neural network designed to capture dependencies in data that is structured as graphs, useful in settings where data interactions are non-linear and complex.
Transformers	A model architecture initially developed for natural language processing that has been adapted for analyzing various types of data. Known for its self-attention mechanism, which helps in understanding the importance of different parts of the data.
Machine learning	A method and field in computer science where systems are able to learn without deliberate instructions through mathematical pattern recognition of data.
AI	A broad field describing computer systems which are able to behave in ways that would normally require human intelligence.
Fusion techniques	Methods used to integrate multiple types of data in AI models. These can be early, joint, or late fusion, depending on when data types are combined during the model training process. There are many other statistical integration methods.
Parallel computation	A strategy in computer science where multiple processes or calculations happen simultaneously rather than one at a time.
Non-euclidean	Data that does not fit into traditional Euclidean geometry frameworks, such as graph data, which is essential for certain types of medical analyses where relationships and connections define data structure.
Clinical metadata	Information accompanying medical data that provides context about the health status, treatment, or diagnostics of a patient, crucial for interpreting imaging data in AI models.
Data curation	The process of organizing, integrating, and managing data collected from various sources to ensure it is accurate, complete, and reliable for AI training and analysis.
Self-attention mechanism	A component of neural network architectures that allows the model to weigh the importance of different parts of the input data differently, improving its ability to understand complex patterns.
Sequential data processing	In AI, the handling of data that is organized in a sequence (such as time series data from patient records), which is critical for understanding temporal patterns and dependencies.
Bias mitigation	Strategies and methodologies aimed at reducing bias in AI models to ensure fairness and equity, particularly important in healthcare applications where biased decisions can have serious implications.
Transfer learning	A machine learning method where a model developed for one task is reused as the starting point for a model on a second task, facilitating rapid deployment and reducing the need for large amounts of data.
Model generalizability	The ability of an AI model to perform well across different settings or populations, not just the ones on which it was trained, which is crucial for applications in diverse clinical environments.

AI, artificial intelligence; GNN, graph neural network.

Main points

- As multimodal artificial intelligence (AI) becomes increasingly integrated into the field of radiology, it is imperative that radiologists become familiar with the existing frameworks, applications, and analyses of such tools.
- Conventional approaches to multimodal AI integration have shown improvement over unimodal approaches in their ability to translate accurately to the clinic.
- Cutting-edge approaches for multimodal biomedical AI applications, such as transformers and graph neural networks, can integrate time series and non-Euclidean biomedical data.
- Key pitfalls of the multimodal biomedical AI landscape include inconsistent taxonomy, a lack of foundational models using varied large-scale representative data sources, and a mismatch between the healthcare arena and the necessary curation of data for AI models.

tional difference has led to transformers being the basis for large language models, such as BERT²⁵ and ChatGPT, but their application in medicine remains largely unexplored.^{25,26} Literature using transformer-based multimodal predictions consistently finds that transformer models outperform typical recurrent or unimodal models.²⁷⁻³⁰

Despite the success of transformers, most literature features single-case applications, where a particular transformer architecture is optimized for a single clinical outcome.³¹ A good example of an impactful application of transformers by Yu et al.³² presents a framework to learn from imaging, clinical, and genetic information to set a new benchmark for diagnosing Alzheimer's disease (area under the receiver operator characteristic curve of 0.993). This work shows how transformers may be able to aid in unifying information across modalities for comprehensive learning in a specific disease space.

The literature on their broader optimization for various clinical or radiology tasks is limited. Khader et al.³³ propose a transferrable large-scale transformer approach, showing that it outperforms existing multimodal approaches leveraging convolutional neural networks (CNNs). They attribute their improvement to a novel technical approach, which selectively limits interactions between data inputs. They demonstrate the generalizability of their model by showing improvement across various decisions, including heart failure and respiratory disease prediction, and domains, including funduscopy

images and chest radiographs paired with non-imaging data.³³

With the increasing popularity of multimodal data and models, there is a need for technical approaches that are transferrable and widely applicable for clinical use.

Graph neural networks

Although transformer-based models excel at capturing dependencies in sequential data,³⁴ their architecture does not inherently account for non-Euclidean structures present in multimodal healthcare data.²³ This gap has led to significant interest in GNNs, which model the data in a graph-structured format. This is particularly relevant to multimodal imaging data, where the relationships and dependencies between data points, such as between an anatomical structure in imaging and a genetic marker or clinical parameters, are not inherently grid-like and could be more accurately represented by graphical connections (Figure 2).

GNNs extend the concept of convolution from regular grids to graphs, with convolutional operations that aggregate feature information from a node's neighbors.³⁵ This approach captures global structural information. Unlike CNNs, where the same filter is applied uniformly across an image or matrix, GNNs adaptively learn how to weight the influence of neighboring nodes, making them adept at handling irregular data that does not conform to a fixed grid.³⁶

This novelty is rooted in the ability of GNNs to learn from non-Euclidean data, which is

crucial for integrating different types of medical information.³⁷ They can explicitly model the complex relationships between modalities, rather than attempting to map them in grid-like structures, such as CNNs, which may not fully take the structure into account³⁸ and could introduce biases related to artificial adjacency in grid formatting. Although exciting work has been taking place recently in medical imaging with GNNs, the bulk of multimodal literature continues to focus on CNNs, requiring tabular fusion in many cases.³⁹ There are several methodologies for fusing modalities.⁴⁰ However, without a graphical approach, there is potential for misinterpretation of the data's relationship when arbitrarily fused in a tabular format. For example, appending an image with a clinical parameter could falsely imply that parameters are adjacent to the imaging features. In contrast, with a GNN, this relationship can be modeled via nodes in a graphical representation, rather than being appended.

Despite the potential and applicability of GNNs, literature leveraging them in the medical space is scarce, likely due to their novelty and the varying custom methods for graphical construction posing a challenge. One study in the oncologic radiology space used a GNN to predict regional lymph node metastasis in esophageal squamous cell carcinoma patients.⁴¹ In their work, Ding et al.⁴¹ constructed a graph by mapping learned embeddings across image features and clinical parameters into a feature space, treating them each as nodes. They then used a graphical attention mechanism to learn the weights of the edges connecting the nodes. In another study, Gao et al.²⁰ used a completely different method for construction to predict the survival of cancer patients using gene expression data. They constructed a graph by considering each patient's primary modality encoding (which could be imaging, though they did not use imaging) as a node, with each gene also as a node. Edge weights were then determined by the level of gene expression for each patient and connected to the primary nodes. In a third study, Lyu et al.⁴² demonstrate a successful GNN for predicting drug interactions by building graphs drawing edges between drugs and drug-related entities (such as targets or transporters). These three examples illustrate the complexity of graph construction and the custom nature of GNN methodology, which may explain the scarcity of literature on the topic despite its promise for relating multimodal data and encodings.

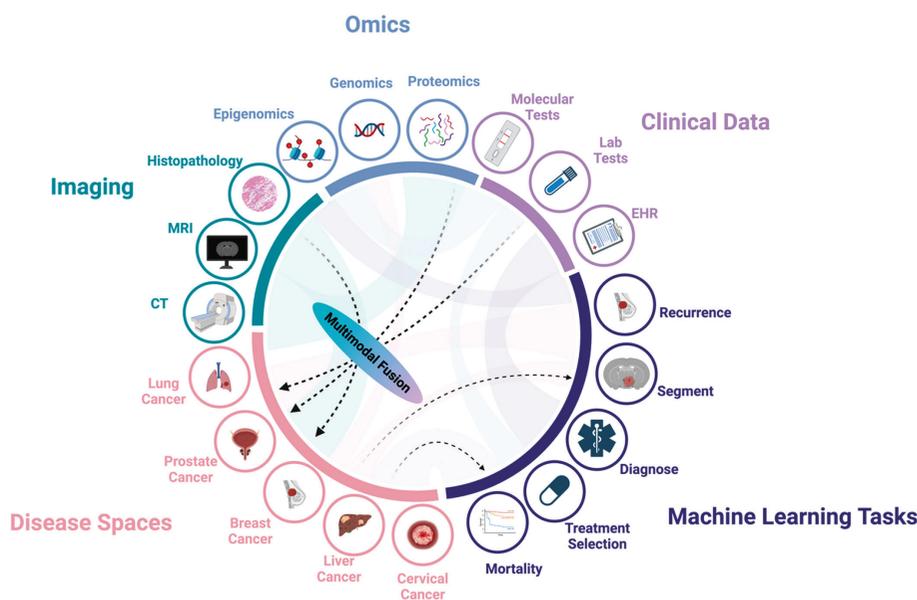


Figure 1. Multimodal medical artificial intelligence (AI) applications across disease spaces. Simplified schematic of the many applications of multimodal medical AI fusing imaging, omics, and clinical data for various tasks across disease spaces.

Modality fusion techniques

Despite the emergence of architectures such as GNNs, which can more deliberately represent data interactions, almost all medical data, whether imaging, molecular, or other signals, can be tabulated. Thus, various fusion techniques (methods for concatenating signals or information) are far more commonly used in multimodal literature.⁹ Fusion techniques can broadly be categorized as early, intermediate/joint, or late fusion. In simple terms, early fusion means that the information is combined before learning via AI occurs, joint fusion means some learning happens before and after combining the two modalities, and late fusion means no learning happens after combining information. Therefore, it can be considered that late fusion aggregates learned information from the two modalities to make a prediction, whereas joint fusion allows for the modalities to interact, and for components of each to have complex relationships in making a pre-

dition. More technically, early fusion generally involves concatenating input modalities into a single vector before feeding them into a model for training. These input modalities can be extracted features or raw data. Joint or intermediate fusion involves concatenating independently learned features prior to further learning. Late fusion generally refers to complete or almost complete learning occurring independently before concatenating vectors for a final activation and prediction. There has also been an emergence of “sketch” fusion, which is similar to early fusion, but rather than concatenation, modalities are translated to a common space. Schematics of early, joint, and late fusion pipelines are presented in Figure 3.

There is a rich and growing base of multimodal models using fusion to combine tabulated free speech,⁴³ genomic,^{44,45} or clinical covariate data with images for diagnostics. Kumar et al.⁴³ combined X-ray images with audio data consisting of respiratory sounds

and coughs for the diagnosis of coronavirus disease 2019. As a result, they showed that early detection is possible with 98.91% accuracy by fusing chest X-ray and cough models. There is limited consensus on the optimal fusion technique, perhaps due to variations in dataset quality, interactions between data sources, or the learning architectures. With many variables at play, developing a comprehensive approach to machine learning fusion, even for a single data type or disease case, becomes challenging. Each fusion modality may have advantages or disadvantages depending on the application, data set, and model architecture. Often, the best approach is to try all three and compare results. Conceptually, however, the pros and cons primarily depend on the concept of confounding variables. Consider the example of a hypothetical model for lung cancer outcome prediction where there are two modalities, one being clinical risk factors, such as cigarette consumption and obesity, and the second being genomic data. If these two modalities are believed to be additive and independent (non-confounding), the requirement may be for the AI to learn from them separately. In this case, late fusion may be appropriate. If it is believed there is significant crosstalk between the variables (the relationship between them is confounding), early or joint fusion may be more appropriate. Early fusion may be more appropriate when using smaller-scale genomic variant data that checks for a set of known variants that increase risk. Conversely, joint fusion may be more appropriate if the model is expected to learn variants of risk from a large amount of genomic sequencing data. Regardless, it is difficult to determine the optimal fusion strategy from the data alone and often worth exploring multiple approaches.

Although early fusion appears to be the most common fusion type across a variety of fields using imaging or imaging features combined with other modalities,^{9,46-51} there are also numerous studies using joint⁵²⁻⁵⁴ and late fusion.⁵⁵ The optimal fusion technique likely depends on the data source, architecture, and other specifics, making consensus challenging. It is important that researchers explore multiple fusion options when designing a multimodal model because, unfortunately, there are no guidelines for multimodal data fusion at this point in the field’s development.

In addition to these common concatenation techniques, there are many other examples of statistical integration methods. When it comes to GNNs, these integration methods

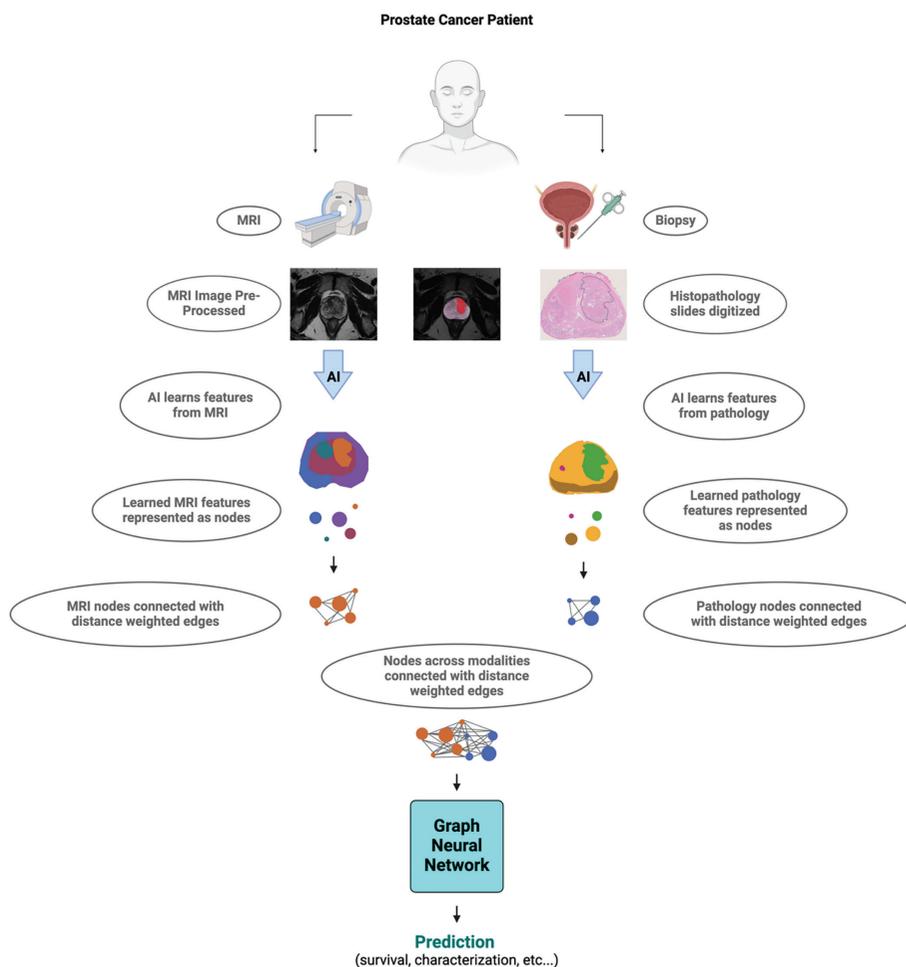


Figure 2. Biomedical data for graph neural networks (GNNs). Example of a hypothetical application of a GNN in the prostate cancer space. Here, typical non-graph neural networks (labeled AI) learn features. Spatial relationships between these features of histopathology data and magnetic resonance images have the potential to be used in graph construction (using distance as the weights of edges and nodes corresponding to structures and features of pathology). AI, artificial intelligence.

can be customized to the relationship between specific modalities and datasets, as previously discussed. There are also many more methods outside the scope of this review, particularly pertaining to other omics data types. For example, mixOmics is a popular package for the integration and analysis of multi-omics data.⁵⁶ Other cutting-edge examples of multi-omics statistical integration frameworks include Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) and xMWAS.^{57,58}

Current status of multimodal imaging work

The existing literature on multimodal AI contains numerous examples of successful multimodal integrations boasting impressive degrees of accuracy and proposed clinical translations.⁵⁹⁻⁶⁹ These publications are promising and show the potential for multi-

modal AI implementation to improve patient outcomes. As the field progresses, there is an increase in highly curated large-scale data sets, paving the way for foundational models.^{29,70} Nevertheless, much of the work in this space and its ability to translate to the clinic is limited by its siloed application, inconsistent taxonomy, and data scarcity.

Multimodal taxonomy

In the broad field of oncology, it is common for physicians to utilize multiple imaging channels to visualize abnormalities and make decisions. It follows that AI models leveraging multiple imaging sequences may be useful for tasks such as detection or segmentation. This raises the question: should combining two images be considered *multimodal*? Here, attention is drawn to the terms *multimodal* and *multichannel*. These

terms are used in different and overlapping contexts across multiple disease spaces. In prostate cancer imaging literature, for example, the detection and segmentation of clinically significant prostate cancer are common goals often labeled as “multimodal” when merely integrating multiple magnetic resonance imaging (MRI) sequences, without incorporating fundamentally different data types.⁷¹⁻⁷⁵ Similar inconsistencies stand across the larger oncology field including, but not limited to, brain cancer,⁷⁶ lung cancer,^{77,78} and breast cancer.⁷⁹

The authors suggest that a multimodal model should combine conceptually different modes of information, whereas multichannel may be more appropriate for technically different (but categorically equivalent or similar) modes, as would be the case in fusing two radiologic images, such as multiple MRI sequences or computed tomography (CT) and MRI. Using this loose idea of “conceptually different images”, one may consider combining digital histopathology images with radiomics as multimodal,⁸⁰ but the examples above (of fusing two radiologic images) would likely be considered multichannel and unimodal. In the authors’ work with deep learning in the prostate cancer space, these image fusion models have been referred to as multichannel rather than multimodal.^{81,82} With this pattern being evident across disease spaces, there is a need to clarify the taxonomy as the term “multimodal” becomes increasingly imprecise.

Generalizable models with transferrable application

The multimodal AI space is rapidly expanding but remains ultra-specific, hindering the transition of findings into general practices. Building models that translate across regions and hospitals without bias may be better explored through foundational models that 1) apply to multiple disease spaces, 2) inform future methodological decision-making by outlining the evidence for engineering decisions or by demonstrating that a method is effective beyond a single isolated case, and 3) prove multicenter validation for clinical use with resistance to bias.

This trend is becoming apparent as the unimodal clinical AI space becomes increasingly saturated, and the most impactful publications focus on foundational models through novel technical innovations, such as with DINO,⁸³ DINOv2,⁸⁴ and iBOT,⁸⁵ increasingly large datasets, and self-supervised learning to leverage unannotated data.⁸⁶ This generalizability has yet to become commonplace in multimodal AI, except for some

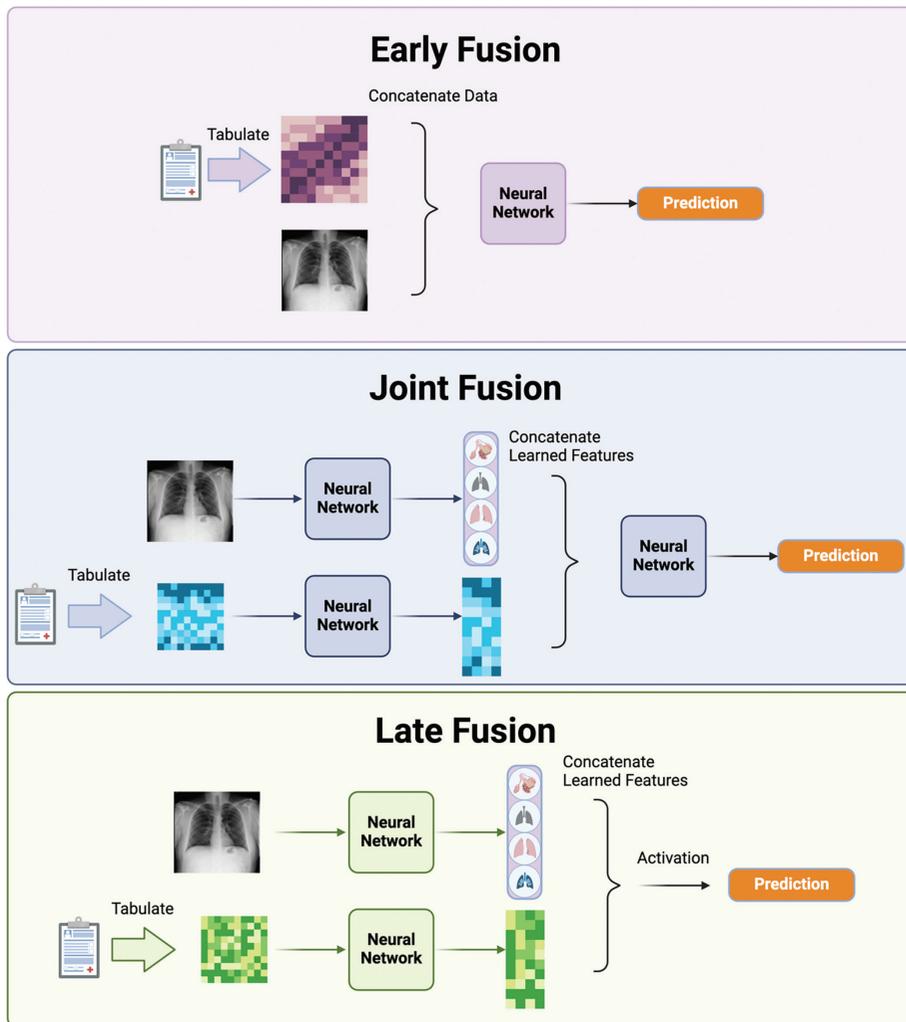


Figure 3. Fusion approaches for multimodal artificial intelligence (AI). Different approaches to training and fusion order are shown with examples in biomedical AI. In early fusion, a lung magnetic resonance image and tabulated electronic health record data are fused before learning. In joint fusion, one or both modalities undergo some learning prior to further learning and prediction. Finally, in late fusion, both modalities undergo all or almost all learning prior to fusion, activation, and prediction. There are many statistical integration methods beyond concatenation not shown.

key examples. For instance, Khader et al.²⁹ provide a compelling case for multimodal transformers by analyzing 25 conditions using imaging and non-imaging patient data from the Medical Information Mart for Intensive Care (MIMIC), instead of evaluating a single disease case. This publication is an impressive example of using up-to-date methodologies (namely, transformers), baseline comparison to alternate approaches for the same dataset, and analysis of various conditions. They observed improvement through multimodal use across all disease cases and reported appropriate statistical evaluation. Unfortunately, it is not common practice for multimodal papers to present statistics compared with a baseline unimodal model or present evidence of value in including both modalities. Rather, such papers often present a means to an end. Instead, Khader et al.²⁹ provided a case for a specific method, informing how future researchers should proceed while highlighting multiple translational impacts.

Another example study pushing towards generalizable multimodal approaches is proposed by Soenksen et al.⁷⁰ They propose and assess a model for Holistic AI in Medicine (HAIM) to support the general development and testing of a variety of multimodal AI systems. Leveraging the MIMIC database, they demonstrate improvement in predicting various healthcare operations including lung lesion detection, 48-hour mortality, and edema. They find that all multimodal inputs improve performance across all predictions. However, there is no statistical analysis presented to inform us which of these tasks shows a statistically significant difference. This work pushes the medical field towards cutting-edge and generalizable multimodal work and emphasizes the need to develop a standard of comparison in the field.⁷⁰

It is noteworthy, but not coincidental, that both models discussed above leverage the same MIMIC database. The MIMIC database is a publicly available repository of EHRs from the Beth Israel Deaconess Medical Center.^{87,88} Though each publication attempts to draw data from multiple sources, this highlights the issue of database bias in designing multimodal algorithms.

Dataset curation

Database bias can manifest in various ways. Based on an analysis of the existing terrain, several examples of bias where the field may be at risk are discussed. As other reviews^{8,89} and even the original MIMIC-IV publication⁸⁸ have stated, data in hospitals

today is typically stored in systems not conducive to or able to support research, especially data science research. Built for security and far behind modern standards for user interface design, storage, and ease of access, it is not uncommon to find scanned versions of electronic medical records as PDF-format files, equivalent information stored in various locations at different hospitals, and logging methods varying between physicians. In other words, there is a significant mismatch between the data format resulting from existing data collection practices across healthcare facilities and the data format necessary for appropriate AI development. These mismatches make it quite challenging to curate datasets such as MIMIC, which require careful planning, financial investment, and an industry-wide shift in how medical data is collected and stored. As a result, models are at risk of being overtrained on the limited existing AI-friendly data.

By using a single center or focusing on training with the handful of carefully curated datasets available, models can “learn” to treat all patients as they would in those specific settings and time periods, regardless of the quality of care one receives at their own institution and the clinical environment of which they are a part. Clinical outcomes can vary significantly depending on the surgeon, environmental exposures, or technology available. For example, patients at the best hospitals in the country may have different outcomes from average hospitals and therefore should be treated differently. Beyond social determinants of health, from a technical perspective, considering that MRI or CT scanners may differ across the country, a model may inadvertently learn that image quality is associated with outcomes or be unable to accurately assess certain images. As with comparing baseline unimodal models, there is a need for guidelines to assess and mitigate bias in AI as it becomes more widespread. Although there are examples of papers identifying or discussing bias,⁹⁰⁻⁹³ few propose analytical frameworks to address or measure bias in AI.⁹⁴ Such publications are varied, and none have become standard practice in the field. Few clinical papers assess bias in clinically specific AI models. Though not multimodal, a machine learning approach was proposed by Chandran et al.⁹⁵ to predict lung cancer risk using the cross-area under the receiver operator characteristic curve to measure disparities in performance by race and ethnicity. They identify key failures in the model’s ability to determine risk for Asian and Hispanic individuals compared

with White and non-Hispanic individuals. The mismatch between the clinical environment and AI-friendly data storage requirements results not only in bias but also makes bias reduction challenging, as curating “representative” data from centers across the entire country is a huge undertaking. The more representative the training data is of the setting in which it is applied, the lower the risk of biased decisions. With evidence that multimodal AI may be more accurate for some AI applications^{27,59-69,80,93} and that multimodal work is more challenging to curate consistently across institutions, researchers and physicians face the decision of how to build and employ AI tools when smaller multimodal sample sizes promise improved overall accuracy, but smaller sample size may increase risk of bias.

This concern is currently pressing and needs to be addressed. One systematic review on GNNs based on EHRs reported that out of 50 papers reviewed, 23 used MIMIC-III and 6 MIMIC-IV.⁹⁶ With the increasing prevalence of AI research and rapid translation of tools to the clinic, there is a need for a change in how data is stored and collected by healthcare providers across the country. Continuing to develop AI tools on the available pool of high-quality curated datasets, such as MIMIC,^{87,88} the UK Biobank,⁹⁷ EMBED,⁹⁸ and the Scottish Medical Imaging Archive,⁹⁹ is risky as tools may be carelessly applied to populations with differing clinical environments or health outcomes. Further, with the medical field being a rapidly changing ecosystem, models and datasets can quickly become less relevant to the current medical system.

Considering the dynamic medical environment and its quickly changing technology and guidelines, AI and the data on which it is trained will have to change as quickly as the clinic. One must be incredibly mindful of the dynamic nature of data when training an AI algorithm. Here, “dynamic” can take on a double meaning. Data can be dynamic in that its surrounding clinical environment changes as knowledge and technology develop. It can also be dynamic in that the information itself changes as a product of aging or biological changes. For example, considering genomics are stable over time, it is unclear what the significance is of their integration with dynamic data, such as an imaging phenotype or proteomics, which can change over a person’s life. Imaging data or radiomics data have been integrated both with stable omics¹⁰⁰ and dynamic omics for multimodal AI.¹⁰¹ Regardless of the biomed-

cal data and if it is dynamic, a change in how the data is collected from, and impacts, the clinic may be just as impactful on the creation of impactful AI as the data itself.

AI tools have the potential to both combat and exacerbate biases by providing evidence-based recommendations. Radiologists and other physicians must understand emerging and existing methods in the field, as well as the importance of data set curation, as they are often the ones making final decisions about how these tools will be used and how they will impact the patient. By being aware of the potential for AI to exacerbate biases, radiologists are relied upon to view these tools as exactly what they are: physician *support* tools. Even if a tool has a proven record of being more accurate than the average physician at, for example, detecting lesions on a certain type of scan, there will still be mistakes, and physicians will need to be able to use these AI tools without catering to their biases. It is difficult to predict exactly what the role of radiologists will be in the future of using and developing AI, but the reality is that it will play a role. The greater the degree to which these tools are understood is, the easier it will be for physicians to interact with them in a way that improves health. On the flip side, a greater understanding among physicians will allow them to conduct their clinic in a way that is conducive to storing data for training strong bias-mitigated models.

Future directions

Multimodal AI will inevitably continue to develop and be explored through the methodologies, foundational models, and translational integrations discussed in this review and beyond. Despite exploring highly developed architectures, methods, and techniques in image processing AI, such as fusion models, transformers, and GNNs, the medical field lags in using up-to-date AI innovations and struggles with consistency in taxonomy, evaluation metrics, and methodology, even within the same disease spaces.

The lack of common practices, which will develop and change as the field matures, severely limits progress and translation. It becomes difficult to generalize conclusions from one publication to the next and across methodologies. Standout publications in the multimodal AI space are characterized by their ability to generalize as foundational models with transferrable applications, incorporate physician perspectives with clear and broad clinical utility, and carefully eval-

uate baseline models using thorough and appropriate evaluation and statistics.

An even more pressing limitation in developing multimodal AI tools with biomedical applications is the lack of comprehensive, high-quality data. As discussed, most reviewed works rely on either a very small set of carefully curated data, which requires extensive time, resources, and funding for AI development, or they draw from a select set of high-quality, open-access datasets. By repeatedly using these same high-quality curated datasets, a suite of AI-based translational tools heavily biased toward the included locations, periods, and patient populations is being developed. With the clinical setting and its outcomes being a constantly changing ecosystem, it is risky to rely on the same datasets. Equitable, bias-free AI will require these systems to be dynamic, constantly updated with new data, and capable of adapting over time with fine-tuning. Technologists and clinicians may have to meet somewhere in the middle, such that technologists will have to build models using less-than-optimal data, and clinicians may have to incorporate certain practices into their data ecosystem to ensure AI models are up to date.

Our narrative review of multimodal AI, combining imaging and other clinical meta-data, aims to propose clarifications for what constitutes “multimodal” AI in imaging, identify up-to-date frameworks with potential for enhanced results in future model research, comment on a shift toward generalizable foundational models, and identify trends and concerns in database curation. As the field progresses from theory to clinic, it is essential for radiologists to stay informed about the latest developments, methodologies, and ethical implications.

The current radiologic landscape is characterized by a transition toward multimodal fusion models, with increasing focus on transformers and GNNs. However, there is a considerable amount of work to be done in terms of scientific due diligence regarding gaps in methodology and model training bias. Moreover, the reliance on the few existing high-quality curated datasets highlights a major risk as AI tools become more common in the clinical setting. There is an urgent need to align the format of data required for training AI with that logged by physicians to curate comprehensive training databases.

In conclusion, while AI in radiology promises significant advancements in the

field, successful and unbiased integration demands a multidisciplinary approach involving continuous education of physicians and AI developers alike. By informing radiologists, we hope to begin bridging the gap between technology and the clinic, guiding future methodologies, practices for dataset curation, and the field as a whole. By harnessing the power of AI, appropriate evaluation, and physician expertise, we hope to save more lives and improve the quality of care for patients worldwide.

Footnotes

Conflict of interest disclosure

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Barış Türkbey receives support as part of a CRADA between NIH and NVIDIA and between NIH and Philips Healthcare, receives royalties from NIH, and is a party to patents or potential patents related to this work. The remaining authors declare that there are no other disclosures relevant to the subject matter of this article. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. All figures were created and licensed using BioRender.com.

References

1. Eberhart RC, Dobbins RW. Early neural network development history: the age of Camelot. *IEEE Eng Med Biol Mag*. 1990;9(3):15-18. [\[CrossRef\]](#)
2. Temsah MH, Altamimi I, Jamal A, Alhasan K, Al-Eyadhy A. ChatGPT Surpasses 1000 Publications on PubMed: Envisioning the Road Ahead. *Cureus*. 2023;15(9):e44769. [\[CrossRef\]](#)
3. Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. *Cureus*. 2023;15(8):e43262. [\[CrossRef\]](#)
4. Li J. Security implications of AI chatbots in health care. *J Med Internet Res*. 2023;25:e47551. [\[CrossRef\]](#)
5. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digit Med*. 2022;5(1):48. [\[CrossRef\]](#)
6. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health*. 2020;2(9):e486-e488. [\[CrossRef\]](#)

7. Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med.* 2021;4(1):5. [\[CrossRef\]](#)
8. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med.* 2022;28(9):1773-1784. [\[CrossRef\]](#)
9. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. *NPJ Digit Med.* 2022;5(1):171. [\[CrossRef\]](#)
10. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. *Inf Fusion.* 2022;77:29-52. [\[CrossRef\]](#)
11. Lyu W, Dong X, Wong R, et al. A Multimodal transformer: fusing clinical notes with structured EHR data for interpretable in-hospital mortality prediction. *AMIA Annu Symp Proc.* 2022;2022:719-728. [\[CrossRef\]](#)
12. Rahman W, Hasan MK, Lee S, et al. Integrating multimodal information in large pretrained transformers. *Proc Conf Assoc Comput Linguist Meet.* 2020;2020:2359-2369. [\[CrossRef\]](#)
13. Rohanian O, Nouriborji M, Jauncey H, et al. Lightweight transformers for clinical natural language processing. *Nat Lang Eng.* 2024. [\[CrossRef\]](#)
14. Keicher M, Burwinkel H, Bani-Harouni D, et al. Multimodal graph attention network for COVID-19 outcome prediction. *Sci Rep.* 2023;13(1):19539. [\[CrossRef\]](#)
15. Zhou HY, Yu Y, Wang C, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat Biomed Eng.* 2023;7(6):743-755. [\[CrossRef\]](#)
16. Nguyen HH, Blaschko MB, Saarakkala S, Tiulpin A. Clinically-inspired multi-agent transformers for disease trajectory forecasting from multimodal data. *IEEE Trans Med Imaging.* 2024;43(1):529-541. [\[CrossRef\]](#)
17. Li B, Nabavi S. A multimodal graph neural network framework for cancer molecular subtype classification. *Bmc Bioinformatics.* 2024;25(1):27. [\[CrossRef\]](#)
18. Wen HZ, Ding JY, Jin W, Wang YQ, Xie YY, Tang JL. Graph neural networks for multimodal single-cell data integration. *Proceedings of the 28th Acm Sigkdd Conference on Knowledge Discovery and Data Mining, Kdd 2022.* 2022:4153-4163. [\[CrossRef\]](#)
19. Zhu JN, Oh JH, Simhal AK, et al. Geometric graph neural networks on multi-omics data to predict cancer survival outcomes. *Comput Biol Med.* 2023;163:107117. [\[CrossRef\]](#)
20. Gao J, Lyu T, Xiong F, Wang J, Ke W, Li Z. Predicting the survival of cancer patients with multimodal graph neural network. *IEEE/ACM Trans Comput Biol Bioinform.* 2022;19(2):699-709. [\[CrossRef\]](#)
21. Alam F, Ananbeh O, Malik KM, et al. towards predicting length of stay and identification of cohort risk factors using self-attention-based transformers and association mining: COVID-19 as a phenotype. *Diagnostics (Basel).* 2023;13(10):1760. [\[CrossRef\]](#)
22. Yang Z, Mitra A, Liu W, Berlowitz D, Yu H. TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nat Commun.* 2023;14(1):7857. [\[CrossRef\]](#)
23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neur In.* 2017:30. [\[CrossRef\]](#)
24. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. *JMLR Workshop Conf Proc.* 2016;56:301-318. [\[CrossRef\]](#)
25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota. Association for Computational Linguistics;* 2019;1:4171-4186. [\[CrossRef\]](#)
26. Nazir A, Wang Z. A Comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. *Meta Radiol.* 2023;1(2):100022. [\[CrossRef\]](#)
27. Khader F, Kather JN, Müller-Franzes G, et al. Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data. *Sci Rep.* 2023;13(1):10666. [\[CrossRef\]](#)
28. Zheng H, Lin Z, Zhou Q, et al. Multi-transSP: Multimodal transformer for survival prediction of nasopharyngeal carcinoma patients. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science.* Springer, Cham; 2022:234-243. [\[CrossRef\]](#)
29. Khader F, Müller-Franzes G, Wang TC, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. *Radiology.* 2023;309(1):e230806. [\[CrossRef\]](#)
30. Li Y, Rao S, Solares JRA, et al. BEHRT: transformer for electronic health records. *Sci Rep.* 2020;10(1):7155. [\[CrossRef\]](#)
31. Liu L, Liu S, Zhang L, To XV, Nasrallah F, Chandra SS. Cascaded multi-modal mixing transformers for alzheimer's disease classification with incomplete data. *Neuroimage.* 2023;277:120267. [\[CrossRef\]](#)
32. Yu Q, Ma Q, Da L, et al. A transformer-based unified multimodal framework for Alzheimer's disease assessment. *Comput Biol Med.* 2024;180:108979. [\[CrossRef\]](#)
33. Khader F, Franzen GM, Wang T, et al. Medical diagnosis with large scale multimodal transformers: leveraging diverse data for more accurate diagnosis. 2022. [\[CrossRef\]](#)
34. Guo J, Jia N, Bai J. Transformer based on channel-spatial attention for accurate classification of scenes in remote sensing image. *Sci Rep.* 2022;12(1):15473. [\[CrossRef\]](#)
35. Meng X, Zou T. Clinical applications of graph neural networks in computational histopathology: a review. *Comput Biol Med.* 2023;164:107201. [\[CrossRef\]](#)
36. Khemani B, Patil S, Kotecha K, Tanwar S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *J Big Data.* 2024;11(1):18. [\[CrossRef\]](#)
37. Zhou Y, Huo H, Hou Z, et al. Co-embedding of edges and nodes with deep graph convolutional neural networks. *Sci Rep.* 2023;13(1):16966. [\[CrossRef\]](#)
38. Zhang XM, Liang L, Liu L, Tang MJ. Graph neural networks and their current applications in bioinformatics. *Front Genet.* 2021;12:690049. [\[CrossRef\]](#)
39. Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. *Nat Biomed Eng.* 2022;6(12):1353-1369. [\[CrossRef\]](#)
40. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med.* 2020;3:136. [\[CrossRef\]](#)
41. Ding M, Cui H, Li B, et al. Integrating preoperative computed tomography and clinical factors for lymph node metastasis prediction in esophageal squamous cell carcinoma by feature-wise attentional graph neural network. *Int J Radiat Oncol Biol Phys.* 2023;116(3):676-689. [\[CrossRef\]](#)
42. Lyu T, Gao J, Tian L, Li Z, Zhang P, Zhang J. MDNN: a multimodal deep neural network for predicting drug-drug interaction events. 2021:3536-3542. [\[CrossRef\]](#)
43. Kumar S, Chaube MK, Alsamhi SH, et al. A novel multimodal fusion framework for early diagnosis and accurate classification of COVID-19 patients using X-ray images and speech signal processing techniques. *Comput Methods Programs Biomed.* 2022;226:107109. [\[CrossRef\]](#)
44. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer.* 2022;3(10):1151-1164. [\[CrossRef\]](#)
45. Subramanian V, Do MN, Syeda-Mahmood T. Multimodal fusion of imaging and genomics for lung cancer recurrence prediction. 2020:arXiv:2002.01982. [\[CrossRef\]](#)
46. Cearns M, Opel N, Clark S, et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl Psychiatry.* 2019;9(1):285. [\[CrossRef\]](#)
47. Cheng B, Liu M, Suk HI, Shen D, Zhang D; Alzheimer's disease neuroimaging initiative.

- Multimodal manifold-regularized transfer learning for MCI conversion prediction. *Brain Imaging Behav.* 2015;9(4):913-926. [\[CrossRef\]](#)
48. Peeken JC, Goldberg T, Pyka T, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med.* 2019;8(1):128-136. [\[CrossRef\]](#)
 49. Zhou H, Chang K, Bai HX, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *J Neurooncol.* 2019;142(2):299-307. [\[CrossRef\]](#)
 50. Brugnara G, Neuberger U, Mahmutoglu MA, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke.* 2020;51(12):3541-3551. [\[CrossRef\]](#)
 51. Qin H, Hu X, Zhang J, et al. Machine-learning radiomics to predict early recurrence in perihilar cholangiocarcinoma after curative resection. *Liver Int.* 2021;41(4):837-850. [\[CrossRef\]](#)
 52. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-point checklist and skin lesion classification using multi-task multi-modal neural nets. *IEEE J Biomed Health Inform.* 2018. [\[CrossRef\]](#)
 53. Spasov SE, Passamonti L, Duggento A, Lio P, Toschi N. A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. *Annu Int Conf IEEE Eng Med Biol Soc.* 2018;2018:1271-1274. [\[CrossRef\]](#)
 54. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology.* 2019;292(1):60-66. [\[CrossRef\]](#)
 55. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol.* 2018;27(11):1261-1267. [\[CrossRef\]](#)
 56. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol.* 2017;13(11):e1005752. [\[CrossRef\]](#)
 57. Singh A, Shannon CP, Gautier B, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics.* 2019;35(17):3055-3062. [\[CrossRef\]](#)
 58. Uppal K, Ma C, Go YM, Jones DP, Wren J. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics.* 2018;34(4):701-702. [\[CrossRef\]](#)
 59. Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci Rep.* 2020;10(1):22147. [\[CrossRef\]](#)
 60. Koutsouleris N, Dwyer DB, Degenhardt F, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry.* 2021;78(2):195-209. [\[CrossRef\]](#)
 61. Rajpurkar P, Park A, Irvin J, et al. AppendixNet: deep learning for diagnosis of appendicitis from a small dataset of CT exams using video pretraining. *Sci Rep.* 2020;10(1):3958. [\[CrossRef\]](#)
 62. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multimodal deep learning on randomized phase III clinical trials. *NPJ Digit Med.* 2022;6(1):71. Erratum in: *NPJ Digit Med.* 2023;6(1):27. [\[CrossRef\]](#)
 63. Janßen C, Boskamp T, LeClerc Arrastia J, et al. Multimodal lung cancer subtyping using deep learning neural networks on whole slide tissue images and MALDI MSI. *Cancers (Basel).* 2022;14(24):6181. [\[CrossRef\]](#)
 64. Khan RA, Fu M, Burbridge B, Luo Y, Wu FX. A multi-modal deep neural network for multi-class liver cancer diagnosis. *Neural Netw.* 2023;165:553-561. [\[CrossRef\]](#)
 65. Steyaert S, Qiu YL, Zheng Y, Mukherjee P, Vogel H, Gevaert O. Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. *Commun Med (Lond).* 2023;3(1):44. [\[CrossRef\]](#)
 66. Yao Y, Lv Y, Tong L, et al. ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. *Brief Bioinform.* 2022;23(6):bbac448. [\[CrossRef\]](#)
 67. Schulz S, Woerl AC, Jungmann F, et al. Multimodal deep learning for prognosis prediction in renal cancer. *Front Oncol.* 2021;11:788740. [\[CrossRef\]](#)
 68. Boehm KM, Aherne EA, Ellenson L, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat Cancer.* 2022;3(6):723-733. [\[CrossRef\]](#)
 69. Dayan I, Roth HR, Zhong A, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med.* 2021;27(10):1735-1743. [\[CrossRef\]](#)
 70. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med.* 2022;5(1):149. [\[CrossRef\]](#)
 71. Nai YH, Teo BW, Tan NL, et al. Evaluation of multimodal algorithms for the segmentation of multiparametric MRI prostate images. *Comput Math Methods Med.* 2020;2020:8861035. [\[CrossRef\]](#)
 72. Bleker J, Kwee TC, Dierckx RAJO, de Jong IJ, Huisman H, Yakar D. Multiparametric MRI and auto-fixed volume of interest-based radiomics signature for clinically significant peripheral zone prostate cancer. *Eur Radiol.* 2020;30(3):1313-1324. [\[CrossRef\]](#)
 73. Chen Q, Xu X, Hu S, Li X, Zou Q, Li Y. A transfer learning approach for classification of clinical significant prostate cancers from mpMRI scans. *Proc SPIE Int Soc Opt Eng.* 2017:10134. [\[CrossRef\]](#)
 74. Mehrtash A, Sedghi A, Ghafoorian M, et al. Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks. *Proc SPIE Int Soc Opt Eng.* 2017:10134. [\[CrossRef\]](#)
 75. Gutierrez Y, Arevalo J, Martinez F. Multimodal contrastive supervised learning to classify clinical significance MRI regions on prostate cancer. *Annu Int Conf IEEE Eng Med Biol Soc.* 2022;2022:1682-1685. [\[CrossRef\]](#)
 76. Nie D, Lu J, Zhang H, et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multimodal neuroimages. *Sci Rep.* 2019;9(1):1103. [\[CrossRef\]](#)
 77. Shim SO, Alkinani MH, Hussain L, Aziz W. Feature ranking importance from multimodal radiomic texture features using machine learning paradigm: a biomarker to predict the lung cancer. *Big Data Res.* 2022;29(28). [\[CrossRef\]](#)
 78. Dong Y, Hou L, Yang W, et al. Multi-channel multi-task deep learning for predicting EGFR and KRAS mutations of non-small cell lung cancer on CT images. *Quant Imaging Med Surg.* 2021;11(6):2354-2375. [\[CrossRef\]](#)
 79. Lu GX, Tian RH, Yang W, et al. Deep learning radiomics based on multimodal imaging for distinguishing benign and malignant breast tumours. *Front Med (Lausanne).* 2024;11:1402967. [\[CrossRef\]](#)
 80. Hu C, Qiao X, Huang R, Hu C, Bao J, Wang X. Development and validation of a multimodality model based on whole-slide imaging and biparametric MRI for predicting postoperative biochemical recurrence in prostate cancer. *Radiol Imaging Cancer.* 2024;6(3):e230143. [\[CrossRef\]](#)
 81. Mehralivand S, Yang D, Harmon SA, et al. A cascaded deep learning-based artificial intelligence algorithm for automated lesion detection and classification on biparametric prostate magnetic resonance imaging. *Acad Radiol.* 2022;29(8):1159-1168. [\[CrossRef\]](#)
 82. Simon BD, Merriman KM, Harmon SA, et al. Automated detection and grading of extraprostatic extension of prostate cancer at MRI via cascaded deep learning and random forest classification. *Acad Radiol.* 2024. [\[CrossRef\]](#)
 83. Caron M, Touvron H, Misra I, et al. Emerging properties in self-supervised vision transformers. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV 2021)*. 2021:9630-9640. [\[CrossRef\]](#)

84. Oquab M, Darcet T, Moutakanni T, et al. DINOv2: learning robust visual features without supervision. 2023;arXiv:2304.07193. Accessed April 01, 2023. [\[CrossRef\]](#)
85. Zhou J, Wei C, Wang H, et al. iBOT: image BERT pre-training with Online Tokenizer. 2021;arXiv:2111.07832. Accessed November 01, 2021. [\[CrossRef\]](#)
86. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-265. [\[CrossRef\]](#)
87. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035. [\[CrossRef\]](#)
88. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. [\[CrossRef\]](#)
89. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR data science institute artificial intelligence survey. *J Am Coll Radiol*. 2021;18(8):1153-1159. [\[CrossRef\]](#)
90. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021;27(12):2176-2182. [\[CrossRef\]](#)
91. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. [\[CrossRef\]](#)
92. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 2020;27(12):2020-2023. [\[CrossRef\]](#)
93. Kocak B, Ponsiglione A, Stanzione A, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)
94. Yang J, Soltan AAS, Eyre DW, Yang Y, Clifton DA. An adversarial training framework for mitigating algorithmic biases in clinical machine learning. *NPJ Digit Med*. 2023;6(1):55. [\[CrossRef\]](#)
95. Chandran U, Reys J, Yang R, Vachani A, Maldonado F, Kalsekar I. Machine learning and real-world data to predict lung cancer risk in routine care. *Cancer Epidemiol Biomarkers Prev*. 2023;32(3):337-343. [\[CrossRef\]](#)
96. Oss Boll H, Amirahmadi A, Ghazani MM, et al. Graph neural networks for clinical risk prediction based on electronic health records: A survey. *J Biomed Inform*. 2024;151:104616. [\[CrossRef\]](#)
97. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. Mar 2015;12(3):e1001779. [\[CrossRef\]](#)
98. Jeong JJ, Vey BL, Bhimireddy A, et al. The EMory BrEast imaging Dataset (EMBED): A racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. *Radiol Artif Intell*. 2023;5(1):e220047. [\[CrossRef\]](#)
99. Baxter R, Nind T, Sutherland J, et al. The scottish medical imaging archive: 57.3 million radiology studies linked to their medical records. *Radiol Artif Intell*. 2024;6(1):e220266. [\[CrossRef\]](#)
100. Huang YJ, Chen CH, Yang HC. AI-enhanced integration of genetic and medical imaging data for risk assessment of Type 2 diabetes. *Nat Commun*. 2024;15(1):4230. [\[CrossRef\]](#)
101. Lai J, Chen Z, Liu J, et al. A radiogenomic multimodal and whole-transcriptome sequencing for preoperative prediction of axillary lymph node metastasis and drug therapeutic response in breast cancer: a retrospective, machine learning and international multicohort study. *Int J Surg*. 2024;110(4):2162-2177. [\[CrossRef\]](#)



Machine learning models for discriminating clinically significant from clinically insignificant prostate cancer using bi-parametric magnetic resonance imaging

Hakan Ayyıldız¹
 Okan İnce²
 Esin Korkut³
 Merve Gülbiz Dağoğlu Kartal⁴
 Atadan Tunacı⁴
 Şükrü Mehmet Ertürk⁴

¹Kars Harakani State Hospital, Clinic of Radiology, Kars, Türkiye

²Rush University Medical Center, Department of Radiology, Division of Vascular and Interventional Radiology, Chicago, Illinois

³University of Health Sciences Türkiye, Bağcılar Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

⁴İstanbul University, İstanbul Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

Corresponding author: Hakan Ayyıldız

E-mail: hakanayildiz77@gmail.com

Received 25 May 2024; revision requested 28 June 2024; last revision received 14 August 2024; accepted 29 August 2024.



Epub: 01.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242856

PURPOSE

This study aims to demonstrate the performance of machine learning algorithms to distinguish clinically significant prostate cancer (csPCa) from clinically insignificant prostate cancer (ciPCa) in prostate bi-parametric magnetic resonance imaging (MRI) using radiomics features.

METHODS

MRI images of patients who were diagnosed with cancer with histopathological confirmation following prostate MRI were collected retrospectively. Patients with a Gleason score of 3+3 were considered to have clinically ciPCa, and patients with a Gleason score of 3+4 and above were considered to have csPCa. Radiomics features were extracted from T2-weighted (T2W) images, apparent diffusion coefficient (ADC) images, and their corresponding Laplacian of Gaussian (LoG) filtered versions. Additionally, a third feature subset was created by combining the T2W and ADC images, enhancing the analysis with an integrated approach. Once the features were extracted, Pearson's correlation coefficient and selection were performed using wrapper-based sequential algorithms. The models were then built using support vector machine (SVM) and logistic regression (LR) machine learning algorithms. The models were validated using a five-fold cross-validation technique.

RESULTS

This study included 77 patients, 30 with ciPCA and 47 with csPCA. From each image, four images were extracted with LoG filtering, and 111 features were obtained from each image. After feature selection, 5 features were obtained from T2W images, 5 from ADC images, and 15 from the combined dataset. In the SVM model, area under the curve (AUC) values of 0.64 for T2W, 0.86 for ADC, and 0.86 for the combined dataset were obtained in the test set. In the LR model, AUC values of 0.79 for T2W, 0.86 for ADC, and 0.85 for the combined dataset were obtained.

CONCLUSION

Machine learning models developed with radiomics can provide a decision support system to complement pathology results and help avoid invasive procedures such as re-biopsies or follow-up biopsies that are sometimes necessary today.

CLINICAL SIGNIFICANCE

This study demonstrates that machine learning models using radiomics features derived from bi-parametric MRI can discriminate csPCa from clinically insignificant PCa. These findings suggest that radiomics-based machine learning models have the potential to reduce the need for re-biopsy in cases of indeterminate pathology, assist in diagnosing pathology-radiology discordance, and support treatment decision-making in the management of PCa.

KEYWORDS

Prostate, magnetic resonance imaging, prostate cancer, radiomics, machine learning, bi-parametric magnetic resonance imaging

You may cite this article as: Ayyıldız H, İnce O, Korkut E, Dağoğlu Kartal MG, Tunacı A, Ertürk ŞM. Machine learning models for discriminating clinically significant from clinically insignificant prostate cancer using bi-parametric magnetic resonance imaging. *Diagn Interv Radiol.* 2025;31(4):313-320.

Prostate cancer (PCa) is the second most common cancer in men, with a rising incidence.¹ The prostate-specific antigen (PSA) test remains a commonly used screening method, although recent studies suggest it has a limited impact on survival outcomes.^{2,3} In the modern medical landscape, the significance of prostate imaging, particularly with magnetic resonance imaging (MRI), has grown. Imaging plays an important role in the diagnosis of PCa, and multiparametric prostate magnetic resonance imaging (mpMRI) is the most commonly used imaging modality for diagnosis. Different versions of Prostate Imaging–Reporting and Data System (PI-RADS[®]) have been published to standardize mpMRI radiology reports.⁴ In cases where the Gleason score is 6, the Gleason grade group (GGG) is 1, and in cases where the Gleason score is 3+4 or higher, the GGG is 2 or above. In PCa, the prognosis is expected to be better if GGG = 1.⁵ However, active surveillance can be applied to patients with GGG <2.⁶ Treatment management varies with the GGG. Although dynamic contrast-enhanced imaging is considered a “safety zone,” bi-parametric magnetic resonance imaging (bpMRI) is increasingly favored due to its speed and, in some studies, comparable diagnostic performance to mpMRI.^{7,8} PI-RADS[®] version 2.1 indicates that bpMRI may be a viable option for decreasing the use of gadolinium contrast media, associated adverse reactions, and examination time.⁴ This

can lead to greater accessibility to prostate MRI for patients. Nevertheless, the PI-RADS[®] version 2.1 suggests mpMRI for patients with a high likelihood of cancer based on factors such as PSA levels, family history, or genetic predisposition. It also recommends mpMRI in cases where image quality may be compromised, such as in patients with hip prostheses.

In PI-RADS[®] version 2.1, for clinically significant prostate cancer (csPCa), at least one of the following must be present: GGG >2, volume >0.5 cc, or extra-prostatic extension. As a result, frequent distinction between clinically significant and clinically insignificant prostate cancer (ciPCa) is achieved by histopathological verification as a result of prostate biopsy, which is an invasive method. Gleason score may need to be re-evaluated by pathology in some cases.⁹ In our study, we aimed to show the role of machine learning algorithms created from radiomics features obtained from T2-weighted (T2W) and apparent diffusion coefficient (ADC) sequences in MRI. A review of the literature reveals numerous machine learning-based studies on PCa detection, particularly csPCa detection.^{10,11} Our study aims to make a significant contribution to the existing literature by providing an easily applicable, reproducible, and more accurate model that facilitates the distinction between csPCa and ciPCa. This model is particularly focused on improving the management of patient populations who may require multiple biopsies over time. In daily practice, this could impact a considerable number of patients.

date: 28/05/2021). Since it was a retrospective study, informed consent was waived. The dataset was obtained by retrospectively scanning the images of patients >18 years of age, whose lesions detected after mpMRI were confirmed histopathologically by systematic core and targeted biopsy at the department of radiology of the institution between 2016 and 2022. Fusion biopsy, which combines MR and ultrasound imaging, was used as the biopsy technique in all patients. The data were obtained on a lesion-by-lesion basis to avoid possible bias during the data collection. The exclusion criteria encompassed the following conditions: 1) elimination of cases with imaging artifacts that hindered the accurate segmentation of cancer lesions and 2) exclusion of instances with incomplete MRI data, including scenarios where essential images were missing.

Magnetic resonance imaging

The primary MRI sequences chosen for radiomic input in prostate imaging included axial T2W images and ADC images. In this study, two distinct MR technologies were employed: the Magnetom Aera 1.5 Tesla (Siemens Healthcare, Germany) and the Achieva 3.0 Tesla (Philips Medical Systems, the Netherlands). ADC images were acquired from diffusion-weighted imaging (DWI) with a b-value of 0 and 1400 sec/mm² on both devices. For the axial T2W images and DWI protocols, the repetition time/echo time parameters were set at 7500/100 and 5000/70 ms for the 1.5 Tesla system and 4200/100 and 3600/70 ms for the 3.0 Tesla system. The field of view (FOV) differed between the devices, with an 18 mm x 18 mm FOV for the 3.0 Tesla system and a 20 mm x 20 mm FOV for the 1.5 Tesla system. A slice thickness of 3.0 mm was maintained consistently on both devices, with no slice gap, to ensure homogeneity in imaging parameters throughout the study. Table 1 provides detailed parameters related to MRI.

Main points

- This study employed machine learning algorithms [support vector machine (SVM) and logistic regression (LR)] to differentiate between clinically significant prostate cancer (csPCa) and clinically insignificant prostate cancer (ciPCa) using radiomics features from bi-parametric magnetic resonance imaging images.
- Feature selection yielded 5 key features from T2-weighted (T2W) images, 5 from apparent diffusion coefficient (ADC) images, and 15 from the combined dataset, which was critical for model accuracy.
- A total of 77 patients were analyzed, with the SVM model achieving area under the curve (AUC) values of 0.64 for T2W, 0.86 for ADC, and 0.86 for combined images, whereas the LR model achieved AUC values of 0.79 for T2W, 0.86 for ADC, and 0.85 for combined images.
- The findings suggest that machine learning models using radiomics can significantly aid in distinguishing csPCa from ciPCa, potentially reducing the need for invasive biopsy procedures.

Methods

Ethics and data source

This study was approved by the Istanbul University, Istanbul Faculty of Medicine Ethics Committee (decision number: 2021/676,

Table 1. Parameters of the magnetic resonance imaging

	T2W (1.5 Tesla system)	DWI (1.5 Tesla system)	T2W (3.0 Tesla system)	DWI (3.0 Tesla system)
TR	7500	5000	4200	3600
TE	100	70	100	70
FOV area	20 mm x 20 mm	20 mm x 20 mm	18 mm x 18 mm	18 mm x 18 mm
Matrix	320 x 320	256 x 256	230 x 180	64 x 64
Voxel size (x, y, z; mm)	0.6 x 0.6 x 3 (mm)	0.8 x 0.8 x 3 (mm)	0.8 x 1 x 3 (mm)	3 x 3 x 3 (mm)
Slice thickness	3 mm	3 mm	3 mm	3 mm
Slice gap	-	-	-	-
Sequence	Turbo spin echo	Echo planar imaging	Turbo spin echo	Echo planar imaging

T2W, T2-weighted; DWI, diffusion-weighted imaging; TR, repetition time; TE, echo time; FOV, field of view.

Image preprocessing and feature extraction

The acquired images underwent normalization through the proprietary algorithm embedded in Olea Sphere® software (Olea Medical, La Ciotat, France). Despite ADC being a computationally derived sequence, it underwent normalization in a manner consistent with the axial T2W series, aligning with the recommendations in radiomics studies.¹² Subsequently, outlier pixels were systematically eliminated using the ± 3 sigma technique.¹³ Following normalization and the removal of outlier pixels, pixel sizes were standardized to a $1 \times 1 \text{ mm}^2$ scale using cubic B-spline interpolation. The gray levels were then discretized uniformly for both series with a bin width of 0.05.¹⁴ Utilizing PyRadiomics, Laplacian of Gaussian (LoG) filter images were extracted from the original images with logarithmic values of 2, 4, and 6. Consequently, four images were derived from a single original image, where one of them represented the original unaltered image.

Segmentations were performed manually by two radiologists using the freehand method, prior to the steps described in the previous paragraph. Each radiologist had 5 and 4 years of experience, respectively, and performed the segmentations independently using axial T2W and ADC images (Figure 1). When necessary for improved tumor orientation, DWI with b-values of 0 and 1400, as well as sagittal T2W images, were incorporated. However, for objective bi-parametric modeling, contrast-enhanced series were intentionally omitted and not reviewed during the segmentation process. During segmentation, the lesion area with high suspicion of tumor was included, whereas areas of uncertainty were excluded. The suspicious lesion underwent volumetric 3D segmentation using Olea Sphere® software. Subsequently, feature extraction was performed from the original image, as well as from three LoG-filtered series within each set, following the steps described in the previous paragraph. The radiomics workflow is summarized in Figure 2.

Data preprocessing and feature selection

To ensure consistency and dependability of machine learning models, meticulous data pre-processing steps were performed.¹⁵ After standardization and discretization were applied uniformly to all data using a consistent bin width, the data were divided into 20 bins. The dataset was randomly split into training and test sets with a 70/30 ratio. To prevent

contamination of the test dataset with the training dataset, data splitting was conducted before any data augmentation. This approach ensured the integrity and independence of the training and test datasets, and T2W and ADC series were combined to construct a unified dataset.

Pearson's correlation coefficient was employed to identify and remove redundant features. Pairs of features with a correlation coefficient exceeding a threshold of 0.80 were identified and subsequently removed.¹⁶ The remaining features, which met these criteria, served as input for the next stage.

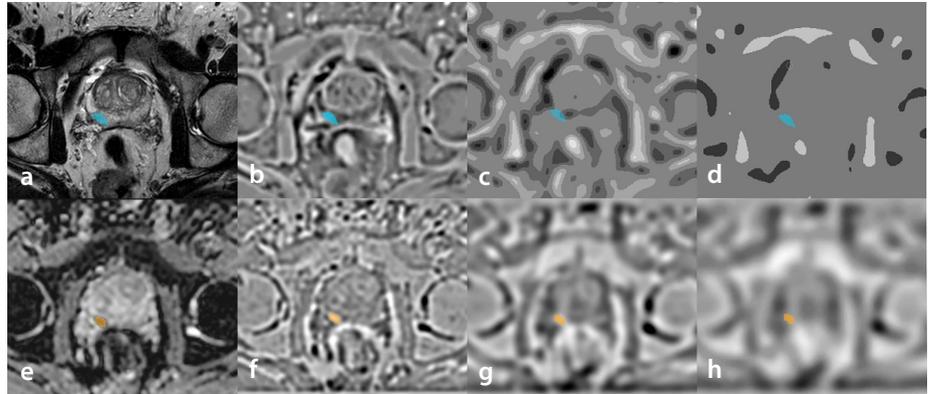


Figure 1. The figure shows examples of the segmentation of the original (a) T2-weighted images with Laplacian of Gaussian (LoG) filters using sigma values of 2 (b), 4 (c), and 6 (d), as well as the original (e) apparent diffusion coefficient images with LoG filters using sigma values of 2 (f), 4 (g), and 6 (h).

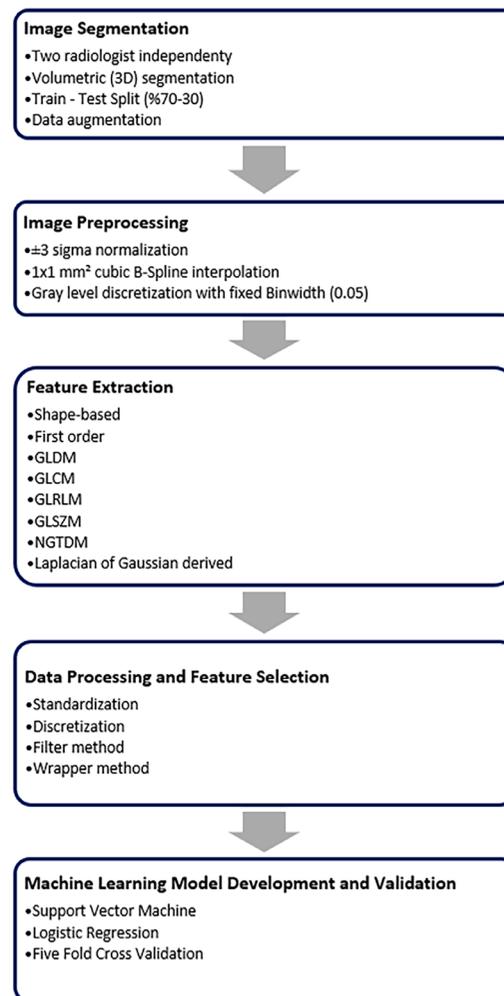


Figure 2. The figure illustrates the radiomics workflow. GLDM, gray-level dependence matrix; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray-tone difference matrix.

A wrapper-based feature selection algorithm was then developed using backward elimination and five-fold cross-validation. Logistic regression (LR) was chosen as the estimator for feature selection. This wrapper method evaluates different models by iteratively including or excluding features to determine the optimal feature combination. Each model was analyzed by iteratively removing one feature at a time. Through multiple evaluations, the most relevant features were identified. Crucial features were selected exclusively from the training folds using cross-validation, thereby avoiding the “double-dipping” phenomenon.¹⁷ As previously indicated, the test set remained untouched throughout the feature selection process due to the prior data division into training and test sets.

Machine learning algorithms

T2W, ADC, and the combined dataset were incorporated into the machine learning modeling. The finalized set of features was used for implementing machine learning algorithms, which were executed using Python (version 3). The first model employed was a support vector machine (SVM) with hyperparameters set to C: 1.0 and kernel: linear. Another model, LR, was used with hyperparameters configured as C: 25, solver: liblinear, and regularization penalty: L2 (Ridge). The performance of the models was evaluated using five-fold cross-validation. Metrics including accuracy, sensitivity, specificity, precision, recall, F1 score, and the area under the curve (AUC) were calculated.

Results

Patients

The study involved a total of 108 patients. However, 14 patients were excluded due to incomplete pathology results, 3 patients had incomplete images, and 14 patients had artifacts in their images (Figure 3). Of the remaining patients, 61% (47 patients) were diagnosed with csPCa, whereas 39% (30 patients) were classified as having ciPCa. Table 2 provides a summary of the patients' characteristics.

Feature extraction and selection

A total of 444 features were extracted from each sequence. These features were categorized as follows: 17 (15.32%) shape, 19 (17.12%) first-order, 24 (21.62%) gray-level co-occurrence matrix, 16 (14.41%) gray-level run-length matrix, 16 (14.41%) gray-level

size-zone matrix, 14 (12.61%) gray-level dependence matrix, and 5 (4.50%) neighboring gray-tone difference matrix features. Subsequently, a combined dataset was generated by concatenating features from both T2W and ADC sequences.

Pearson's correlation coefficient identified 28, 31, and 50 features as non-overlapping in T2W, ADC, and the combined group, respectively. Following the wrapper-based sequential feature selection step, the final feature subsets consisted of 5 features in T2W, 5 in ADC, and 15 in the combined group, details of which are shown in Table 3 and Figure 4.

Models performance

The SVM models demonstrated accuracy scores of 75%, 85%, and 91% in the training group and 64%, 76%, and 72% in the test group for the T2W, ADC, and combined groups, respectively. The corresponding AUC

values with 95% confidence intervals (CI) were 0.75 (0.74–0.76), 0.89 (0.88–0.89), and 0.95 (0.95–0.96) in the training group, and 0.64 (0.62–0.65), 0.86 (0.85–0.88), and 0.86 (0.85–0.88) in the test group for the T2W, ADC, and combined groups, respectively.

The LR models in the T2W, ADC, and combined groups had accuracy scores of 74%, 84%, and 86% in the training group, and 70%, 79%, and 77% in the test group, respectively. The AUC values with 95% CI were as follows: for the T2W, ADC, and combined groups in the training group, 0.83 (0.82–0.83), 0.89 (0.88–0.89), and 0.95 (0.94–0.95); and in the test group, 0.79 (0.78–0.80), 0.86 (0.84–0.88), and 0.85 (0.83–0.87), respectively. Detailed performance analyses for the training group and the test group are presented in Table 4, and Figure 5 shows the receiver operating characteristic curves for all models.

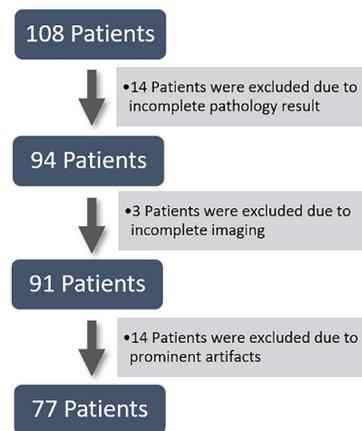


Figure 3. The figure demonstrates the patient selection algorithm. (F, feature; as listed in Table 2).

Table 2. Demographic and patient characteristics for both groups

	csPCa	ciPCa	P value
Age (mean ± SD) (95% CI)	65.22 ± 8.85 (62.59–67.84)	61.61 ± 6.8 (58.97–64.24)	0.086
PSA level (median) (min–max)	7.46 (1.22–38.67)	5.95 (2.0–45.0)	0.044*
Localization (n)			
Peripheral zone	40	24	0.560
Transitional zone	7	6	
MRI technology			
1.5 T scanner	31	19	0.009†
3.0 T scanner	16	11	
Gleason score (n)			
Gleason 3+3	NA	30	
Gleason 3+4	24	NA	
Gleason 4+3	9	NA	
Gleason 4+4	9	NA	
Gleason 4+5	3	NA	
Gleason 5+4	1	NA	
Gleason 5+5	1	NA	

*A significant difference was found between both groups by Mann–Whitney U test revealing a higher value in the csPCa group. †1.5 Tesla scanners have a higher number of patients and a significant difference was found in the Pearson's chi-square test. csPCa, clinically significant prostate cancer; ciPCa, clinically insignificant prostate cancer; SD, standard deviation; CI, confidence interval; PSA, prostate-specific antigen; min–max, minimum–maximum; MRI, magnetic resonance imaging; NA, not available.

Discussion

In our investigation, the efficacy of machine learning models employing prostate bpMRI radiomics analysis for predicting csPCa was explored, revealing promising predictive capabilities. As the two different algorithms work on different principles, an attempt was made to minimize the possibility of overfitting by using them in the algorithms created and to evaluate the usability of the different models. The reasonable and comparable accuracy rates of these algorithms in this study demonstrate the feasibility of using machine learning algorithms to identify csPCa.

In the literature, radiomics studies conducted using ultrasonography and computed tomography in prostate imaging are

available.^{18,19} Nevertheless, the popularity of radiomics studies in prostate MRI is notably increasing. The field of radiomics studies conducted in MR is expansive, encompassing endeavors to differentiate extraprostatic extension, discern normal tissue from cancer, identify recurrence post-radical prostatectomy, and distinguish recurrence after treatment.²⁰ Notably, the treatment approaches for csPCa and ciPCa can vary significantly.²¹⁻²³ Hence, there is a growing significance in conducting studies aimed at differentiating csPCa and ciPCa. Some of these studies have been performed with mpMRI and some with bpMRI. Our study was conducted with bpMRI, which is more accessible, has a shorter duration, and is considered suitable for acquisition with certain criteria in PI-RADS[®] version 2.1, and studies are being conducted to disseminate it.⁴

Zhang et al.²⁴ used GGG 1 and GGG >1 groups in their nomogram study of 159 patients with radiomics, similar to our study. Similar to our study, only bpMRI images were used, and segmentation was performed on DWI, ADC, and T2W. Although the use of internal validation was the advantage of the study, this study was performed with a single 3.0T MR technology. In addition, this study was performed with a radiomic nomogram, and machine learning modeling was not applied. In a retrospective radiomics study of 489 patients, Gong et al.²⁵ derived models from bpMRI data (T2W and DWI). They incorporated clinical modeling by including PSA data in the study. Performed on a single 3.0T MRI machine, they reported an AUC of 0.811 in the training group and 0.788 in the test group for the combined model, which was created without integrating clinical

Table 3. Selected features and their classifications for T2W, ADC, and combined datasets

Selected features					
T2W		ADC		Combined dataset	
Image type	Feature name (feature class)	Image type	Feature name (feature class)	Image type	Feature name (feature class)
Original	Original shape surface area to volume ratio (shape)	Original	Original shape mesh volume (shape)	T2W - original	Original shape surface area to volume ratio (shape)
Original	Original shape sphericity (shape)	Original	Original shape surface area to volume ratio (shape)	T2W - original	Original shape sphericity (shape)
Original	Original first order root mean squared (first order)	Original	<i>Original first order entropy (shape)</i>	T2W - original	Original shape elongation (shape)
Original	Original GLCM correlation (GLCM)	Original	Original first order skewness (first order)	T2W - original	Original shape flatness (shape)
LoG (Sigma: 4)	Original GLCM informal measure of correlation 1 (GLCM)	LoG (Sigma: 4)	Original first order kurtosis (first order)	T2W - original	Original GLCM correlation (GLCM)
				T2W - laplacian of gaussian (Sigma: 2)	Original first order 90 th percentile (first order)
				ADC - original	<i>Original first order entropy (first order)</i>
				ADC - original	Original first order minimum (first order)
				ADC - original	Original GLCM inverse variance (GLCM)
				ADC - original	Original GLSZM small area emphasis (GLSZM)
				ADC - original	Original GLDM large dependence low gray level emphasis (GLDM)
				ADC - LoG (Sigma: 4)	Original GLCM inverse variance (GLCM)
				ADC - LoG (Sigma: 6)	Original GLCM informal measure of correlation 1 (GLCM)
				ADC - LoG (Sigma: 6)	Original GLCM maximal correlation coefficient (GLCM)
				ADC - LoG (Sigma: 6)	Original GLSZM variance (GLSZM)

The same features included in the combined group and T2W are shown in bold; the combined group and ADC are shown in italic. T2W, T2-weighted; ADC, apparent diffusion coefficient; GLCM, gray-level co-occurrence matrix; LoG, Laplacian of Gaussian; GLSZM, gray-level size-zone matrix; GLDM, gray-level dependence matrix.

modeling. However, in this study, PCa was separated into low-grade and high-grade, and patients with a Gleason score <8 were considered low-grade. Li et al.²⁶ used T2W and ADC series in their retrospective study of 381 patients to differentiate csPCa, but 199 patients were selected from the benign patient group. Clinical modeling was also included in the study, and they reported the AUC value obtained without clinical modeling as 0.99 in the training group and 0.98 in the test group. However, this study used a single MR scanner, and lesion segmentation was supported by pathological data and dynamic contrast-enhanced images.²⁶

In current clinical practice, almost all patients with suspected PCa require a biopsy. The advantage of conducting our study only with patients who have a Gleason score of 6 or higher is to avoid the possibility that these patients, diagnosed with cancer, might require re-biopsies during follow-up under current conditions or even immediately after the initial biopsy. Thus, the aim is to create a decision support system to aid the pathology result or to identify patients who need re-biopsy. The use of two MR scanners, 3.0T and 1.5T, and the modeling of both peripheral and transitional zone lesions are advantageous in our study. In addition to its rapid applicability, another advantage of our model for bpMRI over other studies is that it relies solely on ADC series and does not use contrast-enhanced sequences. Furthermore, the significance of our study lies in the high reproducibility of the technique, along with its favorable accuracy rates and AUC values, which are relatively high compared to other studies.²⁷ Other studies in the literature used more images than T2 and ADC and achieved similar accuracy rates to those in our study.¹¹ Additionally, some studies with high accuracy rates focused only on the peripheral or transitional zones. For instance, Fehr et al.²⁸ reported high accuracy rates but performed segmentation in conjunction with pathological results. Chen et al.²⁹ also reported high accuracy rates, but their study did not perform an interobserver analysis.

Our study has several limitations. First, as a retrospective study sourcing data from past registries, it may introduce selection bias. Second, although segmentation was performed independently by two radiologists, the manual nature of this process can introduce subjectivity. Third, the patient population was relatively small, raising concerns about a potential imbalance between groups. Class imbalance can challenge many machine learning algorithms, which typically

assume that all classes are equally distributed.¹⁵ In cases of imbalanced classes, models tend to favor predictions for the majority class. To address class imbalance and reduce the risk of overfitting, especially with a limited number of samples, data augmentation is a

validated technique. The use of different synthetic over-sampling methods can provide a more efficient and effective approach.^{30,31} However, this would result in a substantial portion of the data being synthetic. Furthermore, despite employing a systematic and

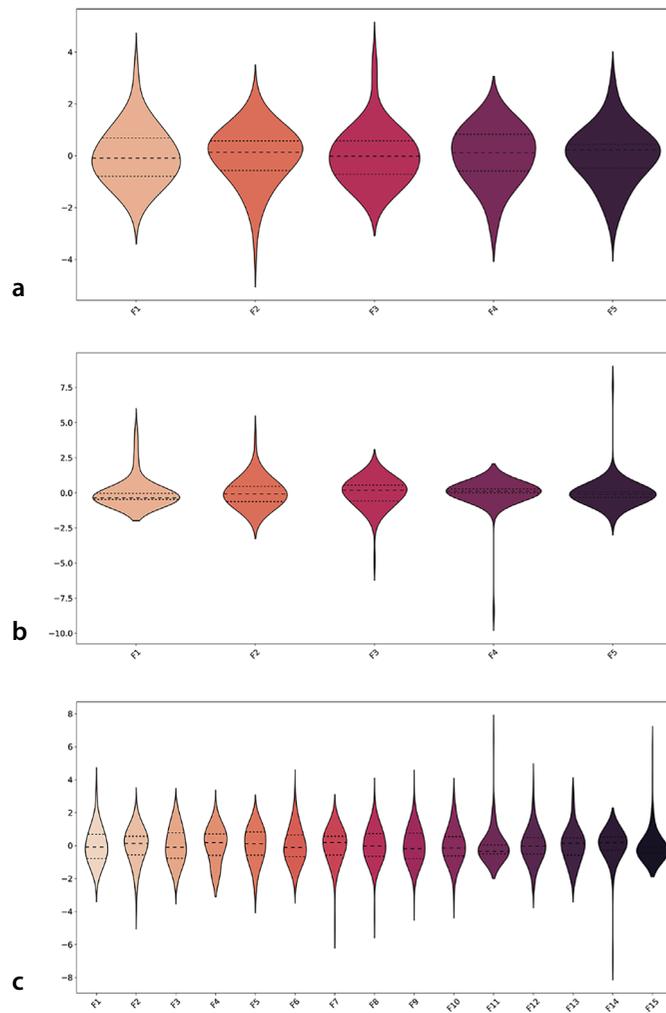


Figure 4. The selected features for T2-weighted images (a), apparent diffusion coefficient images (b), and the combined dataset (c) are shown.

Table 4. Detailed performance statistics for machine learning algorithms on T2W, ADC, and combined dataset

	Group	Accuracy	Sensitivity	Specificity	Recall	F1	AUC (95% CI)
LR-T2W	Train	74%	77%	57%	84%	0.80	0.83 (0.82–0.83)
	Test	70%	76%	56%	79%	0.76	0.79 (0.78–0.80)
SVM-T2W	Train	75%	77%	52%	85%	0.81	0.75 (0.74–0.76)
	Test	64%	69%	46%	75%	0.71	0.64 (0.62–0.65)
LR-ADC	Train	84%	85%	69%	90%	0.87	0.89 (0.88–0.89)
	Test	79%	82%	67%	87%	0.84	0.86 (0.84–0.88)
SVM-ADC	Train	85%	85%	65%	90%	0.88	0.89 (0.88–0.89)
	Test	76%	80%	63%	85%	0.82	0.86 (0.85–0.88)
LR-combined	Train	86%	89%	73%	95%	0.92	0.95 (0.94–0.95)
	Test	77%	85%	70%	81%	0.80	0.85 (0.83–0.87)
SVM-combined	Train	91%	90%	68%	95%	0.93	0.95 (0.95–0.96)
	Test	72%	75%	66%	75%	0.75	0.86 (0.85–0.88)

T2W, T2-weighted image; ADC, apparent diffusion coefficient; F1 score, the harmonic mean of precision and recall; AUC, area under the curve; CI, confidence interval; LR, logistic regression; SVM, support vector machine.

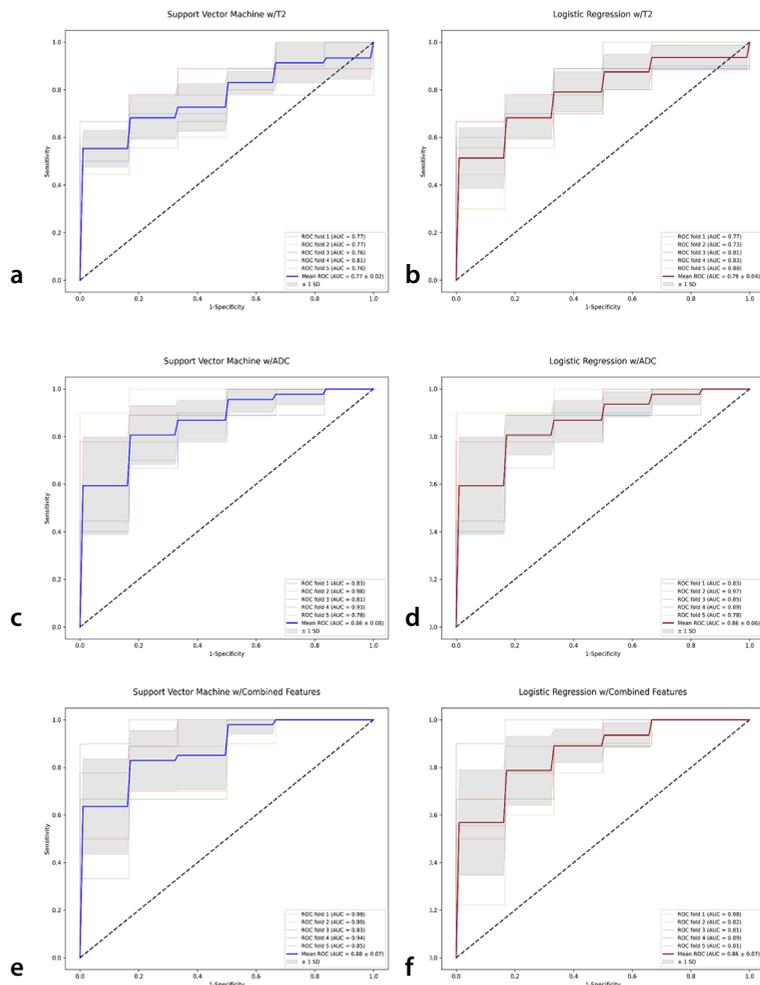


Figure 5. The figure presents the ROC curves, illustrating the ability of the models to differentiate between clinically significant and clinically insignificant prostate cancer across T2W, ADC, and combined datasets. ROC, receiver operating characteristic; T2W, T2-weighted; ADC, apparent diffusion coefficient; AUC, area under the curve; SD, standard deviation.

targeted biopsy approach, the accuracy of the PCa score may be underestimated due to potential limitations in puncture pathology, which might not accurately reflect the true pathological status. Additionally, in PI-RADS® version 2.1, the criteria for csPCa include extraprostatic extension and volume criteria, in addition to the Gleason score. Although patients were selected retrospectively, care was taken to exclude those meeting this criterion from the ciPCa group.

In conclusion, machine learning models utilizing radiomics extracted from prostate bpMRI show promising results in distinguishing between csPCa and ciPCa. However, additional studies with larger datasets are needed to validate these models across external centers before considering their clinical implementation. Incorporating clinical data, such as PSA levels, into these models could lead to the development of more robust tools for clinical practice. The integration of radiomics with artificial intelligence

methodologies, including machine learning, holds significant potential for future advancements in prostate imaging.

Footnotes

Conflict of interest disclosure

Şükrü Mehmet Ertürk, MD, is Publication Coordinator in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int J Cancer*. 2015;136(5):359-386. [\[Crossref\]](#)
2. Ilic D, Djulbegovic M, Jung JH, et al. Prostate cancer screening with prostate-specific antigen (Psa) test: a systematic review

and meta-analysis. *BMJ*. 2018;5:362:k3519.

[\[Crossref\]](#)

3. Eldred-Evans D, Burak P, Connor MJ, et al. Population-based prostate cancer screening with magnetic resonance imaging or ultrasonography: the Ip1-prostagram study. *JAMA Oncol*. 2021;7(3):395-402. [\[Crossref\]](#)
4. 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. [\[Crossref\]](#)
5. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (Isup) consensus conference on gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol*. 2016;40(2):244-252. [\[Crossref\]](#)
6. Evans AJ. Treatment effects in prostate cancer. *Mod Pathol*. 2018;31(Suppl 1):110-121. [\[Crossref\]](#)
7. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Head-to-head comparison between biparametric and multiparametric MRI for the diagnosis of prostate cancer: a systematic review and meta-analysis. *AJR Am J Roentgenol*. 2018;211(5):226-241. [\[Crossref\]](#)
8. Niu XK, Chen XH, Chen ZF, Chen L, Li J, Peng T. Diagnostic performance of biparametric MRI for detection of prostate cancer: a systematic review and meta-analysis. *AJR Am J Roentgenol*. 2018;211(2):369-378. [\[Crossref\]](#)
9. van Santvoort BWH, van Leenders GJLH, Kiemeny LA, et al. Histopathological re-evaluations of biopsies in prostate cancer: a nationwide observational study. *Scand J Urol*. 2020;54(6):463-469. [\[Crossref\]](#)
10. T JMC, Arif M, Niessen WJ, Schoots IG, Veenland JF. Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers (Basel)*. 2020;12(6):1606. [\[Crossref\]](#)
11. Cuocolo R, Cipullo MB, Stanzione A, et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. *Eur Radiol*. 2020;30(12):6877-6887. [\[Crossref\]](#)
12. Becker AS, Wagner MW, Wurnig MC, Boss A. Diffusion-weighted imaging of the abdomen: impact of b-values on texture analysis features. *NMR Biomed*. 2017;30(1). [\[Crossref\]](#)
13. 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. [\[Crossref\]](#)
14. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep*. 2018;8(1):10545. [\[Crossref\]](#)
15. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a

- practical guide for beginners. *Diagn Interv Radiol*. 2019;25(6):485-495. [\[Crossref\]](#)
16. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 2013;36(1):27-46. [\[Crossref\]](#)
 17. Ball TM, Squeglia LM, Tapert SF, Paulus MP. Double dipping in machine learning: problems and solutions. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(3):261-263. [\[Crossref\]](#)
 18. Wildeboer RR, Mannaerts CK, van Sloun RJG, et al. Automated multiparametric localization of prostate cancer based on B-mode, shear-wave elastography, and contrast-enhanced ultrasound radiomics. *Eur Radiol*. 2020;30(2):806-815. [\[Crossref\]](#)
 19. Osman SOS, Leijenaar RTH, Cole AJ, et al. Computed tomography-based radiomics for risk stratification in prostate cancer. *Int J Radiat Oncol Biol Phys*. 2019;105(2):448-456. [\[Crossref\]](#)
 20. Ferro M, De Cobelli O, Musi G, et al. Radiomics in prostate cancer: an up-to-date review. *Ther Adv Urol*. 2022;14:175628722211090. [\[Crossref\]](#)
 21. Miyake H, Sakai I, Inoue TA, Hara I, Fujisawa M. The limited significance of a longer duration of neoadjuvant hormonal therapy prior to radical prostatectomy for high-risk prostate cancer in Japanese men. *Urol Int*. 2006;77(2):122-126. [\[Crossref\]](#)
 22. Carroll PH, Mohler JL. NCCN Guidelines Updates: prostate cancer and prostate cancer early detection. *J Natl Compr Canc Netw*. 2018;16(5s):620-623. [\[Crossref\]](#)
 23. Hassan O, Han M, Zhou A, et al. Incidence of extraprostatic extension at radical prostatectomy with pure gleason score 3 + 3 = 6 (grade group 1) cancer: implications for whether gleason score 6 prostate cancer should be renamed "not cancer" and for selection criteria for active surveillance. *J Urol*. 2018;199(6):1482-1487. [\[Crossref\]](#)
 24. Zhang Y, Chen W, Yue X, et al. Development of a novel, multi-parametric, MRI-based radiomic nomogram for differentiating between clinically significant and insignificant prostate cancer. *Front Oncol*. 2020;10:888. [\[Crossref\]](#)
 25. Gong L, Xu M, Fang M, et al. Noninvasive prediction of high-grade prostate cancer via biparametric MRI radiomics. *J Magn Reson Imaging*. 2020;52(4):1102-1109. [\[Crossref\]](#)
 26. Li M, Chen T, Zhao W, et al. Radiomics prediction model for the improved diagnosis of clinically significant prostate cancer on biparametric MRI. *Quant Imaging Med Surg*. 2020;10(2):368-379. [\[Crossref\]](#)
 27. Li H, Lee CH, Chia D, Lin Z, Huang W, Tan CH. Machine Learning in Prostate MRI for prostate cancer: current status and future opportunities. *Diagnostics (Basel)*. 2022;12(2):289. [\[Crossref\]](#)
 28. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci U S A*. 2015;112(46):6265-6273. [\[Crossref\]](#)
 29. Chen T, Li M, Gu Y, et al. Prostate cancer differentiation and aggressiveness: assessment with a radiomic-based model vs. PI-RADS V2. *J Magn Reson Imaging*. 2019;49(3):875-884. [\[Crossref\]](#)
 30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: Synthetic minority over-sampling technique. *J Artif Intell*. 2002;16:321-357. [\[Crossref\]](#)
 31. He H, Bai Y, Garcia EA, Li S. Adasyn: adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence); 2008: leee. [\[Crossref\]](#)



Reproducibility and interpretability in radiomics: a critical assessment

Aydın Demircioğlu

University Hospital Essen, Institute of Diagnostic and Interventional Radiology and Neuroradiology, Essen, Germany

ABSTRACT

Radiomics aims to improve clinical decision making through the use of radiological imaging. However, the field is challenged by reproducibility issues due to variability in imaging and subsequent statistical analysis, which particularly affects the interpretability of the model. In fact, radiomics extracts many highly correlated features that, combined with the small sample sizes often found in radiomics studies, result in high-dimensional datasets. These datasets, which are characterized by containing more features than samples, have different statistical properties than other datasets, thereby complicating their training by machine learning and deep learning methods. This review critically examines the challenges of both reproducibility issues and interpretability, beginning with an overview of the radiomics pipeline, followed by a discussion of the imaging and statistical reproducibility issues. It further highlights how limited model interpretability hinders clinical translation. The discussion concludes that these challenges could be mitigated by following best practices and by creating large, representative, and publicly available datasets.

KEYWORDS

Computer-aided detection, deep learning, interpretability, radiomics, reproducibility, statistics

Radiomics is a recent field that uses “an automated high-throughput extraction of large amounts of quantitative features of medical images.”¹⁻³ The method “converts imaging data into a high dimensional mineable feature space using a large number of automatically extracted data-characterization algorithms.”⁴

The above definition may seem complex, but it can be succinctly summarized. Similar to how clinical routine involves characterizing a patient using parameters such as age, weight, and hemoglobin levels, radiological images can be analyzed to extract analogous parameters (also called features) that ideally describe the pathology of interest. For example, in the case of a tumor lesion, features such as its volume and diameter can be measured. A critical aspect of radiomics is the extraction of not only morphological features but also the distribution of intensity and texture. This includes, for instance, assessing whether the lesion has high brightness and a homogeneous or coarse texture, and identifying the presence of bright spots. Radiomics involves the extraction of hundreds to thousands of such features to accurately represent the lesion. These features are subsequently used to train a classifier, that, based on the characteristics of a new lesion, can determine, for example, whether the lesion is benign.

The main expectation of radiomics is that these features can serve as surrogates for biomarkers, and thus aid clinical decision making. Radiological imaging could reflect the underlying biological processes, allowing for indirect conclusions. For example, while necrotic cells are not directly observable in computed tomography (CT) scans, their presence may result in the appearance of a hypodense lesion (Figure 1). Thus, measuring the overall intensity of a lesion could be used as an indicator of cell necrosis.

Although radiomics as a field only emerged in the 2010s,^{1,5} the idea can be traced back much further. In a seminal paper published in 1978, Harlow et al.⁶ introduced concepts that are strikingly similar. Later, specifically in the 1990s, similar techniques were introduced as texture analysis.⁷ This is no coincidence, since the underlying idea of applying machine learn-

Corresponding author: Aydın Demircioğlu

E-mail: aydin.demircioglu@uk-essen.de

Received 02 July 2024; revision requested 13 August 2024; accepted 12 September 2024.



Epub: 21.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242719

ing to imaging is the same and dates back to the 1960s.

The primary purpose of radiomics is to support clinical decisions. Ideally, the extracted features provide insights that humans cannot see or systematically process, allowing clinicians to answer questions using this hidden information. Radiomics has also been used to non-invasively identify genetic alterations or gene expression patterns that can be used to predict the outcome or survival risk of patients with cancer.⁸⁻¹⁰

In this review, the basic concepts of radiomics are first introduced, followed by a detailed discussion of the two major reproducibility issues that persist in the current field. Subsequently, radiomics based on deep neural networks is briefly outlined and the issues involved in their application examined. Finally, strategies for avoiding these issues are discussed.

The radiomics pipeline

As with any study, the first step in a radiomics model is to define patient cohorts, applying reasonable inclusion and exclusion criteria that reflect the target population, and defining an outcome of clinical interest.

The application of radiomics to data is technical but relatively straightforward (Figure 2).¹¹ Images are first acquired and the region of interest (ROI) is segmented. This can be a tumor lesion or an entire organ, such as the whole prostate. The ROI plays a critical role in directing the analysis to relevant areas, thereby preventing other unrelated regions from potentially confounding the analysis.

The images are then pre-processed depending on the use case. For example, magnetic resonance imaging may require a normalization step, and CT may be thresholded to a Hounsfield units range of interest. In ad-

dition, preprocessing filters are applied. For instance, smoothing filters can reduce noise that may adversely affect features, whereas wavelet filters can decompose the image into high-frequency and low-frequency components that may carry different information, aiding subsequent analysis.

Next, features are extracted from the ROI. This is a central step, and there are three main types of generic features that are extracted: morphological features, such as volume or sphericity; intensity features, which measure the distribution of values, such as mean brightness; and texture features that reflect the co-occurrence of intensity values.

However, feature extraction will often generate large numbers of features, and

many of them will be irrelevant (i.e., they will not help to solve the problem). Many will also be redundant, that is, their information is already present in other features. Therefore, a feature selection step is applied that retains only the relevant features; for example, a t-test can be used to filter out those that are not significant.

These features are then fed into a classifier, which functions in terms of making a prediction after receiving a set of features. This classifier is trained on the data using machine learning techniques. In other words, following the input of data, the algorithm identifies relevant patterns to make accurate predictions on new data. This model can then be tested and applied to new data, such as routine clinical data.

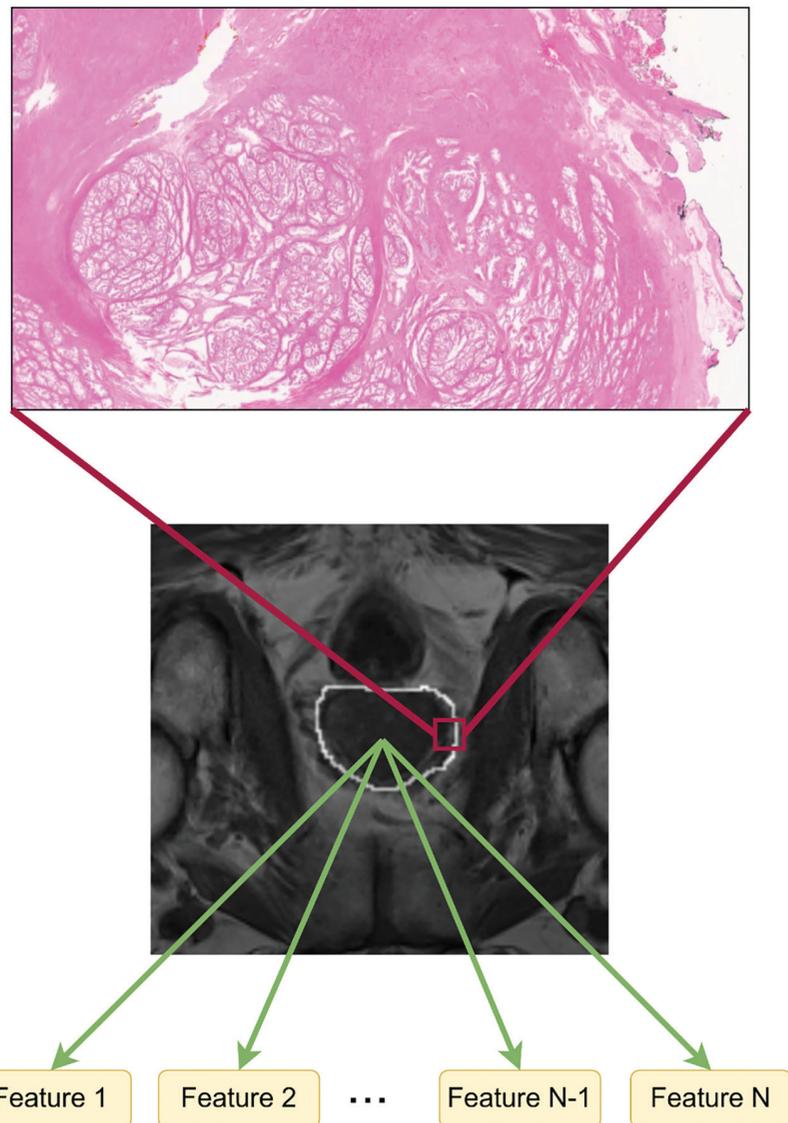


Figure 1. Radiomics aims to identify biomarkers by measuring them indirectly through radiological imaging. Much of the information in the pathological scan (top) is lost in the radiological image (bottom). Features are extracted from the segmented region-of-interest to recover the information of interest (the pathology image is part of the PROSTATE-MRI dataset).⁷⁵ MRI, magnetic resonance imaging.

Main points

- Radiomics is impeded by imaging and statistical reproducibility issues.
- Machine and deep learning modeling are complicated and require extensive validation.
- Radiomic features found to be predictive in modeling often do not correspond to biomarkers due to high correlation, limiting their interpretability.
- Standardization practices and larger, more diverse datasets are important to improve reproducibility.

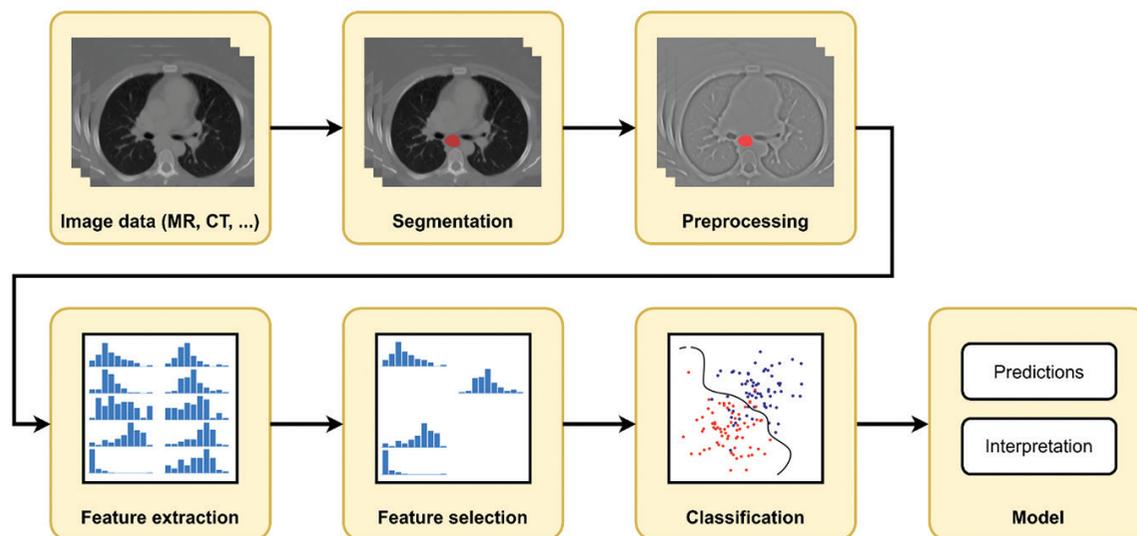


Figure 2. Brief overview of the radiomics pipeline. MR, magnetic resonance; CT, computed tomography.

The radiomics pipeline appears pretty straightforward, but in each step, good practices must be maintained to avoid biased or false-positive results.¹²

Reproducibility issues

Although the pipeline may seem fairly rigid, the key issue is reproducibility. This term describes the requirement that similar findings should be observed when conditions do not change significantly. For example, scanning the same patient twice within a very short time frame should yield similar radiomic features and lead to similar predictions. Non-reproducible studies are essentially random and erratic and cannot be trusted. They may also lead to false positives, which would prevent clinical use.

Reproducibility in radiomics can be divided into two areas: imaging reproducibility and statistical reproducibility. The term “imaging reproducibility” refers to the acquisition of scans and the extraction of features, whereas “statistical reproducibility” refers to modeling using machine learning. Of course, if the imaging is not reproducible, no modeling can correct it (following the well-known “garbage in, garbage out” rule).^{13,14} Nonetheless, the focus will be mainly on statistical reproducibility.

Imaging reproducibility

Imaging reproducibility refers to issues in the acquisition process resulting from variations in imaging parameters and techniques, vendor differences, and similar factors.¹⁵ Since radiomic features are extracted from the acquired images, parameters such as voxel size and reconstruction techniques

can have a significant impact on these features.^{16,17} The effect is also non-linear, which can render images highly non-reproducible.¹⁸ Post-hoc harmonization can mitigate the problem, but only to a limited extent.^{19,20}

Even if the imaging were reproducible, the segmentations are usually sensitive to intra- and inter-rater variability, and these differences can also have a large impact on the extracted features,²¹ making them partially non-reproducible. The same is true for the definition of the features themselves. Even simple features, such as sphericity, can show variations depending on the formulas used to calculate them. Accordingly, the Image Biomarker Standardisation Initiative (IBSI) was launched to standardize these features and assess their reproducibility.²² However, not all software programs are IBSI-compliant, and even the standardized features may still exhibit some differences.²³

Another source of variability is the use of preprocessing filters. Although standardization has recently been considered by the IBSI,²⁴ it is unknown whether preprocessing helps at all, and if so, which filters should be applied. Therefore, these preprocessing filters are applied in parallel to increase the predictive power of the resulting features.²⁵ However, this leads to statistical problems.

Statistical reproducibility

The data generated will often have two characteristics that distinguish it from many other datasets: it will be high-dimensional, meaning that there are more features than samples, and it will be highly correlated. In radiomics, there are two main reasons for this. First, the total sample size is often limited

due to the time and resources required for annotation, the rarity of the disease in question, or privacy concerns. Second, the numerous preprocessing filters extract information that is highly similar. For example, two levels of smoothing will produce features that are very alike. This results in the generation of highly correlated features.

The presence of such data presents significant challenges, as the search for predictive features and patterns becomes exponentially more difficult and resembles “finding a needle in a haystack.”²⁶ Therefore, the risk of identifying spurious patterns and producing false-positive results is significantly increased in such data. While methods such as regularization can help overcome this problem, the issue remains unresolved.

Therefore, radiomics often employs a feature selection step, where the goal is to retain only the relevant features and remove all others, thereby reducing the dimensionality of the data. However, several methods of varying complexity are currently in use.^{11,27,28} Simpler methods, such as Spearman correlation or t-tests, typically operate by considering each variable on its own. These methods are computationally efficient but may overlook dependencies between variables, potentially leading to suboptimal feature selection. More complex methods, such as the least absolute shrinkage and selection operator method,²⁹ the minimum redundancy maximum relevancy method,³⁰ or the Boruta method,³¹ are able to account for such dependencies but are more computationally demanding. While it may be intuitive to assume that more complex methods perform better, it has been shown that for many data-

sets, the differences may not be significant. However, simpler methods tend to be more robust, and therefore more reproducible.²⁷ In addition, many of the feature selection methods do not select relevant features but merely score them, leaving open the decision regarding how many of the highest-scoring features to retain, which reduces their reproducibility.

Accordingly, feature selection is not a complete solution to the problem since the task of dealing with the high-dimensional space is merely transferred from the classifier. Feature selection is also subject to failure and may even underperform, especially given the inherent instability of selection methods and their dependence on the specific data sample.²⁷ For example, the removal of a few samples can have a significant impact on the set of features considered relevant.

Subsequent classifiers are also affected by high dimensionality, either directly or indirectly, if irrelevant features have been selected. Furthermore, many classifiers make assumptions about the data that may not be true, regardless of whether feature selection has been applied. These assumptions are often controlled by hyperparameters; for example, a regularization variable may reflect the amount of noise present in the data. Therefore, the only option is to test many different parameters, which is extremely expensive in terms of computational resources. As a result, studies only test a limited number of parameters, and it remains unclear whether a significantly more effective model could have been obtained by optimizing the hyperparameters.

Validation issues

Any model requires extensive testing, the main reason for this being that models could either memorize the data or find spurious instead of predictive patterns. Such a model would perform well during training, but worse on test data and would not generalize. This problem is called overfitting, and the risk is higher for high-dimensional data, where more patterns can fit the given data.

To avoid this problem, validation is performed first. Unlike testing, validation is mainly used for model selection, specifically to determine good values for the hyperparameters, or to identify which feature selection or classifier method performs better on the given data. Ideally, validation should be performed on a second independent dataset, but alternatively, a portion of the data can be set aside. Certain common schemes

are often employed in radiomics, including simple splitting, cross-validation, and bootstrapping. In simple splitting, a portion of the data (e.g., 70%) is used for training, whereas the remainder is used exclusively for validation. While this method is conceptually simple and computationally fast, it does not utilize all available data for training. Additionally, the results can be highly dependent on the specific split, leading to potential variability; that is, there is a risk that results may be good, or bad, by chance. To mitigate this, the method can be repeated several times and the results averaged. Cross-validation provides a more systematic approach by splitting the data into k subsets and iteratively training on $k-1$ subsets while using the remaining subset for validation. Although computationally more expensive, this method ensures that all data is used for both training and validation, providing a more reliable estimate of the performance. Nested cross-validation further refines this by applying cross-validation twice: once to the entire data for performance estimation and once on the training data for hyperparameter tuning. This scheme provides an unbiased evaluation and is considered a gold standard. Bootstrapping, on the other hand, uses resampling with replacement to create training and validation sets. Since samples can occur multiple times in the training set, this approach simulates different weights for each sample and can thus lead to better estimates. However, to obtain these estimates, a large number of repetitions (e.g., 1,000) is generally required, making it computationally highly expensive.

However, in all cases, the golden rule of machine learning must be followed: training and test sets must be kept strictly separate. Failure to follow this rule will lead to data leakage, meaning that the classifier has already seen some aspects of the test data and could adapt to it, leading to false positives.^{32,33}

Another issue is the variability of the data. Choosing a homogeneous cohort (e.g., from a single scanner) increases the likelihood of obtaining a working model since the predictive patterns seen during training are likely to be present in the test data. At the same time, however, the model will be highly specific and may not generalize beyond the collected data. The opposite, collecting heterogeneous data, is also critical, because the classifier may not be able to identify any predictive patterns at all, especially with small sample sizes, and there will be no relevant model to

test. However, if such a model is successful, its clinical applicability will be much higher, which is the ultimate goal.³⁴

Deep radiomics

Deep learning has recently shown great success in other fields,³⁵ and it is natural to apply deep learning to radiomics. Deep learning is based on artificial neural networks, which, in a simplistic way, try to mimic the human brain, and date back to the early days of machine learning in the 1950s. Conceptually, in the simplest case, a network consists of multiple layers, each of which can be understood as a feature generation step. Layer by layer, the input is transformed into the desired output, and the training data is used to determine the parameters of the layers (Figure 3).

Applying deep learning to radiomics, which is termed deep radiomics, can, in contrast to the generic radiomics discussed above, mitigate two major drawbacks. First, it can potentially reduce the need for segmentation because the network can, at least potentially, determine the ROI itself. Equally important, the network can extract optimal features that are specific to the problem at hand. It can also consider more global features of the data, whereas most generic features are based on local textures. Both can lead to models that perform much better than generic models. While deep learning has only recently gained importance, neural networks have been applied to radiological data since the 1990s.^{36,37}

Issues with deep radiomics

Deep radiomics does not magically bypass the reproducibility problems. For example, changes in acquisition parameters have been shown to have a strong effect on predictive performance, thus affecting generalizability.³⁸ Much is unknown about the stability of deep radiomics models, such as whether a different training sample will yield different features, or whether features from different networks are highly correlated. Robustness to image noise and slightly different segmentations has also not been systematically investigated, which is complicated by the fact that many different architectures exist.

Sample size is an even bigger issue in deep radiomics. Learning directly from data usually requires many more samples to be successful.³⁹ As a result, deep radiomics is currently not as successful as it could be.

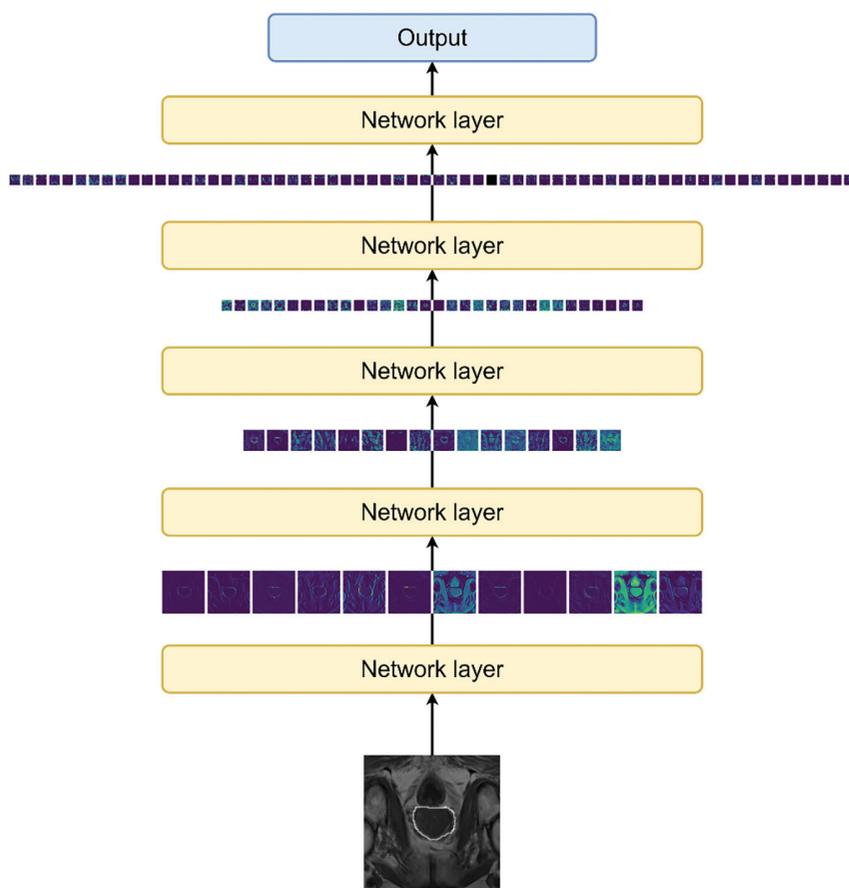


Figure 3. In simple terms, the network can be thought of as a set of layers that transform an input image into a set of output images. Each layer of the network has many parameters that are optimized using the training data. Networks usually do not use segmentation, but can be modified to use it. The network can be used as a feature generator by extracting features from the output of an appropriate layer. For example, in the figure, each of the 64 small images output by the second-to-last layer at the top could be averaged, resulting in 64 numerical features for the given input.

Consequently, several mitigation strategies have been developed.^{40,41} However, they all have their own drawbacks. For example, studies often resort to using image slices for training, which not only increases the sample size but also allows for the use of smaller networks.^{42,43} Nonetheless, this approach partially loses the spatial information, which reduces the potential benefit.

A more common strategy is transfer learning. Here, the network is first trained on a dataset from another domain, most commonly ImageNet, a collection of photographs.⁴⁴ This pre-trained network is then fine-tuned (i.e., it is trained on the radiomic data, often at lower learning rates) to slightly adjust the network. This approach can work because there is a remarkable similarity between the low-level features of the human eye and the network; at lower levels, both appear to operate with filters comparable to Gabor filters.³⁹ Thus, fine-tuning can focus on training the higher layers and perform-

ing better with fewer samples. However, the use of non-medical data for pre-training is again suboptimal, and larger medical data corpora have been introduced only recently, although the extent to which these can help in radiomics remains unclear, as they are usually far smaller than ImageNet.⁴⁵

Since training a deep network involves many hyperparameters (e.g., learning rate, learning schedule, choice of loss function) and can be relatively complicated, another alternative is to bypass any training and instead use only pre-trained networks as feature extractors (Figure 3),⁴⁶ which allows more versatile classifiers, such as boosting, to perform better, especially with smaller sample sizes.⁴⁷ However, since no training is performed in this approach, the disadvantage is again that the features may be less optimal, although fusing them with generic radiomics can still prove helpful.^{48,49}

Finally, the hope that deep radiomics can dispense with segmentation may be in vain

due to the small sample size. In addition, without a proper validation method, deep radiomics is also prone to bias due to over-engineering. In fact, a recent review found no clear advantage of deep radiomics.⁵⁰

Interpretability issues

A key point in radiomics is to identify features that can potentially serve as biomarkers, just as the volume of a lesion indicates its malignancy. However, radiomics attempts to establish such a correspondence “in reverse,” using the coarser and noisier radiological images, where much information is already lost during acquisition. Radiomics seeks to capture the underlying information by making multiple measurements (in the form of different features). These are often correlated, as they can be understood as noisy and incomplete versions of the inaccessible information. There is no guarantee that the information can be recovered from the extracted features, nor that the observed predictivity of a feature actually corresponds to a biomarker.

Given a set of features, radiomics can only identify those that are statistically associated with the outcome. Such an association is not causal and could only be the basis of a subsequent statistically sound test. This problem is exacerbated by the high-dimensionality of the data, where the intuition from the low-dimensional setting that features have a clear meaning and their importance can be easily measured fails.⁵¹ In fact, the very concept of distance becomes somewhat incomprehensible in higher dimensions, often termed the curse of dimensions, and is demonstrated by the fact that in higher dimensions, most of the volume of a unit sphere is near its surface.⁵²

In fact, the use of feature importance as a surrogate has been shown to be questionable because essentially every step in the radiomics pipeline affects the importance of features in the resulting model. Even seemingly unimportant preprocessing steps, such as the choice of discretization method²³ and data normalization, which is performed to obtain the data on a uniform scale, can strongly influence the set of features and thus the interpretability.⁵³ This influence is more evident in the feature selection step, where different methods will emphasize different aspects and thus gain different importance.²⁷ Not only does the subsequent classifier affect interpretation but the selection of the final model can also have a great impact, as often several models will perform very

closely but will select different sets of features as important.^{51,54} In a systematic review, Tohidinezhad et al.⁵⁵ identified 23 models that predict the effect of radiation on brain health. None of these models used exactly the same features, and the models differed widely in the factors that were significantly associated with outcome.

Moreover, even if such an identification were possible, most radiomic features are not interpretable by themselves. For example, it is unclear what semantic meaning a feature such as `wavelet-LHL_glrIm_GrayLevelNonUniformityNormalized` carries, and how to see the difference from a highly correlated feature that is slightly less predictive. It is unlikely that a radiologist would be able to relate the measured values of such a feature to the scan. Feature maps may be helpful for visualization,⁵⁶ but they are currently only a tool and cannot be used to base an interpretation on. In addition, radiomic models are rarely based on a single feature, and a meaningful interpretation of a model using multiple features is barely possible. Paradoxically, radiomics was invented precisely because humans cannot describe textural patterns well.

The potential for highly correlated features to cause interpretation problems is illustrated by a recent study by Welch et al.,⁵⁷ who reexamined the model that Aerts et al.⁴ used in their seminal work on patients with non-small cell lung cancer. The authors showed that volume alone is as predictive as the radiomic model, and moreover, that three of the four texture features found by Aerts et al.⁴ are highly correlated with volume.

Recently, post-hoc interpretations, such as Explainable AI (XAI) methods, have been applied.⁵⁸ However, these are also problematic. Since there are several different XAI methods, it is likely that the resulting meanings will also differ.⁵⁹ Alternatively, explainable classifiers could be used, which generally involves a trade-off between the complexity (and thus interpretability) of the classifier and its predictive performance.⁶⁰ However, even if these methods are successful, they only address the classifier and do not mitigate the problems in the overall pipeline.

The situation is similar for deep radiomics. While the pipeline itself is less complex, training is more difficult, and there are many more choices regarding the architecture. It is highly likely that different choices will lead to vastly different features. In addition, the deep features do not have a mathematical

formula, making any direct interpretation difficult. To remedy this situation, Cho et al.⁶¹ correlated deep features with radiomic features. However, since radiomic features are not fully interpretable by themselves, this approach is limited in scope.

Discussion

Currently, radiomics suffers from both imaging and statistical reproducibility issues, both of which affect the interpretability and applicability of the models. This affects the entire radiomics pipeline, and even feature normalization can lead to reproducibility issues.

Neither of these problems can be easily avoided. Image reproducibility could possibly be mitigated by strict standardization of imaging protocols, but this is all but impossible to implement in practice across multiple centers. Statistical reproducibility is also not easily mitigated. Methodological differences aside, different research groups will often reach different conclusions given the same data.⁶² Although such studies have not been conducted in radiomics, the impact is expected to be even greater, as there is generally less code and data sharing in the health domain.⁶³

One major problem is small sample sizes. Radiomics studies need to include larger and more diverse datasets to have a chance of success. This is illustrated by current models that use deep learning to diagnose chest X-rays, or mammograms that have been shown to perform especially well.^{64,65} These models are often trained on datasets that reach tens of thousands of scans. However, they are not radiomic in the sense that they do not require segmentations. The abundance of data makes segmentations unnecessary, as the network can identify the relevant regions on its own. Although it is virtually impossible to obtain such large sample sizes for rare cancers, more data would potentially reduce the dimensionality of the data and thus increase reproducibility. Nonetheless, radiomics seems to have made no progress since the seminal work of Harlow et al.⁶ in 1976, where sample sizes of around 300 are reported. Small sample sizes are generally unable to reflect heterogeneity. This is even true for within-patient heterogeneity. For example, suppose two features are measured in a single patient at two time points, as in a test-retest scenario, and their sum is predictive. Then, the two features may vary greatly between the two time points such that neither is reproducible; but provided

their sum remains the same, this would not pose any problem for their predictive value. However, if the model was not trained on such data, it would not find that pattern and would fail on new data. Nevertheless, large sample sizes are useless if the images do not carry the necessary information and such predictive patterns do not exist. Hence more data is not always helpful.

Non-reproducible studies may also result from a failure to follow best practices, which can be ensured by adhering to proper guidelines.^{66,67} For example, the study must be described in full detail in a manner that enables replication by others. Code should always be shared, and data should be shared if possible. Best practices encompass every step of the study; for example, it must be ensured that the data selection is appropriate and unbiased relative to the study's objective.^{12,68} The outcome should also be compared with current standards where applicable, for example, if a clinical scoring system is in current use (e.g., the Prostate Imaging Reporting and Data System), the radiomics model should be compared against it.⁶⁹ Statistical tests (e.g., permutation tests) can be used to ensure that the resulting model is different from a random guess, which is crucial when sample sizes are small. While statistical significance should be computed, the clinical significance should also be considered to evaluate the impact of the model. Furthermore, the overall study design must be methodologically sound to avoid reporting false-positive results. In addition, reporting must be clear and complete to ensure reproducibility.⁷⁰

In a seminal paper, Ioannidis argued that around 60% of all medical studies contain false-positive results.⁷¹ Studies with such obvious false positives should therefore be retracted, but this almost never happens in radiomics. On the contrary, such studies are frequently cited.⁷² In addition, methodologically correct studies will fare relatively worse and may appear as "negative" studies that may not be considered for publication.⁷³ To mitigate this, a far more rigorous review process with mandatory code or data sharing would be required, as it could help in identifying potentially biased results before their publication. Currently, such studies are often only identified following publication, making it difficult to address the issue. Ensuring that publications rigorously follow reporting guidelines could be another way to reduce the problem.^{66,67,70}

It is easy to overlook the fact that image processing has gone through a similar evo-

lution in the past. The field started with the manual extraction of many features (which is the origin of the texture features used today), progressed to the extraction of more complicated features such as Fisher vectors,⁷⁴ before the advent of deep learning made these steps obsolete. In fact, the interpretability of deep networks is at the semantic level of images, not features, for example, to answer the question of whether the network takes the tail of a dog into account when predicting its race. This is not easily possible in radiomics, where a visualization of the important areas of a tumor lesion would not help a radiologist understand what the network is doing. Furthermore, in current machine learning, a model is accepted if it generalizes well, not necessarily if the model is interpretable. A similar strategy may be viable for radiomics, where the applicability of models is validated on large datasets.

In conclusion, radiomics currently faces substantial challenges related to imaging and statistical reproducibility that severely impact interpretability and clinical applicability. These problems are difficult to mitigate because imaging standardization is largely impractical and statistical variability is inherent in high-dimensional datasets. As a result, the potential for clinical integration remains uncertain and questionable. A shift toward rigorous data and code sharing practices and the development of large, representative datasets would be required to partially address these challenges.

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

References

- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446. [\[CrossRef\]](#)
- McCague C, Ramlee S, Reinius M, et al. Introduction to radiomics for a clinical audience. *Clin Radiol*. 2023;78(2):83-98. [\[CrossRef\]](#)
- Haneberg AG, Pierre K, Winter-Reinhold E, et al. Introduction to radiomics and artificial intelligence: a primer for radiologists. *Semin Roentgenol*. 2023;58(2):152-157. [\[CrossRef\]](#)
- Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. [\[CrossRef\]](#)
- Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30(9):1234-1248. [\[CrossRef\]](#)
- Harlow CA, Dwyer SJ, Lodwick G. On radiographic image analysis. In: *Digital Picture Analysis*. Springer; 1976:65-150. [\[CrossRef\]](#)
- Freeborough PA, Fox NC. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. *IEEE Trans Med Imaging*. 1998;17(3):475-478. [\[CrossRef\]](#)
- Cromb e A, Bertolo F, Fadli D, et al. Distinct patterns of the natural evolution of soft tissue sarcomas on pre-treatment MRIs captured with delta-radiomics correlate with gene expression profiles. *Eur Radiol*. 2023;33(2):1205-1218. [\[CrossRef\]](#)
- Qi Y, Zhao T, Han M. The application of radiomics in predicting gene mutations in cancer. *Eur Radiol*. 2022;32(6):4014-4024. [\[CrossRef\]](#)
- Xia TY, Zhou ZH, Meng XP, et al. Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. *Radiology*. 2023;307(4):e222729. [\[CrossRef\]](#)
- Koak B, Durmaz EŐ, AteŐ E, Kılıkesmez  . Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol*. 2019;25(6):485-495. [\[CrossRef\]](#)
- Guiot J, Vaidyanathan A, Deprez L, et al. A review in radiomics: making personalized medicine a reality via routine imaging. *Med Res Rev*. 2022;42(1):426-440. [\[CrossRef\]](#)
- Koak B, Cuocolo R, Santos DPD, Stanzione A, Ugga L. Must-have qualities of clinical research on artificial intelligence and machine learning. *Balkan Med J*. 2023;40(1):3-12. [\[CrossRef\]](#)
- Teno JM. Garbage in, garbage out-words of caution on big data and machine learning in medical practice. *JAMA Health Forum*. 2023;4(2):e230397. [\[CrossRef\]](#)
- Zhao B. Understanding sources of variation to improve the reproducibility of radiomics. *Front Oncol*. 2021;11:633176. [\[CrossRef\]](#)
- Ibrahim A, Barufaldi B, Refaee T, et al. MaasPenn radiomics reproducibility score: a novel quantitative measure for evaluating the reproducibility of CT-based handcrafted radiomic features. *Cancers (Basel)*. 2022;14(7):1599. [\[CrossRef\]](#)
- Schwier M, Griethuysen J van, Vangel MG, et al. Repeatability of multiparametric prostate MRI radiomics features. *Sci Rep*. 2019;9(1):9441. [\[CrossRef\]](#)
- Emaminejad N, Wahi-Anwar MW, Kim GHJ, Hsu W, Brown M, McNitt-Gray M. Reproducibility of lung nodule radiomic features: multivariable and univariable investigations that account for interactions between CT acquisition and reconstruction parameters. *Med Phys*. 2021;48:2906-2919. [\[CrossRef\]](#)
- Marcadent S, Hofmeister J, Preti MG, Martin SP, Van De Ville D, Montet X. Generative adversarial networks improve the reproducibility and discriminative power of radiomic features. *Radiol Artif Intell*. 2020;2(3):e190035. [\[CrossRef\]](#)
- Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology*. 2019;291(1):53-59. [\[CrossRef\]](#)
- Granzier RWY, Verbakel NMH, Ibrahim A, et al. MRI-based radiomics in breast cancer: feature robustness with respect to inter-observer segmentation variability. *Sci Rep*. 2020;10(1):14163. [\[CrossRef\]](#)
- Zwanenburg A, Valli eres M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338. [\[CrossRef\]](#)
- Bettinelli A, Marturano F, Avanzo M, et al. A novel benchmarking approach to assess the agreement among radiomic tools. *Radiology*. 2022;303(3):533-541. [\[CrossRef\]](#)
- Whybra P, Zwanenburg A, Andrearczyk V, et al. The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. *Radiology*. 2024;310(2):e231319. [\[CrossRef\]](#)
- Demirciođlu A. The effect of preprocessing filters on predictive performance in radiomics. *Eur Radiol Exp*. 2022;6(1):40. [\[CrossRef\]](#)
- Pappu V, Pardalos PM. High-dimensional data classification. In: Aleskerov F, Goldengorin B, Pardalos PM, eds. *Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday*. Springer New York; 2014:119-150. [\[CrossRef\]](#)
- Demirciođlu A. Benchmarking feature selection methods in radiomics. *Invest Radiol*. 2022;57(7):433-443. [\[CrossRef\]](#)
- Ge G, Siddique A, Zhang J. Inconsistent CT NSCLC radiomics associated with feature selection methods, predictive models and related factors. *Phys Med Biol*. 2023;68(12):125004. [\[CrossRef\]](#)
- Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol*. 1996;58(1):267-288. [\[CrossRef\]](#)
- Peng H, Ding C. Minimum redundancy and maximum relevance feature selection and recent advances in cancer classification: 8. [\[CrossRef\]](#)
- Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. 2010;36(11):1-13. [\[CrossRef\]](#)
- Krawczuk J, Łukaszuk T. The feature selection bias problem in relation to high-dimensional gene data. *Artif Intell Med*. 2016;66:63-71. [\[CrossRef\]](#)
- Demirciođlu A. Measuring the bias of incorrect application of feature selection when using

- cross-validation in radiomics. *Insights Imaging*. 2021;12(1):172. [\[CrossRef\]](#)
34. Doran SJ, Kumar S, Orton M, et al. "Real-world" radiomics from multi-vendor MRI: an original retrospective study on the prediction of nodal status and disease survival in breast cancer, as an exemplar to promote discussion of the wider issues. *Cancer Imaging*. 2021;21(1):37. [\[CrossRef\]](#)
 35. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. [\[CrossRef\]](#)
 36. Lo SB, Lou SA, Lin JS, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Trans Med Imaging*. 1995;14(4):711-718. [\[CrossRef\]](#)
 37. Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans Med Imaging*. 1996;15(5):598-610. [\[CrossRef\]](#)
 38. Blazis SP, Dickerscheid DBM, Linsen PVM, Martins Jarnalo CO. Effect of CT reconstruction settings on the performance of a deep learning based lung nodule CAD system. *Eur J Radiol*. 2021;136:109526. [\[CrossRef\]](#)
 39. Qin Z, Yu F, Liu C, Chen X. How convolutional neural network see the world - a survey of convolutional neural network visualization methods. Published online May 31, 2018. Accessed June 26, 2024. [\[CrossRef\]](#)
 40. Banerjee S, Mitra S, Masulli F, Rovetta S. Glioma classification using deep radiomics. *SN Comput Sci*. 2020;1(4):209. [\[CrossRef\]](#)
 41. Zhang X, Zhang G, Qiu X, et al. Radiomics under 2D regions, 3D regions, and peritumoral regions reveal tumor heterogeneity in non-small cell lung cancer: a multicenter study. *Radiol Med*. 2023;128(9):1079-1092. [\[CrossRef\]](#)
 42. Müller D, Kramer F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Med Imaging*. 2021;21(1):12. [\[CrossRef\]](#)
 43. 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. [\[CrossRef\]](#)
 44. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009:248-255. [\[CrossRef\]](#)
 45. Demircioğlu A. Deep features from pretrained networks do not outperform hand-crafted features in radiomics. *Diagnostics (Basel)*. 2023;13(20):3266. [\[CrossRef\]](#)
 46. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. 2017;7(1):5467. [\[CrossRef\]](#)
 47. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? Published online July 18, 2022. [\[CrossRef\]](#)
 48. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys*. 2017;44(10):5162-5171. [\[CrossRef\]](#)
 49. Han W, Qin L, Bay C, et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *Am J Neuroradiol*. 2020;41(1):40-48. [\[CrossRef\]](#)
 50. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. *Eur Radiol Exp*. 2023;7(1):11. [\[CrossRef\]](#)
 51. Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci*. 2001;16(3):199-231. [\[CrossRef\]](#)
 52. Bellman R. Dynamic programming. *Science*. 1966;153(3731):34-37. [\[CrossRef\]](#)
 53. Demircioğlu A. The effect of feature normalization methods in radiomics. *Insights Imaging*. 2024;15(1):2. [\[CrossRef\]](#)
 54. Demircioğlu A. Evaluation of the dependence of radiomic features on the machine learning model. *Insights Imaging*. 2022;13(1):28. [\[CrossRef\]](#)
 55. Tohidinezhad F, Di Perri D, Zegers CML, et al. Prediction models for radiation-induced neurocognitive decline in adult patients with primary or secondary brain tumors: a systematic review. *Front Psychol*. 2022;13:853472. [\[CrossRef\]](#)
 56. Costa G, Cavinato L, Fiz F, et al. Mapping tumor heterogeneity via local entropy assessment: making biomarkers visible. *J Digit Imaging*. 2023;36(3):1038-1048. [\[CrossRef\]](#)
 57. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: the need for safeguards. *Radiother Oncol*. 2019;130:2-9. [\[CrossRef\]](#)
 58. Manikis GC, Ioannidis GS, Siakallis L, et al. Multicenter DSC-MRI-based radiomics predict IDH mutation in gliomas. *Cancers*. 2021;13(16):3965. [\[CrossRef\]](#)
 59. Amparore E, Perotti A, Bajardi P. To trust or not to trust an explanation: using LEAF to evaluate local linear XAI methods. *PeerJ Comput Sci*. 2021;7:e479. [\[CrossRef\]](#)
 60. Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '13. Association for Computing Machinery; 2013:623-631. [\[CrossRef\]](#)
 61. Cho HH, Lee HY, Kim E, et al. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. *Commun Biol*. 2021;4(1):1286. [\[CrossRef\]](#)
 62. Breznau N, Rinke EM, Wuttke A, et al. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proc Natl Acad Sci*. 2022;119(44):e2203150119. [\[CrossRef\]](#)
 63. McDermott MBA, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: Still a ways to go. *Sci Transl Med*. 2021;13(586):eabb1655. [\[CrossRef\]](#)
 64. Yoon JH, Strand F, Baltzer PAT, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. *Radiology*. 2023;307(5):e222639. [\[CrossRef\]](#)
 65. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE Conference on Computer Vision and Pattern Recognition*. 2017:3462-3471. [\[CrossRef\]](#)
 66. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METHodological RadiomIcs score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMI. *Insights Imaging*. 2024;15(1):8. [\[CrossRef\]](#)
 67. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. [\[CrossRef\]](#)
 68. Moskowitz CS, Welch ML, Jacobs MA, Kurland BF, Simpson AL. Radiomic analysis: study design, statistical analysis, and other bias mitigation strategies. *Radiology*. 2022;304(2):265-273. [\[CrossRef\]](#)
 69. Bao J, Qiao X, Song Y, et al. Prediction of clinically significant prostate cancer using radiomics models in real-world clinical practice: a retrospective multicenter study. *Insights Imaging*. 2024;15(1):68. [\[CrossRef\]](#)
 70. Kocak B, Borgheresi A, Ponsiglione A, et al. Explanation and elaboration with examples for CLEAR (CLEAR-E3): an EuSoMI radiomics auditing group initiative. *Eur Radiol Exp*. 2024;8:72. [\[CrossRef\]](#)
 71. Ioannidis JP. Why most published research findings are false. *PLOS Med*. 2005;2(8):e124. [\[CrossRef\]](#)
 72. Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. *Sci Adv*. 2021;7(21):eabd1705. [\[CrossRef\]](#)
 73. Kocak B, Bulut E, Bayrak ON, et al. NEgative results in radiomics research (NEVER): a meta-research study of publication bias in leading radiology journals. *Eur J Radiol*. 2023;163:110830. [\[CrossRef\]](#)
 74. Sánchez J, Perronnin F, Mensink T, Verbeek J. Image classification with the Fisher vector: theory and practice. *Int J Comput Vis*. 2013;105(3):222-245. [\[CrossRef\]](#)
 75. Choyke P, Turkbey B, Pinto P, Merino M, Wood B. Data from PROSTATE-MRI. The cancer imaging archive. *Cancer Imaging Arch TCIA*. Published online 2016. [\[CrossRef\]](#)



Letter to the editor re: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging

Hinpetch Daungsupawong¹

Viroj Wiwanitkit²

¹Private Academic Consultant, Phonhong, Lao People's Democratic Republic

²Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Dear Editor,

The study titled "Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging" presents a novel approach to medical image analysis.¹ This research aims to evaluate the effectiveness of Microsoft Bing, enhanced with ChatGPT-4 technology, in interpreting abdominal computed tomography (CT) and magnetic resonance imaging (MRI) data. Eighty abdominal images, including 44 CT and 36 MRI scans, were examined, and Bing's assessment was compared with that of a professional radiologist. The results showed that Bing could correctly identify CT scans with 95.4% accuracy and MRIs with 86.1% accuracy. However, Bing experienced some problems: wrongly identifying some images and poorly detecting anatomical regions, imaging planes, MRI sequences, and contrast agents. Bing discovered anomalies in only 35% of the images, with a 10.7% accuracy rate.

Bing's analysis suffers from inaccuracies in detecting imaging types, as evidenced by wrongly labeled CT and MRIs. The identification of MRI sequences and contrast agents was also poor, with success rates of 68.75% and 64.2%, respectively. Furthermore, Bing's low correct interpretation rates for anomalies underscore the difficulties of obtaining therapeutically useful information. Such limitations highlight its reliance on massive datasets and complex algorithms, which may not detect the tiny diagnostic signals found in medical imaging.

The study's comparative and descriptive design may limit its ability to address modest changes in image context or patient pathology. The sample size, although large, may be insufficient to draw broad conclusions. Bing's performance is context-dependent, and using only 80 photos may limit insights into its suitability for a wide range of clinical circumstances. Furthermore, the absence of real-time adaptive learning from feedback may impede the tool's progress, reducing its long-term relevance in radiology.

While Microsoft Bing incorporates ChatGPT-4 technology, there is evidence to suggest that its performance may not be as accurate or contextually aware as the standalone ChatGPT platform. This variation could be due to variances in how each system is taught and optimized for specific tasks. The standalone ChatGPT platform benefits from tailored training on various datasets, which improves performance in delivering nuanced and contextually relevant responses. OpenAI recently added memory features to its ChatGPT platform, allowing it to remember information between sessions for specific users.² As a result, when examining each system's usefulness in medical image analysis and other complicated domains, it is critical to consider its distinct strengths and limits.

To enhance Bing's diagnostic capabilities, future initiatives should focus on integrating more comprehensive datasets, encompassing a wider array of diseases, imaging modalities, and patient demographics. Language disparities in patient demographics dependent on the study location may have a major impact on the interpretation of the results.³ Continuous training with advanced deep learning techniques could further improve its ability to distinguish between various types of images and detect subtle anomalies. Investing in a

KEYWORDS

AI, audit, Bing, evaluation, radiology

Corresponding author: Hinpetch Daungsupawong

E-mail: hinpetchdaung@gmail.com

Received 20 August 2024; revision requested 09 September 2024; accepted 25 September 2024.



Epub: 21.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242994

real-time feedback loop in which Bing learns from radiologists' accurate diagnoses can help improve diagnostic accuracy. As Elek⁴ points out, the way a question is phrased to models like ChatGPT is critical to improving answer accuracy. Enabling web access in ChatGPT or seeking references from the PubMed database after asking queries may improve the model's accuracy.³ Finally, collaboration with medical practitioners might lead to improvements that address specific clinical needs. This will eventually make artificial intelligence systems like Bing more

reliable as a supplement to medical image analysis.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Elek A, Ekizaliođlu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed

tomography and magnetic resonance images. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)

2. Zhong W, Guo L, Gao Q, Ye H, Wang Y. Memorybank: enhancing large language models with long-term memory. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024;38(17):19724-19731. [\[CrossRef\]](#)
3. Elek A. The Role of large language models in radiology reporting. *AJR Am J Roentgenol*. 2023;221(5):707. [\[CrossRef\]](#)
4. Elek A. Improving accuracy in ChatGPT. *AJR Am J Roentgenol*. 2023;221(5):705. [\[CrossRef\]](#)



Reply: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images

Alperen Elek¹
 Duygu Doğa Ekizalioğlu²
 Ezgi Güler²

¹Ege University Faculty of Medicine, İzmir, Türkiye

²Ege University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

Dear Editor,

We would like to express our gratitude to Daungsupawong and Wiwanitkit¹ for their insightful comments on our manuscript, "Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images"². Their feedback is invaluable, and we appreciate the opportunity to engage in this discussion.

Our original study² demonstrated that as tasks become more complex, the model's performance significantly declines, particularly in predicting pathology. Although we acknowledge the current limitations of Bing in distinguishing imaging modalities and detecting contrast agents, we believe these challenges are inherent in the early stages of integrating artificial intelligence (AI) into medical imaging. Expecting AI technologies such as Bing with ChatGPT-4 to match the accuracy of radiologists from the outset may be overly ambitious. Nevertheless, it is worth considering that the fine-tuning or development of models specifically designed for diagnostic purposes might lead to improved results. Furthermore, enabling the upload of DICOM-format files, rather than single slices, could enhance performance.

Regarding the study's sample size, although we agree that a larger dataset would provide more robust insights, it is also important to recognize the value of small-scale, targeted studies in shaping future research directions. Preliminary evaluations can identify key areas for improvement and refine methodologies for larger trials, potentially leading to more focused and impactful studies.

The noted differences between Bing's integrated ChatGPT-4 and the standalone version further highlight the complexity of deploying AI models across various platforms. This variation reflects not only differences in implementation but also the broader challenge of ensuring consistent performance in diverse clinical environments. Addressing these nuances is critical to the safe and effective integration of AI into medical practice.

In conclusion, although our study explores the potential applications of Bing with ChatGPT-4, we stress the importance of using such tools as supplementary aids under expert supervision. They are far from replacing radiologists or being independently reliable for complex medical tasks. We hope this reply contributes to the ongoing discourse on the role of AI in diagnostic radiology and encourages the development of more robust, domain-specific solutions.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

KEYWORDS

Abdomen, artificial intelligence, diagnostic imaging, large language models, magnetic resonance imaging, multidetector computed tomography

Corresponding author: Ezgi Güler

E-mail: gulerezgi@yahoo.com

Received 02 November 2024; revision requested 10 November 2024; accepted 16 December 2024.



Epub: 13.01.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.243123

References

1. Daungsupawong H, Wiwanitkit V. Letter to the editor re: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)
2. Elek A, Ekizalioglu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)



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

Cihan Özgür¹
 Baran Serdar Sunal¹
 Savaş Hereklioğlu¹
 Meltem Öznur²
 Sibel Özkan Gürdal³

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

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

³Tekirdağ Namık Kemal University Faculty of Medicine, Department of Surgical Oncology, Tekirdağ, Türkiye

PURPOSE

This study aimed to evaluate the diagnostic performance of breast magnetic resonance imaging (MRI) in predicting extracapsular extension (ECE) and axillary nodal status in the axillary metastatic lymph nodes of patients with breast cancer.

METHODS

The preoperative MRI scans of 92 patients with breast cancer and axillary metastases who did not receive neoadjuvant treatment between January 2018 and January 2024 were retrospectively examined. The presence of an increased signal in the axillary fatty tissue surrounding the lymph node (flare sign) on T2-weighted images, irregular nodal contour (shaggy margin), axillary asymmetry (difference in the number and size of lymph nodes compared with the unaffected axilla), loss of the fatty hilum in the most suspicious lymph node, and morphological features on T1-weighted images were assessed. Each dissected axillary lymph node was examined for ECE, and the histopathological results were recorded.

RESULTS

Axillary flare sign was significantly associated with the presence of ECE ($P < 0.001$), number of lymph nodes with ECE ($P < 0.001$), the presence of ≥ 4 axillary metastatic lymph nodes ($P < 0.001$), size of the primary tumor ($P = 0.033$), lymphovascular invasion in the primary tumor ($P < 0.001$), and presence of perineural invasion ($P = 0.001$). The flare sign exhibited 65.7% sensitivity, 96% specificity, 97.8% positive predictive value, 51.1% negative predictive value, and 73.9% accuracy in predicting ECE. Additionally, the receiver operating characteristic curve analysis revealed an area under the curve of 0.808 (95% confidence interval: 0.719–0.898).

CONCLUSION

The flare sign has high performance in predicting ECE and axillary nodal status and is associated with primary tumor aggressiveness, indicating its potential utility in preoperative evaluation.

CLINICAL SIGNIFICANCE

The flare sign on breast MRI may play a crucial role in preoperative planning, surgical decision-making, and axillary status assessment by accurately predicting ECE.

KEYWORDS

Axilla, breast neoplasms, extranodal extension, lymphadenopathy, lymph node, magnetic resonance imaging

Corresponding author: Cihan Özgür

E-mail: cihanozгурitf@gmail.com

Received 25 June 2024; revision requested 30 July 2024; accepted 10 October 2024.



Epub: 07.11.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242906

The clinical evaluation of patients with breast cancer involves determining the presence of axillary node metastases as a prognostic indicator, in addition to tumor size.¹ Assessing the axilla using radiology is essential for managing breast cancer, as it provides crucial insights into locoregional staging for surgical procedures. The Breast Imaging Reporting and

Data System lexicon for preoperative breast magnetic resonance imaging (MRI) assessment evaluates axillary lymph nodes within the field of view as “associated features,” even though ultrasound remains essential for determining axillary status.² Extracapsular extension (ECE) occurs when malignant cells invade the surrounding tissue by breaching the capsule of a lymph node.³ There is still a lack of understanding regarding the specific biological mechanisms responsible for ECE. Increased vessel permeability, inflammation around the lymph nodes, and the obstruction of lymphatic channels may indicate that the tumor is aggressive in lymph nodes and lymphatic vessels.⁴ Unlike the previous tumor–node–metastasis classification, the current classification excludes the evaluation of ECE. The most recent reporting protocol from the College of American Pathologists recommends documenting the presence of ECE in metastatic axillary lymph nodes.⁵ Yang et al.⁶ linked the presence of ECE in sentinel lymph nodes (SLNs) to advanced tumor stage, progesterone receptor (PR) status, lymphovascular invasion, increased metastasis rates in non-SLNs, nodal burden, the number of affected non-SLNs, and the total number of positive lymph nodes. Patients with ECE exhibit a significantly higher incidence of pN2 disease. In SLNs, ECE significantly decreases disease-free and overall survival.⁶ Furthermore, ECE in metastatic SLNs is a strong indicator of residual axillary disease.^{7,8} Based on these results, it may be more beneficial to forgo SLN biopsy and opt for axillary lymph node dissection (ALND) if ECE is detected.

An irregular shape, unclear margins, and infiltration into nearby tissues observed on

sonography are indicative of ECE. Only a few studies have reported MRI findings, defined as increased T2 signal intensity surrounding the lymph nodes.^{4,9} Researchers have also used these findings to predict lymph node malignancy.¹⁰ Baltzer et al.¹⁰ proposed the term ‘perifocal edema’ to describe this finding, and researchers have reported that it has a sensitivity of 29.4% and a specificity of 100% in accurately discriminating between malignant and benign lymph nodes.

Given the challenges of directly detecting metastatic spread beyond the lymph node capsule with current imaging techniques, this study aimed to determine the utility of breast MRI in correlating changes in fat signals around the axillary lymph nodes with ECE.

Methods

Patients

This retrospective study was approved by the Non-Interventional Clinical Research Ethics Committee of Tekirdağ Namık Kemal University (protocol number: 2024.36.02.20, date: 27.02.2024). The requirement for informed consent was waived because of the retrospective nature of the study. The study initially included 352 patients with pathologically confirmed axillary metastases from breast cancer who underwent axillary dissection (SLN biopsy and/or ALND) between January 2018 and January 2024. Patients without a preoperative MRI examination (n = 52), those with low image resolution or artifacts (n = 30), and those receiving neoadjuvant chemotherapy (NAC) or hormone therapy (n = 178) were excluded. A total of 92 women who underwent breast MRI within 3 months preoperatively were included in the study (Figure 1).

The evaluation of cases also considered age; postoperative tumor and node stages; maximum tumor size; tumor histology; hormone status; Ki67 index; human epidermal growth factor receptor 2 (HER2) positivity; lymphovascular invasion in the primary tumor; the presence of perineural invasion; the number of total, metastatic, and non-metastatic lymph nodes detected in the axilla; and the number of metastatic lymph nodes with ECE.

Histopathological assessment

A pathologist with 16 years of experience in breast pathology conducted the histopathological evaluation of the surgical specimens. The tumor, histological, and molecular subtypes were determined. Hormone (estrogen and progesterone) receptor status, HER2 status, histology (modified Richardson–Bloom score), and nuclear grade were recorded. Patients with estrogen receptor and PR levels >1% were considered positive for hormone receptors. In cases where the HER2 receptor level was +2 (equivocal), the receptor level was verified using fluorescence *in situ* hybridization analysis. Cases in which both receptors were detected as level 0 and +1 were considered negative, and cases in which level +3 was detected were considered positive. The tumor and node stages were evaluated according to the 8th edition of the American Joint Committee on Cancer staging manual.¹¹

In addition, SLNs were identified using 1% isosulfan blue dye solution on biopsies of dye-retaining nodes. During surgery, imprint cytology and the sequential sectioning of SLNs were performed. Tissue blocks were examined for metastatic cells, and positive cases underwent additional ALND. The final diagnosis was determined using paraffin-em-

Main points

- The flare sign is characterized by increased signal intensity surrounding the lymph nodes observed on fat-suppressed T2-weighted images.
- Breast magnetic resonance imaging, particularly when assessing the presence of the flare sign, demonstrated a sensitivity of 65.7%, specificity of 96%, positive predictive value of 97.8%, negative predictive value of 51.1%, and accuracy of 73.9% in detecting extracapsular extension during preoperative evaluation.
- The presence of the axillary flare sign was associated with several factors indicating the aggressiveness of the primary tumor, including the presence of ≥4 axillary metastatic lymph nodes ($P < 0.001$), larger primary tumor size ($P = 0.033$), lymphovascular invasion in the primary tumor ($P < 0.001$), and perineural invasion ($P = 0.001$).

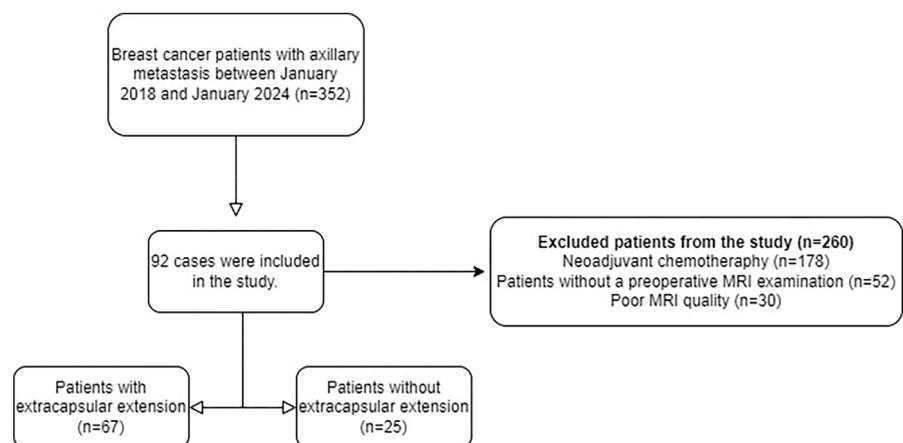


Figure 1. Study participant flow diagram. MRI, magnetic resonance imaging.

bedded tissues stained with hematoxylin and eosin, and metastasis was classified as isolated tumor cells, micrometastasis, and macrometastasis.⁶ Patients with macrometastases in paraffin sections but negative imprint cytology underwent further ALND. Histopathology helped estimate the number of excised benign and metastatic lymph nodes. ECE was defined as positive if a metastatic tumor had spread to the extranodal fat, with or without a desmoplastic stromal response (Figure 2).⁶ The number of metastatic lymph nodes on ECE was also recorded. The length of the ECE, presence of extranodal blood vessel tumor emboli, and extranodal tumor deposits in the metastatic lymph nodes were not evaluated.

Magnetic resonance examination and image acquisition

MRI examinations were performed in the prone position using a 1.5T MRI device (Ingenia; Philips Healthcare, Amsterdam, Netherlands) with a seven-channel dedicated breast coil. The MRI protocol was as follows: axial T2-weighted fat-suppressed sequence [repetition time (TR): 4,317 ms, echo time (TE): 70 ms, slice thickness 3.5 mm, matrix 300 × 258] and T1-weighted turbo spin echo sequence (TR: 490 ms; TE: 8 ms; slice thickness 3.5 mm, matrix 300 × 364); for contrast-enhanced MRI, the gadobutrol dose was 0.1 mmol/kg, and images were obtained six times after saline was injected at a rate of 20 mL/s and then at 2 mL/s. The first imaging was performed 60 s after the contrast injection. Post-contrast sagittal reformatted images were obtained. Images were evaluated using a Picture Archiving Communication System (ISD7, Sectra, Linköping, Sweden). The presence of an increased signal in the axillary fatty tissue surrounding the lymph node (flare sign) on the T2-weighted images, irregular nodal contour (shaggy margin), axillary asymmetry (diagnosed when lymph nodes in the affected axilla differed in number or size compared with the opposite side), loss of the fatty hilum in the most suspicious lymph node, and morphological features (long- and short-axis diameters) were examined on T1-weighted images (Figure 3). Two radiologists with 7- and 6-years' experience in breast radiology retrospectively evaluated the results. The radiologists were blinded to patients' ECE status. Images were reassessed, and a consensus was obtained if the results varied. In case of discrepancies, the two radiologists reached a final decision through discussion.

Statistical analysis

Descriptive analysis was used to determine the frequency and distribution of patient age; postoperative tumor and node stages; maximum tumor size; tumor histology; hormone status; Ki67 index; Cerb2 positivity; lymphovascular invasion in the primary tumor; presence of perineural invasion; number of total, metastatic, and non-metastatic lymph nodes detected in the axilla; and number of metastatic lymph nodes with ECE.

Data analysis was performed using the Statistical Package for Social Sciences (v.25.0, IBM, Armonk, NY, USA). The suitability of continuous variables for a normal distribution was examined using the Shapiro–Wilk test. Receiver operating characteristic (ROC)

curve analysis was performed for parameters that had a significant effect on ECE, and results are presented as area under the curve (AUC), sensitivity and specificity, and 95% confidence intervals (CIs). Sensitivity and specificity were calculated using the Youden index. A logistic regression analysis was performed to determine the effects of various parameters on ECE. Pearson's chi-square test, Fisher's exact test, the Fisher–Freeman–Halton exact test, and Yates correction were used to analyze independent categorical variables. The Mann–Whitney U test was used in two independent group analyses because the data did not exhibit a normal distribution. The statistical significance level was accepted at 0.05.

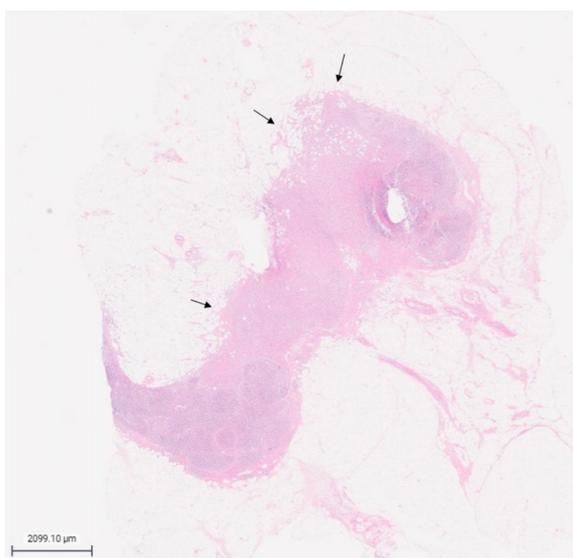


Figure 2. Photomicrograph of a lymph node revealing metastatic cells spreading into extranodal fat (arrows).

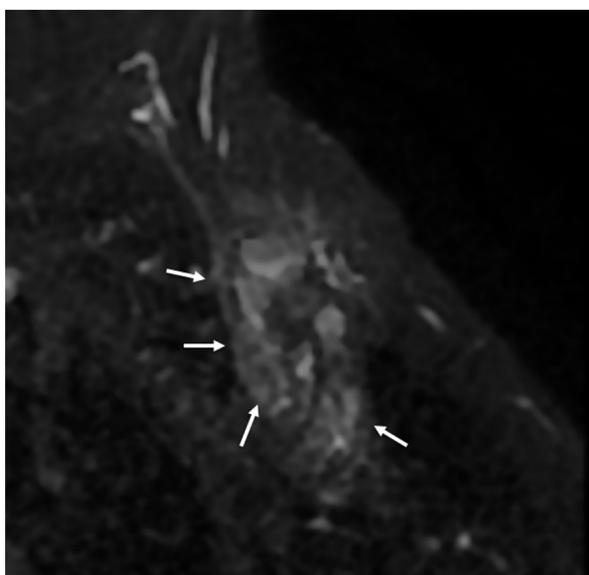


Figure 3. Axial fat-suppressed T2-weighted magnetic resonance image of the axilla in a 65-year-old female patient with invasive ductal carcinoma revealing an increased signal around the metastatic lymph nodes (flare sign) (arrows). Extracapsular extension was detected in five out of eight metastatic lymph nodes as a result of axillary lymph node dissection.

Results

Patient, clinical, and histopathological data

The average age of the 92 women was 55.9 (range: 32–75) years. The most common histological subtype was invasive ductal carcinoma (68.5%). Of the 92 patients, 67 exhibited ECE, whereas the remaining 25 did not exhibit any signs of ECE. The clinical and histopathological data are presented in Table 1.

The mean interval between MRI and surgery was 48 days, with a range of 4–92 days (± 25.5 days). Overall, 35 patients (38%) underwent SLN biopsy, 24 (26.1%) underwent both SLN biopsy and ALND, and 33 (35.9%) underwent ALND. The median number of harvested SLNs was 9 (range: 3–25). Table 2 summarizes the axillary lymph node data.

Statistical analysis

The flare sign, shaggy margin, and presence of at least one of these (flare sign or shaggy margin) were identified as distinguishing variables for ECE in the ROC analysis. The flare sign had an AUC of 0.808 (95% CI: 0.719–0.898; $P < 0.001$); shaggy margin, 0.731 (95% CI: 0.623–0.840; $P < 0.001$); and presence of at least one of these findings, 0.823 (95% CI: 0.723–0.923; $P < 0.001$). The sensitivity of the flare sign was 65.7%, with a specificity of 96.0%, whereas the sensitivity of the shaggy margin was 58.2%, with a specificity of 88.0%. We calculated the sensitivity and specificity for the presence of at least one of the findings (flare sign or shaggy margin) to be 80.6% and 84.0%, respectively (Table 3).

Logistic regression analysis revealed that the existence of a flare sign was associated with an odds ratio (OR) of 45.913 for ECE, whereas a shaggy margin was associated with an OR of 10.214 ($P < 0.001$). The OR was 21.808 when there was either a flare sign or a shaggy margin ($P < 0.001$). The presence of flare signs and/or shaggy margins in cases with ECE was much more frequent than in cases without ECE (Supplementary Table 1).

Patients with flare signs had a greater number of lymph nodes with ECE ($P < 0.001$) and a larger maximal tumor diameter ($P = 0.004$) than those without flare signs. Patients exhibiting flare signs had increased rates of lymphovascular invasion, perineural invasion, and the presence of ≥ 4 metastatic lymph nodes in the axilla. The presence of a flare sign was not associated with the Ki67 index or histological grade (Table 4).

Discussion

Multiple studies have examined the edema surrounding breast masses, while research on the axilla remains limited. This finding is associated with malignancy¹² and is a key indicator of breast cancer aggressiveness, recurrence, and prognosis.^{13,14}

Prediction of the presence of ECE has been frequently studied in head and neck malignancies in the radiology literature. Kimura et al.¹⁵ defined the characteristics that they investigated for the existence of

ECE as flare sign and shaggy margin. Researchers determined that the flare sign has the following diagnostic performance metrics for detecting ECE: sensitivity, 77%; specificity, 93%; accuracy, 88%; positive predictive value, 83%; and negative predictive value, 90%. We determined that some diagnostic metrics, specifically the positive predictive value, sensitivity, and specificity, were similar in our study, despite focusing on completely different anatomical locations.

The literature contains only a limited number of publications aimed at detecting ECE

Table 1. Clinical and histopathological features of the patients

	Patients (n = 92)
Median age (years) (range)	55.9 (32–75)
Median tumor size (mm) (range)	28.6 (7–160)
Receptor status	
HR+/HER2-	77 (83.7%)
HR+/HER2+	7 (7.6%)
HR-/HER2+	3 (3.3%)
HR-/HER2-	5 (5.4%)
Histologic subtype	
Ductal	63 (68.5%)
Lobular/mixed	15 (16.3%)
Other	14 (15.2%)
Lymphovascular invasion	
Absent	39 (42.4%)
Present	53 (57.6%)
Perineural invasion	
Absent	34 (37%)
Present	58 (63%)
Necrosis	
Absent	79 (85.9%)
Present	13 (14.1%)
Calcification	
Absent	62 (67.4%)
Present	30 (32.6%)
Ki67 (%)	
<14	7 (7.6%)
≥ 14	85 (92.4%)
Clinical T stage	
T1 (≤ 2 cm)	40 (43.5%)
T2 (>2–5 cm)	42 (45.7%)
T3 (>5 cm)	8 (8.7%)
T4	2 (2.2%)
Tumor grade (Scarff–Bloom–Richardson)	
1	4 (4.3%)
2	57 (62%)
3	20 (21.7%)

HER2, human epidermal growth factor receptor 2; HR, hormone receptor.

using breast MRI. In a review by Gupta et al.¹⁶, perinodal edema was identified as a specific ECE indicator. It manifests as a hyperechoic halo surrounding the lymph nodes on ultrasound, blurring of the lymph node margins on computed tomography, and areas of T2 hyperintensity in the perinodal fat on MRI.¹⁶ However, these findings are not supported by specific studies. In the study by Kim et al.⁴, ECE was defined as “the presence of strand-like or circumferential T2 high-signal intensities surrounding the nodes.” A similar methodology was used in the present study. The study reported strong agreement in the assessment of perinodal infiltration (k: 0.74; 95% CI: 0.64–0.85), emphasizing that cases with perinodal infiltration were associated with sentinel node identification. However, that study did not conduct a histopathological examination of the axillary lymph nodes in the presence of ECE, and it included patients with a history of NAC. The perinodal infiltration area also has the potential to respond to NAC. By contrast, our study did not include patients who received NAC, and we assessed the presence of ECE individually in each dissected lymph node. The study by Loiselle et al.¹⁷, which involved small patient groups, reported the sensitivity of the peak enhancement level on MRI for detecting ECE as 60% and specificity as 100%. However, this study included only five patients with ECE. Misselt et al.¹⁸ discovered on ultrasonography that unclear margins, node matting, and perinodal edema had high specificity (87%, 84%, and 75%, respectively) but low sensitivity (34%, 52%, and 64%, respectively) in detecting ECE. In our study, we considered the presence of perinodal edema to be equivalent to a flare sign. Notably, MRI has comparable sensitivity but greater specificity. When the presence of a shaggy margin was evaluated alongside a flare sign, sensitivity increased considerably, although this was accompanied by a decrease in specificity.

Our findings suggest that the flare sign is useful in nodal staging, as we found a significant association between its presence and N2 or higher axillary involvement. These results are consistent with those of other studies that have reported the utility of breast MRI in predicting advanced axillary involvement.^{19–21} The American College of Surgeons Oncology Group Z0011 study revealed that ALND may not be necessary in early-stage breast cancer with metastases detected in one or two SLNs.²² However, the study did not establish specific exclusion criteria for patients with ECE. Our findings indicate that ECE may influence axillary management. The

high number of metastatic axillary lymph nodes observed in patients with ECE may support the consideration of ALND without

prior SLN biopsy. Supporting this, a study involving 655 patients demonstrated that ECE detected in SLNs was associated with

Table 2. Features of axillary lymph nodes

	All patients (n = 92)	Patients with ECE (n = 67)	Patients without ECE (n = 25)
Axillary surgery			
Sentinel lymph node biopsy	35	17	18
Axillary lymph node dissection	33	27	6
Conversion to axillary lymph node dissection	24	23	1
Number of lymph nodes removed (median, range)	9 (3–25)	10 (3–25)	5 (3–20)
Number of positive lymph nodes removed			
0	0	0	0
1–3	56	32	24
≥4	36	35	1
N stage			
N0i+	1	0	1
N1mi	3	0	3
N1 (1–3)	54	34	20
N2 (4–9)	24	24	0
N3 (≥10)	10	9	1
Number of examined lymph nodes (mean)	9.6	10.5	6.9
Number of metastatic lymph nodes (mean, range)	4.1 (1–22)	5 (1–22)	1.6 (1–11)
ECE in metastatic lymph nodes			
Number of patients	67	67	0
Number of lymph nodes (mean, range)	3.13 (1–12)		0
Asymmetry	55/92 (63%)	46	9
Loss of fatty hilum	20/92 (21.7%)	16	4
Long-axis diameter (mm) (range)	18.3 (6–44)	18.7 (6–44)	17.3 (8–35)
Short-axis diameter (mm) (range)	9.9 (4–23)	10.4 (4–23)	8.4 (5–22)
Long/short axis ratio	1.9 (1.1–3.8)	1.8 (1.1–3.3)	2.1 (1.3–3.4)
Cortical thickness (mm) (range)	4.4 (1.9–11) (72/92)	4.7 (2.1–11) (51/67)	3.6 (1.9–6.3) (21/25)
Cortical thickness type			
Homogeneous	26/92 (28.2%)	17	9
Non-homogeneous	58/92 (63%)	44	14
Nodular	8/92 (8.7%)	6	2
Flare sign	45/92 (48.9%)	44/67 (65.7%)	1/25 (4%)
Shaggy margin	42/92 (45.7%)	39/67 (58.2%)	3/25 (12%)

ECE, extracapsular extension.

Table 3. Receiver operating characteristic analysis results for parameters that have a significant effect on extracapsular extension

Variables	AUC (95% CI)	P*	Sensitivity (%)	Specificity (%)	PPV	NPV	Accuracy
Flare sign	0.808 (0.719–0.898)	<0.001	65.7	96	97.8	51.1	73.9
Shaggy margin	0.731 (0.623–0.840)	0.001	58.2	88	92.9	44	66.3
Flare sign and/or shaggy margin	0.823 (0.723–0.923)	<0.001	80.6	84	93.1	61.8	81.5

*Receiver operating characteristic analysis, AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

Table 4. Relationship of the flare sign to various parameters

Variables	β	P^*	OR	95% CI	
Number of metastatic lymph nodes with ECE	1.019	<0.001	2.771	1.746	4.399
Tumor size	0.037	0.033	1.038	1.003	1.074
Lymphovascular invasion	1.954	<0.001	7.059	2.752	18.108
Perineural invasion	1.514	0.001	4.545	1.796	11.501
Four or more metastatic lymph nodes	2.338	<0.001	10.357	3.777	28.398
Tumor ki67 index	1.034	0.232	2.812	0.517	15.307
Grade	0.057	0.897	1.059	0.448	2.5

*Logistic regression analysis; OR, odds ratio; CI, confidence interval; ECE, extracapsular extension.

increased axillary nodal burden, disease recurrence, and overall mortality.²³ Additionally, we noted a significantly higher number of non-SLN-positive metastatic axillary lymph nodes in these cases. Another study involving 221 patients indicated that the presence of ECE was directly proportional to the number of involved axillary lymph nodes and disease stage,²⁴ highlighting its potential implications in clinical decision-making.

Our study demonstrated that the flare sign was significantly associated with perineural and lymphovascular invasion in the primary tumor. Reports suggest that perineural and lymphovascular invasion play a significant role in predicting survival outcomes.^{25,26} Therefore, the flare sign, which indicates tumor aggressiveness, has the potential to serve as a prognostic biomarker.

This study has some limitations. Owing to the prolonged interval between MRI and surgery, the findings may change, potentially affecting the results of the study. Furthermore, the length of ECE was not assessed in our study. The length of the ECE may be valuable in certain cases in which false negativity is reported. In their study of 11,730 patients, Gooch et al.²⁷ identified ECE in 2.8% of cases. The study revealed a higher rate of detection of ≥ 4 metastatic lymph nodes in cases with ECE measuring >2 mm than in cases with ECE measuring <2 mm. Extranodal tumor blood vessel embolisms (ENBTVE) and deposits can occur concurrently or independently in the axilla, but ENBTVE was not assessed during pathological examinations conducted in our study. Some cases in which the axillary flare sign is observed may be attributed to these factors. Further studies are required to address this issue. We performed only SLN surgery in some cases. The literature reports a false negative rate of 8.3% for SLN biopsy, and pathological sampling may not have included all cases with possible ECE.²⁸ Similarly, the pathology specimens did not reveal any extranodal tumor deposits. The increased

signal intensity in the axillary fatty tissue may also be caused by isolated tumor deposits. We excluded certain cases from the evaluation due to MRI artifacts and inadequate axillary inclusion in the field of view. This may have also affected our results.

In conclusion, this study demonstrates that the flare sign on breast MRI is a highly valuable diagnostic marker for predicting ECE and axillary nodal status in patients with breast cancer. It exhibits high specificity and positive predictive value. Furthermore, its association with tumor aggressiveness, including lymphovascular and perineural invasion, highlights its prognostic value. Incorporating flare sign into preoperative assessment shows promise in determining the status of metastatic axillary lymph nodes. Further studies with larger patient cohorts are warranted to validate these findings.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Kuerer HM. *Kuerer's Breast Surgical Oncology*. New York, NY: The McGraw-Hill Companies; 2010. [CrossRef]
- Morris E, Comstock C, Lee C, Lehman C, Ikeda D, Newstead G. ACR BI-RADS® magnetic resonance imaging. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology. 2013. [CrossRef]
- Tang P, Moravek M, Oprea-Ilie G, Mon KS, Pambuccian SE. Extranodal extension, an international survey on its evaluation and reporting in breast cancer patients. *Pathol Res Pract*. 2022;237:154070. [CrossRef]
- Kim WH, Kim HJ, Park CS, et al. Axillary Nodal Burden Assessed with Pretreatment Breast MRI is associated with failed sentinel lymph node identification after neoadjuvant chemotherapy for breast cancer. *Radiology*. 2020;295(2):275-282. [CrossRef]

- Fitzgibbons PL, Connolly JL, Bose S, et al. Protocol for the examination of resection specimens from patients with invasive carcinoma of the breast. Northfield, IL: College of American Pathologists; Version: InvasiveBreast 4.4.0.0. Protocol Posting Date: February 2020 [Accessed June 20, 2024]. [CrossRef]
- Yang X, Ma X, Yang W, Shui R. Clinical significance of extranodal extension in sentinel lymph node positive breast cancer. *Sci Rep*. 2020;10(1):14684. [CrossRef]
- Fujii T, Yanagita Y, Fujisawa T, Hirakata T, Iijima M, Kuwano H. Implication of extracapsular invasion of sentinel lymph nodes in breast cancer: prediction of nonsentinel lymph node metastasis. *World J Surg*. 2010;34(3):544-548. [CrossRef]
- Schwentner L, Dayan D, Wockel A, et al. Is extracapsular nodal extension in sentinel nodes a predictor for nonsentinel metastasis and is there an impact on survival parameters?—A retrospective single center cohort study with 324 patients. *Breast J*. 2018;24(4):480-486. [CrossRef]
- Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary nodal evaluation in breast cancer: state of the art. *Radiology*. 2020;295(3):500-515. [CrossRef]
- Baltzer PA, Dietzel M, Burmeister HP, et al. Application of MR mammography beyond local staging: is there a potential to accurately assess axillary lymph nodes? evaluation of an extended protocol in an initial prospective study. *AJR Am J Roentgenol*. 2011;196(5):W641-647. [CrossRef]
- Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. 2017;67(2):93-99. [CrossRef]
- Baltzer PA, Yang F, Dietzel M, et al. Sensitivity and specificity of unilateral edema on T2w-TSE sequences in MR-mammography considering 974 histologically verified lesions. *Breast J*. 2010;16(3):233-239. [CrossRef]
- Chen Y, Wang L, Luo R, Liu H, Zhang Y, Wang D. Focal breast edema and breast edema score on T2-weighted images provides valuable biological information for invasive breast cancer. *Insights Imaging*. 2023;14(1):73. [CrossRef]
- Cheon H, Kim HJ, Kim TH, et al. Invasive breast cancer: prognostic value of peritumoral edema identified at preoperative mr imaging. *Radiology*. 2018;287(1):68-75. [CrossRef]
- Kimura Y, Sumi M, Sakihama N, Tanaka F, Takahashi H, Nakamura T. MR imaging criteria for the prediction of extranodal spread of metastatic cancer in the neck. *AJNR Am J Neuroradiol*. 2008;29(7):1355-1359. [CrossRef]

16. Gupta A, Metcalf C, Taylor D. Review of axillary lesions, emphasising some distinctive imaging and pathology findings. *J Med Imaging Radiat Oncol.* 2017;61(5):571-581. [\[CrossRef\]](#)
17. Loisselle CR, Eby PR, Peacock S, Kim JN, Lehman CD. Dynamic contrast-enhanced magnetic resonance imaging and invasive breast cancer: primary lesion kinetics correlated with axillary lymph node extracapsular extension. *J Magn Reson Imaging.* 2011;33(1):96-101. [\[CrossRef\]](#)
18. Misselt PN, Glazebrook KN, Reynolds C, Degnim AC, Morton MJ. Predictive value of sonographic features of extranodal extension in axillary lymph nodes. *J Ultrasound Med.* 2010;29(12):1705-1709. [\[CrossRef\]](#)
19. Bode M, Schrading S, Masoumi A, et al. Abbreviated MRI for comprehensive regional lymph node staging during pre-operative breast MRI. *Cancers (Basel).* 2023;15(6):1859. [\[CrossRef\]](#)
20. Hyun SJ, Kim EK, Moon HJ, Yoon JH, Kim MJ. Preoperative axillary lymph node evaluation in breast cancer patients by breast magnetic resonance imaging (MRI): can breast MRI exclude advanced nodal disease? *Eur Radiol.* 2016;26(11):3865-3873. [\[CrossRef\]](#)
21. Kuijs VJ, Moosdorff M, Schipper RJ, et al. The role of MRI in axillary lymph node imaging in breast cancer patients: a systematic review. *Insights Imaging.* 2015;6(2):203-215. [\[CrossRef\]](#)
22. Giuliano AE, McCall L, Beitsch P, et al. Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: the American College of Surgeons Oncology Group Z0011 randomized trial. *Ann Surg.* 2010;252(3):426-432. [\[CrossRef\]](#)
23. Choi AH, Surrusco M, Rodriguez S, et al. Extranodal extension on sentinel lymph node dissection: why should we treat it differently? *Am Surg.* 2014;80(10):932-935. [\[CrossRef\]](#)
24. Ilknur GB, Hilmi A, Tülay C, et al. The importance of extracapsular extension of axillary lymph node metastases in breast cancer. *Tumori.* 2004;90(1):107-111. [\[CrossRef\]](#)
25. Narayan P, Flynn J, Zhang Z, et al. Perineural invasion as a risk factor for locoregional recurrence of invasive breast cancer. *Sci Rep.* 2021;11(1):12781. [\[CrossRef\]](#)
26. Schoppmann SF, Bayer G, Aumayr K, et al. Prognostic value of lymphangiogenesis and lymphovascular invasion in invasive breast cancer. *Ann Surg.* 2004;240(2):306-312. [\[CrossRef\]](#)
27. Gooch J, King TA, Eaton A, et al. The extent of extracapsular extension may influence the need for axillary lymph node dissection in patients with T1-T2 breast cancer. *Ann Surg Oncol.* 2014;21(9):2897-2903. [\[CrossRef\]](#)
28. Wong SL, Edwards MJ, Chao C, et al. Sentinel lymph node biopsy for breast cancer: impact of the number of sentinel nodes removed on the false-negative rate. *J Am Coll Surg.* 2001;192(6):684-689. [\[CrossRef\]](#)

Supplementary Table 1. Parameters significantly affecting extracapsular extension

Variables	β	P^{h}	OR	95% CI	
Flare sign	3.827	<0.001	45.913	5.834	361.304
Shaggy margin	2.324	0.001	10.214	2.783	37.491
Flare sign and/or shaggy margin	3.082	<0.001	21.808	6.382	74.519

^hLogistic regression analysis; OR, odds ratio; CI, confidence interval.



Revolutionizing cardiac imaging: how photon-counting computed tomography is redefining coronary artery stent assessment

Çağdaş Topel¹
 Furkan Ufuk²

¹Northwestern University, McGaw Medical Center,
Department of Radiology, Chicago, USA

²The University of Chicago, Department of Radiology,
Chicago, USA

Photon-counting computed tomography (PCCT) represents a groundbreaking advancement in medical imaging, offering unprecedented improvements in spatial resolution, contrast enhancement, and artifact reduction compared with conventional energy-integrating detector (EID) CT scanners.¹ Unlike traditional EID CT, which integrates energy from multiple X-ray photons, PCCT detects and classifies individual photons based on their energy levels.² This capability enables superior tissue differentiation, reduced image noise, and substantially enhanced image clarity—critical advantages in cardiovascular imaging, where precise visualization of coronary arteries and stents is essential for accurate diagnosis and treatment planning.^{1,2}

Coronary artery disease remains a leading global cause of morbidity and mortality. Accurate assessment of coronary anatomy, particularly in patients with percutaneous coronary interventions and stent placement, is crucial for evaluating treatment efficacy and disease progression.¹⁻³ While coronary CT angiography (CCTA) is a widely adopted non-invasive imaging technique for this purpose, its accuracy is often compromised by artifacts, most notably, blooming effects that obscure stent lumens, particularly in small-diameter stents.³ PCCT overcomes these limitations by offering ultra-high-resolution (UHR) imaging, which sharpens stent visualization and enhances the precision of non-invasive assessments.^{3,4}

Recent research by Stein et al.⁵ highlights the transformative potential of PCCT in stent evaluation. Their study, which assessed the performance of dual-source PCCT in imaging small coronary stents using different acquisition modes in a controlled phantom model, revealed striking results. The UHR mode, featuring an ultra-thin 0.2 mm collimation, achieved a median image quality score of 4.0 [interquartile range (IQR): 3.67–4.00], with 37.5% of cases rated as “excellent.” In contrast, the high-pitch mode performed substantially worse, yielding a median score of 2.0 (IQR: 1–3) and proving non-diagnostic in 6.3% of cases.⁵ These findings underscore the unparalleled ability of UHR PCCT to delineate stent boundaries, minimize blooming artifacts, and provide diagnostic confidence comparable to invasive angiography.

Current CCTA guidelines limit its application to stents ≥ 3 mm due to the resolution constraints of conventional CT.⁶ Metallic struts in smaller stents often produce pronounced blooming artifacts, obscuring the lumen and reducing diagnostic reliability.^{3,4,6} However, recent studies suggest that PCCT may redefine these standards.^{5,7,8} In a study by Hagar et al.,⁷ UHR PCCT demonstrated high diagnostic performance in 44 coronary stents, achieving a sensitivity of 100%, specificity above 87%, and an inter-reader agreement of 90.1%. Similarly, Qin et al.⁸ assessed 131 stents and found that UHR images substantially improved in-stent lumen visualization compared with standard-resolution images, with UHR achieving an accuracy of 88.0% compared with 78.3% for standard resolution. Moreover, future advancements in PCCT hardware, including next-generation detectors with <0.1 mm spatial resolution, are expected to further refine its ability to visualize stent strut thickness and detect microcalcifications with unmatched precision.

Despite these promising developments, challenges remain. Motion artifacts in patients with arrhythmias can affect image quality, and global access to PCCT technology is limited, with fewer than 5% of imaging centers currently equipped with this advanced modality.⁹

Corresponding author: Çağdaş Topel

E-mail: cagdas.topel@northwestern.edu

Received 5 February 2025; accepted 25 February 2025.



Epub: 26.03.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2025.253274

You may cite this article as: Topel Ç, Ufuk F. Revolutionizing cardiac imaging: how photon-counting computed tomography is redefining coronary artery stent assessment. *Diagn Interv Radiol.* 2025;31(4):340-341.

In conclusion, PCCT is poised to revolutionize cardiac imaging by surpassing the 3 mm imaging barrier, enhancing tissue characterization, and integrating AI-driven automation into diagnostic workflows. As imaging guidelines evolve and costs decline, PCCT has the potential to democratize high-precision cardiovascular imaging, making state-of-the-art diagnostics more widely accessible. The future of cardiology is becoming less invasive—driven by insight, innovation, and the transformative potential of photon-counting technology.

Conflict of interest disclosure

Furkan Ufuk, MD, is Section Editor in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

References

1. Meloni A, Maffei E, Positano V, et al. Technical principles, benefits, challenges, and applications of photon counting computed tomography in coronary imaging: a narrative review. *Cardiovasc Diagn Ther.* 2024;14(4):698-724. [\[CrossRef\]](#)
2. <http://doi.org/10.21037/cdt-24-52> [\[CrossRef\]](#)
3. Sharma SP, Lemmens MDK, Smulders MW, Budde RPJ, Hirsch A, Mihal C. Photon-counting detector computed tomography in cardiac imaging. *Neth Heart J.* 2024;32(11):405-416. <http://doi.org/10.1007/s12471-024-01904-5> [\[CrossRef\]](#)
4. Sharma A, Cerdas MG, Reza-Soltani S, et al. A Review of Photon-Counting Computed Tomography (PCCT) in the Diagnosis of Cardiovascular Diseases. *Cureus.* 2024;16(11):e73119. Published 2024 Nov 6. <http://doi.org/10.7759/cureus.73119> [\[CrossRef\]](#)
5. Fahrni G, Boccalini S, Prieur C, Douek P, Si-Mohamed SA. Comprehensive Imaging of Coronary Stent Using Ultra-High Resolution Spectral Photon Counting CT: A Multimodality Validation. *JACC Cardiovasc Interv.* 2023;16(19):2466-2468. <http://doi.org/10.1016/j.jcin.2023.07.035> [\[CrossRef\]](#)
6. Stein T, Muhlen Cvz, Verloh N et al. Evaluating small coronary stents with dual-source photon-counting computed tomography: effect of different scan modes on image quality and performance in a phantom. *Diagn Interv Radiol.* 2025 Jan;31(1):29-38. <http://doi.org/10.4274/dir.2024.242893> [\[CrossRef\]](#)
7. Narula J, Chandrashekar Y, Ahmadi A, et al. SCCT 2021 Expert Consensus Document on Coronary Computed Tomographic Angiography: A Report of the Society of Cardiovascular Computed Tomography. *J Cardiovasc Comput Tomogr.* 2021;15(3):192-217. <http://doi.org/10.1016/j.jcct.2020.11.001> [\[CrossRef\]](#)
8. Hagar MT, Soschynski M, Saffar R, et al. Ultra-high-resolution photon-counting detector CT in evaluating coronary stent patency: a comparison to invasive coronary angiography. *Eur Radiol.* 2024;34(7):4273-4283. <http://doi.org/10.1007/s00330-023-10516-3> [\[CrossRef\]](#)
9. Qin L, Zhou S, Dong H, et al. Improvement of coronary stent visualization using ultra-high-resolution photon-counting detector CT. *Eur Radiol.* 2024;34(10):6568-6577. <http://s> [\[CrossRef\]](#)



Multidisciplinary approach to diagnostic radiology education: a novel educational intervention for Turkish medical students

Parth Patel¹
 Emre Altınmakas²
 Görkem Ayas³
 Rachel Stanietzky⁴
 Madeline L. Stewart⁵
 Abdelrahman Elshikh⁶
 Disha Ram⁴
 Hrishika Bhosale⁴
 Mohamed Eltaher⁴
 Serageldin Kamel⁴
 Munevver N. Duran⁷
 Umut Yücel⁸
 Mohamed Badawy⁴
 Scott Rohren⁴
 Khaled M. Elsayas⁴

¹Baylor College of Medicine, Department of Radiology, Texas, USA

²Icahn School of Medicine at Mount Sinai, Department of Diagnostic, Molecular and Interventional Radiology, New York, USA

³Koç Üniversitesi, Graduate School of Health Sciences, İstanbul, Türkiye

⁴The University of Texas MD Anderson Cancer Center, Department of Abdominal Radiology, Texas, USA

⁵New Orleans School of Medicine, Louisiana State University Health, Louisiana, USA

⁶Alexandria University Faculty of Medicine, Department of Radiology, Alexandria, Egypt

⁷University of Texas, Southwestern Medical Center, Texas, USA

⁸AMEOS Klinikum Am Bürgerpark Bremerhaven, Clinic for Pulmonology and Intensive Care Medicine, Bremerhaven, Germany

Corresponding author: Khaled M. Elsayas

E-mail: kmelsayas@mdanderson.org

Received 25 April 2024; revision requested 17 May 2024; accepted 03 June 2024.



Epub: 02.07.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242828

ABSTRACT

Teleconferencing can facilitate a multidisciplinary approach to teaching radiology to medical students. This study aimed to determine whether an online learning approach enables students to appreciate the interrelated roles of radiology and other specialties during the management of different medical cases. Turkish medical students attended five 60–90-minute online lectures delivered by radiologists and other specialists from the United States and Canada through Zoom meetings between November 2020 and January 2021. Student ambassadors from their respective Turkish medical schools recruited their classmates with guidance from the course director. Students took a pretest and posttest to assess the knowledge imparted from each session and a final course survey to assess their confidence in radiology and the value of the course. A paired t-test was used to assess pretest and posttest score differences. A 4-point Likert-type scale was used to assess confidence rating differences before and after attending the course sessions. A total of 1,458 Turkish medical students registered for the course. An average of 437 completed both pre- and posttests when accounting for all five sessions. Posttest scores were significantly higher than pretest scores for each session ($P < 0.001$). A total of 546 medical students completed the final course survey evaluation. Students' rating of their confidence in their radiology knowledge increased after taking the course ($P < 0.001$). Students who took our course gained an appreciation for the interrelated roles of different specialties in approaching medical diagnoses and interpreting radiological findings. These students also reported an increased confidence in radiology topics and rated the course highly relevant and insightful. Overall, our findings indicated that multidisciplinary online education can be feasibly implemented for medical students by video teleconferencing.

KEYWORDS

Educational intervention, online teaching, radiology education, radiology teaching, remote learning, Turkish medical students

Online education can be effective in augmenting the educational experience of students.¹ This study uses high-impact practice principles of online higher education to conduct online education through a video teleconferencing platform. In Türkiye, as elsewhere, the coronavirus disease-2019 (COVID-19) pandemic interrupted medical students' education, as many in-person classes were canceled or moved online, and clinical clerkships were modified to accommodate social distancing measures.² To help address the limitations in radiology education caused by the pandemic and to provide an opportunity to learn about clinical cases from radiologists and other specialists, we created and implemented an online radiology course for a large cohort of medical students in Türkiye.

At the onset of the COVID-19 pandemic, Bao³ published a case study regarding online education at Peking University that provided insight into five high-impact principles of online teaching: relevance, delivery, support, participation, and contingency plans. To help address the limitations in radiology education caused by the pandemic and to provide an opportunity to learn about clinical cases from radiologists and other specialists, we created and implemented an online radiology course for a large cohort of Turkish medical students. In this study, we seek to determine whether our course enabled students to appreciate the interrelated roles of radiology and other specialties in the management of medical cases. Students

completed pretests, posttests, and a course evaluation survey to assess their knowledge of session topics, confidence about radiology, and the effectiveness of the course. Overall, the data suggests that a multispecialty integrated approach to providing radiology education via teleconferencing was well received by medical students.

Methods

Course structure and development

The course director designed an online multidisciplinary radiology lecture series comprising five sessions (liver disease, neck infections, pediatric headaches, shoulder injuries, and first-trimester pregnancy). Distinguished faculties from teaching institutions were invited to participate, and the topics were selected based on the expertise and availability of the faculty volunteers. Sessions occurred biweekly between November 2020 and January 2021 and were conducted over Zoom (Zoom Video Communications, Inc., San Jose, California). Session lecturers were radiologists and non-radiology specialists from different academic institutions throughout the United States and Canada. Each 60–90-minute session focused on the pathology of a given medical case and the role of imaging in making the diagnosis. The interplay between the presiding radiologist and other specialists demonstrated to the students how different healthcare professionals view and approach a given medical presentation.

At the onset of the COVID-19 pandemic, Bao³ published a case study regarding online education at Peking University that provided insight into five high-impact principles of online teaching. The first principle is that the teaching material be relevant to students. This was incorporated by relating radiology topics to pertinent anatomy, physiology, and pathology for each case to increase students' understanding of the role of radiology amongst other specialties. The second

principle of effective delivery of educational material and the fourth principle of audience participation were achieved via video conferencing involving question-and-answer sessions through audio and chat features. The third principle of sufficient support was achieved by having student ambassadors serve as liaisons between the course director and medical students to help coordinate feedback and share notes. The last principle of contingency planning was effectively implemented, as the course was organized amidst the COVID-19 pandemic to address limitations posed upon medical education at the time.

Participants

Student ambassadors from 10 medical schools in Türkiye were recruited by the course director through social media to help recruit participants and coordinate communication with participants during and after the sessions. Each ambassador recruited participants by sending a message to their schools' group chats and personal emails asking all medical students whether they were interested in participating. The ambassadors sent out the registration form using Google Forms.

Because initial registrants exceeded the participant limit of the available Zoom plan (500 participants), ambassadors helped to divide the students randomly into two groups. Each group attended sessions held on a fixed day of the week; half of the students attended sessions on Saturdays and the other half attended sessions on Sundays. The Saturday sessions were live and recorded and then replayed during the Sunday session. Students from the pre-recorded session could have their questions answered by lecturers, with email coordination from student ambassadors. Student ambassadors shared the course information with the students. In addition to helping recruit participants and transmit announcements from the course director to the participants, ambassadors helped prepare pretests and posttests with the session speakers. They also created session handouts that summarized the main points from each session and emailed these handouts to all participants.

Session structure

At the beginning of each live session, students received a link to the pretest, a Qualtrics survey (Qualtrics, Provo, UT), in the Zoom chat. Students were given 10 minutes to complete the pretest. Next, the lecturers

-usually a radiologist and an accompanying specialist- were introduced to the students. The lecturers discussed the case presentation and possible diagnoses. The radiologist explained the imaging modality of choice and interpreted the relevant radiologic findings. Relevant imaging studies in the form of both static images and videos from different modalities were presented. The clinical interpretation of the case was discussed by the specialists and presented alongside the pertinent normal anatomy and imaging findings.

After the lecture, students were given 10 minutes to complete the posttest. The posttest was a Qualtrics survey that was identical to the pretest and provided via a link in the Zoom chat. Following the completion of the posttest, students from the live sessions participated in a question-and-answer session with the speakers. Students from the pre-recorded sessions could email questions and coordinate a discussion through the student ambassadors. Participants submitted their questions using the question-and-answer feature of Zoom. The lecturers were thus able to interact with the medical students in the live audience and provide valuable insights by directly answering their questions.

Course evaluation

The course evaluation consisted of questions asking students about the effectiveness of course sessions and the multidisciplinary approach. Students were also asked to self-identify their gender and year of study in medical school. Students rated each session and provided feedback on the course meeting their expectations and on the educational value of the session topics. In addition, students rated their confidence in the practice of radiology before and after attending the sessions. This was carried out using a scale of four points: not confident, somewhat confident, moderately confident, and very confident. Students answered questions on the best use of radiology practices, interpreting radiographs, safety in radiology, identifying gross abnormalities on imaging, use of imaging as a diagnostic tool, and choosing between different diagnostic tools.

Statistical analysis

A paired t-test was used to compare students' pretest and posttest scores, which were paired using the IP address as the unique identifier. Only data from students who completed both the pretest and

Main points

- Multidisciplinary online education can be feasibly implemented for medical students by video teleconferencing.
- Educational limitations from measures such as social distancing can be addressed through the implementation of effective online teaching.
- Medical students appreciate the role of radiology integrated alongside other medical specialties.

posttest were used in this analysis; students who completed only one test or whose pretests and posttests could not be paired were excluded. This analysis was conducted using Excel (Microsoft, Redmond, Washington). A *P* value of <0.05 was considered significant.

In the course evaluation survey, students were asked to assess their confidence regarding basic radiology skills before and after completion of the course. A 4-point Likert-type scale that ranged from “not confident at all” to “very confident” was used. The difference in the confidence ratings before and after the course completion was evaluated from the course evaluation survey using the Wilcoxon signed-rank test with a one-tailed hypothesis and an alpha value of 0.05. This analysis was performed using R software (R Core Team, Vienna, Austria). Students were

also asked on the course evaluation survey to rate each session, rate the program meeting their expectations, and rate the clinical importance of presented topics.

Results

Participants

The course had 1,438 registrants across 29 Turkish medical schools. Registrants attended sessions voluntarily with an upper limit of 500 for each session. Of those, 1,256 students (86.1%) attended state universities and 202 (13.9%) attended private universities (Table 1). Students who attended the sessions were asked to self-identify their gender and year of study in school. A total of 343 (62.82%) identified as female, 196 (35.90%) identified as male, and 7 (1.29%) identified as other or

preferred not to answer (Table 1). A total of 85 (15.57%) were in their first year, 116 (21.25%) were in their second year, 93 (17.03%) were in their third year, 113 (20.70%) were in their fourth year, 85 (15.57%) were in their fifth year, 52 (9.52%) were in their sixth year, and 2 (0.37%) were in their preparatory year or were alumni (Table 1).

Session pretest and posttest scores

The numbers of participants who completed the pre- and posttest for each session are given in Table 2. In all five sessions, students' posttest scores were significantly higher than their pretest scores (*P* < 0.001). All scores for the live and pre-recorded sessions, along with combined results, show significant improvement in posttest scores compared with pretest scores.

Course evaluation

Of the 546 participants who completed the course evaluation survey, approximately 96% strongly or somewhat agreed that the program increased their knowledge of imaging as a diagnostic tool (Table 3). Approximately 48% of students strongly agreed, and 40% somewhat agreed, that the program increased their interest in radiology. In addition, approximately 71% of the participants strongly agreed and approximately 27% somewhat agreed that the program was a worthwhile experience. More than 93% of students strongly or somewhat agreed that the presence of a surgeon, internist, or other non-radiology specialist during the sessions improved the program. Students reported increased confidence ratings after attending the sessions that were statistically significant for all six survey items outlined in the methods section (Table 4).

Student registrants by medical school type	
State	1,256 (86.10)
Private	202 (13.90)
Gender of attendees, self-identified	
Female	343 (62.82)
Male	196 (35.90)
Other	2 (0.37)
Prefer not to say	5 (0.92)
Number of attendees by year of study	
First year	85 (15.57)
Second year	116 (21.25)
Third year	93 (17.03)
Fourth year	113 (20.70)
Fifth year	85 (15.57)
Sixth year	52 (9.52)
Other	2 (0.37)
Values are n, (%)	

Topic and session date	Students who did both tests, n	Live-session group		Taped-session group		All students	
		Pretest score, mean (SD)	Posttest score, mean (SD)	Pretest score, mean (SD)	Posttest score, mean (SD)	Pretest score, mean (SD)	Posttest score, mean (SD)
Imaging of liver disease (11/7/20–11/8/20)	629	4.58 (1.74)	7.68 (1.70)	4.92 (1.80)	8.19 (1.56)	4.76 (1.78)	7.95 (1.65)
Imaging of neck infections (11/21/20–11/22/20)	419	3.05 (1.45)	3.96 (1.86)	2.98 (1.39)	4.54 (1.66)	3.02 (1.42)	4.25 (1.79)
Imaging of pediatric headaches (12/05/20–12/6/20)	399	3.19 (1.89)	5.95 (2.79)	3.10 (1.79)	6.76 (2.73)	3.15 (1.84)	6.36 (2.78)
Imaging of shoulder injuries (12/19/20–12/20/20)	382	3.71 (1.69)	5.88 (2.44)	4.06 (1.60)	6.61 (2.43)	3.90 (1.65)	6.27 (2.46)
Ultra-sonography of first-trimester pregnancy (01/02/21–01/03/21)	360	3.80 (1.49)	5.70 (2.33)	4.01 (1.66)	6.49 (2.35)	3.91 (1.58)	6.12 (2.37)
All <i>P</i> < 0.001. SD, standard deviation.							

Table 3. Participant responses in course evaluation survey

Survey item	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
This program increased your knowledge of using imaging as a diagnostic tool	60.26	36.63	2.93	0.00	0.18
This program has increased your interest in radiology	48.72	40.11	9.52	1.47	0.18
The material presented in this program was relevant to your medical education	40.66	43.22	12.45	2.75	0.92
The presence of a surgeon, internist, or other non-radiology specialist during education sessions added to the program	62.09	31.14	6.78	0.00	0.00
The program was a worthwhile experience	71.06	26.74	2.20	0.00	0.00

All values are percentages

Table 4. Participant confidence rating before and after the course

Survey item	Mean rating (before course)	Mean rating (after course)	Test statistic (Wilcoxon signed-rank test)
I am familiar with best-use radiology practices	1.85	3.01	1,573
I feel comfortable interpreting radiographs	1.68	2.82	1,004.5
I understand safety in radiology	2.25	3.34	1,723
I am able to identify gross abnormalities on imaging	1.99	3.04	1,584
I am familiar with how imaging is used as a diagnostic tool	2.18	3.25	1,581.5
I am familiar with the different imaging modalities and when to use them	2.01	3.12	1,871

All $P < 0.001$.

Discussion

A multidisciplinary approach to radiology education can be feasibly implemented via video teleconference for medical students. Our course for Turkish medical students was well received and served to increase student confidence in their radiology knowledge, as evidenced by the course evaluation survey results. The findings confirm that our approach enabled students to appreciate the interrelated roles of radiology and other specialties while managing different medical cases. Course evaluation survey results affirmed that students found our multidisciplinary approach worthwhile and gained valuable insight into how radiologists and accompanying specialists approach medical cases and interpret radiological findings. At the onset of the COVID-19 pandemic, Bao³ published a case study regarding online education at Peking University that provided insight into five high-impact principles of online teaching (2020). As outlined in the methods section, these principles were implemented in the course for effective online teaching.

Many studies have discussed different approaches to addressing the educational challenges during the COVID-19 pandemic.³ For example, readout sessions could be performed remotely using teleconferencing software with screen-sharing capabilities.³ Such a platform was used to discuss a “case of the day” for every specialty, host didactic conferences, and virtually present cases to

overcome the pandemic-related challenges of medical education.³ Elsayes et al.⁴ showed that an online multidisciplinary approach involving several specialists helped students gain a more accurate understanding of the role of radiology. This approach teaches students basic and clinical science concepts and realistically prepares them for future practice in which interdisciplinary collaboration is common and important. There is a paucity of data assessing the efficiency of a multidisciplinary approach to medical education.

Our course, and similar ones, provided a solution to the challenges students faced because of COVID-19 distancing measures; we brought together physicians from US and Canadian universities to teach Turkish medical students via Zoom. The high-impact principles from Bao and the multidisciplinary teaching structure enabled us to design an effective multidisciplinary radiology course.³ Students’ educational experience was augmented with pretests and posttests, session handouts, and a question-and-answer session. Other studies have indicated that an interdisciplinary course involving radiology and other disciplines increases students’ comprehension of radiology and anatomy and their overall motivation to learn.⁵⁻⁷ In our experience, medical students seemed to appreciate the integration of clinically relevant content with the radiology and anatomy material. A recent meta-analysis of 62 studies assessing health science students’ perception of online learning demonstrated positive

results, and this method of teaching was preferred to the traditional teaching methods.⁸

In our course, leadership from student ambassadors was crucial for successful implementation. Student ambassadors relayed course information and initial registration opportunities, recruited participants from their medical schools, and prepared session handouts, pretests, and posttests with physician instructors. The ambassadors served as contact points for medical students regarding questions about the course or material. They triaged questions and sought input from physician instructors when needed. Therefore, we encourage implementing a liaison role, as fulfilled by student ambassadors in our course, to serve as a bridge between students and instructors when multiple different academic institutions are involved.

Our study was conducted during the COVID-19 pandemic and has its strengths and limitations. The ability to use online video conferencing for education helped break down geographical barriers and fostered global connections as physicians from the United States and Canada were able to augment the educational experience of Turkish medical students. Technical problems with internet connectivity were rare and statistically insignificant but did disrupt the flow of sessions at times. The limit on live attendance on video conferences impeded the ability of live participation for some students. These factors could be overcome in a setting

where online learning is used more broadly, the lectures recorded, and the limits on live participation removed. A bias may have been introduced in the selection process as those students with a pre-existing interest or proficiency in radiology may be more apt to participate. Additionally, there was no control group to compare the results of the online course with those of traditional teaching methods. The study involved only Turkish medical students, and its applicability to other regions without validation is a limitation.

The COVID-19 pandemic has subsided, and many institutions have returned to conducting in-person didactic lectures. As devastating as the COVID-19 pandemic was, according to a White House document,⁹ future pandemics could be far worse. The document concludes, "the next pandemic will likely be substantially different than COVID-19. We must be prepared to deal with any viral threat".⁹ It is important that we learn from the COVID-19 pandemic and have effective educational systems in place for future crises.

In conclusion, follow-up studies have been planned by the authors to assess the re-

tention of knowledge and skills gained from the course over time. Future studies could also be performed to apply this educational model in different cultural and educational settings to enhance and evaluate the study's broader applicability.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Slanetz PJ, Parikh U, Chapman T, Motuzas CL. Coronavirus disease 2019 (COVID-19) and radiology education-strategies for survival. *J Am Coll Radiol*. 2020;17(6):743-745. [\[CrossRef\]](#)
2. Tokuç B, Varol G. Medical education in Turkey in time of COVID-19. *Balkan Med J*. 2020;37(4):180-181. [\[CrossRef\]](#)
3. Bao W. COVID -19 and online teaching in higher education: a case study of Peking University. *Hum Behav Emerg Technol*. 2020;2(2):113-115. [\[CrossRef\]](#)
4. Elsayes KM, Khan ZA, Kamel S, et Al. Multidisciplinary approach in teaching

diagnostic radiology to medical students: the development, implementation, and evaluation of a virtual educational model. *J Am Coll Radiol*. 2021;18(8):1179-1187. [\[CrossRef\]](#)

5. Boon JM, Meiring JH, Richards PA, Jacobs CJ. Evaluation of clinical relevance of problem-oriented teaching in undergraduate anatomy at the University of Pretoria. *Surg Radiol Anat*. 2001;23(1):57-60. [\[CrossRef\]](#)
6. Dettmer S, Tschernig T, Galanski M, Pabst R, Rieck B. Teaching surgery, radiology and anatomy together: the mix enhances motivation and comprehension. *Surg Radiol Anat*. 2010;32(8):791-795. [\[CrossRef\]](#)
7. Dettmer S, Schmiedl A, Meyer S, et al. Radiological anatomy - evaluation of integrative education in radiology. *Rofo*. 2013;185(9):838-843. [\[CrossRef\]](#)
8. Abdull Mutalib AA, Md Akim A, Jaafar MH. A systematic review of health sciences students' online learning during the COVID-19 pandemic. *BMC Med Educ*. 2022;22(1):524. [\[CrossRef\]](#)
9. Lander ES, Sullivan JJ. American Pandemic Preparedness: Transforming Our Capabilities. *The White House*. 2021. [\[CrossRef\]](#)



A diagnostic model based on magnetic resonance imaging for Menière's disease: a multicentre study

Xinyi Chen^{1*}
 Yanfeng Zhao^{2*}
 Yunchong Han³
 Kai Wei¹
 Shufang Cheng⁴
 Yongjun Ye⁴
 Jie Feng⁴
 Xinchun Huang¹
 Jingjing Xu¹

¹Zhejiang University School of Medicine, The Second Affiliated Hospital, Department of Radiology, Hangzhou, China

²Chinese Academy of Medical Sciences and Peking Union Medical College, National Cancer Center, Cancer Hospital, National Clinical Research Center for Cancer, Department of Radiology, Beijing, China

³The Second Affiliated Hospital of Jiaying University, Department of Radiology, Jiaying, China

⁴The Fifth Affiliated Hospital of Wenzhou Medical University, Zhejiang Key Laboratory of Imaging and Interventional Medicine, Lishui, China

*Joint first authors

Corresponding author: Jingjing Xu

E-mail: jingjingxu@zju.edu.cn

Received 19 February 2025; revision requested 17 March 2025; accepted 13 April 2025.



Epub: 29.05.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2025.253293

PURPOSE

To evaluate the diagnostic performance of delayed post-gadolinium enhancement magnetic resonance imaging (DEMRI) in diagnosing Menière's disease (MD) and to establish an effective MRI-based diagnostic model.

METHODS

This retrospective multicenter study assessed DEMRI descriptors in patients presenting with Ménièreiform symptoms who were examined consecutively between May 2022 and May 2024. A total of 162 ears (95 with MD, 67 controls) were included. Each ear was randomly assigned to either a training set (n = 98) or a validation set (n = 64). In the training cohort, diagnostic models for MD were developed using logistic regression. The area under the curve (AUC) was used to evaluate the diagnostic performance of the different models. The Delong test was applied to compare AUC estimates between models.

RESULTS

The proposed DEMRI diagnostic model demonstrated strong diagnostic performance in both the training cohort (AUC: 0.907) and the validation cohort (AUC: 0.887), outperforming the clinical diagnostic model ($P = 0.01231$; 95% confidence interval: 0.033–0.269) in the validation cohort. The AUC of the DEMRI model was also higher than that of the combined DEMRI-clinical model (AUC: 0.796), although the difference was not statistically significant ($P = 0.054$). In the training set, the sensitivity and specificity of the DEMRI model were 78.9% and 88.5%, respectively.

CONCLUSION

A diagnostic model based on DEMRI features for MD is more effective than one based solely on clinical variables. DEMRI should, therefore, be recommended when MD is suspected, given its significant diagnostic potential.

CLINICAL SIGNIFICANCE

This model may improve the accuracy and timeliness of MD diagnosis, as it is less influenced by the attending physician's level of inquiry or the patient's self-reporting ability. It may also contribute to more effective disease management in patients with MD.

KEYWORDS

Post-gadolinium enhancement magnetic resonance imaging, Menière's disease, endolymphatic hydrops, diagnosis, model

Menière's disease (MD) is a multifactorial condition in which the combined effect of genetic and environmental factors may determine its onset.¹ The main clinical symptoms include idiopathic fluctuating sensorineural hearing loss (SNHL), spontaneous vertigo, aural fullness, and tinnitus. Prosper Ménière first described the disease in 1861, proposing that the pathological site was in the labyrinth rather than the brain.^{2,3} However, diagnosis has been challenging, especially when the initial symptoms are subtle, resulting in limited studies on the epidemiology of MD. The American Academy of Otolaryngology–Head

and Neck Surgery developed guidelines for the diagnosis and therapeutic evaluation of MD in 1972, which were revised in 1985 and 1995.⁴ In 2015, the Barany Society updated and established consensus diagnostic criteria for MD, partly to distinguish migraine-related vertigo from MD.^{5,6} However, these updated criteria still relied on patient self-reports rather than objective medical tests. In addition, an insufficient understanding of MD in some clinical departments has led to delayed diagnosis and treatment.

In 1937, British and Japanese researchers discovered endolymphatic hydrops (EH) in the human temporal bone and provided a pathological description of Menière's syndrome.^{7,8} In 2007, Nakashima et al.⁹ successfully demonstrated EH in a patient with MD using delayed inner ear imaging with a three-dimensional fluid-attenuated inversion recovery (3D-FLAIR) sequence after intratympanic gadolinium injection. Since then, a series of magnetic resonance imaging (MRI) studies on EH have emerged.^{8,10-14} 3D-FLAIR and three-dimensional inversion recovery with real reconstruction (3D-real IR) are the most commonly used imaging sequences for EH.¹¹ With these newer imaging techniques, EH can be visualized in vivo and used to support diagnosis. In addition to EH, several other signs¹⁻²⁰ can also be observed on MRI. As a non-invasive tool, the diagnostic performance of delayed post-gadolinium enhancement MRI (DEMRI) remains to be fully clarified.

The aim of this study is to establish an intuitive and objective diagnostic model for MD, providing an effective diagnostic pathway for patients, improving the efficiency and accuracy of diagnosis, and offering a reference for clinical treatment planning.

Methods

Patients

This multicenter retrospective study followed the principles outlined in the Declaration of Helsinki, including all amendments and revisions. The research was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Zhejiang University School of Medicine IRB-2024-0048 (date: 13.05.2024). Informed written consent was obtained from all participants after an explanation of the nature of the study, as approved by the same ethics committee.

This retrospective study included data from consecutive patients who first visited the otology departments of three medical centers with Menièreform symptoms such as vertigo, hearing loss, tinnitus, and aural fullness and who underwent DEMRI of the inner ear labyrinth between May 2022 and May 2024. A total of 136 patients (272 ears) were retrospectively analyzed. Ultimately, 85 patients (162 ears: 95 MD ears, 67 control ears; mean age: 55.2 ± 13.6 years) were enrolled in the study based on the exclusion criteria (Figure 1). Each ear was treated as a single unit and randomly assigned to either a training set (n = 98 ears) or a validation set (n = 64 ears) in a 6:4 ratio (Figure 1).

Baseline clinical data, including sex, age, affected side, inner ear symptoms (vertigo, hearing loss, tinnitus, aural fullness), and pure tone audiometry (PTA), were extracted from the medical record management

system. Given variations in clinical inquiry, symptoms such as vertigo, hearing loss, tinnitus, and aural fullness were recorded as either "yes" or "no," excluding frequency and duration as specified in the diagnostic criteria for MD. Based on the average hearing threshold from PTA at 0.5 kHz, 1 kHz, and 2 kHz, hearing loss was classified into four stages: stage I = average hearing threshold ≤ 25 dB HL; stage II ≥ 25–40 dB HL; stage III ≥ 40–70 dB HL; and stage IV = average hearing threshold > 70 dB HL.

Magnetic resonance imaging examinations

Patients underwent DEMRI using 3T scanners at the participating centers (Center A: uMR 790, UIH, Shanghai, China; Ceners B and C: Ingenia CX, Philips Healthcare, Netherlands) with a standard 32-channel head and neck coil. Prior to gadolinium administration, a 3D-T2-sampling perfection with application-optimized contrasts using different flip angle evolutions (3D-T2-SPACE) sequence was performed with the following parameters: repetition time (TR): 1300 ms; echo time (TE): 196.68 ms; slice thickness: 0.5 mm; matrix size: 380 × 100; field of view (FOV): 220 × 180 mm; acceleration factor: 2 (2D); scan time: 1 min 47 s. This scan was used to exclude patients with organic brain syndromes, inner ear malformations, or acoustic neuroma.

A 3D-FLAIR sequence was performed 4 hours after administration of a double dose of intravenous gadobutrol (7.5 mL/vial, 1.0 mmol/mL; Bayer AG) to ensure maximum

Main points

- Delayed post-gadolinium enhancement magnetic resonance imaging (DEMRI) of the inner ear enables visualization of endolymphatic hydrops and perilymphatic spaces in patients with Menière's disease (MD), which is critical for diagnosis.
- In the DEMRI-based diagnostic model, the most substantial features were "Cochlea_EH_Grad," "Cochlea_Apex_EH_Score," "VA," and "Vestibule_EH."
- The diagnostic performance of DEMRI for MD is superior to that of clinical information alone.

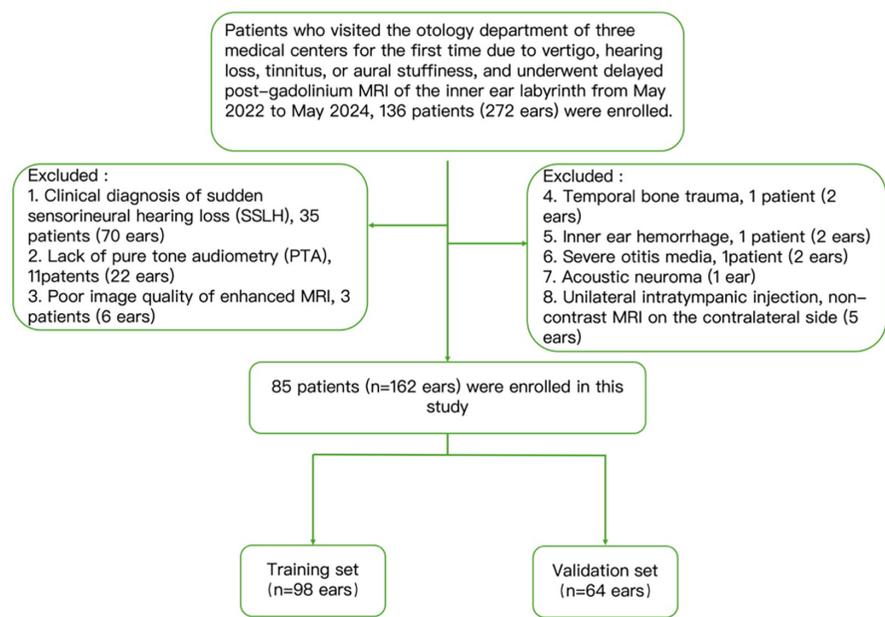


Figure 1. Flowchart of the patient recruitment pathway. MRI, magnetic resonance imaging.

perilymphatic enhancement (PLE). Imaging parameters were: FOV: 220 × 190 mm; section thickness: 0.7 mm; TR: 6500 ms; TE: 426 ms; number of excitations = 1; inversion time = 1935 ms; flip angle = 54°; matrix: 256 × 100; bandwidth: 500 Hz/pixel; turbo factor: 5 (acs); voxel size: 0.86 × 0.86 × 1 mm; scan time: 2 min 56 s. Previous studies^{21,22} have shown that gadobutrol offers advantages over other macrocyclic gadolinium contrast agents in MRI for diagnosing MD due to its higher concentration and greater relaxivity.

Extraction of qualitative and quantitative magnetic resonance imaging features

The MR images were qualitatively analyzed by three experienced radiologists (with 15, 15, and 20 years of experience in head and neck imaging diagnosis, respectively), all blinded to the clinical findings and symptoms.

The degree of EH was indicated by a widening of the negative signal gap within the labyrinth. In this study, the cochlea and vestibule were dichotomized as EH-positive or EH-negative based on the presence or absence of hydrops. Cochlear and vestibular EH grades were evaluated using the visual four-grade method proposed by Gürkov and Bernaerts.^{23,24}

Cochlea

- Normal (grade 0): The scala media (SM) appeared as a vaguely visible dark area with a relatively straight border separating it from the scala vestibuli and scala tympani (Figure 2a).

- Mild hydrops (grade 1): The SM exhibited a distinct hypointense area surrounded by a clear and continuous hyperintense perilymphatic ring (Figure 2b).

- Moderate hydrops (grade 2): The hyperintense perilymphatic ring was substantially interrupted (Figure 2c).

- Severe hydrops (grade 3): The surrounding hyperintense perilymphatic area became a clear, straight line (Figure 2d).

Vestibule

- Normal (grade 0): The saccule and utricle were separated, and their combined area occupied less than half of the vestibular space (Figure 3a).

- Mild hydrops (grade 1): The saccule was equal to or larger than the utricle, and the two could still be distinguished (Figure 3b).

- Moderate hydrops (grade 2): The saccule and utricle were fused, but peripheral perilymph remained visible (Figure 3c).

- Severe hydrops (grade 3): No PLE was observed in the vestibule (Figure 3d).

In addition, a new weighted visual scoring system based on the Inner Ear Structural Assignment Method^{25,26} was employed (Table 1). The signal intensity ratio of PLE to the ipsilateral middle cerebellar peduncle was measured. The semicircular canals and vestibular aqueduct (VA) were graded as 0, 1, or 2, depending on whether they were continuously developed. In total, six clinical variables and 17 MRI features were included in the analysis (Supplementary Table 1).

Statistical analysis

To analyze all data, IBM SPSS (version 27.0) and R software (version 4.2.1) were used. Continuous variables were presented as mean ± standard deviation or median with interquartile range. Measurement data conforming to a normal distribution were compared using the independent sample t-test. The Mann-Whitney U test was used to compare measurement data that did not conform to a normal distribution. Categorical data were compared using the χ^2 test or Fisher's exact test. Kendall's W test was used to assess inter-observer agreement.

Multivariable logistic regression analysis was applied to select MD-related features. Variables with $P < 0.05$ were included in the

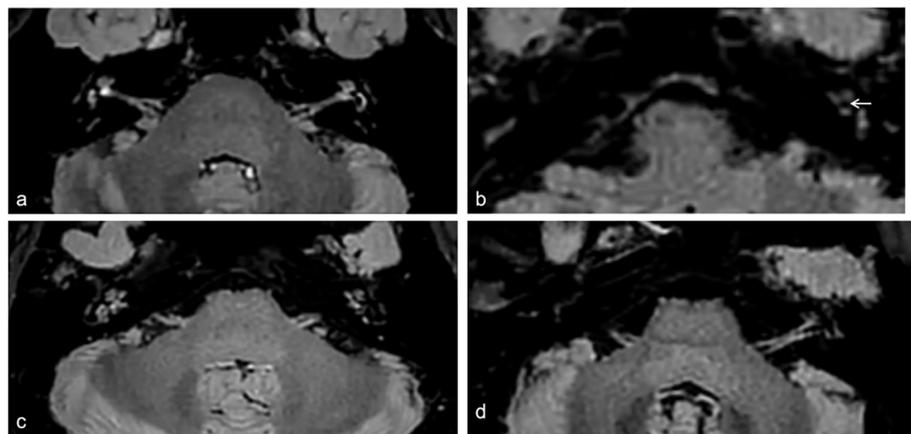


Figure 2. Grading of cochlear hydrops on axial 3D-FLAIR delayed-enhancement images. Grade 0 (normal): The scala media (SM) appears as a faint dark area (arrow) with a relatively straight border between the scala vestibuli and scala tympani (a). Grade 1 (mild hydrops): The SM shows a distinct nodular low signal area (white arrow), surrounded by a clear, continuous high-signal perilymphatic ring (b). Grade 2 (moderate hydrops): The high-signal perilymphatic ring is substantially interrupted (left arrow) (c). Grade 3 (severe hydrops): The surrounding high-signal perilymphatic area appears as a thin, straight line (left arrow) (d). 3D-FLAIR, three-dimensional fluid-attenuated inversion recovery.

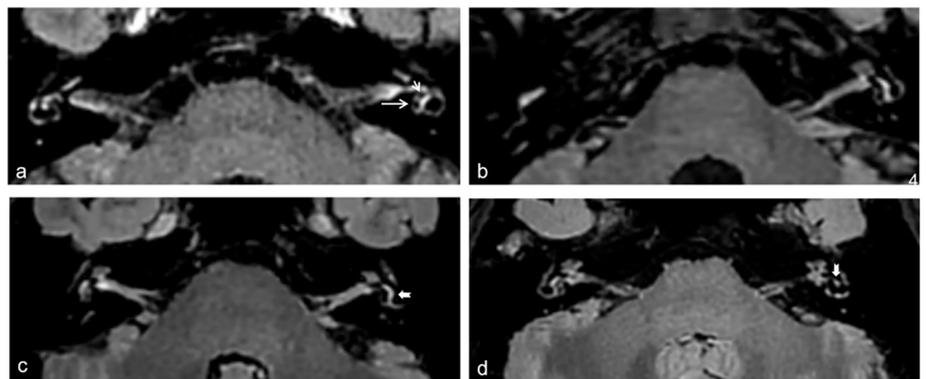


Figure 3. Grading of vestibular hydrops on axial 3D-FLAIR delayed-enhancement images. Grade 0 (normal): The saccule (short arrow) and utricle (long arrow) remain separated; the combined area is less than half of the vestibule (a). Grade 1 (mild hydrops): The saccule is equal to or larger than the utricle (arrow), but the two structures remain distinct (b). Grade 2 (moderate hydrops): The saccule and utricle are fused; peripheral perilymph remains visible (swallow-tail arrow, (c)). Grade 3 (severe hydrops): Complete loss of perilymphatic enhancement in the vestibule (swallow-tail arrow, (d)). 3D-FLAIR, three-dimensional fluid-attenuated inversion recovery.

multivariate logistic regression model using the backward stepwise method to develop three models for MD diagnosis in the training cohort: DEMRI signature, clinical variables, and combined DEMRI-clinical parameters. The validation set was used to validate these models. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to evaluate the diagnostic performance of the different models. The DeLong test was used to compare the AUC values between the models. A two-tailed *P* value less than 0.05 was considered statistically significant. The equation of the multivariate logistic regression model was as follows:

$$P(y = 1|x_1, x_2, \dots, x_m) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)}}$$

Results

Patient characteristics

A total of 85 patients (162 ears; mean age: 53.2 ± 13.6 years; age range: 17–86 years) were included. The detailed clinical and DEMRI characteristics of all ears in the

MD group (n = 95) and the control group (n = 67) are presented in Supplementary Table 1. Except for sex, body mass index, and VA visualization degree, all other observed indicators differed significantly between the MD and control groups (*P* < 0.05). The detailed characteristics of ears in the training set (n = 98) and validation set (n = 64) are shown in Supplementary Table 2.

Diagnostic model development and validation

In the training set, 17 DEMRI-independent descriptors were analyzed using multivariate logistic regression with backward stepwise selection. Four descriptors with *P* < 0.05 (Table 2) were selected to construct the DEMRI diagnostic model, which showed strong diagnostic performance for MD, with an AUC of 0.907 [95% confidence interval (CI): 0.848–0.966] in the training cohort and 0.887 (95% CI: 0.802–0.971) in the validation cohort (Figures 4a, 4b). The same approach was used to build a clinical diagnostic model based on two independent descriptors (PTA

stage, *P* < 0.001; tinnitus fullness, *P* < 0.001). The AUCs of the clinical model in the training and validation cohorts were 0.915 (95% CI: 0.860–0.970) and 0.736 (95% CI: 0.617–0.855), respectively (Figures 4a, 4b).

Using multivariable logistic regression, four independent descriptors—Cochlea_EH_Grade, Vestibule_EH, PTA Stage, and Tinnitus fullness—were identified for the combined DEMRI-clinical model (Table 3). The AUCs of the DEMRI-clinical model for diagnosing MD were 0.947 (95% CI: 0.903–0.990) in the training cohort and 0.796 (95% CI: 0.689–0.902) in the validation cohort (Figures 4a, 4b). DeLong's test was used to compare the correlated ROC curves. In the training set, the AUC of the DEMRI model was nearly equal to that of the clinical model. However, in the validation cohort, the DEMRI model had a significantly higher AUC (*P* = 0.012; 95% CI: 0.033–0.269). The DEMRI-clinical model also outperformed the clinical model in diagnosing MD (*P* = 0.027). Although the DEMRI model had a slightly higher AUC than the DEMRI-clinical model, the difference was not statistically significant (*P* = 0.054) (Table 4).

The weights of the four independent risk factors used in the DEMRI model are illustrated in a nomogram (Figure 5a). The calibration curves of the DEMRI nomogram demonstrated good agreement in both the training and validation sets (Figures 5b, 5c).

Inter-observer agreement on the four magnetic resonance imaging features of the delayed post-gadolinium enhancement magnetic resonance imaging model

Inter-observer agreement for the four MRI features included in the DEMRI model was assessed using Kendall's *W* test. The features "Cochlea_EH_Grade," "Cochlea_Apex_EH_

Table 1. A new weighted visual scoring criteria based on the Inner Ear Structural Assignment Method for inner ear 3D-FLAIR images

Appearance	Cochlea			Vestibule	Semicircular canals		
	Base	Middle	Apex		Superior	Horizontal	Posterior
Not visible ^a	0	0	0	0	0	0	0
Partially visible [*]	2	1	- ^a	3 ^b	1	1	1
Completely visible [†]	3	2	1	6 ^c	2	2	2

Data represent scores awarded based on 3D-FLAIR images. ^aIndicates the absence of a high-signal contrast medium. ^{*}Refers to failure to show a high-signal image of the entire cochlear canal, a high-signal image limited to the tympanic or vestibular scala, interrupted high-signal images of the semicircular canals, or an incomplete high-signal image of the vestibule. [†]Denotes that all labyrinth structures are completely visible. ^aThis option is not applicable, as the apex of the cochlea is very small; only a score of 0 or 1 is assigned. If visible, a score of 1 is given without distinguishing between partial and complete visibility. ^bThe hypointensity zone in the vestibule extends below the lower margin of the horizontal semicircular canal and is scored as 3. ^cThe hypointensity zone in the vestibule is located entirely above the plane of the horizontal semicircular canal and is scored as 6. 3D-FLAIR, three-dimensional fluid-attenuated inversion recovery.

Table 2. Risk factors of DEMRI for MD in the training cohort

Variable	B	Wald	SE	<i>P</i>	OR (95% CI)
(Intercept)	30,073	0	2955.414	0.992	1.15E+13 (0–NA)
Cochlea_EH_Grad	3.19	10,347	0.992	0.001*	24,292 (5,058–297.868)
Cochlea_Apex_EH_Score	3,698	4,298	1,784	0.038*	40,384 (1,906–3014.737)
Vestibule_EH_Score	0.631	1,915	0.456	0.166	1,879 (0.852–5,211)
Horizontal semicircular canal	-21,236	0	1477.708	0.989	0 (NA–1.91E+27)
VA	1,116	4,579	0.522	0.032*	3,053 (1,196+9.632)
Vestibule_EH	3,729	6,663	1,445	0.010*	41,631 (3.44–1172.845)
PE/MCPE	1,612	2,332	1,056	0.127	5,014 (0.607–46,332)

*Statistically significant (*P* < 0.05). Data show multivariable regression results.

Cochlea_EH_Grad, Endolymphatic hydrops (EH) severity in the cochlea (0–3 grade); Cochlea_Apex_EH_Score, EH in the cochlear apex (scored per Table 1); Vestibule_EH_Score, EH in the vestibule (scored per Table 1); Horizontal Semicircular Canal, Development (0 = absent, 1 = partial, 2 = complete); Vestibule_EH, Presence/absence of vestibular EH (binary); VA, Vestibular aqueduct development (0 = absent, 1 = partial, 2 = complete); PE/MCPE, Perilymph-to-middle cerebellar peduncle signal intensity ratio; SE, standard error; OR, odds ratio; CI, confidence interval; DEMRI, delayed post-gadolinium enhancement magnetic resonance imaging; MD, Menière's disease.

Score, "Vestibule_EH," and "VA" all demonstrated very good consistency, with Kendall's coefficients of $W = 0.954, 0.985, 0.967,$ and $0.951,$ respectively. All associated P values were less than 0.001 (Supplementary Table 3).

Discussion

In this study, we developed and validated three models to diagnose MD. The results showed that both the DEMRI model and the combined DEMRI-clinical model had better

clinical diagnostic performance than the clinical model alone (AUC: 0.736; sensitivity: 55.3%; specificity: 92.3%). The DEMRI model demonstrated excellent predictive performance in the validation set (AUC: 0.887; sensitivity: 78.9%; specificity: 88.5%). Although the AUC value of the combined DEMRI-clinical model was slightly lower than that of the DEMRI model, there was no significant difference in diagnostic performance. In the DEMRI model, the most substantial features

were "Cochlea_EH_Grad," "Cochlea_Apex_EH_Score," "VA," and "Vestibule_EH."

MD is associated with a variety of comorbidities, such as migraine, anxiety, allergies, and immune disorders, but its pathogenesis remains unknown.²⁷ EH, characterized by an increase in endolymphatic fluid within the membranous labyrinth of the inner ear, has been identified as the histopathological hallmark of MD. EH is thought to result from disrupted endolymph homeostasis caused by increased production, impaired absorption, or both.²⁸ In EH, excess endolymph volume leads to longitudinal flow from the cochlea to the endolymphatic sac (ES) to restore balance. Gibson²⁹ proposed that when the ES and endolymphatic duct (ED) are functional, they can remove excess endolymph. However, in patients with MD and dysfunctional ES and ED, endolymph may accumulate in the sinus of the ED, leading to substantial overflow. Various methods have been proposed to assess the endolymphatic space both qualitatively and quantitatively.^{23,24,26} Studies have shown that the relationship between MD and EH is strong enough to consider EH a hallmark of MD and a sensitive target for diagnostic detection.¹²

In this study, significant differences were found in all MRI features related to EH between the MD and control groups (Supplementary Table 1). Among these, three EH-related MRI features—"Cochlea_EH_Grad," "Cochlea_Apex_EH_Score," and "Vestibule_EH"—were included in the DEMRI model. It appears that cochlea-related EH carries greater diagnostic weight in MD and that the presence or absence of hydrops in the cochlear apical turn is of particular diagnostic value.

It has been shown that cochlear hydrops follows a reliable pattern of hydropic pro-

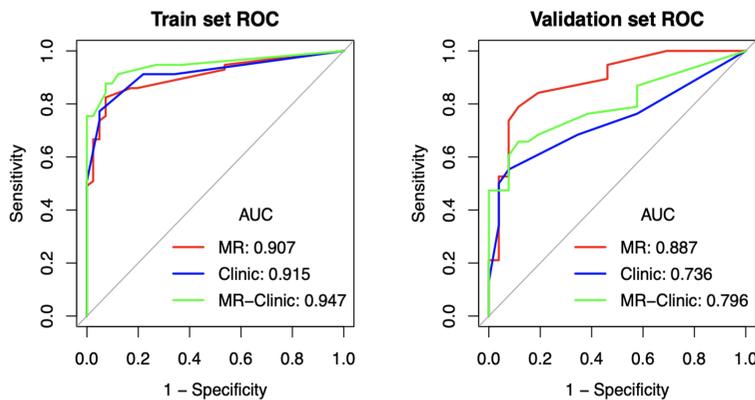


Figure 4. Receiver operating characteristic (ROC) curves. Performance of the three models in both the training and validation cohorts. AUC, area under the curve; MR, magnetic resonance.

Table 3. Risk factors of the DEMRI-clinical model for Menière's disease in the training cohort

Variable	B	Wald	SE	P	OR (95% CI)
(Intercept)	-2.711	5.258	1.182	0.022	0.066 (0.005–0.531)
Cochlea_EH_Grad	1.252	6.755	0.482	0.009*	3.4989 (1.449–10.256)
Vestibule_EH	1.377	2.996	0.796	0.083*	3.964 (0.845–20.353)
PTA Stage	1.581	5.911	0.65	0.015*	4.861 (1.603–20.68)
Tinnitus fullness	-1.821	5.988	0.744	0.014*	0.162 (0.034–0.673)

*Statistically significant ($P < 0.05$). Multivariable regression results show:
 Cochlea_EH_Grad: Cochlear endolymphatic hydrops severity grade (0–3)
 Vestibule_EH: Presence or absence of vestibular endolymphatic hydrops (binary)
 PTA Stage: Hearing loss classification based on pure tone audiometry (0.5 kHz, 1 kHz, 2 kHz).
 SE, standard error; OR, odds ratio; CI, confidence interval; DEMRI, delayed post-gadolinium enhancement magnetic resonance imaging.

Table 4. Diagnostic performance of the three models in the training and validation cohorts

Model	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
DEMRI						
Training cohort	0.907(0.848–0.966)	0.825	0.927	0.940	0.792	0.867
Validation cohort	0.887(0.802–0.971)	0.789	0.885	0.909	0.742	0.828
Clinical						
Training cohort	0.915(0.860–0.970)	0.772	0.951	0.957	0.75	0.847
Validation cohort	0.736(0.617–0.855)	0.553	0.923	0.913	0.585	0.703
DEMRI-clinical						
Training cohort	0.947(0.903–0.990)	0.877	0.927	0.943	0.844	0.898
Validation cohort	0.796(0.689–0.902)	0.658	0.885	0.893	0.639	0.750

AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

Model Features:

DEMRI model: Cochlea_EH_Grad, Cochlea_Apex_EH_Score, VA, Vestibule_EH

Clinical model: PTA Stage, Tinnitus fullness

DEMRI-clinical model: Cochlea_EH_Grad, Vestibule_EH, PTA Stage, Tinnitus fullness.

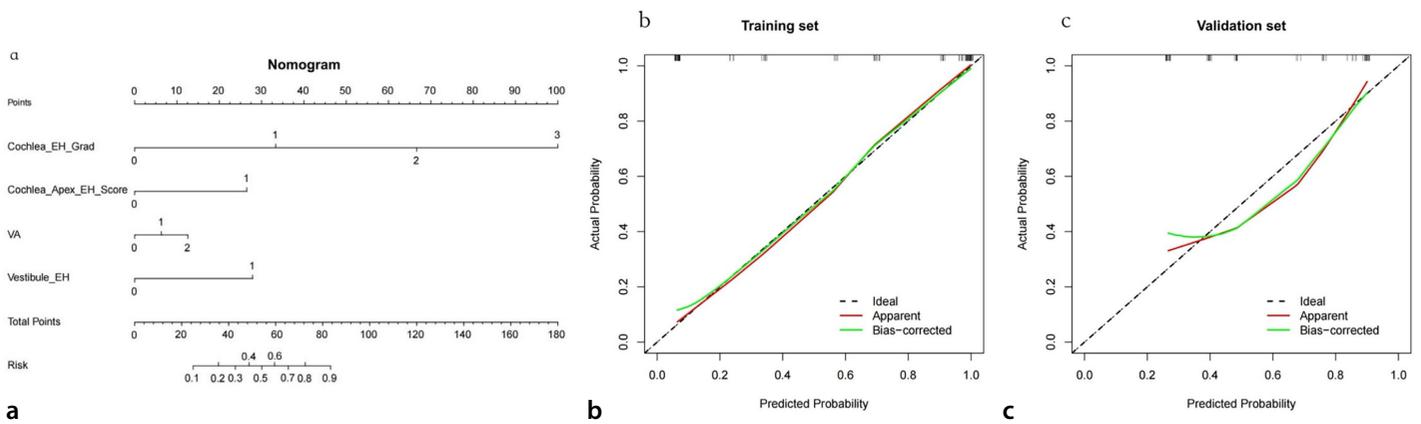


Figure 5. The DEMRI nomogram constructed in the training cohort using Cochlea_EH_Grad, Cochlea_Apex_EH_Score, VA, and Vestibule_EH (a). Calibration curves of the nomogram in the (b) training and (c) validation cohorts. DEMRI, delayed post-

gression over time, typically originating in the apex and proceeding toward the base, tonotopically resembling the progression of hearing loss.^{12,30,31} A hydrodynamic pressure shunt in the pars superior stimulates the utricle and the saccule of the vestibule, resulting in “Vestibule_EH”.^{28,29} This longitudinal hydrops process may explain why the presence or absence of EH in the apical turn of the cochlea is diagnostically substantial for MD. Additionally, experimental studies have suggested that cytochemical and ultrastructural disruption of the hair cells, afferent neurons, and fibrocytes of the lateral cochlear wall are involved in the pathogenesis of EH and occur prior to its development.^{10,32,33} These findings support the conclusion that “Cochlea_EH_Grad” and “Vestibule_EH” are important risk factors for diagnosing MD. It is important to consider both the grade of EH in the cochlea and the presence or absence of EH in the vestibule. The more severe the cochlear EH, the higher the likelihood of diagnosing MD when accompanied by vestibular EH, regardless of the severity of the vestibular component.

However, EH is not pathognomonic for MD, as it has also been observed in vestibular migraine (VM), isolated SNHL, and even in healthy individuals. This limits its diagnostic specificity for MD.^{5,34} VM is a leading cause of recurrent vertigo and is often misdiagnosed as MD despite being 5–10 times more prevalent.³⁵ The clinical overlap between MD and VM presents substantial diagnostic challenges. Emerging evidence suggests that differences in EH patterns may help distinguish the two conditions: MD typically presents with both cochlear and vestibular EH (as seen on Gd-enhanced MRI), whereas EH in VM is rare and usually limited to the cochlea.³⁵⁻³⁷ Thus, inner ear imaging (e.g., Gd-DEMRI) may assist in differential diagnosis. Isolated SNHL

may represent a prodromal phase of MD and warrants further investigation.

Furthermore, this study identified a relatively novel finding: the VA appears to be a substantial risk factor in diagnosing MD. A study by Steve Connor et al.¹⁵ demonstrated that all VA descriptors showed excellent reliability for MD diagnosis and that incomplete VA visualization adds diagnostic value. Mainnemarre et al.¹⁶ further suggested that evaluating the VA on temporal bone computed tomography (CT) could predict the presence of EH on MRI with a high positive predictive value. Attyé et al.³⁸ proposed that discontinuous VA may correlate with MD. A non-visible or partially visible VA may result from bony abnormalities or central fibrosis, leading to endolymphatic stenosis. Although VA performance was included in our model, there was no statistically significant difference in VA between the MD and control groups (Supplementary File). This may be due to the low detection rate of VA on MRI, highlighting the need for clearer imaging techniques or combining MRI with other modalities, such as CT, for more comprehensive evaluation.

Following large-scale validation, our diagnostic model could be incorporated into clinical practice to generate structured radiology reports with probability scores. These reports could support the following: (1) risk stratification, (2) identification of high-risk patients needing specialist referral, and (3) long-term post-treatment management.

Limitations

Limited sample size: Although this is a multicenter study, the sample size (85 patients, 162 ears) is relatively small, which may limit the generalizability of the findings. Future studies with larger cohorts are needed

to validate these results.

Retrospective design: The retrospective nature of the study introduces potential biases in patient selection and data collection. Additionally, some asymptomatic patients with early MD may have been misclassified into the control group. Future research should include normal participants and other differential diagnoses (e.g., VM, benign positional vertigo) for more robust comparisons.

Lack of external validation: Although internal validation was performed, external validation using an independent dataset would further strengthen the reliability of the model.

Imaging feature selection: This study primarily relied on conventional MRI features. Further exploration of advanced imaging biomarkers may improve diagnostic accuracy.

In conclusion, we developed and validated a new DEMRI model for diagnosing MD, which demonstrated higher diagnostic value than clinical inquiry information alone. A combination of a high degree of cochlear EH, invisible cochlear apical turn, vestibular hydrops, and incomplete VA visualization suggests a high risk of MD. Therefore, we recommend DEMRI when MD is suspected due to its substantial diagnostic potential. Further studies are needed to explore the broader applicability of our model and support its clinical implementation.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was supported by the National Natural Science Foundation of China (Grant ID: 82202089).

References

1. Rizk HG, Mehta NK, Qureshi U, et al. Pathogenesis and etiology of Meniere disease: a scoping review of a century of evidence. *JAMA Otolaryngol Head Neck Surg.* 2022;148(4):360-368. [\[Crossref\]](#)
2. Minor LB, Schessel DA, Carey JP. Ménière's disease. *Curr Opin Neurol.* 2004;17(1):9-16. [\[Crossref\]](#)
3. Sajjadi H, Paparella MM. Meniere's disease. *Lancet.* 2008;372(9636):406-414. [\[Crossref\]](#)
4. Committee on Hearing and Equilibrium guidelines for the diagnosis and evaluation of therapy in Ménière's disease. American Academy of Otolaryngology-Head and Neck Foundation, Inc. *Otolaryngol Head Neck Surg.* 1995;113(3):181-185. [\[Crossref\]](#)
5. Hoskin JL. Meniere's disease: new guidelines, subtypes, imaging, and more. *Curr Opin Neurol.* 2022;35(1):90-97. [\[Crossref\]](#)
6. Lopez-Escamez JA, Carey J, Chung WH, et al. Diagnostic criteria for Meniere's disease. *J Vestib Res.* 2015; 25(1):1-7. [\[Crossref\]](#)
7. Hallpike CS, Cairns H. Observations on the pathology of Ménière's syndrome: (section of otology). *Proc R Soc Med.* 1938;31(11):1317-36. [\[Crossref\]](#)
8. Gurkov R, Pyyko I, Zou J, Kentala E. What is Meniere's disease? A contemporary re-evaluation of endolymphatic hydrops. *J Neurol.* 2016;263(Suppl 1):71-81. [\[Crossref\]](#)
9. Nakashima T, Naganawa S, Sugiura M, et al. Visualization of endolymphatic hydrops in patients with Meniere's disease. *Laryngoscope.* 2007;117(3):415-420. [\[Crossref\]](#)
10. Fiorino F, Pizzini FB, Beltramello A, Barbieri F. Progression of endolymphatic hydrops in Ménière's disease as evaluated by magnetic resonance imaging. *Otol Neurotol.* 2011;32:1152-1157. [\[Crossref\]](#)
11. Connor SEJ, Pai I. Endolymphatic hydrops magnetic resonance imaging in Meniere's disease. *Clin Radiol.* 2021;76(1):76. [\[Crossref\]](#)
12. Gluth MB. On the relationship between Meniere's Disease and endolymphatic hydrops. *Otol Neurotol.* 2020;41(2):242-249. [\[Crossref\]](#)
13. Li J, Jin X, Kong X, et al. Correlation of endolymphatic hydrops and perilymphatic enhancement with the clinical features of Meniere's disease. *Eur Radiol.* 2024;34(9):6036-6046. [\[Crossref\]](#)
14. van Steekelenburg JM, van Weijnen A, de Pont LMH, et al. Value of endolymphatic hydrops and perilymph signal intensity in suspected Meniere Disease. *AJNR Am J Neuroradiol.* 2020;41(3):529-534. [\[Crossref\]](#)
15. Connor S, Pai I, Touska P, McElroy S, Ourselin S, Hajnal JV. Assessing the optimal MRI descriptors to diagnose Meniere's disease and the added value of analysing the vestibular aqueduct. *Eur Radiol.* 2024;34(9):6060-6071. [\[Crossref\]](#)
16. Mainnemarre J, Hautefort C, Toupet M, et al. The vestibular aqueduct ossification on temporal bone CT: an old sign revisited to rule out the presence of endolymphatic hydrops in Meniere's disease patients. *Eur Radiol.* 2020;30(11):6331-6338. [\[Crossref\]](#)
17. Connor S, Grzeda MT, Jamshidi B, Ourselin S, Hajnal JV, Pai I. Delayed post gadolinium MRI descriptors for Meniere's disease: a systematic review and meta-analysis. *Eur Radiol.* 2023;33(10):7113-7135. [\[Crossref\]](#)
18. Chen W, Yu S, Xiao H, et al. A novel radiomics nomogram based on T2-sampling perfection with application-optimized contrasts using different flip-angle evolutions (SPACE) images for predicting cochlear and vestibular endolymphatic hydrops in Meniere's disease patients. *Eur Radiol.* 2024;34(9):6082-6091. [\[Crossref\]](#)
19. Kobayashi M, Yoshida T, Fukunaga Y, Hara D, Naganawa S, Sone M. Perilymphatic enhancement and endolymphatic hydrops: MRI findings and clinical associations. *Laryngoscope Investig Otolaryngol.* 2024;9(5):e1312. [\[Crossref\]](#)
20. van den Burg EL, van Hoof M, Postma AA, et al. An exploratory study to detect Meniere's disease in conventional MRI scans using radiomics. *Front Neurol.* 2016;7:190. [\[Crossref\]](#)
21. Eliezer M, Poillon G, Gillibert A, et al. Comparison of enhancement of the vestibular perilymph between gadoterate meglumine and gadobutrol at 3-Tesla in Meniere's disease. *Diagn Interv Imaging.* 2018;99(5):271-277. [\[Crossref\]](#)
22. Xie J, Zhang W, Zhu J, Hui L, Li S, Zhang B. Comparison of inner ear MRI enhancement in patients with Meniere's disease after intravenous injection of gadobutrol, gadoterate meglumine, or gadodiamide. *Eur J Radiol.* 2021;139:109682. [\[Crossref\]](#)
23. Gurkov R, Flatz W, Louza J, Strupp M, Krause E. In vivo visualization of endolymphatic hydrops in patients with Meniere's disease: correlation with audiovestibular function. *Eur Arch Otorhinolaryngol.* 2011;268(12):1743-1748. [\[Crossref\]](#)
24. Bernaerts A, Vanspauwen R, Blaivie C, et al. The value of four stage vestibular hydrops grading and asymmetric perilymphatic enhancement in the diagnosis of Meniere's disease on MRI. *Neuroradiology.* 2019; 61(4):421-429. [\[Crossref\]](#)
25. Fang ZM, Chen X, Gu X, et al. A new magnetic resonance imaging scoring system for perilymphatic space appearance after intratympanic gadolinium injection, and its clinical application. *J Laryngol Otol.* 2012;126(5):454-459. [\[Crossref\]](#)
26. Xiao H, Guo X, Cai H, et al. Magnetic resonance imaging of endolymphatic hydrops in Meniere's disease: a comparison of the diagnostic value of multiple scoring methods. *Front Neurol.* 2022;13:967323. [\[Crossref\]](#)
27. Lopez-Escamez JA, Liu Y. Epidemiology and genetics of Meniere's disease. *Curr Opin Neurol.* 2024;37(1):88-94. [\[Crossref\]](#)
28. Mohseni-Dargah M, Falahati Z, Pastras C, et al. Meniere's disease: pathogenesis, treatments, and emerging approaches for an idiopathic bioenvironmental disorder. *Environ Res.* 2023;238(Pt 1):116972. [\[Crossref\]](#)
29. Gibson WP. Hypothetical mechanism for vertigo in Meniere's disease. *Otolaryngol Clin North Am.* 2010;43(5):1019-1027. [\[Crossref\]](#)
30. Wu Q, Dai C, Zhao M, Sha Y. The correlation between symptoms of definite Meniere's disease and endolymphatic hydrops visualized by magnetic resonance imaging. *Laryngoscope.* 2016;126(4):974-979. [\[Crossref\]](#)
31. Li J, Wang L, Hu N, et al. Longitudinal variation of endolymphatic hydrops in patients with Meniere's disease. *Ann Transl Med.* 2023;11(2):44. [\[Crossref\]](#)
32. Nadol JB, Adams JC, Kim JR. Degenerative changes in the organ of corti and lateral cochlear wall in experimental endolymphatic hydrops and human Ménière's disease. *Acta Otolaryngol Suppl.* 1995; 519:47-59. [\[Crossref\]](#)
33. Ichimiya I, Adams JC, Kimura RS. Changes in immunostaining of cochleas with experimentally induced endolymphatic hydrops. *Ann Otol Rhinol Laryngol.* 1994;103(6):457-468. [\[Crossref\]](#)
34. Lingam RK, Connor SEJ, Casselman JW, Beale T. MRI in otology: applications in cholesteatoma and Meniere's disease. *Clin Radiol.* 2018;73(1):35-44. [\[Crossref\]](#)
35. Chen JY, Guo ZQ, Wang J, et al. Vestibular migraine or Meniere's disease: a diagnostic dilemma. *J Neurol.* 2023;270(4):1955-1968. [\[Crossref\]](#)
36. Nakada T, Yoshida T, Suga K, et al. Endolymphatic space size in patients with vestibular migraine and Meniere's disease. *J Neurol.* 2014;261(11):2079-2084. [\[Crossref\]](#)
37. Sun W, Guo P, Ren T, Wang W. Magnetic resonance imaging of intratympanic gadolinium helps differentiate vestibular migraine from Meniere disease. *Laryngoscope.* 2017;127(10):2382-2388. [\[Crossref\]](#)
38. Attyé A, Barma M, Schmerber S, Dumas G, Eliezer M, Krainik A. The vestibular aqueduct sign: magnetic resonance imaging can detect abnormalities in both ears of patients with unilateral Meniere's disease. *J Neuroradiol.* 2020;47(2):174-179. [\[Crossref\]](#)

Supplementary Table 1. Comparison of clinical variables and MRI features between control group and MD group

	Control ears (n = 67)	Menière's ears (n = 95)	P value
PTA stage			
1	52 (77.6%)	27 (28.4%)	<0.001
2	13 (19.4%)	21 (22.1%)	
3	2 (3.0%)	31 (32.6%)	
4	0 (0%)	16 (16.8%)	
Vertigo			
	49 (73.1%)	92 (96.8%)	<0.001
Tinnitus fullness			
	17 (25.4%)	71 (74.7%)	<0.001
Age			
Mean (SD)	49.9 (13.1)	55.5 (13.6)	0.009
Gender			
Male/female	33/34	43/52	0.733
BMI			
Mean (SD)	23.0 (2.40)	22.7 (2.70)	0.426
Cochlea_Base_EH_Grad			
0	63 (94.0%)	41 (43.2%)	<0.001
1	1 (1.5%)	22 (23.2%)	
2	3 (4.5%)	17 (17.9%)	
3	0 (0%)	15 (15.8%)	
Cochlea_Middle_EH_Grad			
0	63 (94.0%)	40 (42.1%)	<0.001
1	2 (3.0%)	14 (14.7%)	
2	2 (3.0%)	16 (16.8%)	
3	0 (0%)	25 (26.3%)	
Cochlea_Apex_EH_Grad			
0	62 (92.5%)	35 (36.8%)	<0.001
1	2 (3.0%)	16 (16.8%)	
2	3 (4.5%)	19 (20.0%)	
3	0 (0%)	25 (26.3%)	
Cochlea_EH_Grad			
0	61 (91.0%)	25 (26.3%)	<0.001
1	3 (4.5%)	24 (25.3%)	
2	3 (4.5%)	21 (22.1%)	
3	0 (0%)	25 (26.3%)	
Cochlea_Base_EH_Score			
0	0 (0%)	2 (2.1%)	<0.001
2	4 (6.0%)	53 (55.8%)	
3	63 (94.0%)	40 (42.1%)	
Cochlea_Middle_EH_Score			
0	0 (0%)	12 (12.6%)	<0.001
1	4 (6.0%)	45 (47.4%)	
2	63 (94.0%)	38 (40.0%)	
Cochlea_Apex_EH_Score			
0	4 (6.0%)	34 (35.8%)	<0.001
1	63 (94.0%)	61 (64.2%)	
Cochlea_EH_Score			

Supplementary Table 1. Continued			
	Control ears (n = 67)	Menière's ears (n = 95)	P value
0	0 (0%)	1 (1.1%)	<0.001
2	0 (0%)	10 (10.5%)	
3	1 (1.5%)	12 (12.6%)	
4	4 (6.0%)	31 (32.6%)	
5	1 (1.5%)	12 (12.6%)	
6	61 (91.0%)	29 (30.5%)	
Vestibule_EH_Score			
0	0 (0%)	15 (15.8%)	<0.001
3	13 (19.4%)	45 (47.4%)	
4	54 (80.6%)	35 (36.8%)	
Vestibule_EH_Grade			
0	57 (85.1%)	30 (31.6%)	<0.001
1	8 (11.9%)	19 (20.0%)	
2	2 (3.0%)	17 (17.9%)	
3	0 (0%)	29 (30.5%)	
Semicircular canal superior			
Non-visualized	0 (0%)	5 (5.3%)	0.035
Incompletely visualized	0 (0%)	4 (4.2%)	
Completely visualized	67 (100%)	86 (90.5%)	
Semicircular canal horizontal			
Non-visualized	0 (0%)	9 (9.5%)	<0.001
Incompletely visualized	0 (0%)	18 (18.9%)	
Completely visualized	67 (100%)	68 (71.6%)	
Semicircular canal posterior			
Non-visualized	0 (0%)	5 (5.3%)	0.023
Incompletely visualized	0 (0%)	5 (5.3%)	
Completely visualized	67 (100%)	85 (89.5%)	
VA			
Non-visualized	23 (34.3%)	41 (43.2%)	0.318
Incompletely visualized	20 (29.9%)	30 (31.6%)	
Completely visualized	24 (35.8%)	24 (25.3%)	
Cochlea_EH			
EH-positive	60 (89.6%)	25 (26.3%)	<0.001
Vestibule_EH			
EH-positive	57 (85.1%)	30 (31.6%)	<0.001
PLE/MCPE			
Mean (SD)	1.15 (0.291)	1.33 (0.368)	<0.001
Group			
Training set	41 (61.2%)	57 (60.0%)	1
Validation set	26 (38.8%)	38 (40.0%)	
Cochlear and vestibular endolymphatic hydrops were evaluated according to Gurkov and Bernaerts' visual 4-grade method.			
Cochlear and vestibular endolymphatic hydrops score were evaluated according to a new weighted visual scoring system (Table 1) based on Inner Ear Structural Assignment Method.			
PLE/MCPE: Measurements of signal intensity were performed by drawing an oval region of interest along the edge of the cochlear basal turn and a circular region of interest at the left middle cerebellar peduncle to calculate the signal intensity ratio.			
PTA, pure tone audiometry; SD, standard deviation; BMI, body mass index; VA, vestibular aqueduct; PLE, perilymphatic enhancement; MCPE, middle cerebellar peduncle.			

Supplementary Table 2. Comparison of clinical variables and MRI features between training set and validation set

	Training set (n = 98)	Validation set (n = 64)	P value
Label			
Control ears	41 (41.8%)	26 (40.6%)	1
Menière's ears	57 (58.2%)	38 (59.4%)	
PTA stage			
1	47 (48.0%)	32 (50.0%)	0.972
2	20 (20.4%)	14 (21.9%)	
3	21 (21.4%)	12 (18.8%)	
4	10 (10.2%)	6 (9.4%)	
Vertigo			
	90 (91.8%)	51 (79.7%)	0.044
Tinnitus fullness			
	41 (41.8%)	33 (51.6%)	0.292
Age			
Mean (SD)	53.4 (13.0)	53.0 (14.6)	0.852
Gender			
Male/female	43/55	33/31	0.425
BMI			
Mean (SD)	23.0 (2.59)	22.4 (2.54)	0.165
Cochlea_Base_EH_Grad			
0	61 (62.2%)	43 (67.2%)	0.757
1	14 (14.3%)	9 (14.1%)	
2	12 (12.2%)	8 (12.5%)	
3	11 (11.2%)	4 (6.3%)	
Cochlea_Middle_EHGrad			
0	60 (61.2%)	43 (67.2%)	0.75
1	9 (9.2%)	7 (10.9%)	
2	12 (12.2%)	6 (9.4%)	
3	17 (17.3%)	8 (12.5%)	
Cochlea_Apex_EH_Grad			
0	56 (57.1%)	41 (64.1%)	0.653
1	10 (10.2%)	8 (12.5%)	
2	15 (15.3%)	7 (10.9%)	
3	17 (17.3%)	8 (12.5%)	
Cochlea_EH_Grad			
0	51 (52.0%)	35 (54.7%)	0.84
1	15 (15.3%)	12 (18.8%)	
2	16 (16.3%)	8 (12.5%)	
3	16 (16.3%)	9 (14.1%)	
Cochlea_Base_EH_Score			
0	2 (2.0%)	0 (0%)	0.498
2	35 (35.7%)	22 (34.4%)	
3	61 (62.2%)	42 (65.6%)	
Cochlea_Middle_EH_Score			
0	10 (10.2%)	2 (3.1%)	0.235
1	28 (28.6%)	21 (32.8%)	
2	60 (61.2%)	41 (64.1%)	
Cochlea_Apex_EH_Score			

Supplementary Table 2. Continued			
	Training set (n = 98)	Validation set (n = 64)	P value
0	26 (26.5%)	12 (18.8%)	0.341
1	72 (73.5%)	52 (81.3%)	
Cochlea_EH_Score			
0	1 (1.0%)	0 (0%)	0.427
2	8 (8.2%)	2 (3.1%)	
3	8 (8.2%)	5 (7.8%)	
4	21 (21.4%)	14 (21.9%)	
5	5 (5.1%)	8 (12.5%)	
6	55 (56.1%)	35 (54.7%)	
Vestibule_EH_Score			
0	11 (11.2%)	4 (6.3%)	0.562
3	34 (34.7%)	24 (37.5%)	
6	53 (54.1%)	36 (56.3%)	
Vestibule_EH_Grade			
0	52 (53.1%)	35 (54.7%)	0.617
1	14 (14.3%)	13 (20.3%)	
2	12 (12.2%)	7 (10.9%)	
3	20 (20.4%)	9 (14.1%)	
Semicircular canal superior			
Non-visualized	4 (4.1%)	1 (1.6%)	0.61
Incompletely visualized	2 (2.0%)	2 (3.1%)	
Completely visualized	92 (93.9%)	61 (95.3%)	
Semicircular canal horizontal			
Non-visualized	6 (6.1%)	3 (4.7%)	0.772
Incompletely visualized	12 (12.2%)	6 (9.4%)	
Completely visualized	80 (81.6%)	55 (85.9%)	
Semicircular canal posterior			
Non-visualized	3 (3.1%)	2 (3.1%)	0.999
Incompletely visualized	3 (3.1%)	2 (3.1%)	
Completely visualized	92 (93.9%)	60 (93.8%)	
VA			
Non-visualized	37 (37.8%)	27 (42.2%)	0.766
Incompletely visualized	30 (30.6%)	20 (31.3%)	
Completely visualized	31 (31.6%)	17 (26.6%)	
Cochlea_EH			
EH-positive	51 (52.0%)	34 (53.1%)	1
Vestibule_EH			
EH-positive	52 (53.1%)	35 (54.7%)	0.967
PE/MCPE			
Mean (SD)	1.21 (0.318)	1.32 (0.386)	0.065

PTA, pure tone audiometry; SD, standard deviation; BMI, body mass index; VA, vestibular aqueduct; PLE, perilymphatic enhancement; MCPE, middle cerebellar peduncle.

Supplementary Table 3. Inter-observer reliability Kendall'W values for the four DEMRI model features

Grade/score	Cochlea_EH_Grade				Cochlea_Apex_Score		Vestibule_EH		VA		
	0	1	2	3	0	1	Yes	No	0	1	2
Observer1	84	29	26	23	38	124	75	87	66	48	48
Observer2	79	38	32	13	40	122	71	91	73	45	44
Observer3	81	32	25	24	38	124	75	87	66	57	39
Kendall's W	0.954				0.985		0.967		0.951		
P	<0.001				<0.001		<0.001		<0.001		

VA, vestibular aqueduct; DEMRI, delayed post-gadolinium enhancement magnetic resonance imaging.



Splenic artery embolization in the treatment of blunt splenic injury: single level 1 trauma center experience

Katelyn Gill¹
 Sarah Aleman¹
 Alexandra H. Fairchild²
 Bahri Üstünsöz²
 Dan Laney²
 Alison A. Smith³
 Hector Ferral²

¹Louisiana State University Health Sciences Center, School of Medicine, New Orleans, Louisiana, United States

²Louisiana State University Health Sciences Center, Department of Interventional Radiology, New Orleans, Louisiana, United States

³Louisiana State University Health Sciences Center, Department of Surgery, New Orleans, Louisiana, United States

PURPOSE

To describe the experience of a single level 1 trauma center in the management of blunt splenic injuries (BSI).

METHODS

This is a retrospective study with Institutional Review Board approval. The medical records of 450 patients with BSI treated between January 2016 and December 2022 were reviewed. Seventy-two patients were treated with splenic artery embolization (SAE), met the study criteria, and were eligible for data analysis. Spleen injuries were graded in accordance with the American Association for the Surgery of Trauma Organ Injury Scale. Univariate data analysis was performed, with $P < 0.05$ considered statistically significant.

RESULTS

The splenic salvage rate was 90.3% ($n = 65/72$). Baseline demographics were similar between the groups ($P > 0.05$). Distal embolization with Gelfoam[®] had similar rates of splenic salvage to proximal embolization with coils (90% vs. 94.1%, $P > 0.05$). There was no significant difference in the rate of splenic infarction between distal embolization with Gelfoam[®] (20%, 4/20) and proximal embolization with coils (17.6%, 3/17) ($P > 0.05$). There was no significant difference in procedure length (68 vs. 75.8 min) or splenic salvage rate (88.5% vs. 92.1%) between proximal and distal embolization ($P > 0.05$). There was no significant difference in procedure length (69.1 vs. 73.6 min) or splenic salvage rate (93.1% vs. 86.4%) between Gelfoam[®] and coil embolization ($P > 0.05$). Combined proximal and distal embolization was associated with a higher rate of splenic abscess formation (25%, 2/8) when compared with proximal (0%, 0/26) or distal (0%, 0/38) embolization alone ($P = 0.0003$). The rate of asymptomatic and symptomatic splenic infarction was significantly higher in patients embolized at combined proximal and distal locations ($P = 0.04$, $P = 0.01$).

CONCLUSION

The endovascular management of BSI is safe and effective. The overall splenic salvage rate was 90.3%. Distal embolization with Gelfoam[®] was not associated with higher rates of splenic infarction when compared with proximal embolization with coils. Combined proximal and distal embolization was associated with a higher incidence of splenic infarction and splenic abscess formation.

CLINICAL SIGNIFICANCE

Distal splenic embolization with Gelfoam[®] is safe and may be beneficial in the setting of blunt splenic trauma.

KEYWORDS

Coil, combined, distal, embolization, Gelfoam[®], proximal, spleen, trauma

Corresponding author: Katelyn Gill

E-mail: kgill1@lsuhsc.edu

Received 02 May 2024; revision requested 19 May 2024; accepted 23 June 2024.



Epub: 11.07.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242789

Blunt splenic injuries (BSI) are a common occurrence following traumatic events such as motor vehicle accidents, boating accidents, falls, or altercations.¹ Splenic injuries can result in massive blood loss and high mortality.² The management of BSI is aimed toward the preservation of splenic parenchyma, as the spleen serves a key role in the defense against

encapsulated pathogens.³ Safe and efficacious treatments to promote spleen salvage are imperative to improve long-term patient outcomes.

Splenic artery embolization (SAE) is a safe and efficacious spleen-salvaging treatment for high-grade splenic injuries.² This treatment successfully achieves hemostasis,⁴ increases the rate of splenic salvage,^{5,6} and preserves both short- and long-term immune function.³ In the absence of comprehensive guidelines for SAE, the techniques used are often determined by the operator's judgment and experience. At our institution, proximal embolization with coils and distal embolization with Gelfoam[®] remain the most common SAE techniques for the treatment of BSI. Based on current literature, distal embolization with Gelfoam[®] is not recommended because embolization with Gelfoam[®] has been associated with inferior clinical outcomes.

Recent studies have suggested that proximal^{7,8} SAE is associated with significantly lower fluoroscopy time⁹ and lower rates of major complications such as post-embolization abscess and splenic infarction.^{1,2,10} The use of Gelfoam[®] has been discouraged¹¹ because of a reported association with an increased risk of recurrent bleeding¹² and infection when compared with coil or plug embolization.^{13,14} These findings, however, are inconsistently supported by the literature, and the optimal technique for SAE remains under debate. The purpose of this study is to describe a single level 1 trauma center experience in the management of blunt splenic trauma. A secondary goal is to determine if distal SAE with Gelfoam[®] is associated with an increased risk of splenic infarction.

Methods

Patient population

A retrospective cohort study of adult patients with BSI treated between January 2016

and December 2022 was performed at a single level 1 trauma center. Patients aged 18 and older with BSI were included. Children, pregnant women, and prisoners were excluded from the study. In total, 450 patients met the study criteria and were included in the medical record review. Of the eligible 450 patients, 72 were treated with SAE and included in the data analysis (Figure 1). This retrospective study was approved by the LSUHSC New Orleans Institutional Review Board (approval: IRB #5040, date: February 28, 2023), and a waiver of consent was obtained.

Collected variables

The medical record review included the collection of demographic information, embolization techniques and outcomes, and clinical outcomes. The embolization technique was defined by the embolization location and embolization material. Patient charts were reviewed for reported complications or additional interventions up to 90 days following the primary treatment. The clinical outcomes of interest included hospital length of stay, intensive care unit (ICU) length of stay, and readmission within 30 days. The embolization outcomes of interest included procedure length, splenic salvage

rate, splenic abscess formation, splenic infarction, and the need for additional intervention. Secondary interventions included SAE following observation, SAE following a previous SAE, or splenectomy following SAE. The diagnosis of splenic abscess formation required clinical symptoms, leukocytosis, and imaging findings on computed tomography (CT). The presence of splenic infarction was evaluated based on imaging findings. Splenic infarction was considered clinically significant if the following symptoms were present: fever, leukocytosis, left upper quadrant pain, nausea, or vomiting.

Injury grading

Spleen injuries identified in surgery or on abdominal CT scans were graded according to the 2018 American Association for the Surgery of Trauma Organ Injury Scale (AAST-OIS).¹⁵ The injury severity score (ISS) was used to provide an overall injury grade for patients with multiple injuries. Each injury is allocated to one of six body regions and rated with an abbreviated injury score (AIS) from 1 to 5, ranging from minor to critical injuries. The ISS is calculated by squaring the AIS of the three most severely injured body regions and combining them, creating a score of 0 to 75.¹⁶

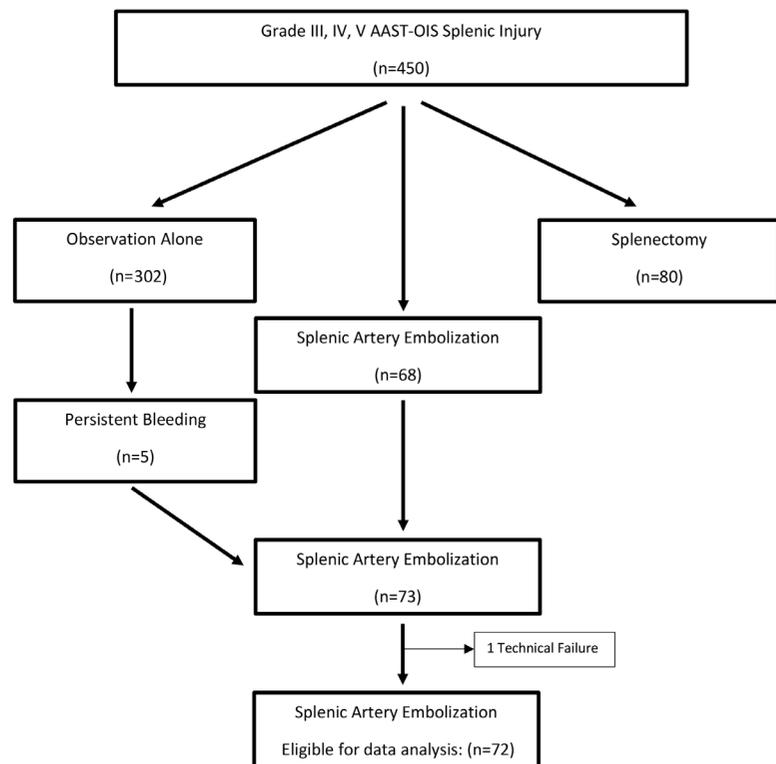


Figure 1. Flowchart. AAST-OIS, American Association for the Surgery of Trauma Organ Injury Scale.

Main points

- For the treatment of blunt splenic injury (BSI), distal embolization with Gelfoam[®] was not associated with an increased risk of splenic infarction or other inferior clinical outcomes when compared with proximal embolization with coils.
- Combined embolization was associated with a higher incidence of splenic infarction and splenic abscess formation.
- Splenic embolization is a safe spleen-salvaging treatment for the management of BSI.

Indications

At our institution, patients with BSI who were hemodynamically unstable were managed with splenectomy. Hemodynamically stable patients were evaluated with contrast-enhanced CT. Patients with AAST-OIS grade III–V BSI, large perisplenic hematoma, contrast extravasation, pseudoaneurysm, or other vascular injuries were referred for SAE. Patients who did not meet the criteria for SAE on their first CT assessment were managed conservatively. These patients were monitored for evidence of ongoing bleeding such as decreasing hematocrit levels or changes in vital signs. Patients with evidence of ongoing bleeding after the initial assessment underwent repeat contrast-enhanced CT scans and were reconsidered for treatment by SAE or splenectomy.

Embolization techniques

All procedures were performed by fellowship-trained interventional radiologists in a state-of-the-art angiography suite. Procedures were performed either under general anesthesia or monitored anesthesia care. Ultrasound guidance was used for arterial access to either the right or left common femoral artery. Selective catheterization of the celiac trunk was performed using standard angiographic catheters. A selective arteriogram of the celiac trunk was performed using a power injector. The anatomy of the splenic artery was delineated, and the angiographic findings were evaluated by the operators (Figure 2). The embolization techniques and materials used were determined during the procedure based on angiographic findings, operator experience, preference, and judgment. Angiographic findings indicating the need for embolization included contrast extravasation, pseudoaneurysm, and contrast blush.

Proximal embolization was defined as embolization of the main splenic artery trunk distal to the dorsal pancreatic artery but proximal to the splenic hilum (Figure 3). Distal embolization was defined as embolization of a splenic artery branch or branches at sites distal to collateral pathways (Figure 4). Embolization was performed using Gelfoam®, particles, coils, and plugs, used either alone or in combination. Embolization was considered complete when contrast extravasation was no longer present. Procedure efficacy was measured using the splenic salvage rate at 30 days. The medical records of patients who underwent embolization were reviewed up to 90 days after the procedure

to determine technical and clinical success rates and incidence of complications.

Statistical analysis

Univariate analysis was performed using ANOVA for continuous variables or χ^2 test for categorical variables. Data were analyzed using GraphPad Prism (version 10.0.0, La Jolla, CA, USA). A *P* value of less than 0.05 was considered statistically significant.

Results

Patient population

A total of 302 (67.1%, 302/450) patients were treated by observation alone, 80 (17.8%, 80/450) required splenectomy, and 68 (15.1%, 68/450) underwent SAE. Of the 302 patients initially treated by observation alone, 5 (1.7%, 5/302) showed signs of persistent bleeding and underwent subsequent SAE. One patient underwent a splenic arteriogram, and embolization was attempted but was technically unsuccessful. This patient was excluded, leaving a total of 72 patients treated by SAE eligible for data analysis (Figure 1, Table 1).

Patients treated by SAE had an average ISS of 26.9 ± 11.8 and spleen AAST-OIS grade of 3.5 ± 0.8 . The patients were hospitalized for an average of 10.9 ± 9.0 days, with an average of 5.9 ± 5.6 days in the ICU. The SAE procedure averaged 71.7 ± 25.8 minutes and the rate of splenic salvage was 90.3% ($n = 65/72$). Seven (9.7%, 7/72) patients exhibited evidence of persistent bleeding after SAE and required splenectomy; one (1.4%, 1/72) splenectomy occurred after a second SAE

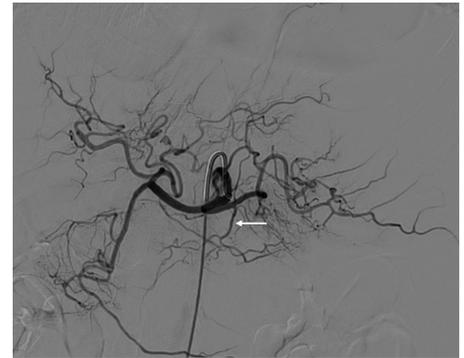


Figure 2. Celiac digital subtraction angiography showing the main splenic artery arising from the celiac trunk. The dorsal pancreatic artery (arrow) arises from the main splenic artery.

Table 1. Splenic artery embolization patient demographics, clinical characteristics, and outcomes

Total	72
Demographics/injury data	
Age, mean (range)	43.2 (18–79)
Male, n (%)	47 (65.3)
Female, n (%)	25 (34.7)
AAST-OIS grade, mean (SD)	3.5 (0.8)
Grade II, n (%)	8 (11.1)
Grade III, n (%)	25 (34.7)
Grade IV, n (%)	34 (47.2)
Grade V, n (%)	5 (6.9)
Injury severity score, mean (SD)	26.9 (11.8)
Embolization outcomes, n (%)	
Procedure length, mean min (SD)	71.7 (25.8)
Splenic salvage, n (%)	65 (90.3)
Failed observation SAE, n (%)	5 (6.9)
Repeat SAE splenectomy, n (%)	1 (1.4)
Post SAE splenectomy, n (%)	7 (9.7)
Splenic abscess, n (%)	2 (2.8)
Splenic infarct, n (%)	11 (15.3)
Symptomatic splenic infarct, n (%)	2 (2.8)
Clinical outcomes	
Hospital length of stay mean days (SD)	10.9 (9.0)
ICU length of stay mean days (SD)	5.9 (5.6)
Readmission within 30 days, n (%)	4 (5.6)

SAE, splenic artery embolization; AAST-OIS, American Association for the Surgery of Trauma Organ Injury Scale; SD, standard deviation; ICU, intensive care unit.

was performed. Two patients (2.8%, 2/72) had imaging findings suggestive of splenic abscess on follow-up CT scans. Splenic infarct was identified in 11 patients (15.3%, 11/72); two of these patients (2.8%, 2/72) had symptomatic splenic infarct. Both patients that presented with symptomatic splenic infarcts were treated with combined embolization using Gelfoam® and coils at proximal and distal locations. One patient presented following a second embolization attempt, later developed a splenic abscess, and was

then treated by splenectomy. One patient was treated for left upper quadrant pain, and no further complication was identified. There were no in-hospital mortalities following SAE.

Embolization technique

Twenty patients (27.8%, 20/72) underwent distal embolization with Gelfoam®, and 17 patients (23.6%, 17/72) underwent proximal embolization with coils. The remaining 35 patients (48.6%, 35/72) were embolized with combinations of embolization material and location and were not included in the data analysis. Baseline demographics, ISS, hospital length of stay, and readmission within 30 days were not significantly different between the groups ($P > 0.05$). There was no significant difference in procedure length or splenic salvage rate between the groups ($P > 0.05$). The mean spleen AAST-OIS grade was significantly higher in patients treated by distal embolization with Gelfoam® (3.75 ± 0.7) than in patients treated by proximal embolization with coils (3.2 ± 0.8) ($P = 0.03$). Splenic infarct was identified on follow-up imaging in four patients (20%, 4/20) treated by distal embolization with Gelfoam® and three patients (17.6%, 3/17) treated by proximal embolization with coils ($P > 0.05$) (Table 2). Two splenic abscesses were reported in patients embolized using a combination of

Gelfoam® distally and coils proximally (5.7%, 2/35).

The patients were additionally stratified by embolization location and material. Twenty-six patients (36.1%, 26/72) underwent proximal embolization, 38 patients (52.8%, 38/72) underwent distal embolization, and 8 patients (11.1%, 8/72) underwent combined proximal and distal embolization (Table 3). Twenty-nine patients (40.3%, 29/72) were embolized with Gelfoam®, 22 patients (30.5%, 22/72) were embolized with coils, and 21 patients (29.2%, 21/72) underwent embolization with two or more embolic agents (Table 4).

Combined embolization was performed with Gelfoam® and coils ($n = 16$, 22.2%), particles and coil ($n = 3$, 4.2%), and plug and coils ($n = 2$, 2.8%). Baseline demographics were similar between the groups ($P > 0.05$). The ISS, hospital length of stay, and readmission within 30 days were also similar between the groups ($P > 0.05$). There was no significant difference in procedure length or splenic salvage rate between the groups ($P > 0.05$). Patients embolized with coils had a significantly lower spleen AAST-OIS (3.01 ± 0.9) than those undergoing Gelfoam® embolization (3.7 ± 0.7) and combined embolization (3.6 ± 0.7) ($P = 0.02$). No splenic abscesses were reported for any technique used in isolation.

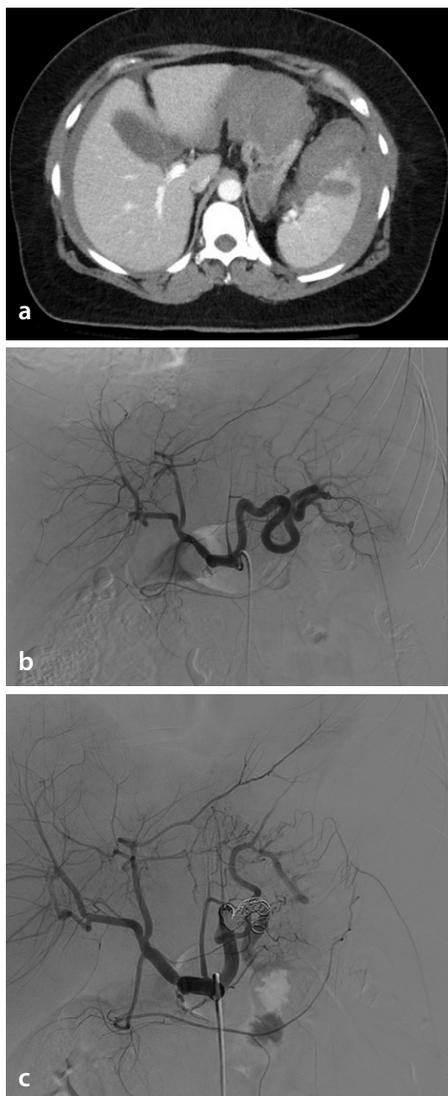


Figure 3. Proximal splenic artery embolization with coils in a 49-year-old woman following a motor vehicle collision. (a) Contrast-enhanced computed tomography showing a grade III splenic laceration and significant hemoperitoneum. (b) Pre-embolization celiac digital subtraction angiography (DSA) with no contrast extravasation or pseudoaneurysm. (c) Post embolization DSA displaying decreased but preserved perfusion to the spleen. Multiple detachable AZUR CX coils (Terumo Interventional Systems, Tokyo, Japan) and pushable Tornado coils (Cook Medical, Bloomington, IN, USA) can be seen in the main splenic artery, distal to the dorsal pancreatic artery.

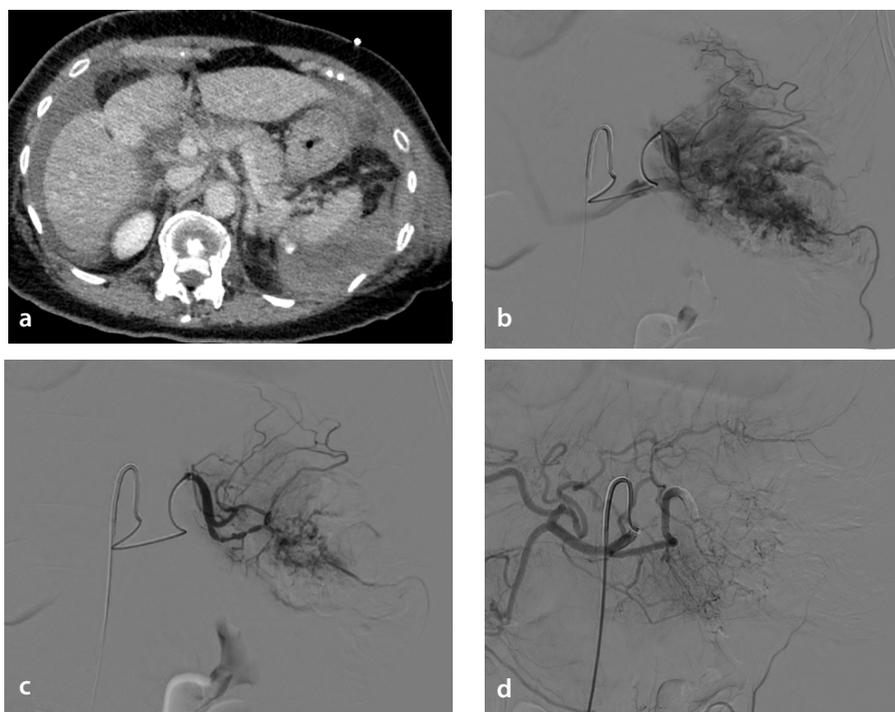


Figure 4. Distal splenic artery embolization with Gelfoam® in a 75-year-old woman following a motor vehicle collision. (a) Contrast-enhanced computed tomography showing multiple splenic lacerations with a perisplenic hemoperitoneum and active contrast extravasation. (b, c) Pre-embolization celiac digital subtraction angiography (DSA) showing multiple splenic arterial blushes without active contrast extravasation. (d) Post-embolization DSA displaying Gelfoam® embolization of splenic artery branches distal to all collateral pathways.

Table 2. Comparison of embolization techniques

	Distal w/Gelfoam® (n = 20)	Proximal w/coil (n = 17)	P value
Demographics/injury data			
Age, mean (SD)	44.5 (20.3)	39.8 (16.5)	0.45
Male, n (%)	16 (80)	11 (64.7)	0.46
AAST grade, mean (SD)	3.75 (0.7)	3.2 (0.8)	0.03
Injury severity score, mean (SD)	30.2 (12.4)	24.4 (11)	0.14
Post-embolization outcomes			
SAE procedure length, mean min (SD)	71.4 (25.5)	68.9 (21.7)	0.75
Post SAE splenectomy, n (%)	2 (10)	1 (5.9)	1
Splenic abscess, n (%)	0	0	1
Splenic infarct, n (%)	4 (20)	3 (17.6)	1
Symptomatic splenic infarct, n (%)	0	0	1
Total units of PRBCs, mean (SD)	1.6 (3.4)	0.9 (1.3)	0.43
Clinical outcomes			
Hospital length of stay, mean days (SD)	12 (10.1)	6.8 (4.5)	0.06
ICU length of stay, mean days (SD)	5.3 (4)	4.5 (4.1)	0.55
Readmission within 30 days, n (%)	1 (5)	1 (5.9)	1

SD, standard deviation; AAST, American Association for the Surgery of Trauma; SAE, splenic artery embolization; PRBCs, packed red blood cells; ICU, intensive care unit.

Table 3. Patient population stratified by embolization location

	Proximal (n = 26)	Distal (n = 38)	Combined (n = 8)	P value
Demographics/injury data				
Age, mean (SD)	41.8 (17.1)	42.6 (18.2)	50.8 (12.4)	0.42
Male, n (%)	14 (53.8)	28 (73.6)	5 (62.5)	0.26
AAST grade, mean (SD)	3.3 (0.8)	3.6 (0.9)	3.6 (0.5)	0.35
Injury severity score, mean (SD)	25.9 (12.4)	28.1 (11.6)	24.5 (11.3)	0.64
Embolization material				
Gelfoam®	9	20	0	
Coil	17	5	0	
Gelfoam® + coil	0	10	6	
Particles + coil	0	2	1	
Plug + coil	0	1	1	
Embolization outcomes				
Procedure length, mean min (SD)	68.0 (22.9)	75.8 (26.3)	64.4 (31.6)	0.35
Post SAE splenectomy, n (%)	3 (11.5)	3 (7.9)	1 (12.5)	0.86
Splenic abscess, n (%)	0	0	2 (25)	0.0003
Splenic infarct, n (%)	3 (11.5)	4 (10.5)	4 (50)	0.04
Symptomatic splenic infarct, n (%)	0	0	2 (25)	0.01
Total units of PRBCs, mean (SD)	2.9 (6.4)	2.4 (4.5)	1.9 (2.4)	0.87
Clinical outcomes				
Hospital length of stay, mean days (SD)	9.2 (9.0)	12.3 (9.7)	10 (5.5)	0.4
ICU length of stay, mean days (SD)	5.9 (9.0)	6.0 (5.3)	5.8 (4)	0.24
Readmission within 30 days, n (%)	1 (3.8)	2 (5.3)	1 (12.5)	0.64

SD, standard deviation; AAST, American Association for the Surgery of Trauma; SAE, splenic artery embolization; PRBCs, packed red blood cells; ICU, intensive care unit.

The rate of splenic abscess formation was significantly higher for patients embolized at both proximal and distal locations than for those who received either proximal or distal embolization alone ($P = 0.0003$). The rate of asymptomatic and symptomatic splenic infarction was significantly higher in patients embolized at combined proximal and distal locations ($P = 0.04$, $P = 0.01$).

Discussion

SAE is a safe and efficacious treatment option for BSI; the current study revealed an overall splenic salvage rate of 90.3%, which is consistent with the current literature.^{2,4,5} At our center, the preferred embolization techniques are distal embolization with Gelfoam® and proximal embolization with coils. There was no significant difference in splenic salvage rates or procedure length between the various embolization techniques. Recent publications have criticized the distal embolization approach, citing longer procedure times and higher complication rates, including splenic abscess formation and splenic infarction, than proximal embolization.^{1,2,10} Gelfoam® embolization has been associated with a higher incidence of life-threatening complications and a lower rate of clinical success than coil embolization.² In addition, Gelfoam® has been reported to have an increased risk of re-bleeding due to its temporary nature.^{12,17,18} These inferior outcomes associated with distal embolization and embolization with Gelfoam® were not found in our study.

Table 4. Patient population stratified by embolization material

	Gelfoam® (n = 29)	Coil (n = 22)	Combined (n = 21)	P value
Demographics/injury data				
Age, mean (SD)	44.9 (18.8)	43.5 (17.6)	40.7 (15)	0.7
Male, n (%)	19 (65.5)	15 (68.2)	13 (61.9)	0.97
AAST grade, mean (SD)	3.7 (0.7)	3.1 (0.9)	3.6 (0.7)	0.02
Injury severity score, mean (SD)	30.7 (13.5)	24 (11.0)	22.7 (47.3)	0.54
Post-embolization outcomes				
SAE procedure length, mean min (SD)	69.1 (24)	73.6 (23.4)	73.3 (30.8)	0.78
Post SAE splenectomy, n (%)	2 (6.9)	3 (13.6)	2 (9.5)	0.72
Splenic abscess, n (%)	0	0	2 (9.5)	0.08
Splenic infarct, n (%)	4 (13.8)	3 (13.6)	4 (19.4)	0.84
Symptomatic splenic infarct, n (%)	0	0	2 (19.4)	0.08
Total units of PRBCs, mean (SD)	2.6 (6.4)	2 (3.2)	3 (9.5)	0.81
Clinical outcomes				
Hospital length of stay, mean days (SD)	11.0 (8.8)	10.5 (8.9)	11.2 (10.1)	0.97
ICU length of stay, mean days (SD)	5.2 (4)	7 (6.2)	5.9 (6.7)	0.52
Readmission within 30 days, n (%)	1 (3.4)	1 (4.5)	2 (9.5)	0.63

SD, standard deviation; AAST, American Association for the Surgery of Trauma; SAE, splenic artery embolization; PRBCs, packed red blood cells; ICU, intensive care unit.

In particular, distal embolization with Gelfoam® was not associated with an increased rate of splenic infarction when compared with proximal embolization with coils. Distal embolization is commonly cited as being associated with a higher incidence of splenic infarction than proximal embolization.^{1,17,18} This increased risk of splenic infarction is thought to be the result of a lack of collateral blood flow to the spleen, which is commonly preserved by proximal embolization.¹ Although splenic infarction often occurs in the absence of clinical consequences, it may result in sequelae such as splenic abscess or rupture.¹⁹ However, our study found no association between distal embolization and an increased risk of splenic infarct or abscess.

Physicians at our institution often select distal embolization with Gelfoam® because of its many perceived benefits. Distal embolization preserves normal blood flow to a larger portion of the spleen, allowing for a more targeted treatment of focal lesions than proximal embolization.⁸ Distal embolization may also be preferentially selected over proximal embolization because of the lack of access for re-intervention distal to the initial site of embolization following proximal embolization.¹ Gelfoam® acts as a temporary agent and may preserve larger portions of splenic parenchyma. This may be especially beneficial for the treatment of trauma patients, who are often younger and

have fewer comorbidities that would hinder the healing process. In addition, Gelfoam® represents an inexpensive embolization material when compared with other options, lowering the cost burden for both hospitals and patients.^{8,10,20}

In our study, two patients were reported to have follow-up CT scans with findings suggestive of splenic abscess formation following embolization. Both patients were embolized with a combination of Gelfoam® distally and coils proximally. Our study also found that patients treated with combined embolization had a significantly higher rate of symptomatic and asymptomatic splenic infarct than patients treated with any embolization technique used in isolation. Combined embolization has previously been associated with higher rates of complications, with one meta-analysis indicating that combined embolization had a complication rate more than double that of proximal or distal embolization alone.² The increase in complications following combined embolization may be attributed to the loss of perfusion to greater portions of the spleen. This is supported by our study, which identified no abscesses following embolization with any technique used in isolation yet identified splenic abscesses in 25% of patients embolized with combined proximal and distal embolization.^{2,11,12,21,22}

This study demonstrates that distal embolization with Gelfoam® and the use of Gelfoam® or distal embolization alone is safe and may be beneficial in the setting of acute blunt trauma. The limitations of the present study are that it is a retrospective, single-center study. However, the results of our study represent a real-life, level 1 trauma center experience representative of the patient population in our region.

In conclusion, the splenic salvage rate in this study was 90.3%. The results of this study suggest that distal SAE with Gelfoam® is safe, and in our experience, it was not associated with an increased risk of splenic infarction compared with proximal SAE with coils.

Footnotes

Conflict of interest disclosure

Bahri Üstünsöz, MD, is Section Editor in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Alison A. Smith is a paid consultant for Aroa Biosurgery and on the advisory board for Prytime Medical Devices. Other authors have nothing to disclose.

References

1. Quencer KB, Smith TA. Review of proximal splenic artery embolization in blunt abdominal trauma. *CVIR Endovasc*. 2019;2(1):11. [CrossRef]
2. Rong JJ, Liu D, Liang M, et al. The impacts of different embolization techniques on splenic artery embolization for blunt splenic injury: a systematic review and meta-analysis. *Mil Med Res*. 2017;4:17. [CrossRef]
3. Lukies M, Zia A, Kavnoudias H, et al. Immune function after splenic artery embolization for blunt trauma: long-term assessment of CD27(+) IgM B-cell levels. *J Vasc Interv Radiol*. 2022;33(5):505-509. [CrossRef]
4. Lin BC, Wu CH, Wong YC, et al. Splenic artery embolization changes the management of blunt splenic injury: an observational analysis of 680 patients graded by the revised 2018 AAST-OIS. *Surg Endosc*. 2023;37(1):371-381. [CrossRef]
5. Cretcher M, Panick CEP, Boscanin A, Farsad K. Splenic trauma: endovascular treatment approach. *Ann Transl Med*. 2021;9(14):1194. [CrossRef]
6. Aoki M, Onogawa A, Matsumoto S, Matsushima K. Recent trends in the management of isolated high-grade splenic injuries: A nationwide analysis. *J Trauma Acute Care Surg*. 2023;94(2):220-225. [CrossRef]
7. Ahuja C, Farsad K, Chadha M. An overview of splenic embolization. *AJR Am J Roentgenol*. 2015;205(4):720-725. [CrossRef]

8. Raikhlin A, Baerlocher MO, Asch MR, Myers A. Imaging and transcatheter arterial embolization for traumatic splenic injuries: review of the literature. *Can J Surg.* 2008;51(6):464-472. [\[CrossRef\]](#)
9. Brahmabhatt AN, Ghobryal B, Wang P, Chughtai S, Baah NO. Evaluation of splenic artery embolization technique for blunt trauma. *J Emerg Trauma Shock.* 2021;14(3):148-152. [\[CrossRef\]](#)
10. Ekeh AP, Khalaf S, Ilyas S, Kauffman S, Walusimbi M, McCarthy MC. Complications arising from splenic artery embolization: a review of an 11-year experience. *Am J Surg.* 2013;205(3):250-254. [\[CrossRef\]](#)
11. Habash M, Ceballos D, Gunn AJ. Splenic artery embolization for patients with high-grade splenic trauma: indications, techniques, and clinical outcomes. *Semin Intervent Radiol.* 2021;38(1):105-112. [\[CrossRef\]](#)
12. Smith HE, Biffi WL, Majercik SD, Jednacz J, Lambiase R, Cioffi WG. Splenic artery embolization: have we gone too far? *J Trauma.* 2006;61(3):545-546. [\[CrossRef\]](#)
13. Abada HT, Golzarian J. Gelatine sponge particles: handling characteristics for endovascular use. *Tech Vasc Interv Radiol.* 2007;10(4):257-260. [\[CrossRef\]](#)
14. Lopera JE. Embolization in trauma: review of basic principles and techniques. *Semin Intervent Radiol.* 2021;38(1):18-33. [\[CrossRef\]](#)
15. Kozar RA, Crandall M, Shanmuganathan K, et al. Organ injury scaling 2018 update: Spleen, liver, and kidney. *J Trauma Acute Care Surg.* 2018;85(6):1119-1122. [\[CrossRef\]](#)
16. Javali RH, Krishnamoorthy, Patil A, Srinivasarangan M, Suraj, Sriharsha. Comparison of injury severity score, new injury severity score, revised trauma score and trauma and injury severity score for mortality prediction in elderly trauma patients. *Indian J Crit Care Med.* 2019;23(2):73-77. [\[CrossRef\]](#)
17. Imbrogno BF, Ray CE. Splenic artery embolization in blunt trauma. *Semin Intervent Radiol.* 2012;29(2):147-149. [\[CrossRef\]](#)
18. Schnüriger B, Inaba K, Konstantinidis A, Lustenberger T, Chan LS, Demetriades D. Outcomes of proximal versus distal splenic artery embolization after trauma: a systematic review and meta-analysis. *J Trauma.* 2011;70(1):252-260. [\[CrossRef\]](#)
19. Entriaken C, Weed Z, Parikh PP, Ekeh AP. Complications following splenic embolization for trauma: have things changed over time? *J Surg Res.* 2022;277:44-49. [\[CrossRef\]](#)
20. Lin BC, Wu CH, Wong YC, et al. Comparison of outcomes of proximal versus distal and combined splenic artery embolization in the management of blunt splenic injury: a report of 202 cases from a single trauma center. *Surg Endosc.* 2023;37(6):4689-4697. [\[CrossRef\]](#)
21. Liu PP, Lee WC, Cheng YF, et al. Use of splenic artery embolization as an adjunct to nonsurgical management of blunt splenic injury. *J Trauma.* 2004;56(4):768-773. [\[CrossRef\]](#)
22. Franco F, Monaco D, Volpi A, Marcato C, Larini P, Rossi C. The role of arterial embolization in blunt splenic injury. *Radiol Med.* 2011;116(3):454-465. [\[CrossRef\]](#)



Safety and efficacy of flow diverter stents in the treatment of bifurcation cerebral aneurysms: single-center experience

Yerbol Makhambetov
 Aiman Maidan
 Chingiz Nurimanov
 Assylbek Kaliyev
 Baurzhan Kunakbayev
 Nurtay Nurakay
 Serik Dyussebayev
 Nursultan Makhambetov

National Centre for Neurosurgery, Department of Vascular and Functional Neurosurgery, Astana, Kazakhstan

PURPOSE

The use of flow diverter (FD) stents is continually expanding. Aneurysms on arterial bifurcation typically have an undesirable anatomical form, are frequently wide-necked, and include one or more side-branch arteries. In recent years, the off-label use of flow diversion in treating intracranial aneurysms beyond the internal carotid artery has become increasingly popular. This study reports our center's initial experience treating bifurcation aneurysms with FD devices, documenting occlusion outcomes using the O'Kelly–Marotta and modified Cekirge–Saatci scales, as well as the safety of FD usage in bifurcation locations.

METHODS

This retrospective, single-center study analyzed a prospectively maintained database of patients with cerebral aneurysms treated endovascularly. The study identified bifurcation aneurysms that were treated between January 2019 and May 2022 by placing an FD device covering the neck of the aneurysm.

RESULTS

Our short series suggests that flow diversion is a viable therapeutic option for bifurcation aneurysms with favorable angiographic outcomes.

CONCLUSION

In highly selective cases, flow diversion may be considered for treating bifurcation aneurysms in patients who will undergo follow-up examinations in the future.

CLINICAL SIGNIFICANCE

Flow diversion has emerged as a valuable technique in the management of bifurcation aneurysms, offering the potential for satisfactory occlusion and long-term outcomes.

KEYWORDS

Intracranial aneurysm, bifurcation cerebral aneurysm, flow diverter, O'Kelly–Marotta Scale, modified Cekirge–Saatci scale

Corresponding author: Aiman Maidan

E-mail: maidanaiman@gmail.com

Received 27 June 2024; revision requested 01 August 2024; accepted 12 September 2024.



Epub: 21.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242903

The majority of cerebral aneurysms can now be successfully treated with flow diverter (FD) stents, and their applications are continually expanding to include distal aneurysm locations.¹ Aneurysms located on arterial bifurcation typically exhibit undesirable anatomical morphology, involving one or more side-branch arteries and often possessing a large neck.² Bifurcation artery aneurysms may sometimes feature wide-neck topologies that incorporate adjacent branches, resulting in a particularly complex morphology.³ Due to the high rate of long-term occlusion and lower surgical morbidity, despite the above-mentioned features, endovascular treatment is often considered the primary option for bifurcation aneurysms in some institutions.⁴

Nonetheless, there is a growing trend in utilizing endovascular methods to treat bifurcation aneurysms, driven by advancements in angiographic imaging, increased operator exper-

You may cite this article as: Makhambetov Y, Maidan A, Nurimanov C, et al. Safety and efficacy of flow diverter stents in the treatment of bifurcation cerebral aneurysms: single-center experience. *Diagn Interv Radiol.* 2025;31(4):366-371.

tise, and the adoption of more sophisticated techniques. Various endovascular techniques, such as stent-assisted coiling, balloon remodeling, Y-stenting, and Woven EndoBridge devices, have been employed for bifurcation aneurysm treatment. However, the outcomes do not provide satisfactory occlusion and have some complication rates.⁵

Flow diversion has emerged as an alternative method for treating challenging bifurcation aneurysms, particularly those involving a single side branch or those with a history of endovascular or surgical failure. The effectiveness and safety of this approach are still under investigation, with ongoing debate regarding the role of flow diversion in bifurcation aneurysms.⁶

Methods

This retrospective single-center study is based on a prospectively maintained database of patients with cerebral aneurysms treated endovascularly. The study identified bifurcation aneurysms treated between January 2019 and May 2022 by placing an FD device covering the neck of the aneurysm, as shown in Figure 1.

The study's FDs were used regardless of the availability of appropriate clips or stents, with dual-trained physicians overseeing the specifics of the treatment plan. The common antiplatelet regimen was acetylsalicylic acid (100 mg daily) and ticagrelor (one 90-mg

tablet twice daily). Postoperative follow-up visits were scheduled at 6, 12, and 24 months for comprehensive neurological assessments. Magnetic resonance imaging (MRI)/MR angiography was accepted if the patient was unable to undergo digital subtraction angiography (DSA).

Operation characteristics

Following the induction of general anesthesia, all procedures were conducted using a biplane flat-panel DSA unit (Artis Zee, Siemens). A long 6-Fr introducer sheath was inserted into the femoral artery. A guiding catheter, either Chaperon (MicroVent) or Asahi (Asahi Intecc), and, in some cases, a more distal intermediate catheter (Fargomax, Balt; or Sofia, MicroVent), were positioned. The appropriate microcatheter (Echelon Medtronic; Headway, MicroVent; Gama-17, Balt Extrusion) was advanced into the chosen bifurcation branch using 0.014 guidewires. When positioning the FD stent, particular attention was given to covering the fewest branches as possible. Following the FD stent's deployment, adjunctive coiling was performed using a jailed microcatheter in aneurysms greater than 15 mm. The Silk FD (Balt Extrusion) was used in 11 patients, and 4 patients were treated with the P48 MW (Phenox). All aneurysms in our series were treated with a single FD stent. Three aneurysms (20%) required the use of coils as an adjunct due to their size (patients #7, #11, and #13). In one case of low-profile FD usage with the P48 stent, additional coiling was used for better occlusion (patient #12). Patient #15 had coiling initially after the rupture, and the FD was used 2 weeks later. Following extubation in the intensive care unit (ICU), the patients spent at least 2 hours in a neurovascular ICU.

Clinical and imaging assessment

At every stage, the patients underwent a clinical evaluation using the modified Rankin scale (mRS). Following treatment, clinical statuses and any neurologic impairments at discharge or follow-up were documented. At least two DSAs, including 3-dimensional selective angiography runs, were performed at 6, 12, and 24 months after the operation. The angiographic results were assessed according to the O'Kelly–Marotta and modified Cekirge–Saatci grading scales. Integrated branch changes and the presence of intimal hyperplasia were also assessed.

Written informed consent was obtained from the patients for publication and any accompanying images. The Ethical Committee of the National Center provided ethical approval for this study (number 2 of ethical approval for neurosurgery on June 19, 2024). In addition, the investigators ensured that the study conformed to the principles of the Declaration of Helsinki (last revised in 2013) and was conducted in accordance with the ICH Guidelines for Good Clinical Practice.

Results

Baseline population characteristics

There was a total of 15 patients: 7 men and 8 women, with a mean age of 54.6 ± 8.1 years (range 46–68 years). The pretreatment mRS scores were 1 for 12 patients, 0 for one patient, 3 for one patient, and 4 for one patient. Four cases involved ruptured aneurysms, two patients had multiple intracranial aneurysms, and five patients presented with headaches. Table 1 summarizes the baseline clinical characteristics.

Main points

- Flow diversion may be a viable treatment option for certain bifurcation aneurysms with complex anatomical features and challenging hemodynamics, particularly in cases when the aneurysm is off-centered or involves small distal vessels, and the patient can commit to ongoing follow-up examinations. However, this approach requires careful patient selection and thorough assessment of the potential risks and benefits in a multidisciplinary setting.
- The modified Cekirge–Saatci classification scale is highly useful in reporting aneurysms treated by flow diverter stents.
- Understanding the development and implications of neointimal hyperplasia is essential for clinicians involved in the management of bifurcation aneurysms treated with flow diversion. As research in this area continues to evolve, ongoing efforts are focused on refining techniques and identifying strategies to minimize the impact of neointimal hyperplasia while optimizing the long-term outcomes for patients undergoing flow diversion for bifurcation aneurysms.

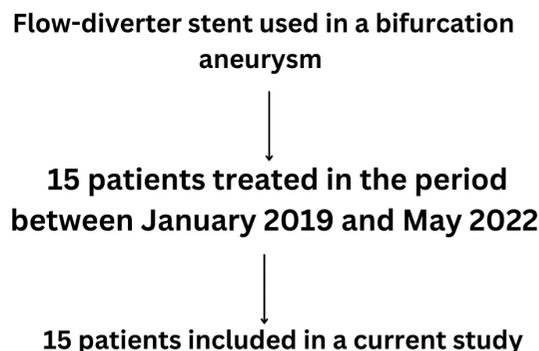


Figure 1. The flowchart of the study representing the criteria and number of patients from initial retrieval to the final study cohort.

Aneurysm characteristics

Seven patients had anterior cerebral artery distal bifurcation aneurysms, four patients had middle cerebral artery bifurcation aneurysms, three had an anterior communicating artery (ACom) aneurysm, and one had a posterior cerebral artery (PCA) bifurcation aneurysm. Two (13%) were giant aneurysms, and six (40%) were large aneurysms; the re-

maining aneurysms were small. We did not treat any aneurysms during the acute phase.

Outcomes

Table 2 summarizes the angiographic outcomes and the immediate post-procedure results. The average follow-up period for all patients in the study was 22 ± 9 months (95% confidence interval for the mean, 19–

24 months). In the immediate results, we observed mostly grade C outcomes (33%) according to the O’Kelly–Marotta grading scale. During the follow-up period, 10 out of 15 patients attended their follow-up appointments. Among them, 80% achieved complete occlusion (class 1 by the modified Cekirge–Saatci grading scale) or had stable, altered angioarchitecture. Intimal hyperplasia was detected in 27% of our patients but

Table 1. Patient, aneurysm, and stent characteristics

Case no.	Presentation	Age/gender	Side/location	Size of neck/dome (mm)	Dome-to-neck ratio*	FD stent, size (mm)	Additional coiling
1	mIAs, arterial hypertension	48/F	Bilobar Small Right A1–A2	2.00 3.66 × 2.15	1.83	Silk Vista 2.50 × 20	No
2	Headache	46/F	Saccular Small Left A1–A2	2.5 3.09 × 2.97	1.23	P48 MW 3.00 × 18	No
3	Speech problems; right hemiparesis SAH in 2014	49/M	Saccular Small Right A2	2 2.5 × 2.5	1.25	Silk Vista 2.75 × 20	No
4	Left hemiparesis SAH in 2022	52/F	Small ACom	2.5 2.5 × 3.3	1.32	Silk Vista Baby 2.5 × 20	No
5	Loss of conciseness	47/F	Giant Right M1–M2	5.31 17.5 × 6.0 × 15.3	3.2	Silk Vista Baby 2.75 × 25	No
6	Headache	63/F	Large Right A1–A2	3.7 5.39 × 3.50 × 4.12	1.45	Silk Vista Baby 2.75 × 20	No
7	Vision decrease in left eye	62/M	Large ACom	6.5 10.7 × 11.4 × 15	1.75	Silk Vista Baby 2.75 × 25	Yes
8	Headache	68/M	Large Right P1–P2	9.1 11.9 × 14.0	1.5	Silk Vista Baby 2.25 × 20	No
9	Headache AH Recanalized	54/F	Large Right M1–M2	3.5 7.5 × 3.0	2.14	Silk 2.25 × 15	No
10	Hemorrhage after stent	50/F	Small Left A3	2.87 2.5 × 2.5 × 2.0	0.87	P48 1.5 × 20	No
11	Ischemic stroke in 1998	64/M	Giant Right M1–M2	3.8 21 × 12.2	5.5	Silk Vista 3.0 × 25	Yes
12	mIAs, headache	47/M	Small Right A2–A3	7.18 4.70 × 3.74	0.6	P48 MW 3.0 × 18	Yes
13	Headache	58/M	Large Left A2–A3	3.39 6.28 × 4.20 × 3.39	1.85	Silk Vista 2.5 × 15	Yes
14	SAH 2016 Recanalized	62/M	Large Left M1–M2	3.5 7.5 × 3.0	2.14	P48 MW 3.0 × 18	No
15	SAH in 2022	50/F	Recanalized Small ACom	3.0 3 × 3	1.0	Silk Vista 2.25 × 15	Yes

FD, flow diverter; F, female; M, male; ACom, anterior communicating artery.

Table 2. Immediate and latest follow-up results

Case no.	mRS before	Immediate results		Last FU result			Integrated branch site changes*	Intimal hyperplasia	Complications		mRS after
		Jailed artery(s)	O'Kelly–Marrotta grading system	O'Kelly–Marrotta grading system	Modified Cekirge–Saatci classification	MRI/MRA			Peri operative	90 days FU	
1	1	Right A2	C2	D (6, 12 months)	1	abs	No change	No	No	No	1
2	0	Left A1, A2	C2	D (6, 12 months)	1	abs	No change	No	No	No	0
3	4	Left A1, A2	A2	D (6, 12 months)	1A	abs	Changing caliber	Yes	No	No	1
4	3	Left A1, A2	A3	-	-	-	-	-	Yes [§]	-	3
5	1	Right parietal MCA	C1	-	-	-	-	-	No	-	1
6	1	Left A1, A2	D	-	-	-	-	-	Yes [†]	-	1
7	1	Right A1, A2	B3	B3	5A	Stagnation	No change	No	No	No	1
8	1	Right PCA	C2	D (6, 12, 24 months)	1	abs	No change	No	No	No	1
9	1	Right M1	A2	D (6, 12 months)	1	abs	No change	No	No	No	3
10	1	Right M3	D	D (6, 12 months)	1	abs	No change	Yes	No	No	1
11	1	Right M1	C2	-	-	abs	-	-	-	-	1
12	1	Left A1	C2	D (12 months)	2	Aneurysm neck filling (26 months)	Changing caliber	Yes	No	No	1
13	1	Left A2, A3	C1	D (6, 12, 24 months)	1	abs	No change	Yes	No	No	0
14	1	Right A2, A3	C2	-	-	-	-	-	-	-	6
15	1	Left A2	A2	D (6, 12, 24 months)	1	abs	No change	No	No	No	0

*No change/occlusion/changing caliber. [§]Spontaneous stent separation in the microcatheter from the delivery system, Echelon-10 replaced by HeadwayDUO; Silk Vista Baby 2.25 × 20 mm changed to Silk Vista Baby 2.5 × 20 mm. [†]Due to the difficult acute angle of detachment of the A1 segment of the left PMA and the rigidity of the distal segment of the microcatheter, gamma17_d was replaced with gamma17_DS. mRS, modified Rankin scale; MRI, magnetic resonance imaging; MRA, magnetic resonance angiography.

was not clinically significant; these patients were advised to continue dual antiplatelet therapy (DAPT).

In this study, adjunctive coiling was performed in one-third of the patient cohort. This was done by coiling the aneurysm dome while sparing the neck and then applying the stent. Some technical complications are also described in Table 2.

Unfortunately, our case series included one death: 6 months after discharge, one patient (#14) passed away from a myocardial infarction, according to the clinical history provided by the family.

One patient (#7) was lost to follow-up, and three patients (#4–6) declined follow-up during the telephone interview. One patient (#11) is currently receiving chemotherapy

for cancer, but MRI shows signs of aneurysm obliteration.

Case examples

Case #1: patient #8

A 68-year-old man presented to our clinic in 2020 with a large incidental aneurysm at the junction of the posterior communicating artery and the PCA, measuring 11.9 × 14.0 mm in diameter, with a neck size of 9.1 mm. The decision was made to treat the aneurysm with FD placement. The operation was performed using a Silk Vista Baby measuring 2.25 × 20 mm. Control angiography showed aneurysm occlusion at the initial follow-up 6 months later (Figure 2). At the 3-year follow-up, the O'Kelly–Marotta grade was D and class 1A according to the modified Cekirge–Saatci classification.

Case #2: patient #7

A 62-year-old man presented with an initial complaint of decreased vision in his left eye. MRI revealed a large ACom aneurysm with maximum dimensions of 10.7 × 11.4 × 15 mm and a neck size of 6.5 mm. The management plan involved the use of an FD device, and the procedure included placing a Silk Vista Baby stent measuring 2.75 × 25 mm in the A1–A2 segment of the right anterior cerebral artery, along with additional coiling. During follow-ups at 12- and 24-months post-procedure, residual aneurysm filling with remodeling was noted. Despite this finding, the decision was made to continue observation with DAPT due to ongoing obliteration. At the 3-year follow-up, the O'Kelly–Marotta grade was 3B and class 3 according to the modified Cekirge–Saatci classification, as shown in Figure 3.

Discussion

FDs are medical devices used in the treatment of intracranial aneurysms, including those located at bifurcations.¹ Bifurcation aneurysms occur at the junction where two

blood vessels divide, creating a Y-shaped structure.² Treating aneurysms in these locations poses specific challenges, and FDs are one of the evolving tools in neuro-interventional procedures.³

Schob et al.⁴ conducted a retrospective analysis of patients treated with indirect flow diversion for off-centered bifurcation aneurysms. The authors found that indirect flow diversion was a safe and effective approach, with reduced perfusion of the aneurysm immediately after implantation and observable occlusion and reduction in size at follow-up.⁴ The pipeline embolization device (Medtronic, Irvine, CA, USA) was granted Food and Drug Administration approval in 2011, making it the first FD approved for use in the United States market.⁵ Initially, approval was intended for use in treating large or giant wide-necked aneurysms in the internal carotid artery, from the petrous to the superior hypophyseal segments, in individuals aged 22 years and older.⁶ For more than 5 years, the off-label use of FD stents for the treatment of distal aneurysms has been reported, but it remains debated.⁶

In our preliminary experience, FD devices were found to be safe for use in bifurcational aneurysms. The outcomes of this assessment were documented using the O'Kelly–Marotta and modified Cekirge–Saatci scales. In our series, among the patients who attended follow-up examinations, the total occlusion rate was 53% (8/15), with 80% of these classified as class 1 by the modified Cekirge–Saatci classification. This outcome is primarily attributed to factors related to the patients, such as missed follow-up MRI appointments and one case of mortality. Some studies have reported high rates of complete occlusion with flow diversion in bifurcation aneurysms, ranging from 62% to 80%.^{7,8} While high rates of complete occlusion with flow diversion in bifurcation aneurysms have been reported, it is important to consider the potential drawbacks and complications associated with this technique. Additionally, there have been reports of ischemic complications (NICE lesions) and procedure-related morbidities in patients treated with flow diversion for bifurcation aneurysms;⁹ however, we did not observe any in our case series.

Some emerging technologies, such as intrasaccular flow disruption devices and intrasaccular FDs, are being investigated as alternatives to traditional flow diversion techniques. These devices aim to address the limitations of standard flow diversion by providing more precise aneurysm occlusion while minimizing the risk of delayed aneurysm rupture.¹⁰ Stent-assisted coiling was initially introduced for wide-neck aneurysms based on the hypothesis that a stent can provide a framework to hold the coils in the aneurysmal cavity, preventing coil migration.¹¹

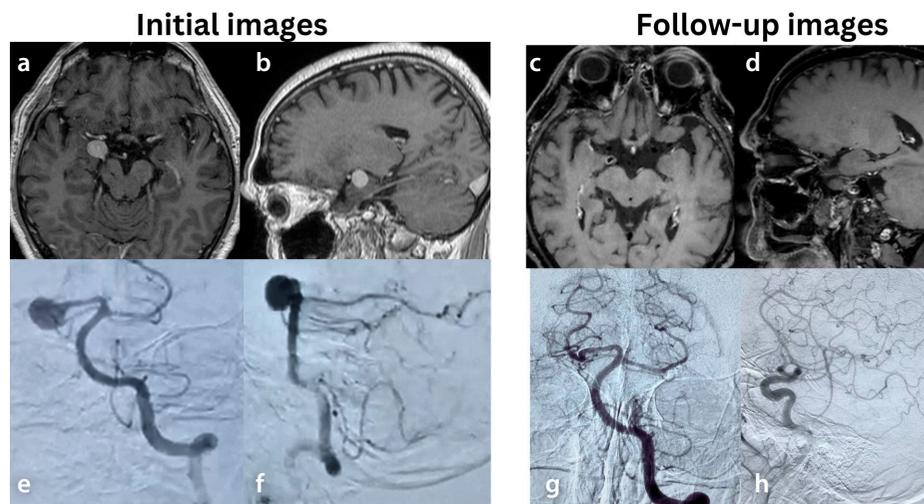


Figure 2. Patient #8. (a) Axial post-contrast images demonstrating a large incidental aneurysm at the junction of posterior communicating artery and posterior cerebral artery (PCA). (b) Sagittal images demonstrating a PCA aneurysm with adjacent brain compression. (c, d) Control magnetic resonance imaging/magnetic resonance angiograph demonstrating the absence of the aneurysm. (e, f) Pre-operative digital subtraction angiography demonstrating aneurysm sizes of 11.9×14.0 mm and a neck size of 9.1 mm. (g, h) At 3 years of follow-up, O'Kelly–Marotta grade D and class 1A according to the modified Cekirge–Saatci classification. The sequence of images illustrates the potential physiological development following flow diverter reconstruction, beginning with mechanical flow diversion and advancing to natural aneurysm thrombosis and complete occlusion. This is followed by internal parent artery repair, leading to complete anatomical restoration with the disappearance of the aneurysm-thrombus mass and a decrease in the regional mass effect.

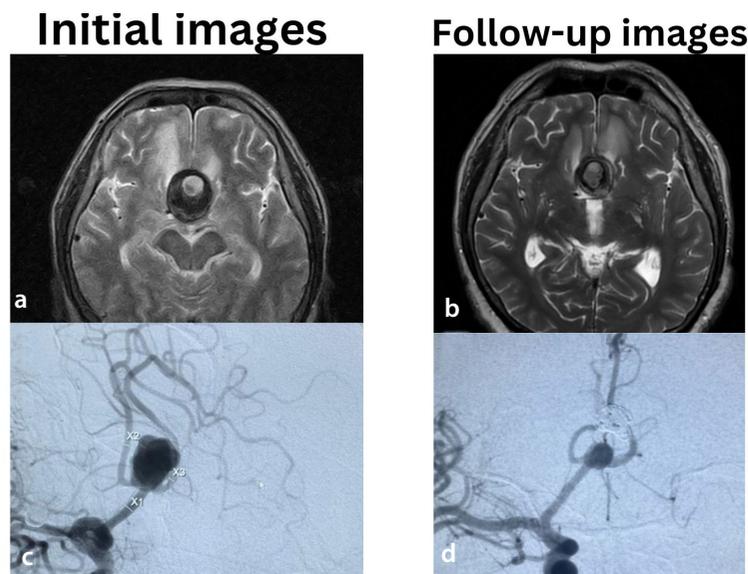


Figure 3. Patient #7. (a) Axial T2 Propeller magnetic resonance imaging showing a large partially thrombosed anterior communicating artery (ACoM) aneurysm measuring $37 \times 33 \times 28$ mm on initial presentation on 09/14/2022, with perianeurysmal edema. (b) Axial T2 image from the control presentation on 04/11/2024, showing signs of aneurysm shrinkage measuring $29 \times 28 \times 26$ mm. (c) Initial angiogram in the working projection demonstrating a large contrast-filled aneurysm measuring $10.7 \times 11.4 \times 15$ mm arising from ACoM. (d) At 3 years of follow-up, O'Kelly–Marotta grade 3B and class 3 according to the modified Cekirge–Saatci classification.

However, some new stents are being developed specifically for this approach.⁷ Patient #12 presented with a filling of the aneurysmal neck despite coiling, and the underlying cause remained unclear.

The modified Cekirge–Saatci classification, first described in 2015,¹² is an FD-specific occlusion classification that allows the subclassification of incorporated branches, aneurysm neck, and aneurysm occlusion changes in patients. However, not many centers report outcomes using this scale, which makes it difficult to assess outcomes in bifurcation or other challenging pathologies. In our series, one patient was classified as modified Cekirge–Saatci class 1A, with a reduced caliber branch due to intimal hyperplasia, although this was not clinically significant. The cause of intimal hyperplasia is the vascular endothelium, which is located at a crucial interface and becomes vasoactive in response to minute changes in hemodynamic conditions.¹³

Intimal hyperplasia is a common physiological response to vascular injury or alterations in blood flow dynamics, and it can be a significant issue when dealing with FDs, especially in the context of treating cerebral aneurysms.^{13,14} Intimal hyperplasia involves the proliferation of smooth muscle cells and the accumulation of extracellular matrix within the intima, the innermost layer of the blood vessel.¹⁴ When an FD is placed, the body may respond to the presence of foreign material and altered flow dynamics by initiating a healing response, which can include the development of intimal hyperplasia.¹⁵

Intimal hyperplasia has been reported with FD usage but not clearly in bifurcation aneurysms.^{13,14} However, neither pore density nor metal coverage has a significant association with aneurysmal occlusion.¹⁵ According to systematic reviews,¹⁶ ticagrelor was associated with better survival¹⁷ and lower neointimal hyperplasia.¹⁸ Nevertheless, the 27% rate of intimal hyperplasia observed at the latest follow-up in our series of patients could be due to non-adherence to therapy.

This study has several limitations that need to be acknowledged. First, the retrospective single-center design introduces inherent bias, presenting descriptive data on a limited and diverse population. The retrospective nature of the research relies on previously collected data, which might introduce several biases and limitations of the single-centered study. For instance, the core lab analyses and follow-up images may be incomplete, and the study cannot control for all variables that might influence outcomes. Additionally, the

retrospective nature means that the study is subject to selection bias, as it depends on cases that were previously selected for treatment or observation. These factors may affect the generalizability and accuracy of the findings. Second, subgroup analysis by FD stent type is not possible due to the limited population size. The rarity of flow diversion usage in bifurcational aneurysms further exacerbates this issue, making it challenging to conduct prospective studies or randomized trials. Consequently, it is crucial to interpret the results of this study while considering the inherent selection bias. The study's main limitation lies in its small sample size, suggesting that surgical outcomes might differ when research is conducted on a larger scale. We recommend that future researchers recruit larger sample sizes through multicenter studies.

In conclusion, flow diversion has emerged as an exceptionally valuable technique in the management of bifurcation aneurysms, providing the potential for satisfactory occlusion and favorable long-term outcomes. However, this technique requires further assessment.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Lv X, Yang H, Liu P, Li Y. Flow-diverter devices in the treatment of intracranial aneurysms: a meta-analysis and systematic review. *Neuroradiol J*. 2016;29(1):66-71. [\[CrossRef\]](#)
2. Lee KS, Zhang JJY, Nguyen V, et al. The evolution of intracranial aneurysm treatment techniques and future directions. *Neurosurg Rev*. 2022;45(1):1-25. [\[CrossRef\]](#)
3. Mortimer AM, Bradley MD, Mews P, Molyneux AJ, Renowden SA. Endovascular treatment of 300 consecutive middle cerebral artery aneurysms: clinical and radiologic outcomes. *AJNR Am J Neuroradiol*. 2014;35(4):706-714. [\[CrossRef\]](#)
4. Schob S, Brill R, Siebert E, et al. Indirect Flow diversion for off-centered bifurcation aneurysms and distant small-vessel aneurysms, a retrospective proof of concept study from five neurovascular centers. *Front Neurol*. 2022;12:801470. [\[CrossRef\]](#)
5. Shin DS, Carroll CP, Elghareeb M, Hoh BL, Kim BT. The evolution of flow-diverting stents for cerebral aneurysms; historical review, modern application, complications, and future direction. *J Korean Neurosurg Soc*. 2020;63(2):137-152. [\[CrossRef\]](#)
6. Liang B, Lesley WS, Robinson TM, Chen W, Benardete EA, Huang JH. Off-label

application of pipeline embolization device for intracranial aneurysms. *Neurointervention*. 2019;14(2):116-124. [\[CrossRef\]](#)

7. Liebig T, Gal G, O Kelly C, et al. Neqstent coil-assisted flow diverter (NQS) for the treatment of bifurcation aneurysms: the coil-assisted flow diversion safety and performance study (CAFI). *J NeuroInterv Surg*. 2024;16(7):721-725. [\[CrossRef\]](#)
8. Bhogal P, AlMatter M, Bázner H, Ganslandt O, Henkes H, Aguilar Pérez M. Flow diversion for the treatment of MCA bifurcation aneurysms—a single centre experience. *Front Neurol*. 2017;8:20. [\[CrossRef\]](#)
9. Michelozzi C, Darcourt J, Guenego A, et al. Flow diversion treatment of complex bifurcation aneurysms beyond the circle of Willis: complications, aneurysm sac occlusion, reabsorption, recurrence, and jailed branch modification at follow-up. *J Neurosurg*. 2018;131(6):1751-1762. [\[CrossRef\]](#)
10. Nelson PK, Lylyk P, Szikora I, Wetzel SG, Wanke I, Fiorella D. The pipeline embolization device for the intracranial treatment of aneurysms trial. *AJNR Am J Neuroradiol*. 2011;32(1):34-40. [\[CrossRef\]](#)
11. Hong Y, Wang YJ, Deng Z, Wu Q, Zhang JM. Stent-assisted coiling versus coiling in treatment of intracranial aneurysm: a systematic review and meta-analysis. *PLoS One*. 2014;9(1):e82311. [\[CrossRef\]](#)
12. Ravindran K, Salem MM, Alturki AY, Thomas AJ, Ogilvy CS, Moore JM. Endothelialization following flow diversion for intracranial aneurysms: a systematic review. *AJNR Am J Neuroradiol*. 2019;40(2):295-301. [\[CrossRef\]](#)
13. Cekirge HS, Saatci I. A new aneurysm occlusion classification after the impact of flow modification. *AJNR Am J Neuroradiol*. 2016;37(1):19-24. [\[CrossRef\]](#)
14. Sindeev S, Prothmann S, Frolov S, et al. Intimal hyperplasia after aneurysm treatment by flow diversion. *World Neurosurg*. 2019;122:e577-e583. [\[CrossRef\]](#)
15. Wang CB, Shi WW, Zhang GX, Lu HC, Ma J. Flow diverter treatment of posterior circulation aneurysms. A meta-analysis. *Neuroradiology*. 2016;58(4):391-400. [\[CrossRef\]](#)
16. Podlasek A, Al Sultan AA, Assis Z, Kashani N, Goyal M, Almekhlafi MA. Outcome of intracranial flow diversion according to the antiplatelet regimen used: a systematic review and meta-analysis. *J Neurointerv Surg*. 2020;12(2):148-155. [\[CrossRef\]](#)
17. Hodis S, Ding YH, Dai D, et al. Relationship between aneurysm occlusion and flow diverting device oversizing in a rabbit model. *J Neurointerv Surg*. 2016;8:94-98. [\[CrossRef\]](#)
18. Kim HK, Jeong MH, Lim KS, et al. Effects of ticagrelor on neointimal hyperplasia and endothelial function, compared with clopidogrel and prasugrel, in a porcine coronary stent restenosis model. *Int J Cardiol*. 2017;240:326-331. [\[CrossRef\]](#)



Computed tomography-guided cryoablation in treating adrenal metastases: a retrospective single-center study

Claudio Pusceddu¹
 Eliodoro Faiella²
 Claudio Cau¹
 Pierluigi Rinaldi¹
 Luca Melis³
 Salvatore Marsico⁴

¹Mater Olbia Hospital, Department of Radiology, Olbia, Italy

²Campus Bio Medico, Department of Radiology, Rome, Italy

³Businco Hospital, Department of Nuclear Medicine, Cagliari, Italy

⁴Hospital del Mar, Department of Radiology, Barcelona, Spain

PURPOSE

To assess the effectiveness and safety of computed tomography (CT)-guided cryoablation for treating adrenal metastases (AMs).

METHODS

This study included 12 patients treated with 13 CT-guided cryoablation procedures for AMs between 2016 and 2020. Patients were selected based on specific criteria, including tumor size ≤ 5 cm and suitability for surgery. Procedures were performed by expert radiologists, with comprehensive monitoring for complications and regular post-treatment evaluations.

RESULTS

The primary technical success rate was 91.7%, with a secondary success rate of 100% following repeat procedures. Over an 8–24-month follow-up period, local tumor recurrence was observed in 16.7% of patients, and systemic progression occurred in five (41.6%) patients. The average overall survival duration was 26.4 ± 5.6 months.

CONCLUSION

CT-guided cryoablation is a feasible and effective treatment option for AMs, demonstrating high technical success rates and manageable complications.

CLINICAL SIGNIFICANCE

This study highlights CT-guided cryoablation as a promising treatment for AMs, offering a minimally invasive alternative to surgery with good local control and safety profile. Further research, including multi-center studies, is needed to confirm these findings.

KEYWORDS

Adrenal glands, cryotherapy, computed tomography, local treatment, oncology

The increasing occurrence of metastatic tumors in the adrenal glands, originating from various types of cancers, such as lung, breast, colorectal, hepatocellular carcinoma, and melanoma, presents substantial challenges in the field of oncology.¹⁻⁴

Studies indicate that adrenal metastatic presence can be as high as 27% in patients with widespread cancer.^{5,6}

While the efficacy of direct adrenal interventions is yet to be established through randomized studies, the role of surgical resection, particularly for isolated adrenal metastases (AMs), has gained recognition in observational studies.⁷⁻¹¹

However, surgical procedures such as adrenalectomy are often hampered by individual health concerns and the intricacies of the operations, resulting in prolonged hospital admissions.¹⁻⁴

In contrast, computed tomography (CT)-guided methods, such as radiofrequency and microwave ablation, have demonstrated encouraging 1-year survival rates without local re-

Corresponding author: Salvatore Marsico

E-mail: salvatore.marsico@hotmail.it

Received 06 August 2024; revision requested 02 September 2024; accepted 25 September 2024.



Epub: 21.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242956

You may cite this article as: Pusceddu C, Faiella E, Cau C, Rinaldi P, Melis L, Marsico S. Computed tomography-guided cryoablation in treating adrenal metastases: a retrospective single-center study. *Diagn Interv Radiol.* 2025;31(4):372-376.

currence, ranging from 70.5% to 82% in the treatment of AMs.¹⁻³

As a CT-guided technique for treating AMs, cryoablation stands out among ablation methods due to its unique advantages. Unlike radiofrequency and microwave ablation, cryoablation allows for real-time visualization of the iceball, ensuring precise targeting and minimizing the risk of damaging surrounding tissues. This method also benefits from the cold-induced anesthesia effect, which reduces pain during the procedure and often eliminates the need for general anesthesia, leading to quicker recovery times.¹²⁻¹⁶

As a more recent development, CT-guided cryoablation of AMs offers several benefits, including clear visualization of the treated area, minimized discomfort, and expedited recovery, achieving promising results in recent studies, although its ultimate efficacy remains under evaluation.¹⁷⁻²⁴

The purpose of this study is to evaluate the efficacy and safety of CT-guided cryoablation in the treatment of AMs in comparison with other established ablation techniques.

Methods

This study was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were waived for the study due to its retrospective nature. The study was approved by Armando Businco Oncology Hospital's Ethics Committee (decision no: 53/15, date: 14/12/2015).

The study involved 12 patients who received 13 CT-guided cryoablation procedures

for AMs between January 2016 and December 2020.

Eligibility criteria included patients unsuitable for surgery, tumor size ≤ 5 cm, controlled or absent extra-adrenal tumors, and life expectancy ≥ 3 months. Exclusion criteria were adrenal vein invasion, significant coagulation disorders, active infections, or bleeding (Figure 1, Table 1).

Diagnosis involved patient history, abdominal CT/magnetic resonance imaging, and biopsy results, complemented by positron emission tomography-CT for detecting extra-adrenal tumors.

The procedures were performed by two expert interventional radiologists. Informed consent was secured from parents or guardians, and patients were briefed on possible complications. Prior to the procedure, the

skin entry site was anesthetized using 1% lidocaine, and patients received conscious sedation with midazolam and tramadol. The interventions utilized a multidetector CT system (SOMATOM® go.Top 128, Siemens Healthineers, Erlangen, Germany). An initial non-contrast CT scan was performed to ascertain the morphological features of the lesion.

Following this, the cryoablation was conducted with the patient in the prone position-utilizing a cryoablation system (Visual ICE™, Galil Medical-Boston Scientific, Arden Hills, MN, USA). This system was equipped with a single 17G insulated cryoprobe that could create ablation zones of varying diameters: IceSphere 1.5 (22 × 28 mm at -20°C and 15 × 24 mm at -40°) and IceRod 1.5 (29 × 45 mm at -20°C and 18 × 40 mm at -40°) (Figures 2, 3). Argon was used as the cryogen

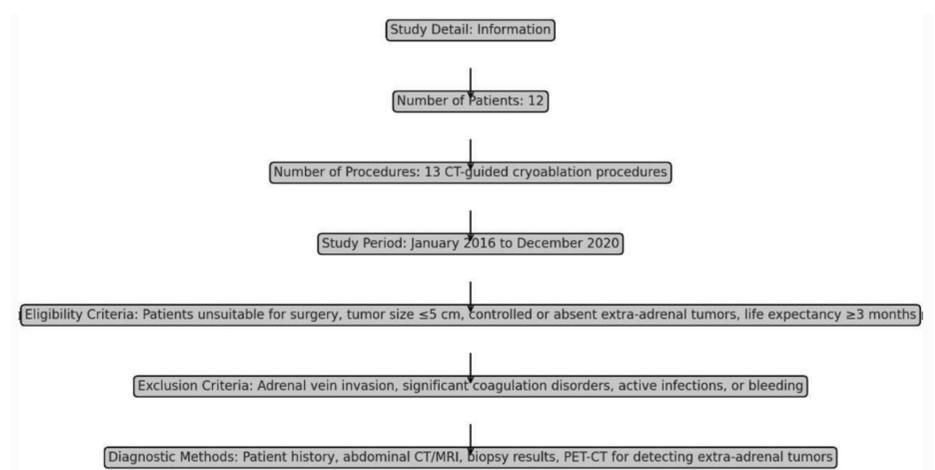


Figure 1. Flowchart of patient selection. CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography.

Table 1. Patient characteristics and treatments

Patient	Age/sex	Primary tumor	Size of lesion in mm	Number of cryoprobes
1	74/M	NSCLC	24 × 18	2
2	78/F	NSCLC	18 × 28	1
3	65/F	RCC	17 × 11	1
4	68/M	NSCLC	32 × 22	2
5	64/F	RCC	2.5 × 2	2
6	67/M	SCLC	28 × 22	2
		SCLC	20 × 24*	2
7	45/F	Gastric cancer	30 × 24	3
8	57/M	NSCLC	20 × 26	2
9	58/M	NSCLC	30 × 22	2
10	63/F	NSCLC	28 × 21	1
11	54/M	SCLC	22 × 18	1
12	61/F	Synovial sarcoma	30 × 25	3

*Local recurrence treated with a second cryoablation. M, male; F, female; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; RCC, renal cell carcinoma.

Main points

- Computed tomography-guided cryoablation achieved a high primary technical success rate of 91.7% and a secondary success rate of 100% in treating adrenal metastases (AMs).
- The procedure demonstrated good local control with a low recurrence rate, comparable to other ablation techniques.
- The safety profile was favorable, with no major complications reported, making cryoablation a viable option for patients unsuitable for surgery.
- The study suggests that cryoablation can effectively manage AMs while minimizing patient discomfort and recovery time.
- The findings support the need for further research to establish cryoablation as a standard treatment for AMs.

to create the iceball during the cryoablation process.

The cryoprobes were coaxially inserted into the adrenal lesions, as shown in Figure 2d. Scans were taken during the procedure to verify the placement and orientation of the instrument. The procedure involved a single 10-minute freezing cycle, followed by an 8-minute passive thawing phase. The freezing cycle was repeated twice. Post-freezing phase, non-contrast CT images were captured to evaluate the iceball's coverage and to identify any immediate complications.

To ensure patient safety, particularly in terms of managing the risk of hypertensive crises that can occur during interventional procedures on the adrenal gland, an anesthesiologist was present to continuously monitor vital parameters, including blood pressure. Additionally, an alpha-blocker was readily available in the operating room to promptly address any hypertensive emergencies.

Following the completion of the cryoablation procedures, abdominal CT images were acquired in all patients to detect early complications.

Patients were monitored during the first 24 hours and then discharged 1 day after the procedure if they were experiencing no discomfort.

Regular physical and laboratory evaluations, comprising blood cell count analysis, adrenal hormone measurements, and tumor marker monitoring based on the histological characteristics of the primary tumor, were conducted monthly.

Following adrenal cryoablation treatments and during the entire clinical and radiologic follow-up period, all the patients continued to receive their usual systemic therapies.

Regularly scheduled CT scans of the chest, abdomen, and pelvis, both with and without contrast, were conducted for patients at intervals of 1, 3, 6, and 12 months post-cryoablation, followed by biannual scans. The scanning protocols mirrored those used initially to evaluate the adrenal tumor lesions prior to the cryoablation.

Two months after the procedure, a complete disappearance of contrast enhancement in CT imaging was interpreted as a full response to the ablation treatment. By the 6th month, any alterations in contrast enhancement or size increases of the treated areas seen in CT scans were indicative of potential recurrence or progression of the disease. The schedule for these radiological evaluations was largely influenced by the understanding that early post-procedure CT scans (within the first 30 days) can show intense enhancement of the treated areas due to the inflammatory response triggered by the cryoablation process.²⁰

Hormonal analysis, including adrenal hormone levels, was conducted to assess the treatment's impact on adrenal function.

Statistical analysis

Statistical analysis was conducted to evaluate the outcomes of the cryoablation procedures. Descriptive statistics were used to summarize patient demographics, treatment characteristics, and follow-up data, including means, standard deviations, and percentages. For survival analysis, the Kaplan–Meier method was employed to estimate overall survival (OS) and progression-free survival (PFS) rates. The OS was defined as the time from the date of the first cryoablation procedure to the date of death from any cause, whereas the PFS was defined as the time from the first cryoablation procedure to the first documented evidence of disease progression or recurrence. The Kaplan–Meier curves

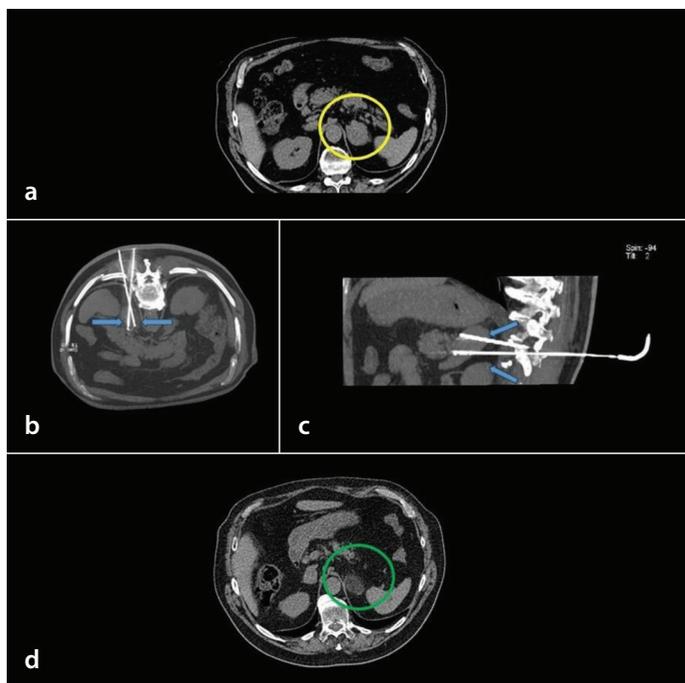


Figure 2. A 78-year-old woman with a history of non-small cell lung cancer with solid left adrenal metastasis [yellow circle in the preoperative computed tomography (CT) scan in a]. Treatment of the lesion with cryoablation using two Icesphere 1.5 cryoprobes (Visual ICE™, Galil Medical-Boston Scientific, Arden Hills, MN, USA), via posterior percutaneous access (blue arrows in the intraoperative axial scan image in b and in the sagittal plane reconstruction in c). Complete necrosis of the lesion in the CT follow-up (green arrow in d).

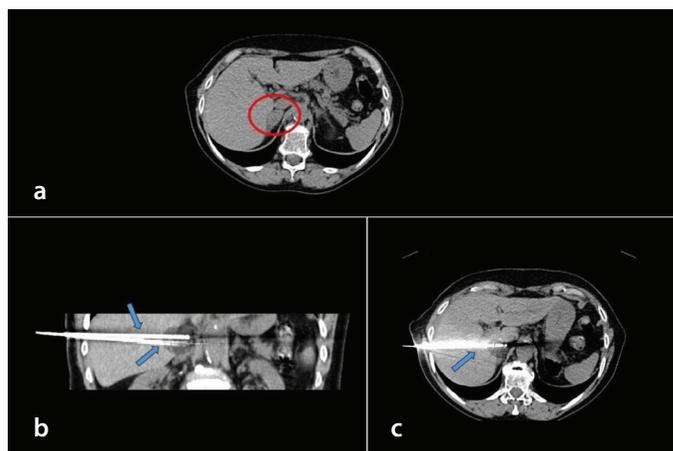


Figure 3. A 68-year-old man with a history of non-small cell lung cancer with solid right adrenal metastasis (red circle in the preoperative computed tomography scan in a). Treatment of the lesion with cryoablation using two Icesphere 1.5 cryoprobes (Visual IC™, Galil Medical-Boston Scientific, Arden Hills, MN, USA) via transhepatic access (blue arrows in the intraoperative reformatted coronal scan image in b and in the axial plane in c).

were used to illustrate survival probabilities over time, and the corresponding 95% confidence intervals (CIs) were calculated.

Results

All cryoablation sessions were successfully completed, and no procedure-related complications were observed. The CT scans performed post-procedure confirmed the expansion of the iceball, extending 1 cm beyond the tumor borders, thus establishing an adequate safety margin.

Patient demographics and treatment overview:

- A total of 12 patients with AMs underwent cryoablation treatment.
- Each patient presented with a single adrenal tumor.
- None of the patients had undergone adrenal resection prior to cryoablation.

Technical success:

- Complete tumor ablation, defined by the disappearance of any intratumoral arterial enhancement according to mRECIST criteria¹⁷ was achieved in 11 out of 12 patients.
- This resulted in a primary success rate of 91.7%.
- One patient exhibited residual tumor activity at follow-up and required a subsequent cryoablation session, which led to complete local tumor control, bringing the overall technical success rate to 100%.

Follow-up and recurrence:

- The follow-up period ranged from 8 to 24 months, with a mean duration of 16.3 ± 5.1 months.
- Six patients (50%) underwent additional systemic therapy post-cryoablation.
- Among this subgroup, three patients developed extra-adrenal tumors.
- Local tumor recurrence was observed in two patients (16.7%), occurring within a span of 6–20 months (median: 13 months).
- One of these patients underwent a second cryoablation session due to local recurrence, after which no further local recurrences were detected in subsequent CT scans.

Systemic progression and mortality:

- Systemic progression was documented in five patients (41.7%):
 - Two cases of renal cancer recurrence
 - Two cases of non-small cell lung cancer recurrence
 - One case of multiple bone metastases
- During the study period, five patients died:
 - Four deaths were due to tumor progression
 - One death was due to heart failure
- The average OS duration was calculated to be 26.4 ± 5.6 months (95% CI: 20.2–32.6).

Discussion

In recent advancements in oncological treatment, cryoablation has emerged as a standout technique for addressing AMs.^{15–19} This minimally invasive procedure has gained prominence due to its unique mechanism of action and significant clinical benefits, particularly when compared with other thermal ablation methods, such as radiofrequency and microwave ablation.

Cryoablation works by inducing cellular destruction through the rapid freezing of tumor tissue, leading to ice crystal formation within cells and subsequent cell death.¹² One of the key advantages of cryoablation is that it allows for visualization of the iceball in real-time using CT imaging. This real-time visualization enables precise monitoring of the ablation process, ensuring that the iceball extends beyond the tumor margins, which is critical for achieving a complete ablation and minimizing the risk of recurrence.^{13,14} Additionally, the cold-induced anesthesia effect often results in a generally painless procedure, reducing or eliminating the need for general anesthesia, which is a significant benefit for patients who may be at higher risk for anesthesia-related complications.^{15,16}

This single-center study contributes valuable insights into the efficacy and safety of CT-guided cryoablation for AMs. With 12 patients undergoing 13 procedures, we recorded a primary success rate of 91.7% and a secondary success rate of 100%, noteworthy when compared with existing literature where success rates for cryoablation, radiofrequency, and microwave ablation for AMs typically range from 70.5% to 82%.^{1,2,19–25}

The high technical success rates observed in this study underscore the efficacy of cryoablation, particularly in achieving high primary complete ablation rates. Furthermore, the ability to perform follow-up cryoablations in cases of residual tumor activity enhances the overall efficacy of this treatment modality.^{11–18}

In a study by Zhang et al.²², CT-guided cryoablation for AMs demonstrated a primary technical success rate of 90.3% and a secondary success rate of 100% among 31 patients. The study reported a 19.4% local progression rate and favorable survival outcomes, with 1-, 3-, and 5-year local PFS rates at 80.6%, 37.8%, and 18.4%, respectively. These results further validate the high efficacy of cryoablation, particularly in cases where initial treatment might not fully eradicate the tumor.²²

In 2021, Aoun et al.²³ assessed the technical feasibility and safety of percutaneous cryoablation for AMs in 34 patients. The local recurrence rate was 10% over 1.8 years. Recurrence was higher in tumors >3 cm. Major complications occurred in 5% of cases, with one directly linked to the procedure. Blood pressure increases were more significant in patients with residual adrenal tissue, but were manageable, especially with pre-treatment using alpha blockade.

Comparative studies further support the efficacy and safety of cryoablation. For example, in a meta-analysis by Pan et al.²⁵, the efficacy and safety of image-guided percutaneous thermal ablation for AMs were evaluated. The study revealed a 1-year local control rate of 80% and a 1-year OS rate of 77%, with severe adverse events occurring in 16.1% of cases and intraprocedural hypertensive crises in 21.9%. These findings align closely with the outcomes of the present study, where we observed that cryoablation effectively controlled local tumor progression with minimal complications.²⁵

An added benefit of cryoablation is the reduced risk of major complications, such as hypertensive crises, which are often associated with other thermal ablation methods.^{26,27}

In the present study, minor complications, primarily mild increases in blood pressure, were effectively managed without any reports of severe hypertensive crises. This aspect positions cryoablation as a potentially more favorable option for specific patient groups, particularly those who may be at higher risk for complications associated with other forms of thermal ablation.

However, it is important to acknowledge the limitations of this study. The retrospective design may have introduced selection bias, and the single-center nature, coupled with a small patient cohort, limits the generalizability of the findings. Consequently, while our results are promising, they highlight the need for multi-center randomized controlled trials to further validate the efficacy and safety of cryoablation in treating AMs and to explore its broader application across diverse patient populations.

In summary, this study reinforces the clinical efficacy of cryoablation for the treatment of AMs, particularly in achieving high primary complete ablation rates with a favorable safety profile. As the field of oncological treatment continues to evolve, cryoablation stands out as a compelling option, offering both precision and safety in the management of AMs.

In conclusion, CT-guided cryoablation is safe and effective in treating AMs, equating to other ablation techniques in terms of technical success, local tumor control, and the handling of complications.

These findings support the method's consideration as a practical alternative in oncological treatments, given its versatility in addressing various cancer types and its efficacy in both localized and systemic disease management.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Hasegawa T, Yamakado K, Nakatsuka A, et al. Unresectable adrenal metastases: clinical outcomes of radiofrequency ablation. *Radiology*. 2015;277(2):584-593. [\[CrossRef\]](#)
- Frenk NE, Daye D, Tuncali K, et al. Local control and survival after image-guided percutaneous ablation of adrenal metastases. *J Vasc Interv Radiol*. 2018;29(2):276-284. [\[CrossRef\]](#)
- Zhong H, Wang Z, Liu Y, et al. Efficacy and safety of ultrasound-guided percutaneous ablation for adrenal metastases: a meta-analysis. *J Ultrasound Med*. 2023;42(8):1779-1788. [\[CrossRef\]](#)
- Moinzadeh A, Gill IS. Laparoscopic radical adrenalectomy for malignancy in 31 patients. *J Urol*. 2005;173(2):519-525. [\[CrossRef\]](#)
- Strong VE, D'Angelica M, Tang L, et al. Laparoscopic adrenalectomy for isolated adrenal metastasis. *Ann Surg Oncol*. 2007;14(12):3392-3400. [\[CrossRef\]](#)
- Abrams HL, Spiro R, Goldstein N. Metastases in carcinoma; analysis of 1000 autopsied cases. *Cancer*. 1950;3(1):74-85. [\[CrossRef\]](#)
- Gunjur A, Duong C, Ball D, Siva S. Surgical and ablative therapies for the management of adrenal 'oligometastases' - a systematic review. *Cancer Treat Rev*. 2014;40(7):838-846. [\[CrossRef\]](#)
- Muth A, Persson F, Jansson S, Johanson V, Ahlman H, Wängberg B. Prognostic factors for survival after surgery for adrenal metastasis. *Eur J Surg Oncol*. 2010;36(7):699-704. [\[CrossRef\]](#)
- Howell GM, Carty SE, Armstrong MJ, et al. Outcome and prognostic factors after adrenalectomy for patients with distant adrenal metastasis. *Ann Surg Oncol*. 2013;20(11):3491-3496. [\[CrossRef\]](#)
- Vazquez BJ, Richards ML, Lohse CM, et al. Adrenalectomy improves outcomes of selected patients with metastatic carcinoma. *World J Surg*. 2012;36(6):1400-1405. [\[CrossRef\]](#)
- Moreno P, de la Quintana Basarrate A, Musholt TJ, et al. Adrenalectomy for solid tumor metastases: results of a multicenter European study. *Surgery*. 2013;154(6):1215-1222. [\[CrossRef\]](#)
- Gage AA, Baust J. Mechanisms of tissue injury in cryosurgery. *Cryobiology*. 1998;37(3):171-186. [\[CrossRef\]](#)
- Littrup PJ, Jallad B, Vorugu V, et al. Lethal isotherms of cryoablation in a phantom study: Effects of heat load, probe size, and number. *J Vasc Interv Radiol*. 2009;20(10):1343-1351. [\[CrossRef\]](#)
- Patel N, King AJ, Breen DJ. Percutaneous image-guided cryoablation of small renal masses. *Abdom Radiol (NY)*. 2016;41(4):754-766. [\[CrossRef\]](#)
- Baust JG, Gage AA. The molecular basis of cryosurgery. *BJU Int*. 2005;95(9):1187-1191. [\[CrossRef\]](#)
- He C, Zhao L, Yu HL, et al. Pneumothorax after percutaneous CT-guided lung nodule biopsy: a prospective, multicenter study. *Quant Imaging Med Surg*. 2024;14(1):208-218. [\[CrossRef\]](#)
- Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis*. 2010;30(1):52-60. [\[CrossRef\]](#)
- Bang HJ, Littrup PJ, Goodrich DJ, et al. Percutaneous cryoablation of metastatic renal cell carcinoma for local tumor control: feasibility, outcomes, and estimated cost-effectiveness for palliation. *J Vasc Interv Radiol*. 2012;23(6):770-777. [\[CrossRef\]](#)
- Rong G, Bai W, Dong Z, et al. Long-term outcomes of percutaneous cryoablation for patients with hepatocellular carcinoma within Milan criteria. *PLoS One*. 2015;10(4):e0123065. [\[CrossRef\]](#)
- Sun L, Zhang W, Liu H, Yuan J, Liu W, Yang Y. Computed tomography imaging-guided percutaneous argon-helium cryoablation of muscle-invasive bladder cancer: initial experience in 32 patients. *Cryobiology*. 2014;69:318-322. [\[CrossRef\]](#)
- Welch BT, Atwell TD, Nichols DA, et al. Percutaneous image-guided adrenal cryoablation: procedural considerations and technical success. *Radiology*. 2011;258(1):301-307. [\[CrossRef\]](#)
- Zhang W, Sun LJ, Xu J, Fu YF, Zhuang ZX. Computed tomography-guided cryoablation for adrenal metastases: local control and survival. *Medicine (Baltimore)*. 2018;97(51):e13885. [\[CrossRef\]](#)
- Aoun HD, Littrup PJ, Nahab B, et al. Percutaneous cryoablation of adrenal metastases: technical feasibility and safety. *Abdom Radiol (NY)*. 2021;46(6):2805-2813. [\[CrossRef\]](#)
- Welch BT, Atwell TD, Nichols DA, et al. Percutaneous image-guided adrenal cryoablation: procedural considerations and technical success. *Radiology*. 2011;258(1):301-307. [\[CrossRef\]](#)
- Pan S, Baal JD, Chen WC, et al. Image-guided percutaneous ablation of adrenal metastases: a meta-analysis of efficacy and safety. *J Vasc Interv Radiol*. 2021;32(4):527-535. [\[CrossRef\]](#)
- Aoun HD, Littrup PJ, Nahab B, et al. Percutaneous cryoablation of adrenal metastases: technical feasibility and safety. *Abdom Radiol (NY)*. 2021;46(6):2805-2813. [\[CrossRef\]](#)
- Littrup PJ, Jallad B, Vorugu V, et al. Lethal isotherms of cryoablation in a phantom study: Effects of heat load, probe size, and number. *J Vasc Interv Radiol*. 2009;20(10):1343-1351. [\[CrossRef\]](#)



Evaluating the prognostic impact of inflammatory markers on treatment outcomes in patients with intrahepatic cholangiocarcinoma undergoing radioembolization

Sinan Sözütok
 Ferhat Can Pişkin
 Hüseyin Tuğsan Ballı
 Berkay Dik

Çukurova University Faculty of Medicine, Balcalı Hospital, Department of Radiology, Adana, Türkiye

PURPOSE

Intrahepatic cholangiocarcinoma (iCCA) is a rare and aggressive malignancy with limited treatment options, often diagnosed at advanced stages. Radioembolization has emerged as a promising therapy, but its efficacy varies among patients, necessitating reliable biomarkers to predict treatment response. This study evaluates the prognostic impact of systemic inflammatory response markers on treatment outcomes in patients with iCCA undergoing radioembolization.

METHODS

This retrospective study included 70 patients with iCCA treated with radioembolization between January 2016 and December 2023. Inflammatory markers, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII), were measured from peripheral blood samples. Treatment response was assessed using the modified RECIST criteria, and survival analyses were performed using the Kaplan–Meier method and Cox proportional hazards regression.

RESULTS

Patients with lower NLR, PLR, and SII values exhibited significantly higher objective response rates ($P = 0.032$, $P = 0.016$, and $P = 0.001$, respectively). High levels of NLR, PLR, and SII were associated with shorter overall survival (12 vs. 16 months, $P = 0.007$; 12 vs. 16 months, $P = 0.004$; and 10 vs. 22 months, $P < 0.001$, respectively) and progression-free survival (3 vs. 7 months, $P = 0.046$ for SII). Multivariate analysis identified high SII ($P = 0.040$), lymph node metastasis ($P = 0.042$), and high serum total bilirubin ($P = 0.013$) as significant independent prognostic factors.

CONCLUSION

Systemic inflammatory markers such as NLR, PLR, and SII are valuable prognostic indicators for patients with iCCA undergoing radioembolization. These markers can aid in identifying patients likely to benefit from personalized treatment strategies, potentially improving clinical outcomes.

CLINICAL SIGNIFICANCE

The clinical significance of this study lies in its demonstration that systemic inflammatory markers (NLR, PLR, and SII) serve as valuable prognostic indicators for predicting treatment outcomes in patients with iCCA undergoing radioembolization, thus aiding in the identification of patients who may benefit from personalized treatment strategies and potentially improving clinical outcomes.

KEYWORDS

Intrahepatic cholangiocarcinoma, radioembolization, systemic inflammatory markers, treatment response, outcomes

Corresponding author: Ferhat Can Pişkin

E-mail: ferhatcpiskin@gmail.com

Received 14 July 2024; revision requested 02 September 2024; accepted 17 October 2024.



Epub: 25.11.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242929

You may cite this article as: Sözütok S, Pişkin FC, Ballı HT, Dik B. Evaluating the prognostic impact of inflammatory markers on treatment outcomes in patients with intrahepatic cholangiocarcinoma undergoing radioembolization. *Diagn Interv Radiol*. 25 November 2024 DOI: 10.4274/dir.2024.242929 [Epub Ahead of Print]. 2025;31(4):377–383.

Intrahepatic cholangiocarcinoma (iCCA) is a rare and highly aggressive malignancy originating from the intrahepatic bile ducts. The global incidence of iCCA is projected to increase tenfold over the next 20–30 years.¹ This rise is attributed to various risk factors, including chronic liver diseases such as hepatitis B and C, cirrhosis, and lifestyle factors such as alcohol consumption and obesity.^{2,3} Despite advancements in medical science, treatment options for iCCA remain limited, and most patients are diagnosed at advanced stages, minimizing the benefits of available systemic therapies.⁴ Consequently, there is a pressing need to explore novel and more effective therapeutic strategies for iCCA.

Radioembolization has emerged as a promising treatment option for patients with iCCA. It is a form of locoregional therapy that involves the injection of radioactive microspheres directly into the tumor's blood supply, effectively delivering high doses of radiation while sparing surrounding healthy tissues.⁵ Clinical studies have demonstrated that radioembolization, either as a monotherapy or in combination with chemotherapy, can significantly improve overall survival (OS) and progression-free survival (PFS) in patients with iCCA.⁶ However, the therapeutic efficacy of radioembolization varies considerably among patients, underscoring the necessity for reliable biomarkers to predict treatment response and optimize patient selection.

Main points

- Systemic inflammatory markers, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII), serve as significant prognostic indicators for patients with intrahepatic cholangiocarcinoma (iCCA) undergoing radioembolization.
- Lower levels of NLR, PLR, and SII are significantly associated with an improved treatment response, as well as enhanced overall survival and progression-free survival.
- Elevated SII, the presence of lymph node metastasis, and increased serum total bilirubin levels are identified as independent prognostic factors correlating with reduced survival in patients with iCCA.
- The integration of inflammatory markers into clinical decision-making processes has the potential to guide personalized treatment strategies, thereby optimizing outcomes for patients receiving radioembolization.

In recent years, the systemic inflammatory response (SIR) has garnered significant attention for its role in influencing cancer treatment outcomes. Biomarkers such as the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII) have been identified as indicators reflecting the impact of the SIR on the tumor microenvironment. Elevated SIR levels are known to substantially affect tumor growth, invasion, and metastasis, thereby influencing treatment outcomes.⁷ Emerging evidence suggests that the relationship between inflammation and cancer progression involves complex signaling pathways, including cytokine release, immune cell recruitment, and changes in the extracellular matrix, which collectively create a tumor-promoting microenvironment.⁸

The SIR has been demonstrated to significantly impact treatment response in liver tumors. Notably, in patients with hepatocellular carcinoma (HCC), elevated SIR levels are associated with poorer outcomes following radioembolization therapy.^{9,10} However, the literature on the effects of the SIR in patients with iCCA undergoing radioembolization remains limited.¹¹ By evaluating multiple inflammatory markers, this study aims to fill this gap by providing crucial insights into how these markers impact treatment outcomes in patients with iCCA. Our findings highlight the potential of implementing personalized therapies for this malignancy, which has a generally poor prognosis. To our knowledge, this is one of the first studies to comprehensively evaluate the impact of these markers in iCCA, and it includes the largest patient cohort studied to date. These contributions are critical for optimizing treatment strategies and facilitating more personalized approaches for patients with iCCA.

Therefore, this study aims to evaluate the prognostic impact of inflammatory responses on treatment outcomes in patients with iCCA undergoing radioembolization.

Methods

Study population

The study was performed in accordance with the ethical standards of the 1964 Declaration of Helsinki, and signed informed consent forms were obtained from all patients. The Institutional Clinical Research Ethical Committee (Çukurova University Faculty of Medicine Research Ethics Committee, meeting dated: 14.06.2024, decision number: 62/145) approved this single-center observational study.

This retrospective study reviewed the clinical records of patients diagnosed with iCCA who underwent radioembolization therapy at the radiology department between January 2016 and December 2023. The inclusion criteria were as follows: patients aged ≥ 18 years, with histopathologically confirmed iCCA, suitable for radioembolization therapy, and with complete clinical and imaging data. Patients with a minimum follow-up period of 3 months were included in the study. The exclusion criteria were as follows: patients who had previously undergone other local treatments such as transarterial chemoembolization or thermal ablation, those who did not achieve a response to treatment, those with incomplete medical records, and those who underwent surgery after radioembolization (Figure 1).

Pre-treatment clinical and imaging assessment

All patients underwent clinical, laboratory, and radiological evaluation before treatment. The Eastern Cooperative Oncology Group (ECOG) performance scores of the patients were assessed. Patients with ECOG performance scores between 0 and 2 were planned for treatment. Laboratory tests were conducted for the complete blood count, biochemistry, and hormone marker values of the patients before treatment. Dynamic contrast-enhanced magnetic resonance imaging (MRI) was performed, and the obtained images were evaluated by two abdominal radiologists with over 5 years' experience. The evaluation included an assessment of tumor size, number, location, presence of macrovascular invasion, and lymph node involvement.

Inflammatory markers measurement

Peripheral blood samples were collected from all patients within 1 week before radioembolization therapy. The NLR was calculated by dividing the absolute neutrophil count by the absolute lymphocyte count. The PLR was calculated by dividing the absolute platelet count by the absolute lymphocyte count. The aspartate aminotransferase (AST) to lymphocyte ratio (ALRI) was calculated by dividing the AST value by the lymphocyte count. The SII was calculated as platelet count \times neutrophil count/lymphocyte count.

Radioembolization procedure

All patients underwent a detailed pretreatment evaluation, including contrast-enhanced MRI to assess the extent of the disease, along with liver function tests. The radioembolization procedure was performed using resin-

glass-based yttrium-90 (90Y)-loaded microspheres in accordance with the Cardiovascular and Interventional Radiological Society of Europe Standards of Practice in Transarterial Radioembolization.¹² The choice of glass or resin microspheres was based on supply or logistical factors, regardless of tumor characteristics. Post-procedural imaging with 90Y positron emission tomography/computed tomography was performed to confirm the distribution of the microspheres within the liver and to detect any extrahepatic shunting.

Follow-up and assessment

Patients were followed up at regular intervals of 1, 3, and 6 months post-treatment and every 3 months thereafter, with clinical examination, laboratory tests, and imaging studies. OS and PFS were the primary endpoints. Treatment response was evaluated using the modified Response Evaluation Criteria in Solid Tumors 1.1 criteria based on imaging studies.¹³ Patients who demonstrated a complete response, partial response, or stable response to treatment were classified as having an objective response, whereas those with tumor progression were defined as having no objective response.

Statistical analysis

Data were analyzed using SPSS software version 24.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean \pm standard deviation or median (interquartile range), and categorical variables as frequencies and percentages. The laboratory values of the patients did not show a homogeneous distribution; therefore, the median value was used to divide the patients into groups. The groups were formed as "below median" and "above median." The chi-square test was employed to compare categorical variables between two groups. The Mann-Whitney U test was used to compare parametric data that did not show a homogeneous distribution. Survival curves for OS and PFS were plotted using the Kaplan-Meier method and compared using the log-rank test. Cox proportional hazards regression analysis was employed to identify independent prognostic factors for survival. A *P* value of less than 0.05 was considered statistically significant.

Results

In this study, a total of 70 patients were included, 34 (48.57%) of whom were men. The age range was 34–89 years, with a mean age of 59.29 \pm 11.89 years. The demographic and clinicopathological characteristics of the patients are presented in Table 1.

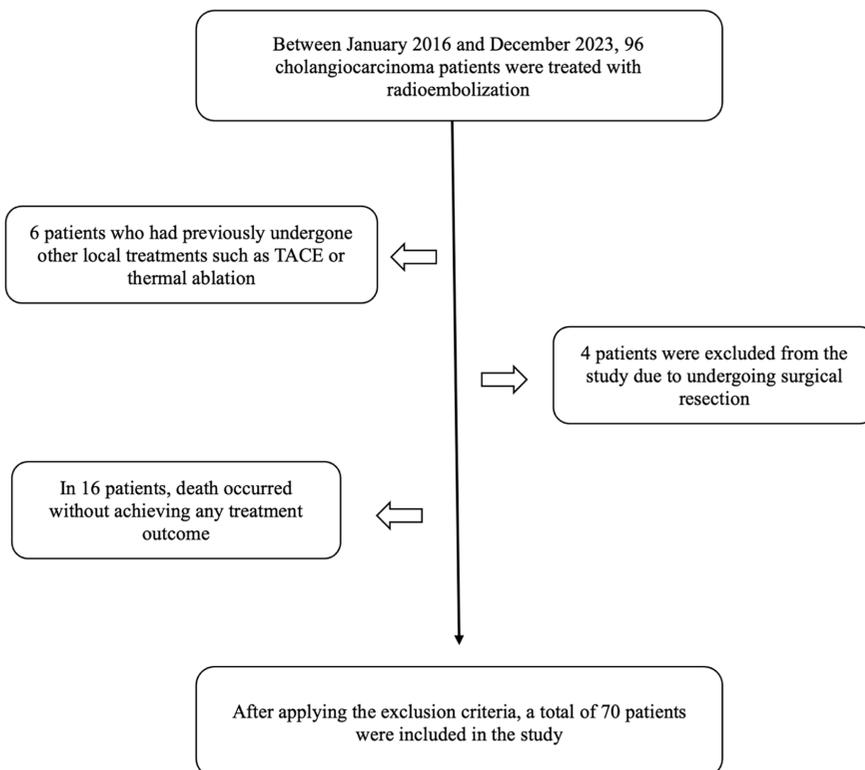


Figure 1. Study flowchart. TACE, transarterial embolization.

Table 1. Baseline demographic and clinical characteristics of the patients (n = 70)

Variables	n (%)
Age, years	59.2 \pm 11.8
Gender	
Men	34 (48.5%)
Women	36 (51.4%)
ALT, U/L	22.5 (27.2–38.3)*
AST, U/L	30 (33.3–46.5)*
Albumin, g/L	3.9 (3.5–3.8)*
TBIL, mg/dL	0.6 (0.5–1.1)*
CA19-9, U/mL	95 (770–3516)*
Largest tumor size	
\leq 6 cm	48 (68.5%)
>6 cm	22 (31.4%)
Lymph node metastasis	
Yes	41 (58.5%)
No	29 (41.4%)
Tumor number	
1	27 (38.5%)
>1	43 (61.4%)
Macrovascular invasion	
Yes	31 (55.7%)
No	39 (55.7%)
NLR	3.3 (3.6–5.7)*
PLR	143 (158–237)*
SII	771 (901–1398)*
ALRI	22.3 (24.6–41.7)*

*Values in parentheses represent median (lower limit–upper limit). ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBIL, total bilirubin; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; SII, systemic immune-inflammation index; ALRI, aspartate aminotransferase-to-lymphocyte ratio index.

Response to treatment

In the follow-up evaluation at 3 months to assess response to treatment, 6 (8.5%) patients achieved a complete response, 31 (44.2%) patients had a partial response, 1 (1.4%) patient had a stable response, and 38 (54.2%) patients exhibited progressive disease. An objective response was observed in 32 (45.7%) patients. Patients with low NLR, PLR, and SII values had a significantly higher objective response rate ($P = 0.032$, $P = 0.016$, and $P = 0.001$, respectively). No significant difference was found between the groups for ALRI ($P > 0.05$) (Table 2). The median NLR, PLR, and SII values of patients who obtained an objective response were significantly lower than those of the other group ($P = 0.037$, $P = 0.002$, and $P = 0.002$, respectively) (Table 3).

Overall survival analysis

In this study, 65 (92.8%) patients died. The median survival time was 14 months [95% confidence interval (CI): 10–16]. The mean survival time for the surviving patients was 28.4 months (95% CI: 11.1–45.6). Kaplan–Meier analysis revealed that patients with high NLR, PLR, and SII values had significantly shorter OS times than other patients (12 vs. 16 months, $P = 0.007$, 12 vs. 16 months, $P = 0.004$, and 10 vs. 22 months, $P < 0.001$,

respectively). No significant difference was found between the groups for ALRI (11 vs. 16, $P = 0.071$) (Figure 2 a-d). In the univariate analysis, the demographic information of the patients, tumor burden, liver function tests, tumor hormone markers, and inflammatory scores were evaluated. Tumor burden, liver function tests, and inflammatory scores were identified as prognostic factors. In the multivariate analysis, high SII values ($P = 0.040$), the presence of lymph node metastasis ($P = 0.042$), and high serum total bilirubin (TBIL) values ($P = 0.013$) were identified as significant independent prognostic factors (Table 4).

Progression/recurrence analysis

In this study, 44 (62.8%) patients experienced progression/recurrence. The median PFS was 7 months (95% CI: 6–9). Kaplan–Meier analysis showed that patients with high SII values had significantly shorter PFS than other patients (3 months, 95% CI: 3–6 vs. 7 months, 95% CI: 7–22, $P = 0.046$). No significant difference was found between the groups for NLR, PLR, and ALRI ($P > 0.05$) (Figure 3). In the univariate analysis for PFS, the demographic information of the patients, tumor burden, liver function tests, tumor hormone markers, and inflammatory scores

were evaluated. However, no prognostic factors were identified ($P > 0.05$).

Discussion

This study assessed the prognostic impact of inflammatory markers on treatment outcomes in patients with iCCA undergoing radioembolization. Our findings indicate that patients with higher NLR, PLR, and SII values had lower treatment response rates, shorter OS times, and higher progression/recurrence rates. These results align with the existing literature, underscoring the critical role of the SIR in predicting treatment outcomes.

Neutrophils suppress anti-tumor immunity through immunosuppressive cytokines, fostering a pro-tumorigenic environment, whereas platelets promote tumor progression by protecting circulating tumor cells from immune surveillance and facilitating their adhesion to the endothelium, thereby promoting metastasis. Conversely, lymphocytes support anti-tumor immunity by targeting and destroying cancer cells; hence, elevated NLR, PLR, and SII levels indicate a shift toward a pro-tumorigenic state, reflecting tumor biology and significantly affecting treatment outcomes.¹⁴ Recent studies have examined the impact of the inflammatory response on treatment outcomes in patients with CCA, emphasizing the prognostic value of these markers in this specific cancer type.^{11,14,15}

Yu et al.¹⁵ demonstrated that lower SII values in patients with liver cancer undergoing interventional therapy were associated with improved treatment outcomes, including higher response rates and reduced rates of recurrence and metastasis. Similarly, studies on patients with HCC undergoing radioembolization have shown that lower NLR and PLR values correlate with better treatment responses and OS.¹⁶ In our study, we found that patients with low NLR, PLR, and SII values had significantly higher objective response rates ($P = 0.032$, $P = 0.016$, and $P = 0.001$, respectively). These findings align with the existing literature, suggesting that lower systemic inflammation is also associated with better treatment outcomes in patients with CCA.

Filippi et al.¹¹ reported that elevated NLR levels were significantly associated with shorter OS times in patients with iCCA treated with 90Y-radioembolization, with a median OS of 7.5 months for patients with high NLR compared with 17.5 months for those with low NLR. In this study, Kaplan–Meier analysis showed that patients with higher SII

Table 2. Comparison of treatment response among patients grouped by inflammatory markers

	Objective response n (%)	Non-objective response n (%)	<i>P</i>
NLR			
Below median	24 (34.3%)	12 (17.1%)	0.032
Above median	14 (20.0%)	20 (28.6%)	
PLR			
Below median	24 (34.3%)	11 (15.7%)	0.016
Above median	14 (20.0%)	21 (30.0%)	
SII			
Below median	19 (27.1%)	16 (22.9%)	0.001
Above median	19 (27.1%)	16 (22.9%)	
ALRI			
Below median	21 (30.0%)	14 (20.0%)	0.472
Above median	17 (24.3%)	18 (25.7%)	

NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; SII, systemic immune-inflammation index; ALRI, aspartate aminotransferase-to-lymphocyte ratio.

Table 3. Comparison of inflammatory markers of values among patients grouped by treatment response

	Objective response	Non-objective response	<i>P</i>
NLR	2.7 (1.36–28)	3.9 (1.3–16.2)	0.037
PLR	125 (27.2–1180)	215 (32.5–665)	0.002
SII	605 (119–5503)	1166 (331–4322)	0.002
ALRI	17.9 (6.1–230)	23.9 (6.1–135)	0.120

NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; SII, systemic immune-inflammation index; ALRI, aspartate aminotransferase-to-lymphocyte ratio.

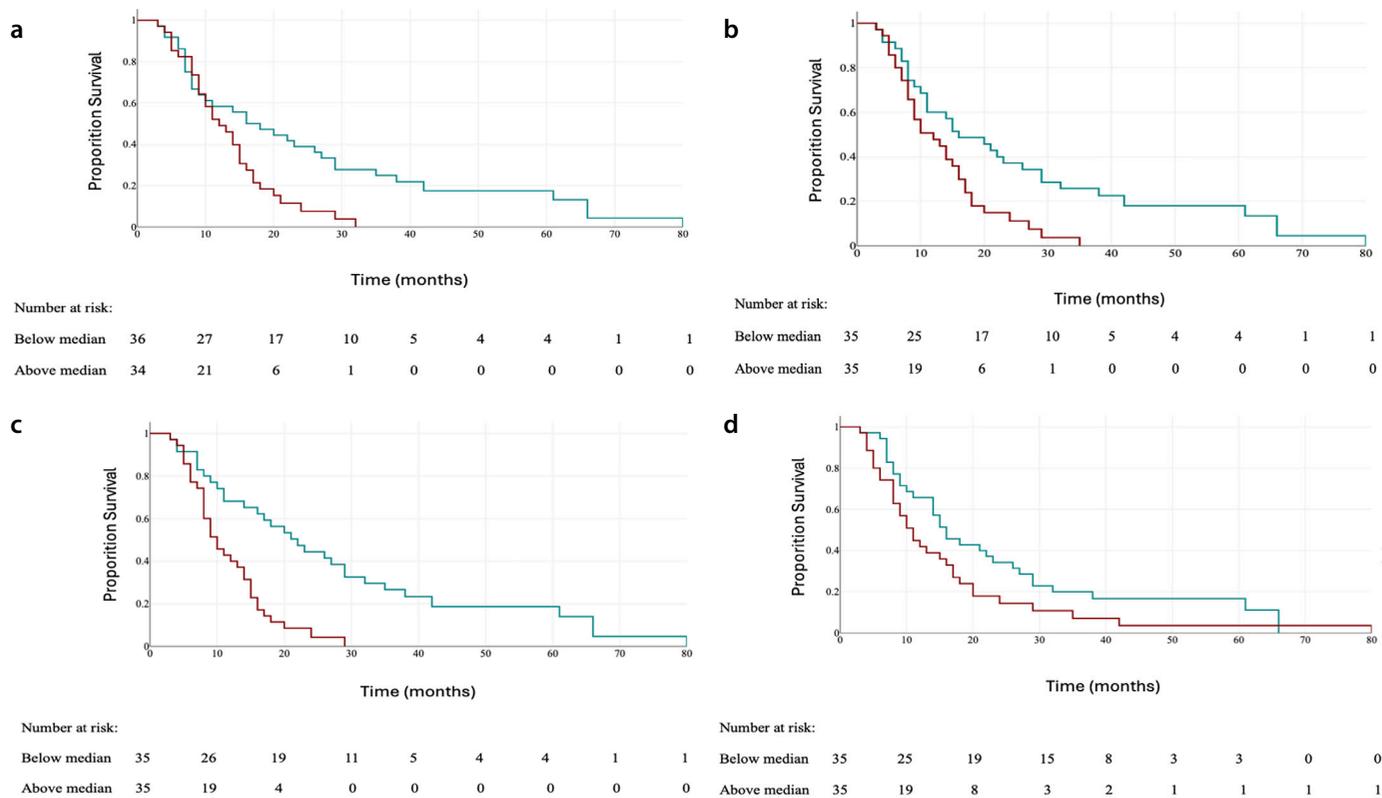


Figure 2. (a-d) Kaplan-Meier curves comparing overall survival for those with an above or below median neutrophil-to-lymphocyte ratio (NLR) (a), platelet-to-lymphocyte ratio (PLR) (b), systemic immune-inflammation index (SII) (c), and aspartate aminotransferase-to-lymphocyte ratio (ALRI) (d). High NLR, PLR, and SII were associated with significantly shorter survival times, whereas no significant difference was observed for ALRI.

Table 4. Univariate and multivariate Cox regression models examining risk factors for overall survival

Variables	Univariate			Multivariate		
	HR	95% CI	P	HR	95% CI	P
Age, years	0.87	0.53–1.42	0.572			
Gender, men	1	0.97–1.02	0.754			
ALT, U/L	1.01	1.01–1.02	0.057			
AST, U/L	1.01	0.92–1.02	0.070			
Albumin, g/L	0.08	0.6–1.26	0.447			
TBIL, mg/dL	1.35	1.11–1.66	0.004	1.3	1.06–1.59	0.013
CA19-9, U/mL	1.49	0.81–2.71	0.197			
Largest tumor size, cm (>6/≤6)	2.17	1.19–3.94	0.011	1.49	0.74–3	0.260
Lymph node metastasis (yes/no)	2.36	1.37–4.07	0.002	1.85	1.02–3.34	0.042
Tumor number (>1/1)	1.45	0.87–2.41	0.156			
Macrovascular invasion (yes/no)	1.46	0.89–2.4	0.132			
NLR	2.04	1.19–3.51	0.010	1.21	0.61–2.43	0.584
PLR	2.11	1.24–3.61	0.006	0.99	0.48–2.02	0.975
SII	3.17	1.79–5.51	<0.001	2.36	1.04–5.34	0.04
ALRI	1.55	0.94–2.55	0.082			

HR, hazard ratio; CI, confidence interval; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBIL, total bilirubin; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; SII, systemic immune-inflammation index; ALRI, aspartate aminotransferase-to-lymphocyte ratio.

values had significantly shorter OS times ($P < 0.001$). Specifically, patients with high SII had a median OS of 10 months compared with 22 months for those with low SII. Cox regression analysis in our study identified several independent prognostic factors for OS in patients

with iCCA undergoing radioembolization. Specifically, high SII values [hazard ratio (HR): 2.36, 95% CI: 1.04–5.34, $P = 0.04$], the presence of lymph node metastasis (HR: 1.85, 95% CI: 1.02–3.34, $P = 0.042$), and elevated serum TBIL levels (HR: 1.3, 95% CI: 1.06–1.59,

$P = 0.013$) were significant independent predictors of poor survival. These findings are consistent with previous studies, such as the work by Li et al.¹⁷, which demonstrate that elevated SII levels are associated with poorer survival outcomes in patients with perihilar

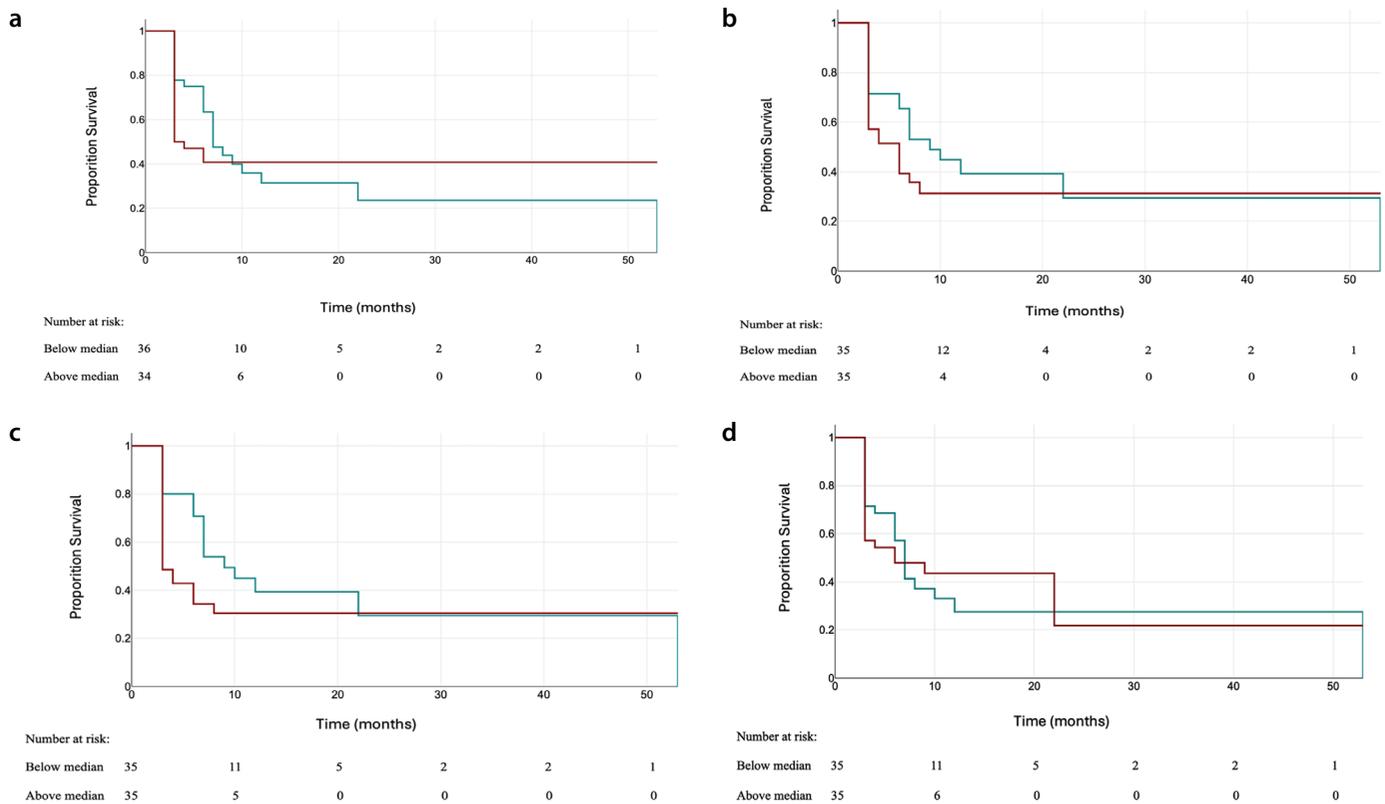


Figure 3. Kaplan–Meier curves comparing progression-free survival for those with an above or below median neutrophil-to-lymphocyte ratio (NLR) (a), platelet-to-lymphocyte ratio (PLR) (b), systemic immune-inflammation index (SII) (c), and aspartate aminotransferase-to-lymphocyte ratio (ALRI) (d). High SII were associated with significantly shorter progression-free survival times, whereas no significant difference was observed for NLR, PLR, and ALRI.

CCA (HR: 1.57, 95% CI: 1.17–2.10, $P = 0.003$). Furthermore, the presence of lymph node metastasis has been widely recognized as a significant prognostic factor in iCCA, as highlighted by Li et al.¹⁸, who reported similar associations with OS (HR: 1.88, 95% CI: 1.32–2.68, $P < 0.001$). Elevated serum TBIL, indicative of impaired liver function, also aligns with findings from multiple studies that underscore its role in predicting survival outcomes in hepatobiliary malignancies.¹⁵

Several studies have shown that the SIR significantly impacts tumor progression and recurrence. For example, Yu et al.¹⁵ demonstrated that patients with liver cancer and elevated SII levels had a shorter PFS (HR: 1.152, 95% CI: 1.878–5.329, $P < 0.001$). Similarly, Li et al.¹⁸ found that patients with high preoperative SII levels and NLR were significantly correlated with a shorter PFS (HR: 1.385, 95% CI: 1.005–1.909, $P = 0.046$). In our study, patients with iCCA and higher SII values had significantly shorter PFS, with a median PFS of 3 months compared with 7 months for those with lower SII values, indicating a higher risk of tumor progression and recurrence.

This study has several strengths. It includes a comprehensive retrospective analysis of a large patient cohort and a detailed

evaluation of inflammatory markers, providing robust data on the prognostic significance of these markers in patients with iCCA undergoing radioembolization. The findings from our study contribute to the growing body of literature by suggesting that systemic inflammatory markers, such as SII, NLR, and PLR, could serve as valuable prognostic indicators in this patient population. These markers may help identify patients who are more likely to benefit from the addition of radioembolization to their treatment regimen, potentially guiding more personalized and effective therapeutic strategies.

This study has several limitations. Its retrospective nature may introduce selection bias and limit the ability to establish causal relationships. The sample size, although sufficient for initial findings, is relatively small and may not fully represent the broader population of patients with iCCA. This study relies on data from a single institution, which may limit the generalizability of the results. Variability in radioembolization techniques and the lack of standardized protocols for measuring inflammatory markers could also influence the outcomes. Future prospective studies with larger, multicentric cohorts and standardized methodologies are necessary to validate these findings and better explain

the role of systemic inflammatory markers in predicting treatment outcomes for patients with iCCA.

In conclusion, this study underscores the prognostic significance of systemic inflammatory markers (NLR, PLR, and SII) in patients with iCCA undergoing radioembolization. Elevated levels of these markers correlate with poorer treatment response, shorter OS, and increased progression. Incorporating these biomarkers into clinical practice can aid in patient stratification and personalized treatment planning, potentially improving outcomes. Further prospective studies are needed to validate these findings and enhance the use of inflammatory markers in guiding iCCA treatment.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Qurashi M, Vithayathil M, Khan SA. Epidemiology of cholangiocarcinoma. *Eur J Surg Oncol.* 2023;9:107064. [CrossRef]

2. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. [\[CrossRef\]](#)
3. Valle JW, Kelley RK, Nervi B, Oh DY, Zhu AX. Biliary tract cancer. *Lancet*. 2021;397(10272):428-444. [\[CrossRef\]](#)
4. Labib PL, Davidson BR, Sharma RA, Pereira SP. Locoregional therapies in cholangiocarcinoma. *Hepat Oncol*. 2017;4(4):99-109. [\[CrossRef\]](#)
5. Salem R, Thurston KG. Radioembolization with 90Yttrium microspheres: a state-of-the-art brachytherapy treatment for primary and secondary liver malignancies. Part 1: technical and methodologic considerations. *J Vasc Interv Radiol*. 2006;17(8):1251-1278. Erratum in: *J Vasc Interv Radiol*. 2006;17(10):1594. [\[CrossRef\]](#)
6. Elvevi A, Laffusa A, Elisei F, et al. Any role for transarterial radioembolization in unresectable intrahepatic cholangiocarcinoma in the era of advanced systemic therapies? *World J Hepatol*. 2023;15(12):1284-1293. [\[CrossRef\]](#)
7. Dolan RD, Lim J, McSorley ST, Horgan PG, McMillan DC. The role of the systemic inflammatory response in predicting outcomes in patients with operable cancer: systematic review and meta-analysis. *Sci Rep*. 2017;7(1):16717. [\[CrossRef\]](#)
8. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*. 2009;30(7):1073-1081. [\[CrossRef\]](#)
9. Young S, Rubin N, D'Souza D, et al. Inflammatory scores: correlation with clinical outcomes in hepatocellular carcinoma patients undergoing transarterial radioembolization. *Cardiovasc Intervent Radiol*. 2022;45(4):461-475. [\[CrossRef\]](#)
10. Öcal O, Kupčinskas J, Morkunas E, et al. Prognostic value of baseline interleukin 6 levels in liver decompensation and survival in HCC patients undergoing radioembolization. *EJNMMI Res*. 2021;11(1):51. [\[CrossRef\]](#)
11. Filippi L, Di Costanzo GG, Tortora R, et al. Prognostic value of neutrophil-to-lymphocyte ratio and its correlation with fluorine-18-fluorodeoxyglucose metabolic parameters in intrahepatic cholangiocarcinoma submitted to 90Y-radioembolization. *Nucl Med Commun*. 2020;41(1):78-86. [\[CrossRef\]](#)
12. Mahnken AH, Spreafico C, Maleux G, Helmberger T, Jakobs TF. Standards of practice in transarterial radioembolization. *Cardiovasc Intervent Radiol*. 2013;36(3):613-622. [\[CrossRef\]](#)
13. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis*. 2010;30(1):52-60. [\[CrossRef\]](#)
14. Yang Z, Zhang D, Zeng H, et al. Inflammation-based scores predict responses to PD-1 Inhibitor treatment in intrahepatic cholangiocarcinoma. *J Inflamm Res*. 2022;15:5721-5731. [\[CrossRef\]](#)
15. Yu L, Zheng S, Yang J, Fu Z, Zhu X, Su K. Association between the systemic immune inflammation index and recurrence or metastasis after interventional therapy in patients with primary liver cancer- a retrospective cohort study. *J Gastrointest Oncol*. 2023;14(2):780-788. [\[CrossRef\]](#)
16. Ren A, Li Z, Cheng P, Zhang X, Deng R, Ma Y. Systemic immune-inflammation index is a prognostic predictor in patients with intrahepatic cholangiocarcinoma undergoing liver transplantation. *Mediators Inflamm*. 2021;2021:6656996. [\[CrossRef\]](#)
17. Li J, Gao L, Liu T, Feng D. Association of systemic inflammation index with survival in patients with advanced perihilar cholangiocarcinoma treated with interventional therapy. *Front Oncol*. 2022;12:1038759. [\[CrossRef\]](#)
18. Li H, Wang JJ, Zhang M, et al. Prognostic significance of systemic immune-inflammation index in patients with intrahepatic cholangiocarcinoma undergoing hepatic resection. *World J Gastrointest Oncol*. 2020;12(4):467-482. [\[CrossRef\]](#)



Efficacy of endovascular circulating false lumen occlusion in chronic aneurysmal descending aortic dissections

Emeric Gremen^{1,2}
 Mathieu Finas²
 Elliott Mathieu²
 Frédéric Thony²
 Mathieu Rodiere²
 Julien Ghelfi^{1,2}

¹Grenoble-Alpes University Faculty of Medicine, Department of Radiology, Grenoble, France

²Grenoble Alpes University Hospital, Department of Radiology, Grenoble, France

PURPOSE

To evaluate the efficacy of endovascular circulating false lumen occlusion (CFLO) in inducing positive aortic remodeling in chronic aneurysmal descending aortic dissection (AD).

METHODS

This retrospective monocentric study included patients treated by CFLO between 2003 and 2022 in the context of chronic AD with progressive descending aneurysmal evolution and persistent circulating false lumen (FL). The procedure was achieved with coils, plugs, and/or glue at the entry tear or in the FL and/or with covered stenting in the supra-aortic trunk. The primary endpoint evaluated the positive aortic remodeling, defined as stabilization or a decrease in the aortic diameter on a computed tomography scan at the 1-year follow-up after the procedure. The FL circulating status, safety, and occurrence of aneurysm events during follow-up were also evaluated.

RESULTS

Twenty patients [median age: 65.4 years, interquartile range (IQR): 58.4–69.9; 13 men] were included, with a median duration from an acute AD of 32.5 months (IQR: 8.8–76.5). Twelve patients (60%) achieved complete FL thrombosis after CFLO, whereas 8/20 patients (40.0%) experienced partial thrombosis. Additionally, positive aortic remodeling was observed in 13 patients (65%). Following the procedure, the aneurysmal aortic diameter decreased in 8/20 patients (40.0%) and remained stable in 5/20 patients (25.0%). Two patients (10%) had complications related to the procedure. Two patients (10%) had secondary aneurysm events during follow-up.

CONCLUSION

CFLO is a feasible and efficient method to induce FL thrombosis and reduce aneurysmal progression in chronic AD.

CLINICAL SIGNIFICANCE

The positive outcomes observed highlight the potential of this technique to improve patient management in complex aortic pathologies. This approach offers a valuable option in the management of chronic AD and emphasizes the importance of endovascular interventions in enhancing patient outcomes.

KEYWORDS

Chronic aortic dissection, false lumen, true lumen, entry tear, endovascular occlusion, circulating false aortic lumen, embolization, coils, plugs, glue, stent, thoracic endovascular aortic repair, false lumen thrombosis, aortic aneurysm

Corresponding author: Emeric Gremen

E-mail: egremen@chu-grenoble.fr

Received 28 August 2024; revision requested 19 September 2024; last revision received 10 October 2024; accepted 17 October 2024.



Epub: 17.03.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2025.242986

Aortic dissection (AD) is a serious pathology with a high mortality rate of 11%–25% in the acute phase.¹ One third of dissections develop an aneurysm requiring reintervention within 4 years.^{2,3} A persistent patent false lumen (FL) fed by an entry tear (ET) is associated with a relatively high risk of long-term mortality and late aortic events in patients with AD;⁴ however, complete FL thrombosis limits this risk.⁵ The development of chronic diastolic aortic hypertension^{6,7} contributes to elevated pressure within the expanding FL, increasing the risk of aortic rupture.⁸

You may cite this article as: Gremen E, Finas M, Mathieu E, Thony F, Rodiere M, Ghelfi J. Efficacy of endovascular circulating false lumen occlusion in chronic aneurysmal descending aortic dissections. *Diagn Interv Radiol.* 2025;31(4):384-391.

In chronic dissection, optimal antihypertensive therapy is recommended,⁹ but the chronic aneurysmal evolution may require open surgical repair¹⁰ or thoracic endovascular aortic repair (TEVAR).¹¹ These two approaches do not always result in FL thrombosis, which continues to circulate in 80%–90% of patients after isolated proximal surgical repair¹² and in up to 43% of patients treated using TEVAR.¹³ By targeting the ET feeding the circulation of the FL, endovascular circulating false lumen occlusion (CFLO) using embolization material (coils, glue, and plugs) or an uncovered stent can be an effective approach. This other endovascular method may be a valuable alternative or a complementary option when the FL remains patent and is supplied by an accessible ET. This treatment enables personalized targeted embolization for each patient to achieve FL thrombosis, thereby limiting aneurysmal evolution and reducing the risk of aortic rupture. Although TEVAR is now considered the gold standard for this indication,¹⁴ the risk of spinal cord ischemia, although low (approximately 8%), remains a serious concern with serious consequences (approximately 1.5% incidence of paraplegia).¹⁵ Currently, no comparative efficacy studies exist regarding CFLO and traditional TEVAR. The advantage of CFLO treatment may lie in the fact that it is considered a less invasive option than the placement of an aortic stent graft and can avoid the associated complications. Moreover, CFLO could

reduce the risk of spinal cord ischemia associated with TEVAR by maintaining the patency of certain intercostal arteries.

The main objective of this study was to evaluate the efficacy of CFLO on mid-term positive aortic remodeling in chronic aneurysmal descending AD. In addition, particular attention has been given to the safety of this treatment by investigating the associated adverse effects associated with the intervention.

Methods

Patients

This monocentric retrospective study was conducted between 2003 and 2022 at the University Hospital of Grenoble Alpes (France). The inclusion criteria (Figure 1) were patients who underwent CFLO for chronic (i.e., >90 days)¹⁶ persistent descending AD (type A after surgery or B with or without previous TEVAR) in the context of pejorative aneurysmal evolution with a patent FL fed by an ET unable to be treated by TEVAR and the following: 1) an aortic descending diameter >55 mm; and/or 2) aortic descending diameter progress >5 mm/year; and/or 3) TEVAR failure with a persistent untreated punctiform entry tear (PET) far from the prosthesis and retrograde false lumen flow (RFLF). The exclusion criteria were no imaging follow-up >12 months, no imaging at baseline, traumatic AD, or aneurysm without AD or from infectious etiology. Aneurysmal evolution treated by isolated TEVAR or the treatment of a perigraft leak in the context of TEVAR were also excluded. All procedures performed were in accordance with the ethical standards of the national research committee and with the 1964 Helsinki Declaration and

its later amendments or comparable ethical standards. This study was approved by the Comité d’Ethique pour la Recherche en Imagerie Médicale (approval no: CRM-2107-194, date: 09/2021) For this type of retrospective study, informed consent is not required. A declaration of informed non-opposition was required for each enrolled participant.

Pathologic aortic evaluation

Imaging follow-up was performed using computed tomography (CT) scans to explore the aorta from the supra-aortic trunks (SATs) to the common femoral arteries, performed with and without contrast material enhancement. Aortic diameters were evaluated using the latest available imaging before CFLO (baseline CT scan) and the latest available imaging after CFLO (follow-up CT scan), which was defined as the end of follow-up.

The Ishimaru classification¹⁷ was utilized to localize the aortic aneurysm level and the ETs (Table 1) on CT scans. This classification divides the aorta into several zones, allowing for the precise identification of lesions: zone 0 involves the origin of the innominate artery; zone 1, the origin of the left common carotid artery; zone 2, the origin of the left subclavian artery; zone 3, the proximal descending thoracic aorta down to the T4 vertebral body (aortic isthmus); and zone 4, the remainder of the thoracic aorta.

To increase reproducibility and avoid cross-sectional effects, the aortic diameter was measured by the largest short-axial diameter of the outer contour of the aneurysm and, in the aortic arch, by the largest diameter perpendicular to the curvature.¹⁸ Measurement was performed by a single observer and controlled by intraobserver reproducibility measurements.

Main points

- Chronic aortic dissection (AD) often leads to persistent patent false lumen (FL) circulation, which increases the risk of aneurysmal progression and aortic rupture.
- Thoracic endovascular aortic repair is now the standard treatment for AD but does not always achieve complete thrombosis of the FL (43% remains patent) and may fail in cases with unfavorable anatomy.
- This study evaluates the efficacy of endovascular circulating false lumen occlusion (CFLO), using coils, glue, plugs, and an uncovered stent, in achieving FL thrombosis and inducing positive aortic remodeling, aiming to promote these techniques within the therapeutic arsenal.
- Complete FL thrombosis was achieved in 60% of patients, and positive aortic remodeling was obtained in 65% of patients.
- The feared consequences and complications were minimal compared with the underlying pathology, with no reported cases of spinal cord ischemia associated with CFLO.

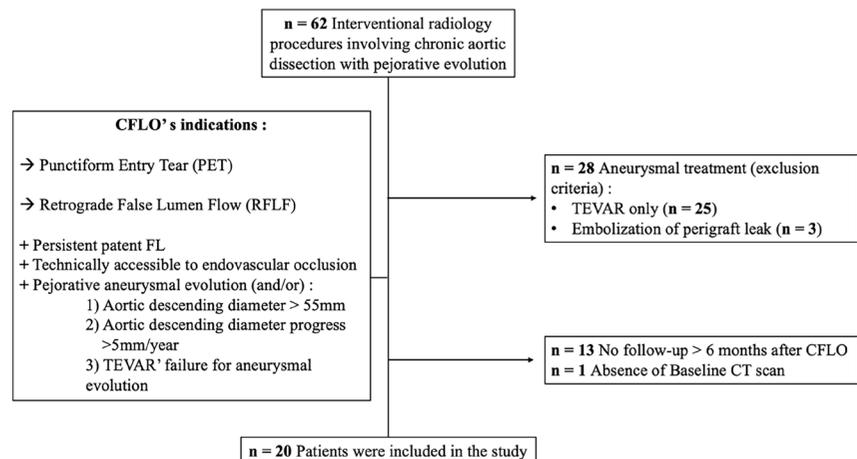


Figure 1. Study flowchart. CFLO, circulating false lumen occlusion; FL, false lumen; TEVAR, thoracic endovascular aortic repair; CT, computed tomography.

Angiography and embolization techniques

The procedures were performed in an angiography department (Allura Integris 15, Philips Medical systems, Best, Netherlands, and Artis Zeego, Siemens, Erlangen, Germany) by three interventional radiologists with 8 (M.F.), 18 (M.R.), and 30 (F.T.) years of experience. The CFLO was performed using different embolization materials, such as coils (AZUR® Peripheral HydroCoil, Terumo, Rueil-Malmaison, France; Nester Embolization Coil, Cook Medical Europe LTD, Limerick, Ireland), glue (Onyx® Liquid Embolic Agent, Medtronic, Paris, France; Hystoacril®, B. Braun Medical SA, Granges-Paccot, Schweiz), and plugs (Amplatzer® Vascular Plug II, Abbot Medical, Issy-les-Moulineaux, France). Covered (WallGraft®, Boston Scientific, Marlborough, MA, USA; AdvantaV12®, Atrium Medical Corporation, Hudson, NH, USA) or uncovered (SmartControl®, Cordis, Issy-les-Moulineaux, France; Zilver®, Cook Medical Europe LTD, Limerick, Ireland) stents were also utilized to prevent reinjection from an ET in the SAT. These materials were used separately (CFLO single technique) or in combination (CFLO combined technique).

The patients' cases were reviewed prior to the procedure to determine the most appropriate strategy. The precise description of the conditions for the use of embolic materials is challenging, as each case requires personalized embolization that considers various parameters, such as the location, size, shape and number of ETs, dimensions of the FL, and blood flow velocity. Similarly, the overall morphology of the aorta and the potential access routes should be considered, as not all types of materials can be implanted. Plugs, for example, can sometimes be difficult to insert, making it necessary to prioritize the remaining tool options. Some of these cases have been illustrated as examples in Figures 2, 3, and 4, and the entire range of embolization tools should be considered, especially the liquid agents, as illustrated by the column relating to glue in Figure 4.

Several key FL embolization scenarios are described here, although it is challenging to be more precise in this complex pathology. For SAT PETs, stenting was preferred to cover the ET, redirecting the flow into the trunk lumen. Punctiform ET occlusion can be performed directly by placing a plug with one fin in the true lumen (TL) and the rest of the materials in the FL. There is a risk of coil or glue migration in the case of rapid flow in the FL, which can be reduced by the combined use of these materials (for example, a

framework of large coils sealed by glue). Any RFLF can be managed using glue, with the natural retrograde flow directed toward the aneurysm.

Under local anesthesia, radiologic percutaneous arterial (femoral, radial, or humeral) access was performed using a 4–6F introducer sheath. Digital subtraction angiography was performed within the aorta to identify the ET targeted on the CT scan. The end of the procedure was characterized by the absence of opacification of the FL on the final angiogram. Sometimes, if the benefit–risk ratio of the technique becomes unfavorable, the FL is allowed to circulate weakly, leading to secondary thrombosis. An additional procedure could be performed in the case of a persistent increase in aortic diameter control imaging.

Study endpoints

The primary objective of this study was to evaluate the efficacy of CFLO in the aortic remodeling at the thoracic level. The primary

outcome was defined by positive aortic remodeling¹⁶ described as a reduction in or the stability of the total aortic maximal diameter on the follow-up CT scan at the 1-year follow-up along with a reduction in the FL diameter and/or expansion of the TL. An increase in the aortic maximal diameter at the 1-year follow-up was considered to represent worsening aortic remodeling. In the follow-up CT scan, measurement of the descending aortic diameter was performed at the same maximum aortic level as at baseline. An aortic diameter decrease was defined as a decrease >2 mm on the follow-up CT scan. A diameter increase was defined as an increase >2 mm. Stability was defined as a constant diameter without any increase or decrease. The maximum aortic diameter, the diameter of the FL, and the diameter of the TL are the main parameters monitored in patients with chronic dissection, regardless of whether the FL is perfused or not.

Secondary objectives were to evaluate the efficacy of CFLO on the circulation of the FL (technical success), the safety of the pro-

Table 1. Demographics and aortic pathologic data

Variables	n = 20
Age (years), median (IQR)	65.4 (58.4–69.9)
Gender (male), n (%)	13 (65.0)
Hypertension, n (%)	12 (60.0)
Type of dissection*, n (%)	
A	16 (80.0)
B	4 (20.0)
Type of initial treatment before CFLO, n (%)	
Bentall	6 (30.0)
Tirone David	1 (5.0)
Aorto-aortic tube	9 (45.0)
TEVAR	7 (35.0)
Surgical SAT reimplantation	4 (20.0)
Aortic aneurysm level, n (%)	
Zone 3	12 (60.0)
Zone 4	8 (40.0)
Location of ET, n (% of total ET)	
SAT	8/32 (25.0)
Distal surgical anastomosis	8/32 (25.0)
Zone 1	2/32 (6.3)
Zone 2	3/32 (9.4)
Zone 3	5/32 (15.6)
Zone 4	6/32 (18.8)
Indications for CFLO, n (%)	
Punctiform entry tear	16 (80.0)
Retrograde false lumen flow	4 (20.0)

*Stanford classification; n, number; IQR, interquartile range; CFLO, circulating false lumen occlusion; TEVAR, thoracic endovascular aortic repair; SAT, supra-aortic trunk; ET: entry tear.

cedure, and the occurrence of aortic events during follow-up. The technical success of CFLO was defined as complete FL thrombosis at the thoracic level, assessed by the absence of opacification of the FL on the follow-up CT scan (delayed contrast enhanced ≥ 80 s) at 3 months, in the two areas described by the Ishimaru classification on the descending aorta: zones 3 and 4. The FL was evaluated as partially thrombosed (neither patent nor thrombosed, defined as a clot within the FL with a residual circulating flow channel) or patent (similar to the baseline CT scan). The safety of CFLO was assessed by monitoring the occurrence of complications during follow-up, with data collected according to the European Society of Cardiovascular and Interventional Radiology.¹⁹

Statistical analysis

Standard descriptive statistics were used for continuous quantitative variables, presented as medians with interquartile ranges (IQRs), and for qualitative variables, presented as numbers and percentages. Regarding aortic measurements, a paired-sample mean comparison analysis was performed using Student's t-test. Standard descriptive statistics were used to describe aortic remodeling, technical success, and safety, with a calculation of the corresponding number and percentages. Regarding intraobserver reproducibility, the intraclass correlation coefficient was analyzed using the Pearson correlation test for quantitative variables. The analyses were conducted using Jamovi® (version 1.6.23.0). A *P* value < 0.05 was considered significant.

Results

Population characteristics

During the study period, 62 patients underwent procedures for aneurysmal pathology of the thoracic descending aorta (Figure 1). Twenty-five patients (40.3%) were excluded due to isolated TEVAR treatment and three patients (4.8%) for the embolization of perigraft leaks related to TEVAR. Thirteen patients (21%) were excluded for having less than 12 months of follow-up, and one patient (1.6%) was excluded due to the lack of a baseline CT scan. Twenty patients (32.3%) were finally included and underwent CFLO to prevent pejorative aortic remodeling.

The study included 7 women (35%) and 13 men (65%), with a median age of 65.4 years (IQR: 58.4–69.9) (Table 1). Sixteen patients (80%) had type A dissection, and four patients (20%) had type B dissection. Additionally, seven patients (35%) had a history of

previous TEVAR. The median duration from AD to CFLO was 32.5 months (IQR: 8.8–76.5), with the majority of aortic aneurysms located in zone 3 (60%), the isthmic segment. A total of 32 ETs were identified in 20 patients, including 8 in the SAT (25%) and 8 at the distal surgical anastomosis level (25%).

The median total number of CFLO sessions was approximately 1.5 (IQR: 1.0–2.3). Ten patients (50%) underwent two or more CFLO sessions, and ten patients (50%) received isolated CFLO without prior or additional TEVAR. The embolization agents used are summarized in Table 2.

Efficacy

False lumen circulatory status

At the 1-year follow-up, 12 patients (60%) achieved technical success with complete FL thrombosis after CFLO (Table 3), whereas 8 patients (40%) had partial thrombosis. No patients had a patent FL on the follow-up CT scan. Complete thrombosis was achieved in

the isthmic segment in zone 3 in 17 patients (85%). Thirteen patients (65%) had thrombosis in zone 4, including 12 patients who also had thrombosis in zone 3. One patient had thrombosis only in the distal third of the thoracic aorta without proximal thrombosis.

Impact on aortic remodeling

After the procedure, 13 patients (65%) exhibited positive aortic remodeling at the 1-year follow-up imaging, including 8 patients with decreased aneurysmal aortic diameter and 5 patients with stabilized diameters (Table 3). In the subgroup with complete FL occlusion, 75% exhibited positive remodeling. The median maximal aortic diameter on the baseline CT scan was 52.5 mm (IQR: 43.8–59.8) compared with 54.0 mm (IQR: 41.8–68.5) on the follow-up CT scan (*P* = 0.115). The median FL diameter significantly decreased at 1 year compared with the baseline CT [16.0 (IQR: 10.3–26.8) vs. 25.0 mm (IQR: 19.5–33.0); *P* = 0.044]. Among the 10 patients with isolated CFLO, 6 achieved

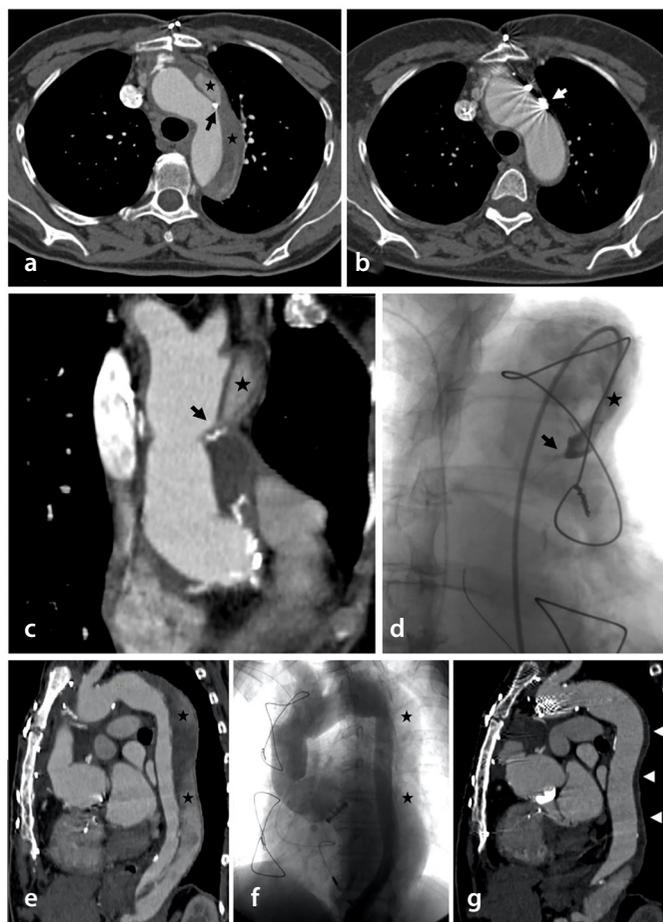


Figure 2. Treatment of a thoracic chronic dissecting aortic aneurysm after Bentall surgery in a 73-year-old patient with a history of chronic aortic dissection by coiling of the proximal portal of entry. The target entry tear (ET, black arrows) is located within segment 2, feeding the false lumen (FL, black stars). The coils (white arrows) indirectly occlude the ET, allowing thrombosis and regression of the FL (white arrowheads). Computed tomography scan with injection at arterial time before (a, c, e) and after (b, g) embolization. Serigraphy with injection and locating the ET by retrograde catheterization of the FL (d) and the true lumen (f).

FL thrombosis and 9 (90%) achieved positive aortic remodeling. The intraclass correlation coefficient for the intraobserver reproducibility of the aortic diameter measurements was 0.997 (95% confidence interval: 0.995, 0.998).

Safety and aortic events

The median follow-up was 4.7 years (IQR: 2.9–12.4). After CFLO, five patients (25%) experienced transient post-embolization syndrome, characterized by increased pain (lasting approximately 10 days) and a temporary increase in C-reactive protein levels, related to large FL thrombosis; these symptoms were resolved within 3 months (grade I). Two patients (10%) had complications associated with CFLO: one transient ischemic stroke (grade II) and one coil migration that could be recaptured (grade I). Four patients (20%) required additional TEVAR during follow-up. Among the seven patients with a persistent patent FL, two had secondary aortic events related to aneurysmal evolution (Table 4), but no patients with complete FL occlusion experienced aortic events during follow-up. No cases of spinal cord ischemia were reported in connection with the CFLO procedure.

Discussion

This study demonstrated satisfactory CFLO results in inducing positive aortic remodeling in chronic aneurysmal descending AD in 13 out of 20 patients (65%), with a decreased aneurysmal aortic diameter observed in 40% of patients. Additionally, for 25% of patients, the diameter remained stable, associated with either a decrease in FL size or an increase in TL size. Moreover, 60% of patients had technical success with complete FL thrombosis after CFLO. These results are encouraging when considering that the blood flow circulation in the FL is the key risk factor for aortic enlargement, associated with a significant and significantly higher aortic mean growth rate, as described by Sueyoshi et al.¹⁸ The natural evolution of a patent FL is not toward thrombosis, and the authors demonstrated that only one third of FL cases progressed to complete thrombosis without the need for interventional treatment. In cases where FL patency persists, abdominal-level embolization may be appropriate but should be considered with caution, as this further complicates treatment given that the thrombus may extend to the mesenteric arteries in cases of extensive dissection.

Direct FL occlusion was first described by Loubert et al.²⁰ as the “cork in the bottle neck”

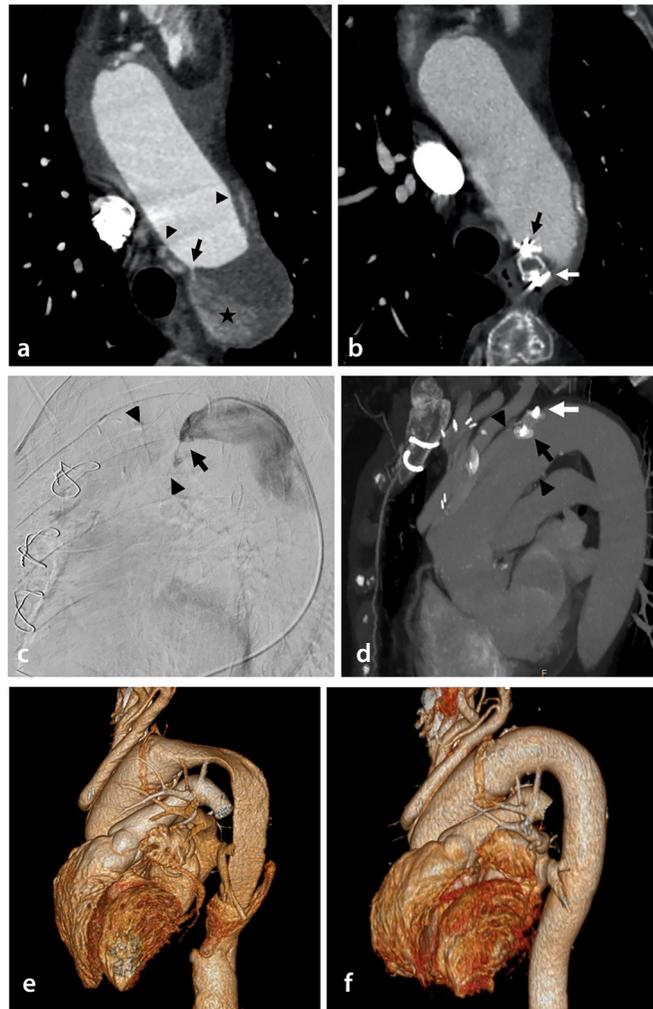


Figure 3. Plug treatment of chronic aortic dissection with aneurysmal progression at the thoracic level in a patient after Bentall surgery. The entry tear (ET, black arrow) is located at the distal anastomosis (arrowheads) of the ascending aortic surgical prosthesis. A type II plug (white arrow) is deployed through the ET. The proximal wing is in the true lumen (TL), and the body and distal wing are in the false lumen (FL). Axial computed tomography scan with injection at arterial time before (a) and after (b) embolization. Serigraphy with injection (c) identifying the ET. Maximum intensity projection reconstruction scans (d) after embolization. Three-dimensional reconstruction at arterial time before (e) and after (f) embolization. After treatment: regression of the FL, re-expansion of the TL, and decrease in the aortic diameter.

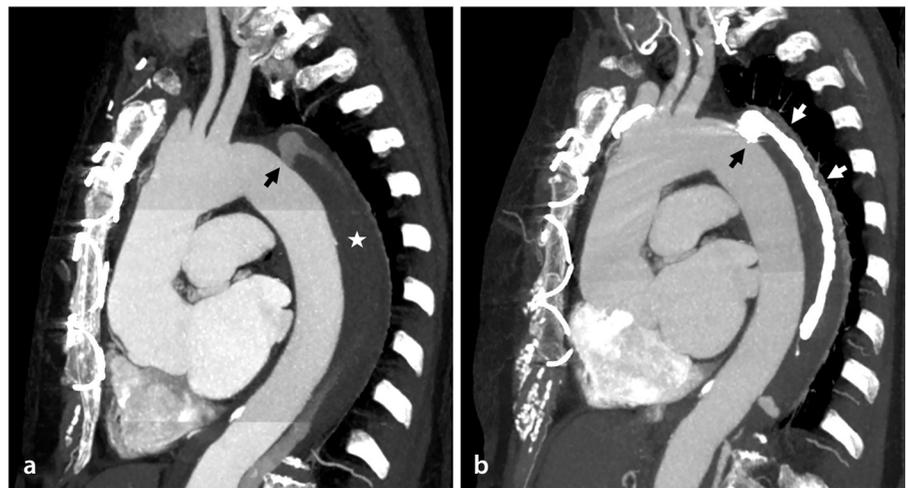


Figure 4. Treatment of a chronic aneurysmal dissection at the thoracic level using glue. Sagittal maximum intensity projection reconstruction scans before (a) and after (b) glue embolization (white arrows). Occlusion of the entry tear (black arrow) allows thrombosis of the false lumen (white star).

strategy. A few studies have been conducted on embolization agents using coils and cyanoacrylate glue^{21,22} or plugs;^{23,24} these have been concomitant or with prior TEVAR but in small cohorts. Even if indications of CFLO differ with TEVAR, aortic diameter control has been reported in 65% of patients managed with isolated TEVAR.²⁵ In the present study, the results suggest that CFLO may improve TEVAR results and that this technique can be an effective option when TEVAR cannot be performed because of anatomical contraindications. The retrograde expansion of dissection often requires extended methods, such as a covering stent,²⁶ to achieve sufficient sealing and avoid inadvertently covering the SAT with the stent graft.

This study did not reveal a significant reduction in aortic diameter ($P = 0.115$). Although the optimal timing for the embolization procedure is unclear, it seems more effective in the subacute period (14 to 90 days),²⁷ as the aorta is less likely to reshape in the chronic phase.³ In this study, the average time from AD to CFLO was 33 months, indicating that earlier intervention might have led to improved outcomes. Additionally, assessing aortic diameters at 1 year may underestimate the long-term effects of this condition. Thrombus reduction in the FL may take time, emphasizing the need for a longer follow-up to observe these effects. In cases with a large thrombotic FL, the thrombus does not decrease in size, which means there is no overall reduction in aortic diameter.

Of course, the retrospective character of this study leads to missing data, and the monocentric design is a limitation, but this study was conducted in a single specialized university hospital, reducing the risk of patients being treated in another close geographical center. The number of patients included allowed inter-individual variability and was sufficient for this type of intervention. Due to the rarity of these cases, our study offers an additional contribution to the literature, which could be valuable for future guidance on the subject. The absence of a control group is a limitation but ethically justifiable, as the natural course of the pathology is well understood. The association of CFLO and TEVAR techniques is common and, unfortunately, limits the demonstration of the effectiveness of the isolated CFLO technique. Nevertheless, the isolated CFLO technique was used with approximately 50% of our study population and produced satisfying results, achieving 60% complete thrombosis and 90% positive aortic remodeling, demonstrating its effectiveness when per-

formed alone. In the literature, CFLO achieves complete FL thrombosis in 60% of patients compared with 40% for isolated TEVAR at the

2-year follow-up²⁸ and 60%–80% for combined CFLO and TEVAR in a review conducted by Spanos et al. involving 101 patients.²⁹

Table 2. Angiographic data

Variables	n = 20
Duration from AD (months), median (IQR)	32.5 (8.8–76.5)
Procedure duration (min), median (IQR)	124.0 (78.5–158.5)
Scopy duration (min), median (IQR)	17.4 (12.1–24.3)
Dose (Gy.cm ²), median (IQR)	175.2 (59.1–319.2)
CFLO single technique, n (%)	7 (35.0)
CFLO combined technique, n (%)	13 (65.0)
Occlusion material used*, n (%)	
Coil	5 (25.0)
Coil + glue	2 (10.0)
Coil + plug	3 (15.0)
Coil + SAT stenting	2 (10.0)
Glue + plug	2 (10.0)
Plug + SAT stenting	2 (10.0)
Plug	2 (10.0)
Coil + glue + plug	2 (10.0)
Total number of CFLO sessions, median (IQR)	1.5 (1.0–2.3)
Patients with ≥2 CFLO sessions, n (%)	10 (50.0)
Isolated CFLO**	10 (50.0)
Additional TEVAR treatment after CFLO, n (%)	4 (20.0)
Interval to CFLO after TEVAR (months), median (IQR)	6.6 (1.5–16.7)
Interval to TEVAR after CFLO (months), median (IQR)	10.2 (4.9–16.9)

*Considering all CFLO sessions; **isolated CFLO = CFLO without prior or additional TEVAR; n, number; AD, aortic dissection; IQR, interquartile range; CFLO, circulating false lumen occlusion; SAT, supra-aortic trunk; TEVAR, thoracic endovascular aortic repair.

Table 3. Imaging follow-up

Variables	n = 20
Duration from baseline CT scan to CFLO (days), median (IQR)	51.0 (110.8–10.3)
Patient follow-up (years), median (IQR)	4.7 (2.9–12.4)
Follow-up aortic diameter evolution, n (%)	
Decrease	8 (40.0)
Stable	5 (25.0)
Increase	7 (35.0)
Baseline CT scan aortic diameter (mm), median (IQR)	52.5 (43.8–59.8)
1-year CT scan aortic diameter (mm), median (IQR)	54.0 (41.8–68.5)
Baseline CT scan FL diameter (mm), median (IQR)	25.0 (19.5–33.0)
1-year CT scan FL diameter (mm), median (IQR)	16.0 (10.3–26.8)
1-year CT scan FL circulatory status, n (%)	
Thrombosed	12 (60.0)
Partially thrombosed	8 (40.0)
1-year CT scan frequencies of complete thrombosis, n (%)	
Zone 3	17 (85.0)
Zone 4	13 (65.0)

n, number; CT, computed tomography; CFLO, circulating false lumen occlusion; IQR, interquartile range; FL, false lumen.

Table 4. Symptoms and complications reported after CFLO	
Variables	n = 20
Complications after CFLO, n (%)	
Grade I, n (%)	
Pre-embolization coil migration	1 (5.0)
Post-embolization syndrome	5 (25.0)
Grade II, n (%)	
Transient ischemic stroke	1 (5.0)
Secondary aneurysm event*, n (%)	
Aneurysm chronic infection	1 (5.0)
Aneurysm infection with aorto-esophageal fistula	1 (5.0)

*Not directly related to CFLO; n, number; according to the European Society of Cardiovascular and Interventional Radiology classification system for complications; CFLO, circulating false lumen occlusion; 1 patient contracted an aortic aneurysm infection complicated by an aortic fissure and esophago-aortic fistula; 16 patients did not present any complications or secondary aneurysm events.

(Cardiology Department, Grenoble Alpes University Hospital) for their invaluable contributions and expert insights into this pathology. Their participation and feedback have been instrumental in the development and refinement of our study.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This study was funded by a research scholarship from the Missions d'Enseignement, de Recherche, de Référence et d'Innovation for manuscript writing.

References

1. Pape LA, Awais M, Woznicki EM, et al. Presentation, diagnosis, and outcomes of acute aortic dissection. *J Am Coll Cardiol*. 2015;66(4):350-358. [\[CrossRef\]](#)
2. Grabenwoger M, Fleck T, Czerny M, et al. Endovascular stent graft placement in patients with acute thoracic aortic syndromes. *Eur J Cardiothorac Surg*. 2003;23(5):788-793. [\[CrossRef\]](#)
3. Sayer D, Bratby M, Brooks M, Loftus I, Morgan R, Thompson M. Aortic morphology following endovascular repair of acute and chronic type B aortic dissection: implications for management. *Eur J Vasc Endovasc Surg*. 2008;36(5):522-529. [\[CrossRef\]](#)
4. Li D, Ye L, He Y, Cao X, Liu J, Zhong W, et al. False lumen status in patients with acute aortic dissection: a systematic review and meta-analysis. *J Am Heart Assoc*. 2016;5(5):3172. [\[CrossRef\]](#)
5. Marui A, Mochizuki T, Koyama T, Mitsui N. Degree of fusiform dilatation of the proximal descending aorta in type B acute aortic dissection can predict late aortic events. *J Thorac Cardiovasc Surg*. 2007;134(5):1163-1170. [\[CrossRef\]](#)
6. Tsai MT, Wu HY, Roan JN, et al. Effect of false lumen partial thrombosis on repaired acute type A aortic dissection. *J Thorac Cardiovasc Surg*. 2014;148(5):2140-2146. [\[CrossRef\]](#)
7. Alimohammadi M, Sherwood JM, Karimpour M, Agu O, Balabani S, Díaz-Zuccarini V. Aortic dissection simulation models for clinical support: fluid-structure interaction vs. rigid wall models. *BioMed Eng OnLine*. 2015;14(1):34. [\[CrossRef\]](#)
8. Oda T, Minatoya K, Sasaki H, et al. Surgical indication for chronic aortic dissection in descending thoracic and thoracoabdominal aorta. *Circ Cardiovasc Interv*. 2017;10(2):4292. [\[CrossRef\]](#)

The anticipated consequences and complications were minimal compared with the underlying pathology. Therefore, CFLO can be considered relatively safe considering the absence of complications in 90% of patients, whereas TEVAR has a 30-day morbidity rate of 36.5%.²⁹ In comparison to TEVAR, no cases of spinal cord ischemia or associated complications have been reported with the use of CFLO, which could be a decisive advantage when choosing between the two techniques.

As mentioned previously, the precise description of the conditions for the use of embolic materials is challenging, and many parameters must be considered in the selection of the embolization material to be used. The small sample size of patients and the study design unfortunately did not allow for additional statistical tests to identify favorable or unfavorable predictive factors that influence the effectiveness of embolization in achieving either complete or partial FL thrombosis. One factor that appears to contribute to the failure of remodeling is the presence of calcifications on the wall of the FL. However, future efforts should focus on identifying these predictive factors, and further studies are needed to support and refine the indications for CFLO.

The CFLO is a technically complicated procedure with multiple parameters to be managed. The success of this treatment depends on several factors, including the experience of the operator, and must be performed in close collaboration with medical-surgical teams. A high level of understanding of aortic pathology and fluid mechanics is required. These results promote the use of CFLO techniques in the future, as they are currently not included in the guidelines.

This study highlights the potential of CFLO as a valuable addition to the therapeutic arsenal for chronic AD, particularly when traditional methods such as TEVAR do not achieve complete FL thrombosis. This technique offers a promising alternative for patients who do not respond fully to existing treatments and could reduce the risk of aneurysmal progression and aortic rupture, ultimately improving long-term patient outcomes. By targeting the ET and using various embolization materials (coils, glue, and plugs) or uncovered stents, CFLO allows for a tailored treatment approach that can be adapted to individual patient anatomy and pathology, enhancing treatment efficacy and safety. The study suggests that CFLO may reduce the incidence of complications related to persistent FL circulation, such as aneurysmal progression, thereby potentially decreasing the need for more invasive procedures. No cases of spinal cord ischemia were reported with this procedure, which may provide a significant advantage when comparing CFLO to TEVAR. It is important to emphasize that this management should be multidisciplinary and performed in close collaboration with cardiac surgery and cardiology teams.

In conclusion, this study supports the efficacy and safety of endovascular occlusion in the management of chronic AD. Further studies are needed to define the exact place and most opportune timing of this procedure in the management algorithm for patients with aneurysmal evolution following AD.

Acknowledgments

We sincerely thank Pr. Chavanon O. (Cardiac Surgery Department, Grenoble Alpes University Hospital) and Pr. Ormezzano O.

9. Estrera AL, Miller CC 3rd, Safi HJ, et al. Outcomes of medical management of acute type B aortic dissection. *Circulation*. 2006;114:384-389. [\[CrossRef\]](#)
10. Roselli EE, Abdel-Halim M, Johnston DR, et al. Open aortic repair after prior thoracic endovascular aortic repair. *Ann Thorac Surg*. 2014;97(3):750-756. [\[CrossRef\]](#)
11. Upchurch GR, Escobar GA, Azizzadeh A, et al. Society for vascular surgery clinical practice guidelines of thoracic endovascular aortic repair for descending thoracic aortic aneurysms. *J Vasc Surg*. 2021;73(1):55-83. [\[CrossRef\]](#)
12. Larsen M, Bartnes K, Tsai TT, et al. Extent of preoperative false lumen thrombosis does not influence long-term survival in patients with acute type A aortic dissection. *J Am Heart Assoc*. 2013;2(4):112. [\[CrossRef\]](#)
13. Brunkwall J, Kasprzak P, Verhoeven E, et al. Endovascular repair of acute uncomplicated aortic type B dissection promotes aortic remodelling: 1 year results of the ADSORB trial. *Eur J Vasc Endovasc Surg*. 2014;48(3):285-291. [\[CrossRef\]](#)
14. Manetta F, Newman J, Mattia A. Indications for thoracic endovascular aortic repair (TEVAR): a brief review. *Int J Angiol*. 2018;27(4):177-184. [\[CrossRef\]](#)
15. Spratt JR, Walker KL, Neal D, et al. Rescue therapy for symptomatic spinal cord ischemia after thoracic endovascular aortic repair. *J Thorac Cardiovasc Surg*. 2024;168(1):15-25. [\[CrossRef\]](#)
16. Lombardi JV, Hughes GC, Appoo JJ, et al. Society for vascular surgery (SVS) and society of thoracic surgeons (STS) reporting standards for type B aortic dissections. *J Vasc Surg*. 2020;71(3):723-747. [\[CrossRef\]](#)
17. Mitchell RS, Ishimaru S, Ehrlich MP, et al. First international summit on thoracic aortic endografting: roundtable on thoracic aortic dissection as an indication for endografting. *J Endovasc Ther*. 2002;9(Suppl 2):98-105. [\[CrossRef\]](#)
18. Sueyoshi E, Sakamoto I, Hayashi K, Yamaguchi T, Imada T. Growth rate of aortic diameter in patients with type B aortic dissection during the chronic phase. *Circulation*. 2004;110(11 Suppl 1):256-261. [\[CrossRef\]](#)
19. Filippiadis DK, Binkert C, Pellerin O, Hoffmann RT, Krajina A, Pereira PL. Cirse quality assurance document and standards for classification of complications: the cirse classification system. *Cardiovasc Intervent Radiol*. 2017;40(8):1141-1146. [\[CrossRef\]](#)
20. Loubert MC, van der Hulst VP, De Vries C, Bloemendaal K, Vahl AC. How to exclude the dilated false lumen in patients after a type B aortic dissection? The cork in the bottleneck. *J Endovasc Ther*. 2003;10(2):244-248. [\[CrossRef\]](#)
21. Hofferberth SC, Nixon IK, Mossop PJ. Aortic false lumen thrombosis induction by embolotherapy (AFTER) following endovascular repair of aortic dissection. *J Endovasc Ther*. 2012;19(4):538-545. [\[CrossRef\]](#)
22. Zhang H, Ge YY, Lu K, et al. Coil embolization for persistent thoracic false lumen of type B aortic dissection after thoracic endovascular aortic repair. *Ann Vasc Surg*. 2019;57:60-68. [\[CrossRef\]](#)
23. Pellenc Q, Roussel A, De Blic R, et al. False lumen embolization in chronic aortic dissection promotes thoracic aortic remodeling at midterm follow-up. *J Vasc Surg*. 2019;70(3):710-717. [\[CrossRef\]](#)
24. Miletic KG, Kindzelski BA, Hodges KE, et al. Impact of endovascular false lumen embolization on thoracic aortic remodeling in chronic dissection. *Ann Thorac Surg*. 2021;111(2):495-501. [\[CrossRef\]](#)
25. Scali ST, Feezor RJ, Chang CK, et al. Efficacy of thoracic endovascular stent repair for chronic type B aortic dissection with aneurysmal degeneration. *J Vasc Surg*. 2013;58(1):10-17. [\[CrossRef\]](#)
26. Rohlfs F, Tsilimparis N, Diener H, et al. Chronic type B aortic dissection: indications and strategies for treatment. *J Cardiovasc Surg (Torino)*. 2015;56(2):231-238. [\[CrossRef\]](#)
27. Mani K, Clough RE, Lyons OT, et al. Predictors of outcome after endovascular repair for chronic type B dissection. *Eur J Vasc Endovasc Surg*. 2012;43(4):386-391. [\[CrossRef\]](#)
28. Guangqi C, Xiaoxi L, Wei C, et al. Endovascular repair of stanford type B aortic dissection: early and mid-term outcomes of 121 cases. *Eur J Vasc Endovasc Surg*. 2009;38(4):422-426. [\[CrossRef\]](#)
29. Spanos K, Kölbel T, Rohlfs F, et al. Intentional targeted false lumen occlusion after aortic dissection: a systematic review of the literature. *Ann Vasc Surg*. 2019;56:317-329. [\[CrossRef\]](#)
30. Georg Y, Schwein A, Lejay A, et al. Practical experience with the TAG and conformable TAG devices- lessons learned in about 100 cases. *J Cardiovasc Surg (Torino)*. 2013;54(5):605-615. [\[CrossRef\]](#)



Amide proton transfer-weighted magnetic resonance imaging for the evaluation of testicular spermatogenic function: a preliminary study

Guanglei Tang^{1*}

Shulin Ma^{1*}

Wenhao Fu¹

Weijian Yun¹

Yang Peng¹

Jian Guan^{1,2}

¹The First Affiliated Hospital, Sun Yat-Sen University, Department of Radiology, Guangdong, China

²Linzhi People's Hospital, Department of Radiology, Tibet, China

*Joint first authors

Corresponding authors: Jian Guan, Yang Peng

E-mails: guanj6@mail.sysu.edu.cn, pengy63@mail.sysu.edu.cn

Received 23 January 2025; revision requested 21 February 2025; last revision received 17 March 2025; accepted 23 March 2025.



Epub: 20.05.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2025.253248

PURPOSE

To determine the amide proton transfer-weighted (APT_w) imaging features in testes with age, and to assess the feasibility of APT_w magnetic resonance imaging (MRI) in assessing testicular spermatogenic function.

METHODS

A total of 23 male patients with clinically confirmed hypospermatogenesis caused by epididymo-orchitis were included in the case group (group A) and another 93 men (age range, 20–80 years) were included in the control group. The control group was divided into four subgroups: group B1 (20–34 years, n = 25), group B2 (35–49 years, n = 23), group B3 (50–64 years, n = 21), and group B4 (65–80 years, n = 24). All participants underwent 3.0T MRI scan, and the APT signal intensity (SI) and apparent diffusion coefficient (ADC) value of each testis were examined. The ADC and APT SI were independently measured by two radiologists blinded to clinical data, and average values were calculated. A power analysis was conducted to determine the required sample size.

RESULTS

APT SI was negatively correlated with age ($r = -0.510$, $P < 0.001$), whereas ADC was positively correlated with age ($r = 0.317$, $P = 0.006$). The APT SI was significantly higher in group A (1.77 ± 0.41) than in group B1 (1.43 ± 0.21), group B2 (1.37 ± 0.31), group B3 (1.30 ± 0.35), and group B4 (1.20 ± 0.35) (all $P < 0.01$). The ADC was significantly higher in group A [$(0.549 \pm 0.091) \times 10^{-3} \text{ mm}^2/\text{s}$] compared with group B1 [$(0.449 \pm 0.047) \times 10^{-3} \text{ mm}^2/\text{s}$], group B2 [$(0.475 \pm 0.022) \times 10^{-3} \text{ mm}^2/\text{s}$], and group B3 [$(0.488 \pm 0.051) \times 10^{-3} \text{ mm}^2/\text{s}$] (all $P < 0.01$), whereas no statistically significant difference was found between group A and group B4 ($P > 0.05$).

CONCLUSION

The APT SI of the normal testes decreased with age, whereas a significant elevation of APT SI was detected in patients with hypospermatogenesis caused by epididymo-orchitis.

CLINICAL SIGNIFICANCE

Hypospermatogenesis caused by degeneration or inflammation can be differentiated by APT quantity combined with ADC value.

KEYWORDS

Testis, oligospermia, orchitis, MRI, amide proton transfer

Orchitis and epididymo-orchitis are important causes of male infertility. A prolonged course of bilateral orchitis may impair spermatogenesis and lead to non-obstructive azoospermia (NOA).¹⁻³ The spermatogenic function of normal testes decreases with increasing age.⁴⁻⁶ Magnetic resonance imaging (MRI) allows for non-invasive assessment of testicular lesions and provides adequate anatomic information, satisfactory tissue contrast, and functional information.^{7,8} Functional MRI methods, including dynamic contrast-enhanced MRI,^{9,10} diffusion-weighted imaging (DWI),¹⁰⁻¹⁵ magnetization transfer imaging (MTI),¹⁶⁻¹⁸ and

MR spectroscopy (MRS),¹⁹⁻²¹ have recently provided useful additional diagnostic data for determining normal and abnormal testes.

Amide proton transfer-weighted (APT_w) imaging, a specific type of chemical exchange saturation transfer MRI, has been introduced as a novel endogenous contrast modality for MRI by detecting low concentration solutes, including mobile proteins and peptides in tissue or tumor samples with abundant amide chemical constituents.^{22,23} APT has been examined as an imaging biomarker in a variety of cancers and non-oncological diseases.²⁴⁻²⁹ Currently, a limited number of studies have evaluated APT in the assessment of human testicular metabolic profile. Whether APT_w imaging can be used to evaluate the spermatogenic potential of the testis remains unknown. The purpose of this study is to examine the APT_w imaging features of the testis with age, and to examine the feasibility of APT_w MRI in assessing testicular spermatogenic function.

Methods

Study population

The current study was approved by the Institutional Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (protocol number: 537/2022, date: 2022/11/25), and signed informed consent was obtained from all participants. The study initially included 28 consecutive men with hypospermatogenesis caused by epididymo-orchitis (age range, 22–57 years; mean age, 35 years) who underwent scrotal MRI with DWI and APT_w between June 2020 and June 2023. The inclusion criteria were as follows: (1) diagnosis of hypospermatogenesis according to the 2021 WHO guidelines³⁰ (semen analysis showing azoospermia or sperm concentration below the lower reference limit of 15 million sperm/mL of ejaculate after centrifugation in at least two tests); (2) clinical diag-

nosis of epididymo-orchitis (diagnostic criteria defined by The 2016 European guideline on the management of epididymo-orchitis³¹) based on medical history, physical examination, and laboratory tests; and (3) seminal plasma biochemistry showing elevated polymorphonuclear elastase concentration (>250 ng/mL), indicating active testicular injury by inflammation.³²⁻³⁴ The exclusion criteria included (1) poor image quality for ADC or APT_w map; (2) other testicular diseases, including testicular tumors; and (3) a history of testicular injury or surgery.

During the same period, another 111 healthy males were referred for scrotal MRI as controls. In the control group, no abnormal findings of the scrotum were found by physical examination, as well as no traumatic history of the scrotum. Volunteers under 50 years old underwent seminal plasma biochemistry and showed normal results, and those over 50 years old had genetically related children with natural insemination. The exclusion criteria were the same as described for the case group.

Among the 139 men enrolled, 15 were excluded because of APT and ADC images of insufficient quality, including motion artifacts (n = 5), poor image quality caused by B0 field inhomogeneity (n = 6), and a combination of motion artifacts and B0 field inhomogeneity (n = 4). In addition, 8 were excluded because of complications with other testis diseases or with a history of testicular injury or surgery. Among the 23 excluded cases, 5 were in the case group and 18 were controls. Thus, a total of 116 men were analyzed, including 23 men (age range, 22–57 years; mean age, 36.4 years) with hypospermatogenesis caused by epididymo-orchitis (case group A) and 93 male volunteers (control group B; age range,

20–80 years). The control group was further divided into four subgroups by age: group B1 (20–34 years, n = 25), group B2 (35–49 years, n = 23), group B3 (50–64 years, n = 21), and group B4 (65–80 years, n = 24). The study flowchart is presented in Figure 1.

Magnetic resonance examinations

All participants underwent non-enhanced MRI, DWI, and APT_w examinations in the supine position after urination using a 3.0T MR scanner (Ingenia CX, Philips Healthcare, Best, The Netherlands) with a 32-channel phased-array torso coil. Axial and coronal fast spin-echo T2-weighted imaging [repetition time/echo time (TR/TE): 2,500/65 ms] and axial spin-echo T1WI (TR/TE: 600/20 ms) images were obtained, with a small towel placed between the thighs to stabilize the testes and the penis taped to the anterior abdominal wall. An additional axial fat saturated T1W sequence was acquired when T1 hyperintense foci were detected in testes.

DWI was performed with free-breathing, with spectral attenuated inversion recovery axial single-shot spin-echo echo-planar imaging (TR/TE, 4,500/107 ms; 2 NSA) with b values of 0, 900, and 4,000 s/mm², a slice thickness of 3 mm, an intersection gap of 0.6 mm, a field of view (FOV) of 120*71 mm², a sampling resolution of 1.8*1.8*3 mm³, and a total scan time of 4 minutes and 55 seconds for each DWI scan. Moreover, APT_w imaging was performed with a three-dimensional TSE-mDixon sequence. Parameters for the APT_w sequence were as follows: saturation power, 2 μT; saturation duration, 2 s; frequency offsets, ± 3.5, ± 3.42, ± 3.58, -1,540 ppm; TR/TE, 5,864/8.8 ms; FOV, 250*346 mm²; sampling resolution, 1.8*1.8*4 mm³; slice thickness, 4 mm; sensitivity encoding factor,

Main points

- Amide proton transfer (APT) signal intensity (SI) of normal testes tended to decrease with increasing age, and apparent diffusion coefficient (ADC) of normal testes was positively correlated with age.
- Patients with hypospermatogenesis caused by epididymo-orchitis exhibited much higher APT SI compared with all age groups of controls.
- An increased APT SI combined with an elevated ADC is more likely to be associated with active or persistent inflammation.

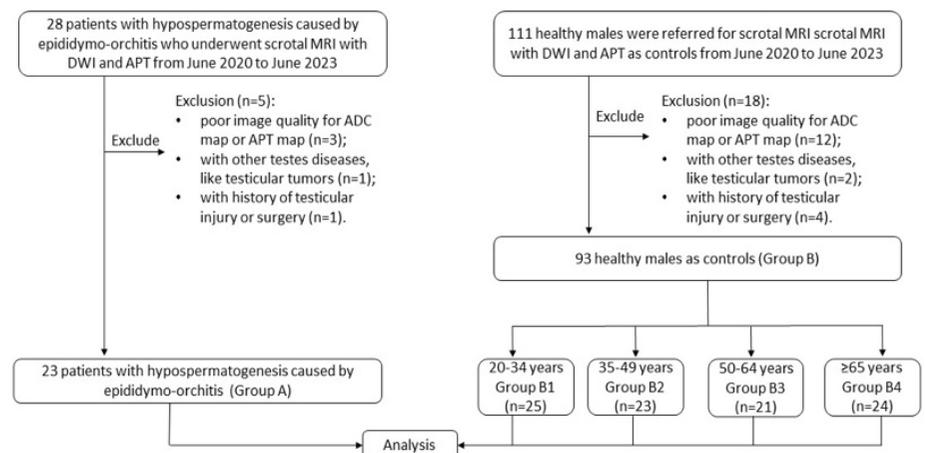


Figure 1. Study flowchart showing the patient inclusion and exclusion criteria. DWI, diffusion-weighted imaging; APT, amide proton transfer; MRI, magnetic resonance imaging; ADC, apparent diffusion coefficient.

2; and total scan time, 4 minutes and 59 seconds for each APTw scan.

Image analysis

The ADC and APT SI were independently measured by two radiologists (P.Y. and J.G., with 6 and 20 years of experience, respectively) blinded to clinical data, and average values were used. Circular regions of interest (ROIs) were placed centrally (with the center at the intersection of the long and short axes) on both the automatically generated ADC maps and the APTw maps to encompass most of the testes without artifacts or margins, with an area of no less than 110 mm². The ROIs are shown in Figure 2. Each testis was measured three times to determine the average ADC and APT SI.

Statistical analysis

Statistical analysis was performed using SPSS version 22 (IBM, Armonk, NY, USA). Data normality was assessed using the Shapiro-Wilk test. Continuous variables were expressed as mean ± standard deviation and categorical variables in terms of count. The ADC and APT SI values of the patients and the controls were compared using the independent samples t-test. Spearman's rank correlation coefficient was used to analyze correlations between age and APT/ADC.

Interobserver reproducibility was evaluated for ADC and APT SI measurements using the intraclass correlation coefficient (ICC). The level of agreement was considered excellent (ICC > 0.74), good (ICC = 0.60–0.74), fair (ICC = 0.40–0.59), or poor (ICC < 0.40). Statistical significance was defined as *P* < 0.05.

A priori power analysis was performed using G-Power software (version 3.1.9.7). To detect the difference in the APT SI and ADC value of testes between patients with epididymo-orchitis and the controls, a sample size of 15 was required for each group based on

these data. The sample size was considered sufficient to draw conclusions in this study.

Results

Average patient age and testicular APT SI and ADC values are listed in Table 1. In group B, Spearman's rank correlation coefficient showed that APT SI was negatively correlated with age (*r* = −0.510, *P* < 0.001) (Figure 3),

whereas ADC was positively correlated with age (*r* = 0.317, *P* = 0.006) (Figure 4). The APT SI was significantly higher in group A (1.77 ± 0.41) (Figure 3) compared with group B and all its subgroups (all *P* < 0.01). The ADC was significantly higher in group A [(0.549 ± 0.091) × 10^{−3} mm²/s] compared with group B [(0.482 ± 0.052) × 10^{−3} mm²/s, *P* < 0.001], including group B1 [(0.449 ± 0.047) × 10^{−3} mm²/s, *P* < 0.001], group B2 [(0.475 ± 0.022) ×

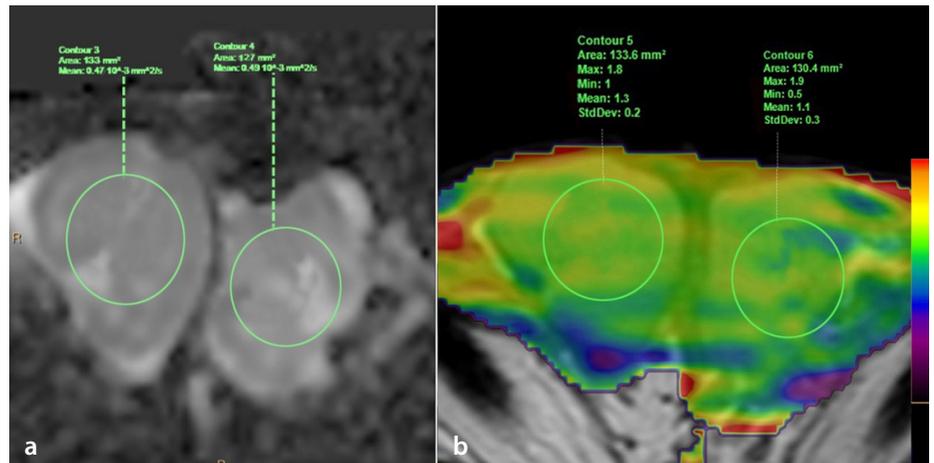


Figure 2. (a, b) Placement of regions of interest in testes.

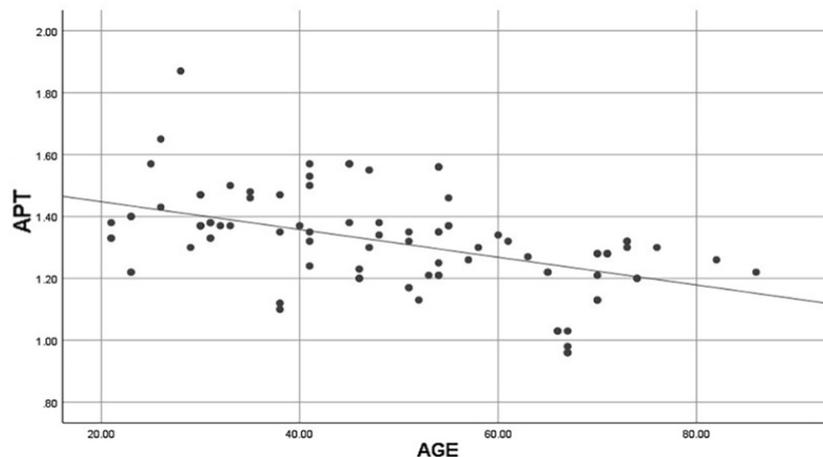


Figure 3. Scatter plot of APT signal intensity and age. APT, amide proton transfer.

Table 1. Testicular APT SI and ADC values in various groups

Group	Age	APT SI	<i>P</i> ^a	ADC (mm ² /s)	<i>P</i> ^b
Group A	36.40 ± 5.43	1.77 ± 0.29	NA	0.549 ± 0.091 × 10 ^{−3}	NA
Group B	48.97 ± 16.54	1.32 ± 0.16	<0.001	0.482 ± 0.052 × 10 ^{−3}	<0.001
Group B1	27.88 ± 3.85	1.43 ± 0.21	0.008	0.449 ± 0.047 × 10 ^{−3}	<0.001
Group B2	41.90 ± 4.15	1.37 ± 0.31	<0.001	0.475 ± 0.022 × 10 ^{−3}	0.002
Group B3	54.94 ± 3.70	1.30 ± 0.35	<0.001	0.488 ± 0.051 × 10 ^{−3}	<0.001
Group B4	71.50 ± 5.63	1.20 ± 0.35	<0.001	0.512 ± 0.089 × 10 ^{−3}	0.111

^a: Comparison with APT SI of group A. ^b: Comparison with ADC value of group A. APT SI, amide proton transfer signal intensity; ADC, apparent diffusion coefficient.

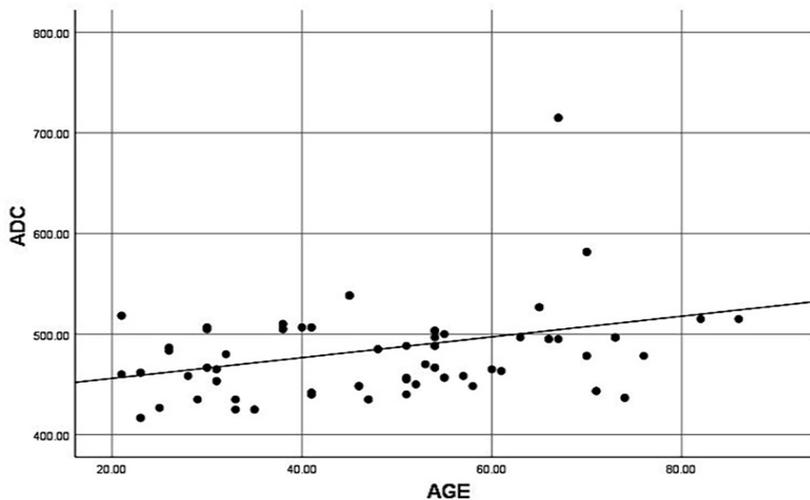


Figure 4. Scatter plot of ADC and age. ADC, apparent diffusion coefficient.

$10^{-3} \text{ mm}^2/\text{s}$, $P = 0.002$], and group B3 [$(0.488 \pm 0.051) \times 10^{-3} \text{ mm}^2/\text{s}$, $P < 0.001$], but there was no statistically significant difference between group A and group B4 [$(0.549 \pm 0.091) \times 10^{-3}$ vs. $(0.512 \pm 0.089) \times 10^{-3} \text{ mm}^2/\text{s}$, $P = 0.111$]. Examples of control groups and patients with epididymo-orchitis are shown in Figure 5.

Interobserver agreement was high for both the ADC (ICC = 0.850, 95% confidence interval (CI): 0.829–0.975) and the APT SI (ICC = 0.820, 95% CI: 0.805–0.853) measurements.

Discussion

Recently published studies have investigated the potential of functional MRI in male infertility.^{13,16,21} Proton MRS (^1H MRS), DWI, and MTI are promising for the evaluation of male infertility.^{13,16,21} Increased testicular ADC and decreased magnetization transfer ratio have been reported in patients with impaired spermatogenic function.^{13–15} Previous studies also reported that ^1H -MRS may be utilized for the evaluation of the testes of patients with NOA, in which decreased levels of choline, myoinositol, and lipids are found.^{20,21} Both ADC and MTI reflect certain histological features of testes, but neither can provide information at the levels of molecules and metabolites. Although ^1H -MRS may be valuable as a tool for quantitative evaluation of metabolites in spermatogenesis, the study of testes has adopted a common approach from brain MRS, which neglects the differences in metabolites between testes and the brain,^{20,21} meaning it is unclear whether the method can accurately reflect the actual metabolic status of the testis. Moreover,

APT_w imaging represents a non-invasive imaging technique, one used as an adjunct tool to conventional MRI that allows for the determination of metabolite concentration changes, such as mobile proteins and peptides in organs or tumors, and is utilized in studies involving brain tumors, hepatocellular carcinoma, bladder cancer, prostate cancer, and endometrioid endometrial adenocarcinoma.^{24–28} Beyond its applications in oncology, APT imaging has also been utilized in non-neoplastic conditions such as renal impairment^{35–36} and multiple sclerosis (MS).³⁷ In these studies, the potential mechanism underlying APT SI is hypothesized to be the increased concentration of mobile proteins and peptides during the progress of chronic kidney disease and within MS-associated chronic inflammatory lesions. Additionally, renal dysfunction will affect ion exchange and disrupt the original acid-base balance. This might change the pH value within tissue to a certain extent, leading to the increase of exchange rate between amide and water protons, consequently elevating the APT values. The concentration changes of a variety of proteins and peptides are involved in the production of spermatozoa by the testes,^{38,39} meaning applying APT in the evaluation of spermatogenic function of the testis is a promising approach. To the best of our knowledge, this is the first study that reports APT SI for the testis and spermatogenic function.

In this study, a negative correlation was found (coefficient: -0.510) between the APT SI of the testis and age in 94 healthy controls. In addition, the ADC value of normal testes increased with age, corroborating the find-

ings obtained by Wang et al.¹³ Aging in men is associated with both functional and structural alterations of the testis. With increasing age, testosterone levels and sperm production progressively decrease.^{5,6,40,41} Circulating testosterone levels are known to decrease by 0.4%–2% each year after the age of 30, which is due to altered Leydig cell number and function.^{4–6} In addition, the function of tubules and the number of Sertoli cells also decrease with age.^{5,6} According to previous reports, the APT SI of brain tumors is positively correlated with cellular density and/or proliferation (i.e., intracellular mobile protein and peptide concentrations).^{24,25} Therefore, decreased number of germ cells and reduced amounts of extracellular mobile proteins and peptides might account for APT SI reduction and ADC increase in testes. As such, elderly men with decreased spermatogenic function should, in theory, exhibit decreased APT and elevated ADC. The above results corroborate those of previous studies.^{13,14}

The testicular ADC values were higher in the patients with hypospermatogenesis caused by epididymo-orchitis (group A) than in all age groups of controls, including the oldest (group B4), which appeared to reflect impaired spermatogenic function in the group A patients. Moreover, the testicular APT SI values were significantly increased in group A compared with all age groups of controls, including the youngest (group B1). The results revealed that both ADC and APT could serve as biomarkers of impaired spermatogenic function caused by epididymo-orchitis. Interestingly, APT SI and ADC values in group A showed opposite trends of changes versus the observed trends with aging. Both parameters are believed to be related to cellular density,^{24,25,42,43} meaning reduced APT SI and elevated ADC should be observed in patients with hypospermatogenesis because of reduced cellular density. The abnormally elevated testicular APT SI in patients with epididymo-orchitis may be caused by the destruction of testes by active inflammation, which induces a series of pathological alterations, including microscopic necrosis, and an accumulation of testicular metabolites rich in mobile proteins and peptides. In addition, the APT SI is influenced by the chemical exchange rate between amide and water protons; this exchange rate depends mainly on the concentration of amide protons and the pH value in tissue.²³ We hypothesize that inflammatory reaction may alter the intracellular environment of testicular cells, thereby disrupting the original acid-base balance and changing the pH value in the testis to

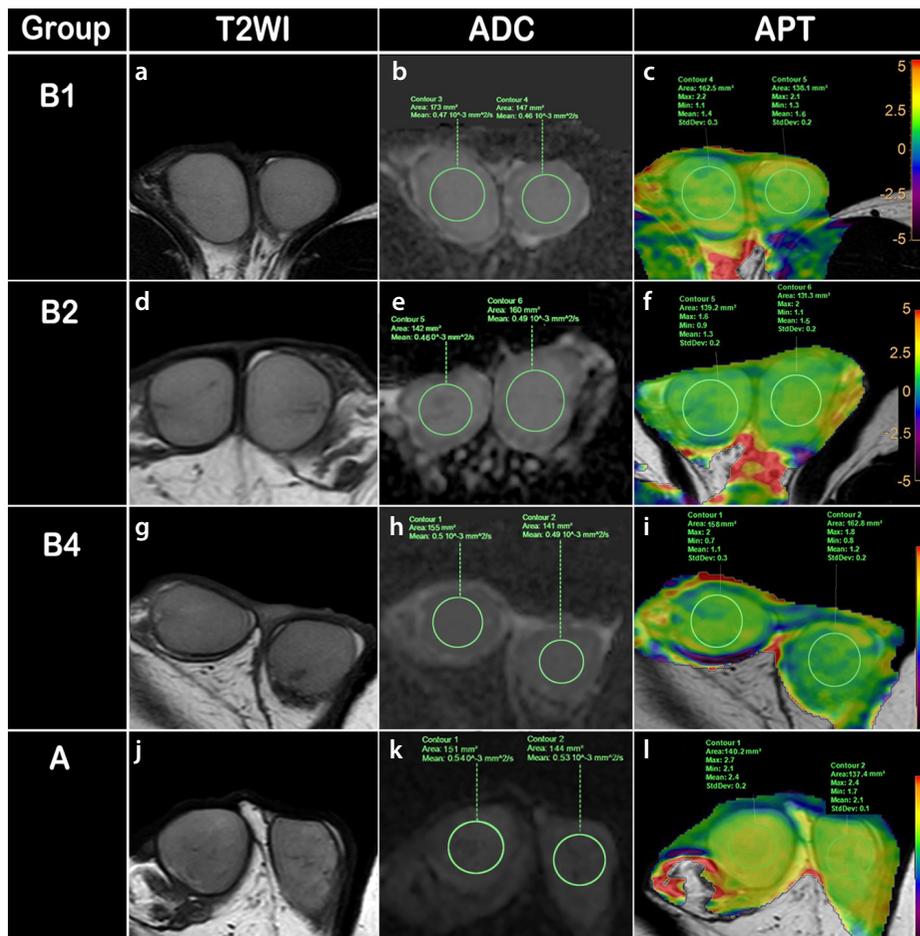


Figure 5. (a-c) A 25-year-old man in group B1 with normal semen analysis. (a) Axial T2-weighted imaging (WI) showing homogeneous hyperintensity of both testes. (b) An ADC map demonstrating average ADC values of $0.457 \times 10^{-3} \text{ mm}^2/\text{s}$ and $0.460 \times 10^{-3} \text{ mm}^2/\text{s}$ for the right and left testes, respectively. (c) Amide proton transfer (APT) map demonstrating low average APT signal intensity (SI) values of 1.5 and 1.45 for the right and left testes, respectively. (d-f) A 38-year-old man in group B2 with normal semen analysis. (d) Axial T2WI showing homogeneous hyperintensity of both testes. (e) ADC map demonstrating average ADC values of $0.460 \times 10^{-3} \text{ mm}^2/\text{s}$ and $0.490 \times 10^{-3} \text{ mm}^2/\text{s}$ for the right and left testes, respectively. (f) APT map demonstrating low average APT SI values of 1.3 and 1.5 for the right and left testes, respectively. (g-i) A 75-year-old man in group B4 married with two children, without traumatic history in testes. (g) Axial T2WI showing homogeneous hyperintensity of both testes. (h) ADC map demonstrating average ADC values of $0.510 \times 10^{-3} \text{ mm}^2/\text{s}$ and $0.490 \times 10^{-3} \text{ mm}^2/\text{s}$ for the right and left testes, respectively. (i) APT map demonstrating low average APT SI values of 1.1 and 1.2 for the right and left testes, respectively. (j-l) A 28-year-old man in group A with hypospermatogenesis caused by epididymo-orchitis who was clinically confirmed. (j) Axial T2WI showing homogeneous hyperintensity of both testes. (k) ADC map demonstrating higher average ADC values of $0.540 \times 10^{-3} \text{ mm}^2/\text{s}$ and $0.530 \times 10^{-3} \text{ mm}^2/\text{s}$ for the right and left testes, respectively, compared with controls. (l) Higher APT SI values of the right and left testes; determined on the APT map (2.4 and 2.1, respectively) compared with controls.

a certain extent. This might accelerate the exchange rate between amide and water protons, leading to increased APT values. Further investigation is required to explore the mechanism underlying the changes of testicular APT SI in patients with epididymo-orchitis. The different trends of changes in APT SI and ADC may help determine the cause of hypospermatogenesis. A decreased APT SI combined with an increased ADC may suggest hypospermatogenesis caused by aging or chronic injuries, whereas an in-

creased APT SI combined with an elevated ADC is more likely to be associated with active or persistent inflammation. This may further improve clinical decision-making.

The limitations of this study are as follows. First, only a portion of the included volunteers (20–50 years old) underwent semen analysis to examine spermatogenic function. In addition, for volunteers over 50 years old, their statements of prior paternity were accepted as evidence of fertility, which might have led to errors in these age groups. Sec-

ond, generally no histopathologic confirmation was available for patients with suspected epididymo-orchitis, with all group A patients clinically confirmed by seminal plasma biochemistry. Third, the failure rate (23/139) of APTw imaging due to artifacts was still considerable and needs to be improved in the future. Finally, the heterogeneity in the severity distribution of hypospermatogenesis within our cohort (predominantly comprising severe cases), limits the generalizability of findings to milder forms of hypospermatogenesis. Future multicenter studies with larger, balanced samples are warranted to validate our findings.

In conclusion, according to the above preliminary results, the APT SI of normal testes tended to decrease with increasing age, whereas patients with hypospermatogenesis caused by epididymo-orchitis exhibited much higher APT SI compared with all age groups of controls. Thus, when combined with the ADC, which generally increases with both age and impaired spermatogenic function, APT SI may provide additional diagnostic information.

Acknowledgments

We thank Mr. Jinhui Zhao and Ms. Jiaxiu He (Department of Radiology, Linzhi People's Hospital, Tibet, PR China), for their support during this research.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

The work was supported by: (1) Science and Technology Projects in Guangzhou (number: SL2022A04J01837); (2) Medical Scientific Research Foundation of Guangdong Province, China (number: A2022008); (3) Natural

Science Foundation of Tibet [number: XZZR202402135(W)].

References

1. Theas MS. Germ cell apoptosis and survival in testicular inflammation. *Andrologia*. 2018;50(11):e13083. [Crossref]
2. Jalal H, Bahadur G, Knowles W, Jin L, Brink N. Mumps epididymo-orchitis with prolonged detection of virus in semen and the development of anti-sperm antibodies. *J Med Virol*. 2004;73(1):147-150. [Crossref]

3. Peng Y, Ouyang L, Lin Z, Zhang F, Wang H, Guan J. MRI findings of nonobstructive azoospermia: lesions in and out of pelvic cavity. *Abdom Radiol (NY)*. 2020;45(7):2213-2224. [\[Crossref\]](#)
4. Vermeulen A. Andropause. *Maturitas*. 2000;34(1):5-15. [\[Crossref\]](#)
5. Kothari LK, Gupta AS. Effect of ageing on the volume, structure and total Leydig cell content of the human testis. *Int J Fertil*. 1974;19(3):140-146. [\[Crossref\]](#)
6. Johnson L, Petty CS, Neaves WB. Influence of age on sperm production and testicular weights in men. *J Reprod Fertil*. 1984;70(1):211-218. [\[Crossref\]](#)
7. Tsili AC, Giannakis D, Sylakos A, Ntorkou A, Sofikitis N, Argyropoulou MI. MR imaging of scrotum. *Magn Reson Imaging Clin N Am*. 2014;22(2):217-238, vi. [\[Crossref\]](#)
8. Woldrich JM, Im RD, Hughes-Cassidy FM, Aganovic L, Sakamoto K. Magnetic resonance imaging for intratesticular and extratesticular scrotal lesions. *Can J Urol*. 2013;20(4):6855-6859. [\[Crossref\]](#)
9. Tsili AC, Argyropoulou MI, Astrakas LG, et al. Dynamic contrast-enhanced subtraction MRI for characterizing intratesticular mass lesions. *AJR Am J Roentgenol*. 2013;200(3):578-585. [\[Crossref\]](#)
10. Manganaro L, Saldari M, Pozza C, et al. Dynamic contrast-enhanced and diffusion-weighted MR imaging in the characterisation of small, non-palpable solid testicular tumours. *Eur Radiol*. 2018;28(2):554-564. [\[Crossref\]](#)
11. Tsili AC, Argyropoulou MI, Giannakis D, Tsampalas S, Sofikitis N, Tsampoulas K. Diffusion-weighted MR imaging of normal and abnormal scrotum: preliminary results. *Asian J Androl*. 2012;14(4):649-654. [\[Crossref\]](#)
12. Maki D, Watanabe Y, Nagayama M, et al. Diffusion-weighted magnetic resonance imaging in the detection of testicular torsion: feasibility study. *J Magn Reson Imaging*. 2011;34(5):1137-1142. [\[Crossref\]](#)
13. Wang H, Guan J, Lin J, et al. Diffusion-weighted and magnetization transfer imaging in testicular spermatogenic function evaluation: preliminary results. *J Magn Reson Imaging*. 2018;47(1):186-190. [\[Crossref\]](#)
14. Tsili AC, Astrakas LG, Goussia AC, Sofikitis N, Argyropoulou MI. Volumetric apparent diffusion coefficient histogram analysis of the testes in nonobstructive azoospermia: a noninvasive fingerprint of impaired spermatogenesis? *Eur Radiol*. 2022;32(11):7522-7531. [\[Crossref\]](#)
15. Han BH, Park SB, Seo JT, Chun YK. Usefulness of testicular volume, apparent diffusion coefficient, and normalized apparent diffusion coefficient in the MRI evaluation of infertile men with azoospermia. *AJR Am J Roentgenol*. 2018;210(3):543-548. [\[Crossref\]](#)
16. Tsili AC, Ntorkou A, Goussia A, et al. Diffusion tensor imaging parameters in testes with nonobstructive azoospermia. *J Magn Reson Imaging*. 2018;48(5):1318-1325. [\[Crossref\]](#)
17. Tsili AC, Ntorkou A, Baltogiannis D, et al. Magnetization transfer imaging of normal and abnormal testis: preliminary results. *Eur Radiol*. 2016;26(3):613-621. [\[Crossref\]](#)
18. Ntorkou A, Tsili AC, Goussia A, et al. Testicular apparent diffusion coefficient and magnetization transfer ratio: can these MRI parameters be used to predict successful sperm retrieval in nonobstructive azoospermia? *AJR Am J Roentgenol*. 2019;213(3):610-618. [\[Crossref\]](#)
19. Firat AK, Uğraş M, Karakaş HM, et al. 1H magnetic resonance spectroscopy of the normal testis: preliminary findings. *Magn Reson Imaging*. 2008;26(2):215-220. [\[Crossref\]](#)
20. Aaronson DS, Iman R, Walsh TJ, Kurhanewicz J, Turek PJ. A novel application of 1H magnetic resonance spectroscopy: non-invasive identification of spermatogenesis in men with non-obstructive azoospermia. *Hum Reprod*. 2010;25(4):847-852. [\[Crossref\]](#)
21. Ntorkou A, Tsili AC, Astrakas L, et al. In vivo biochemical investigation of spermatogenic status: 1H-MR spectroscopy of testes with nonobstructive azoospermia. *Eur Radiol*. 2020;30(8):4284-4294. [\[Crossref\]](#)
22. Zhou J, Lal B, Wilson DA, Laterra J, van Zijl PC. Amide proton transfer (APT) contrast for imaging of brain tumors. *Magn Reson Med*. 2003;50(6):1120-1126. [\[Crossref\]](#)
23. Zhou J, Payen JF, Wilson DA, Traustman RJ, van Zijl PC. Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nat Med*. 2003;9(8):1085-1090. [\[Crossref\]](#)
24. Togao O, Yoshiura T, Keupp J, et al. Amide proton transfer imaging of adult diffuse gliomas: correlation with histopathological grades. *Neuro-oncology*. 2014;16(3):441-448. [\[Crossref\]](#)
25. Zhou J, Blakeley JO, Hua J, et al. Practical data acquisition method for human brain tumor amide proton transfer (APT) imaging. *Magn Reson Med*. 2008;60(4):842-849. [\[Crossref\]](#)
26. Wang HJ, Cai Q, Huang YP, et al. Amide proton transfer-weighted MRI in predicting histologic grade of bladder cancer. *Radiology*. 2022;305(1):E59. [\[Crossref\]](#)
27. Takayama Y, Nishie A, Sugimoto M, et al. Amide proton transfer (APT) magnetic resonance imaging of prostate cancer: comparison with Gleason scores. *Magma*. 2016;29(4):671-679. [\[Crossref\]](#)
28. Takayama Y, Nishie A, Togao O, et al. Amide proton transfer MR imaging of endometrioid endometrial adenocarcinoma: association with histologic grade. *Radiology*. 2018;286(3):909-917. [\[Crossref\]](#)
29. Li C, Peng S, Wang R, et al. Chemical exchange saturation transfer MR imaging of Parkinson's disease at 3 Tesla. *Eur Radiol*. 2014;24(10):2631-2639. [\[Crossref\]](#)
30. Björndahl L, Kirkman Brown J; other Editorial Board Members of the WHO Laboratory Manual for the Examination and Processing of Human Semen. The sixth edition of the WHO laboratory manual for the examination and processing of human semen: ensuring quality and standardization in basic examination of human ejaculates. *Fertil Steril*. 2022;117(2):246-251. [\[Crossref\]](#)
31. Street EJ, Justice ED, Kopa Z, et al. The 2016 European guideline on the management of epididymo-orchitis. *Int J STD AIDS*. 2017;28(8):744-749. [\[Crossref\]](#)
32. Kopa Z, Wenzel J, Papp GK, Haidl G. Role of granulocyte elastase and interleukin-6 in the diagnosis of male genital tract inflammation. *Andrologia*. 2005;37(5):188-194. [\[Crossref\]](#)
33. Collodel G, Nerucci F, Signorini C, Iacoponi F, Moretti E. Associations between biochemical components of human semen with seminal conditions. *Syst Biol Reprod Med*. 2019;65(2):155-163. [\[Crossref\]](#)
34. Zorn B, Sesek-Briski A, Osredkar J, Meden-Vrtovec H. Semen polymorphonuclear neutrophil leukocyte elastase as a diagnostic and prognostic marker of genital tract inflammation--a review. *Clin Chem Lab Med*. 2003;41(1):2-12. [\[Crossref\]](#)
35. Ju Y, Wang Y, Luo RN, et al. Evaluation of renal function in chronic kidney disease (CKD) by mDIXON-quant and amide proton transfer weighted (APT_w) imaging. *Magn Reson Imaging*. 2023;103(9):102-108. [\[Crossref\]](#)
36. Ju Y, Liu A, Wang Y, et al. Amide proton transfer magnetic resonance imaging to evaluate renal impairment in patients with chronic kidney disease. *Magn Reson Imaging*. 2022;87(3):177-182. [\[Crossref\]](#)
37. By S, Barry RL, Smith AK, et al. Amide proton transfer CEST of the cervical spinal cord in multiple sclerosis patients at 3T. *Magn Reson Med*. 2018;79(2):806-814. [\[Crossref\]](#)
38. Schlegel PN, Katzovitz MA. Male Reproductive Physiology. In: Chapple CR, Steers WD, Evans CP, editors. *Urologic Principles and Practice*. Cham: Springer International Publishing; 2020:41-62. [\[Crossref\]](#)
39. Neto FT, Bach PV, Najari BB, Li PS, Goldstein M. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol*. 2016;59:10-26. [\[Crossref\]](#)
40. Tajar A, Forti G, O'Neill TW, et al. Characteristics of secondary, primary, and compensated hypogonadism in aging men: evidence from the European male ageing study. *J Clin Endocrinol Metab*. 2010;95(4):1810-1818. [\[Crossref\]](#)
41. Araujo AB, O'Donnell AB, Brambilla DJ, et al. Prevalence and incidence of androgen

deficiency in middle-aged and older men: estimates from the Massachusetts male aging study. *J Clin Endocrinol Metab.* 2004;89(12):5920-5926. [\[Crossref\]](#)

42. Mazaheri Y, Shukla-Dave A, Hricak H, et al. Prostate cancer: identification with combined

diffusion-weighted MR imaging and 3D 1H MR spectroscopic imaging--correlation with pathologic findings. *Radiology.* 2008;246(2):480-488. [\[Crossref\]](#)

43. Zelhof B, Pickles M, Liney G, et al. Correlation of diffusion-weighted magnetic resonance

data with cellularity in prostate cancer. *BJU Int.* 2009;103(7):883-888. [\[Crossref\]](#)



Validation of R2* magnetic resonance imaging for quantifying secondary iron overload in pediatric patients

Tahani M. Ahmad^{1,2,3}
 Fareed Khdair Ahmad^{4,5}
 Mohamed Abdolell³
 Olfat Ahmad^{6,7}
 Matthew T. Rogers²

¹The University of Jordan School of Medicine,
Department of Radiology, Amman, Jordan

²IWK Health Center, Department of Pediatric Imaging,
Halifax, Canada

³Dalhousie University Faculty of Medicine,
Department of Diagnostic Radiology, Halifax, Canada

⁴Jordan University Hospital, Section of Pediatric
Gastroenterology, Amman, Jordan

⁵The University of Jordan School of Medicine,
Department of Pediatrics, Amman, Jordan

⁶King Hussein Cancer Center (KHCC), Amman, Jordan

⁷Hopp Children's Cancer Center (KiTZ), Division of
Pediatric Neuro-oncology, Heidelberg, Germany

PURPOSE

Non-invasive assessment of iron deposition is the standard of care for guiding chelation therapy in patients with iron overload. Several magnetic resonance imaging (MRI)-based techniques have been developed. This study compares the MRI-based R2* method with the standard R2-based method for quantifying iron levels in the liver and heart in children and young adults with secondary iron overload.

METHODS

A single-center prospective study was conducted over 2.5 years involving 14 patients aged 4–22 years with secondary iron overload. These patients underwent 40 MRI scans using both R2 and R2* methods at same time. A total of 36 scans were analyzed, comparing the two methods using linear regression analysis and Bland–Altman plots.

RESULTS

The study shows a significant correlation between liver iron concentration measurements obtained using the R2* method and those obtained using the R2-based method (adjusted R² = 0.77128). The agreement was even stronger for R2* values in the cardiac septum (adjusted R² = 0.93483).

CONCLUSION

The R2* method for assessing iron deposition in the liver and cardiac septum is comparable to the R2-based method and is suitable for clinical use. However, due to slight differences in measurements between the two techniques, it is advisable to consistently use one method for monitoring treatment in each patient. Further research is needed to refine the calibration equations.

CLINICAL SIGNIFICANCE

This study highlights the MRI-based R2* method as a reliable, non-invasive, and cost-effective alternative to the R2-based method for monitoring iron overload in pediatric patients, with no additional costs for institutions or third parties.

KEYWORDS

Iron deposition, magnetic resonance imaging, chelation, liver

Corresponding author: Tahani M. Ahmad

E-mail: tahani1523@yahoo.com

Received 13 February 2025; revision requested 17 March 2025; accepted 28 April 2025.



Epub: 03.06.2025

Publication date: 08.07.2025

DOI: 10.4274/dir.2025.253282

Quantitative assessment of iron deposition in body organs is critical for the clinical management of patients with iron overload. Excessive iron in primary and secondary hemochromatosis is absorbed through the intestine and accumulates in various body parts. The liver is the main storage site for excess iron. However, when the liver storage capacity is exceeded, the iron overflows to other organs—including the reticuloendothelial system—such as the spleen, bone marrow, cardiac septum, and pancreas, leading to various complications (e.g., diabetes, cardiomyopathy, and, more commonly, liver fibrosis and cirrhosis). Only a small portion of iron is eliminated through the sloughed mucosa, feces, and menstruation. There is little correlation between ferritin levels and iron deposition, particularly in the heart.^{1,2} Liver iron concentration (LIC) reflects body iron content. Therefore, LIC assessment is used as a surrogate to indicate total body iron deposition.^{2,3} Liver biopsy is the gold standard for quantifying LIC. However, it has certain limitations, such as being invasive and inconvenient,

You may cite this article as: Ahmad TM, Khdair Ahmad F, Abdolell M, Ahmad O, Rogers MT. Validation of R2* magnetic resonance imaging for quantifying secondary iron overload in pediatric patients. *Diagn Interv Radiol.* 2025;31(4):399-404.

requiring sedation, and the relatively small cross-sectional sampling volume of the hepatic parenchyma, which can lead to inaccurate results, particularly if the iron deposition is not homogeneous.² Misestimated LIC measurements can lead to the progression of liver fibrosis and cirrhosis and undesired side effects of medications and chelation therapy.

In the last few decades, various non-invasive magnetic resonance imaging (MRI)-based techniques—both qualitative and quantitative—for assessing LIC have emerged, gained popularity, and are now used in clinical practice. Although they cannot provide objective measurements or detect slight iron deposition, qualitative methods, such as gradient echo (in-phase and out-of-phase) sequences, can alert possible iron overload. For example, iron deposition results in a decreased signal on the in-phase sequence. Quantitative methods include R2-based and R2*-based relaxometry techniques and the signal intensity ratio (SIR) method.⁴ Relaxometry techniques depend on MRI signal shortening with increased iron deposition, whereas the SIR method is based on the observation of a drop in liver signal intensity with increased iron deposition in comparison to reference tissues such as the paraspinous muscles. Developed by Clark and St Pierre⁵, the United States Food and Drug Administration-validated, widely accepted, and standardized method is based on R2 measurement and is commercially known as FerriScan® (Resonance Health, Australia). This method requires validation with an external phantom, involves additional cost, and takes several days to return quantitative results to the medical team. More importantly, it involves a longer scanning time (approx. 15 min), which makes it susceptible to breathing and motion artifacts. The R2* (R2* = 1/T2*) method has been proposed as a reliable alternative for assessing LIC. This method is faster (approx. 15 s), and R2* values can be converted to tissue iron content using appropriate calibration curves.

Research groups have published calibration equations based on the T2*/R2* technique; however, there has been little independent validation of these calibration equations in clinical settings, with no consensus on the ideal acquisition protocol. It is noteworthy that LIC measurements vary between institutions due to differences in acquisition parameters and post-processing fitting algorithms. Therefore, internal validation of these parameters is necessary before providing clinical results.

Although images for the R2* relaxometry technique are easy to acquire, the need for post-processing steps limits their adoption in clinical practice.⁶ Several MRI vendors have attempted to promote the use of this method by providing software packages that help analyze the images and convert R2* values into clinically relevant LIC values. This is conducted at the scanner console or local workstation by drawing a few regions of interest (ROI) over the right lobe of the liver and cardiac septum. This software varies according to the vendor and the sequences acquired.⁷ However, there is limited data on the validation of these potentially useful products. At IWK health, we use a 1.5 T GE magnet with STARmap analysis via CardiacVX software Wipro GE Healthcare Pvt. Ltd. In this study, we aimed to validate the results of this software in comparison to the standard FerriScan® method to eliminate the need for third-party processing. We expect that this will benefit our institution, as well as other facilities with a similar setup.

Methods

Patient selection and study design

This prospective study is compliant with Health Insurance Portability and Accountability Act standards and the principles of the Declaration of Helsinki, and was approved by the IWK Institutional Research Ethics Board (# 1021517), with approval obtained in April 2018 and renewed annually until completion of the study. All patients with known secondary iron overload referred to by their primary hematologist, gastroenterologist, or treating physicians for MRI evaluation of iron overload at our institution were recruited over 2.5 years and consent was obtained at the time of the scan. A total of 40 scans were performed for 14 patients between October 2020 and April 2023. Of these, four scans were excluded from the analysis as they were considered technically inadequate, either due to motion artifacts or lack of specific liver or cardiac sequences.

Magnetic resonance imaging studies

FerriScan® studies were performed according to the specified protocol. This included five sequences of axial spin-echo images with variable time to echo (TE) of 6, 9, 12, 15, and 18 ms. The field of view (FOV) was 34 cm, with a 5-mm slice thickness and 5-mm spacing, covering 11 slices. The non-breath-hold (BH) scan duration was 2 min and 28 s for each sequence. Other parameters included repetition time (TR) = 1,000 ms, matrix = 256 × 192, bandwidth (BW) = 62.5 kHz, and number of excitations (NEX) = 1. A bag of normal saline was used as a phantom for imaging to provide a reference signal intensity for measurement correction purposes in case of any potential machine drift.

A single BH technique was used to obtain non-contrast-enhanced T2* axial gradient-echo images at the level of the main portal vein, with increasing TE sequences. Noise correction was applied, but fat correction was not performed. The parameters for this sequence were TR = 170 ms, multi-echo TE = 0.9–7.7 ms at regular intervals of approximately 0.97 ms, flip angle (FA) = 10°, BW = 50 kHz, and matrix = 220 × 220. The inherent phased-array body coil was used as the transmitter coil, and a receive-only coil was used for signal collection. Furthermore, a single mid-ventricular short axis 2D GE (multi-echo fast gradient-recalled echo) cardiac gated slice was imaged at eight TEs ranging from 2.09 to 19.9 ms, with increments of 2.53 ms. The FOV was 40 cm, and the slice thickness was 10 mm (no gap). Other parameters included FA = 20°, frequency matrix = 224, phase matrix = 128, NEX = 1, and BW = 83.33 kHz. Fast imaging employing steady-state acquisition sequences were used to obtain axial, two-chamber, four-chamber, and then short-axis views. All scans were performed on a single 1.5 T MRI scanner (Signa HD Twin-Speed, 2002/hardware update in 2012; GE Healthcare, Milwaukee, WI, USA).

Image analysis

The FerriScan® sequences were sent to the primary company, Resonance Health Center in Australia, for quantitative analysis as per routine clinical practice. The multi-echo gradient images were transferred to the onsite Advantage Workstation (AW, GE HealthCare) for diagnostic imaging processing. Three small ROIs, each at least 1 cm in diameter, were manually drawn in the right lobe of the liver on areas that appeared homogeneous in signal intensity, devoid of vessels or biliary trees, and away from the diaphragm

Main points

- Magnetic resonance imaging is the standard clinical practice for assessing tissue iron deposition, particularly in children.
- The R2* method is comparable to the standard R2-based method for quantifying iron overload in children.
- Adherence to one method is crucial to avoid underestimation or overestimation of iron deposition.

(Supplementary Figure 1a). The mean value of these ROIs was used for analysis. The fitted curve was evaluated, and truncation was occasionally used to remove late outlier points to account for the plateau observed due to the low signal-to-noise ratio at later TE values.⁷⁻⁹ The $R2^*$ value was converted to LIC using the vendor-provided STARmap analysis via CardiacVX software, which is based on Dr. Wood's calibration formula [$\text{Fe (mg/g)} = 0.0254 \times R2^* + 0.202$].¹⁰

For cardiac analysis, a homogeneous full-thickness ROI was drawn on the cardiac septum to avoid the epicardium and blood pool (Supplementary Figure 1b). Studies have shown that mid-ventricular septal iron correlates well with global left ventricular iron concentration.^{7,11}

Magnetic resonance imaging-based quantification of liver iron concentration and T2 relaxometry

The severity of liver iron overload using MRI techniques is categorized as follows according to the literature:

- Normal: $<1.8 \text{ mg/g}$ ($<32 \text{ } \mu\text{mol/g}$)
- Borderline: $1.8\text{--}3.2 \text{ mg/g}$ ($32\text{--}57 \text{ } \mu\text{mol/g}$)
- Mild: $3.2\text{--}7.0 \text{ mg/g}$ ($57\text{--}125 \text{ } \mu\text{mol/g}$)
- Moderate: $7.0\text{--}15.0 \text{ mg/g}$ ($125\text{--}269 \text{ } \mu\text{mol/g}$) (increased risk of complications)
- Severe: $>15.0 \text{ mg/g}$ ($>269 \text{ } \mu\text{mol/g}$) (high mortality risk)¹²

The $T2^*$ thresholds for iron overload severity

- $T2^* > 20 \text{ ms}$: normal
- $T2^* 10\text{--}20 \text{ ms}$: mild iron overload
- $T2^* 5\text{--}10 \text{ ms}$: moderate iron overload
- $T2^* < 5 \text{ ms}$: severe iron overloads^{13,14}

Statistical analysis

The $T2^*$ values were automatically transformed into reciprocal $R2^*$ values for analysis to obtain a positive linear correlation: $R2^* \text{ (Hz)} = 1,000/T2^* \text{ (ms)}$. The mean $R2^*$ value from the three hepatic ROIs was used for analysis. The $R2^*$ was converted to LIC using the vendor-specific software, in line with Dr. Woods's calibration-based formula. The FerriScan® LIC was considered the "gold standard" for comparison. The agreement between LICs calculated by the $R2^*$ and FerriScan® methods was assessed using linear regression analysis and Bland–Altman analysis, the latter of which characterizes both systematic differences (bias) and random fluctuations (variance).

The data followed a normal distribution pattern. Due to the limited sample size and challenges in collecting enough pediatric cases in a reasonable time frame—and as the primary focus of the study was to assess the images' characteristics and not the longitudinal changes or patient-specific effect—we assumed independence of the images and thus used linear regression analysis and Bland–Altman testing.

Results

Overall, 36 MRI scans from 14 patients were included in the analysis (Figure 1). Each

scan included both FerriScan® and STARmap sequences and analyses (Supplementary Figure 2). The age range of our cohort was 4.0–22.0 years, with mean and median ages of 14.56 and 17 years, respectively. The group consisted of nine male patients and five female patients (Table 1). However, the female group underwent more MRI follow-up scans during this period, resulting in an equal distribution of the MRI studies between male and female patients, with each group representing 50% of the total scans analyzed. Among these patients, two had sideroblastic anemia, one had Blackfan–Diamond anemia, one had pyruvate kinase deficiency anemia, and the remainder had B-thalassemia major disease.

The LIC measured by FerriScan® ranged from 2.4 to 39.5 mg/g tissue dry weight (DW), with an average of 19.44 mg/g DW. In comparison, $R2^*$ values ranged from 96.4 to 894.5 Hz, corresponding to LIC values of 2.6–22.9 mg/g DW, with an average of 12.9 mg/g DW. Cardiac $R2^*$ values assessed by FerriScan® ranged from 19.0 to 369.0 Hz, whereas cardiac $R2^*$ values assessed locally ranged from 20.0 to 282.0 Hz (Table 2).

The regression analysis showed a substantial correlation between LIC values calculated by the $R2^*$ method and those measured by FerriScan®, with an adjusted R^2 of 0.77128. This indicates that approximately 77% of the LIC values obtained using the FerriScan® method can be predicted or accounted for by the values obtained using the $R2^*$ method. This finding indicates that the $R2^*$ method is a reliable predictor of LIC values, although not a perfect match with FerriScan® measurements. The analysis yielded a slope of 0.43054 and a y-intercept of 4.539, suggesting a consistent relationship between the two measurement techniques.

The Bland–Altman analysis revealed a bias of 6.53, meaning FerriScan® values were, on average, 6.53 mg/g DW higher than those estimated by the $R2^*$ method. The standard deviation was 7.239, with limits of agreement between -7.658 and 20.719 . This means that in most cases, the difference between the two methods would fall within this range. The confidence intervals ranged from -3.43 to -11.8 and from 16.49 to 24.944 . The highest agreement was observed for LICs below 12 mg/g DW , as shown in the scattergram and regression line in Figure 2a.

The cardiac analysis indicated an excellent correlation between the two methods, with an adjusted R^2 of 0.93483. This suggests that the differences or changes seen in the

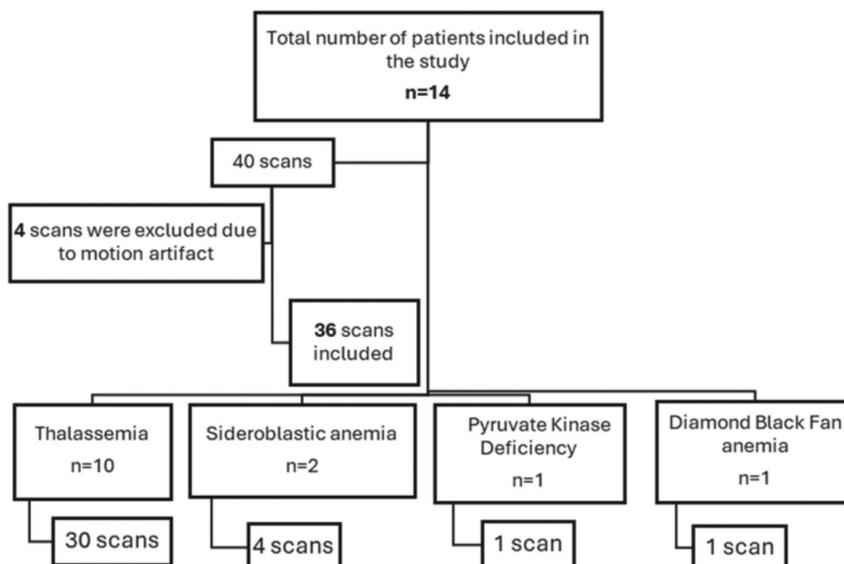


Figure 1. Flow chart summarizing the study population, with the number of scans performed.

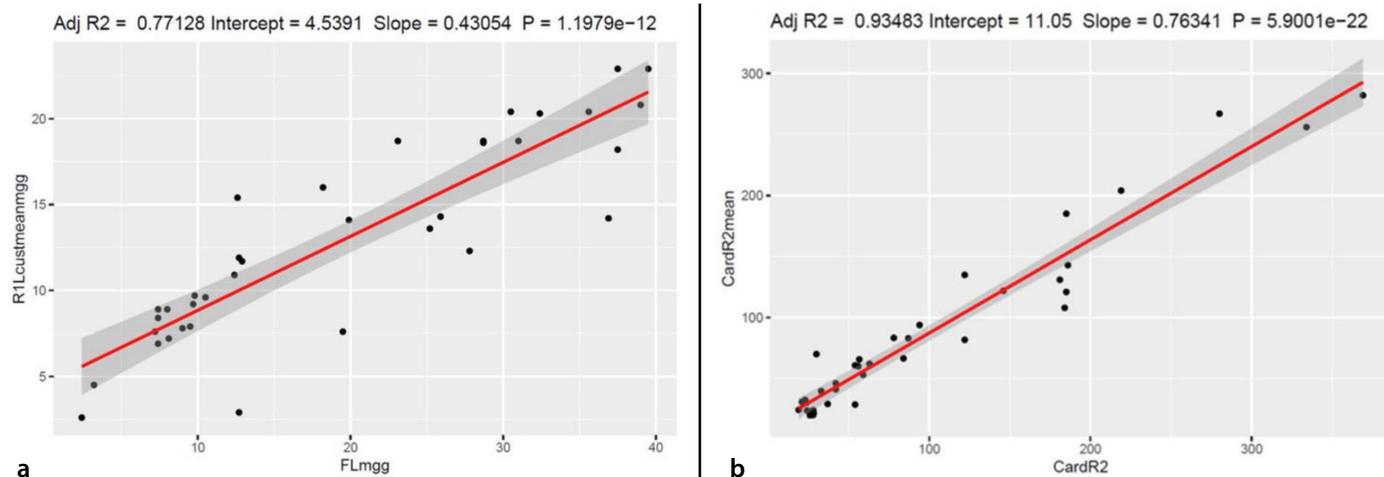


Figure 2. Logistic regression calibration curves illustrating the correlation between FerriScan® readings and R2* STARmap values. The x-axis represents FerriScan® readings in mg/g dry tissue, and the y-axis represents R2* STARmap values in mg/g dry tissue. (a) Liver values; (b) cardiac septum values. Both curves demonstrate substantial correlation.

Table 1. Population characteristics		
Total number of eligible cohorts	14	
Age range (mean) years	4-22 (14.56)	
Male sex (%)	9/14 (64%)	
Underlying clinical conditions of the cohort	Number of patients	Number of the scans Included in the analysis
B-thalassemia major	10	30
Sideroblastic anemia	2	4
Diamond-black fan anemia	1	1
Pyruvate kinase deficiency anemia	1	1

Table 2. Liver and cardiac iron values measured using FerriScan® and R2* methods			
	Liver iron concentration mg/g dry tissue weight	Cardiac iron deposition R2 Hz	Cardiac iron deposition T2 ms
FerriScan®	2.4–39.5 (19.44)	19.0–369.0 (100.11)	2.7–52.80 (20.24)
R2* method	2.6–22.9 (12.9)	20.0–282.0 (87.48)	3.5–50.00 (20.35)

R2* values (measured using FerriScan®) can mostly be predicted using the values measured locally. As such, the local R2* method is almost as good as FerriScan® in measuring these values. The slope was 0.763, with an intercept of 11.05, indicating a strong relationship between the two sets of measurements (Figure 2b).

Discussion

This project evaluates the reliability of the local R2* method for the assessment of the iron overload in the liver and heart compared with the gold standard FerriScan® method. Our findings demonstrated substantial agreement between the R2* method and FerriScan® in estimating LIC, particularly in cases of mild to moderate iron overload

(LIC < 12 mg/g DW). This is consistent with the findings obtained by Meloni et al.¹⁶, who concluded that signal decay models result in clinically acceptable estimations of LIC provided the ROIs are correctly drawn and the proper calibration curve is used to correct for any systematic differences in R2* estimation.^{14,15} The agreement between the two methods was highest at LICs below 12 mg/g DW, in line with the findings by Abou Zahr et al.¹⁷, where the best agreement was observed at LICs below 15 mg/g DW in their cohort. This suggests that both methods perform similarly in the range of mild to moderate liver iron deposition.

In our cohort, FerriScan® values consistently showed higher LICs compared with the R2* method, with a positive bias

of 6.7%. This overestimation of FerriScan® was previously reported in Reeder et al.'s¹² study. We believe that this difference could be partially attributed to the calibration equation used in the CardiacVX software, which is based on Wood et al.'s¹⁰ 2005 study. Although this equation was confirmed by Hankins et al.¹⁸ in 2009, Meloni et al.¹⁶ later demonstrated a 15%-lower scaling coefficient between R2* and LIC. It is worth noting that Dr. Wood's equation was originally formulated on a single-echo sequence, whereas our images were based on multi-echo sequences. However, this may have had a limited contribution, as cross-validation of single-center and multicenter R2* relaxometry methods, including both single and multi-echo sequences, has been performed in earlier studies and has demonstrated no significant difference³ Since FerriScan® is considered the gold standard, the observed overestimation by this method should be considered when interpreting results. Although R2* values might underestimate iron content compared with FerriScan®, this discrepancy could be due to the calibration model used by the latter, rather than an inherent inaccuracy in the R2* method itself. Therefore, while FerriScan® may provide more accurate results in the context of our study, both methods are valuable for assessing LIC; however, discrepancies in calibration should be considered.

For validation, we plan to reanalyze the measured ROIs using another calibration equation, such as Garbowski's et al.¹⁹ 2014 equation, which follows parameters closer to ours, to determine whether better agreement can be reached.²⁰ This would require an update of the software to set the latter equa-

tion as the default one. The wide range of agreement between the two methods have been previously explained by both Wood et al.¹⁰ and Clark and St Pierre⁵, and may be attributed to the spatial variability of iron concentration within the liver. Other confounding factors, such as iron particle size, shape, and local metabolites (fat and fibrosis), that affect signal relaxation might also contribute to this wide range.

The cardiac analysis demonstrated a substantial agreement between our analysis and that of FerriScan[®]. The difference in measurement is not surprising, as variations in R2* values with different ROI sizes and measurements have been reported. Thus, we opted to use the mean values of three ROIs for the analysis.¹⁸⁻²⁰ The intra- and inter-rater, as well as inter-scanner reproducibility of R2* analysis, was assessed in previous work by Hernando et al.³, who further demonstrated that the calibration from different studies can be translated, improving the utilization of R2* mapping. Kirk et al.²¹ also tested the reproducibility of the R2* technique among five different international centers on different scanners and concluded that the measurement of tissue T2* (heart and liver) can be achieved reproducibly between centers across the world, provided appropriate vendor-specified sequences are followed, along with appropriate software analysis packages and calibration curves.^{1,22} The inter-scanner and inter-center reproducibility and transferability were also evaluated and confirmed in multiple other studies.^{8,9}

The strength of our study lies in the prospective collection of cases and the simultaneous acquisition of sequences for both methods, specifically focusing on pediatric patients, which minimizes confounding factors, such as fatty infiltration and cirrhosis, that could contribute to signal alterations.²² However, the study also has several few limitations. First, the study has a small sample size, a common challenge in pediatric research. Second, although Dr. Wood's 2005 equation is the first published calibration formula and still the most widely used equation, in future studies, we aim to analyze ROIs with a calibration equation more aligned with our parameters to potentially improve agreement.

In conclusion, this study successfully validated the use of R2* MRI for assessing iron overload in the heart and liver of children with secondary iron overload. While R2* measurements do not perfectly align with FerriScan[®] results, the R2* is a reliable predic-

tor of LIC. This emphasizes the importance of using a consistent method for assessment and follow-up. Monitoring trends in iron concentration is crucial for adjusting clinical management. A larger multicenter validation of the R2* method is necessary to establish its reliability across various settings and populations.

Acknowledgments

The authors would like to thank Scribendi Inc. for its English language editing service.

Footnotes

Conflict of interest disclosure

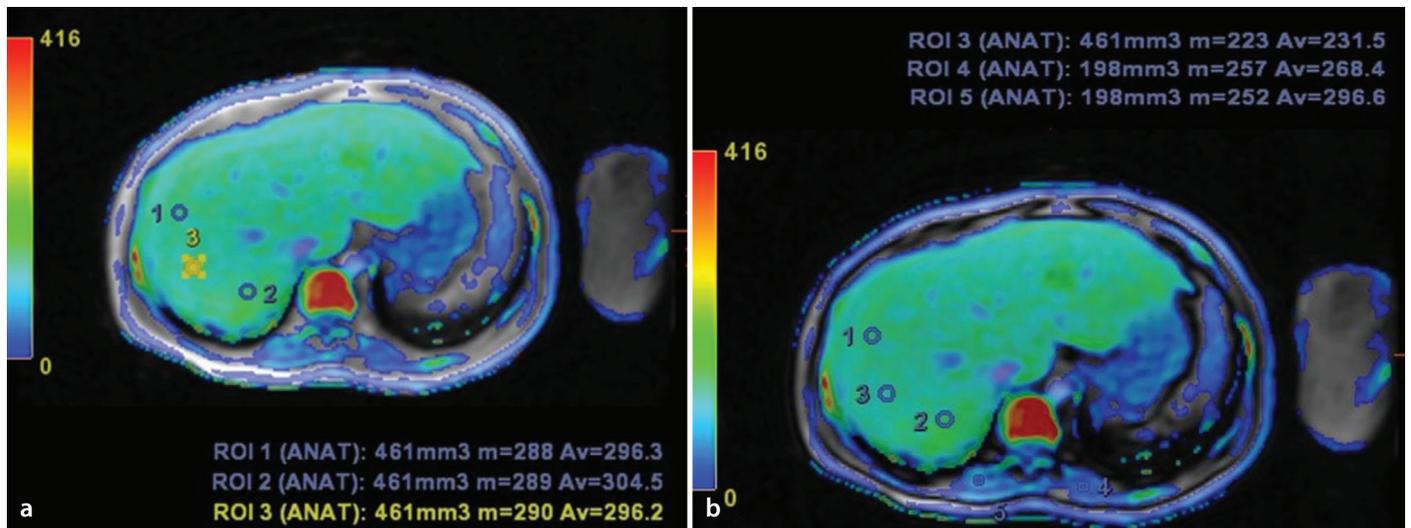
The authors declared no conflicts of interest.

References

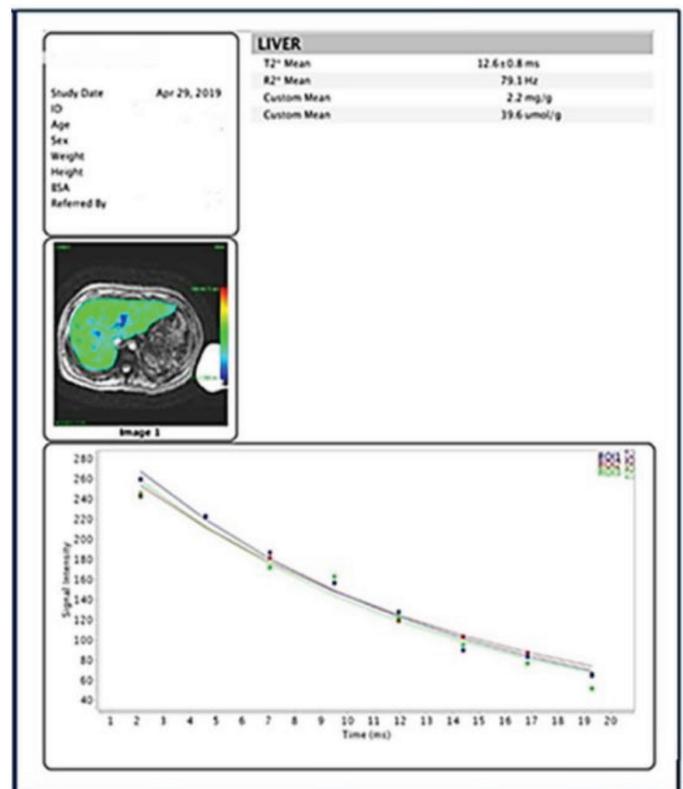
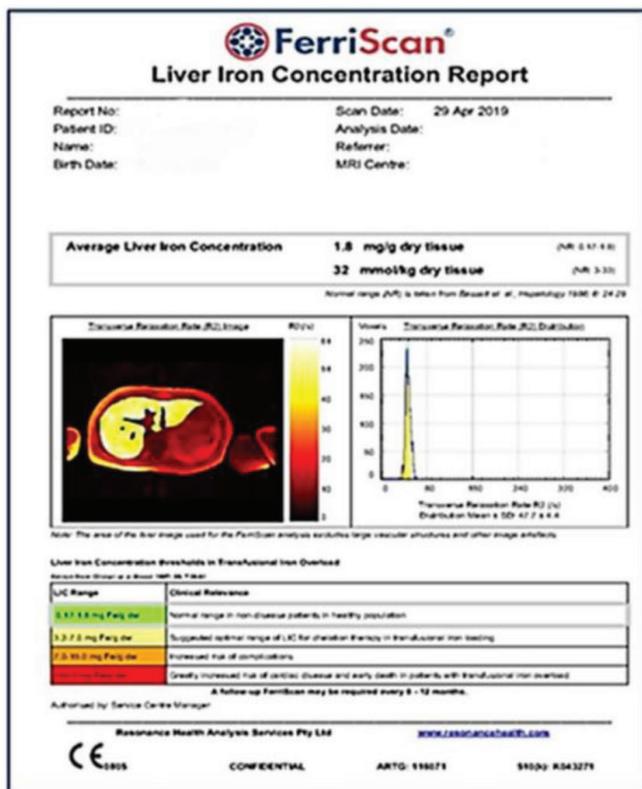
1. Alam MH, He T, Auger D, Smith GC, Drivas P, Wage R, et al. Validation of T2* in-line analysis for tissue iron quantification at 1.5 T. *J Cardiovasc Magn Reson*. 2016;18(1):23. [\[Crossref\]](#)
2. Labranche R, Gilbert G, Cerny M, Vu KN, Soulières D, Olivie D, et al. Liver iron quantification with MR imaging: a primer for radiologists. *Radiographics*. 2018;38(2):392-412. [\[Crossref\]](#)
3. Hernando D, Levin YS, Sirlin CB, Reeder SB. Quantification of liver iron with MRI: state of the art and remaining challenges. *J Magn Reson Imaging*. 2014;40(5):1003-1021. [\[Crossref\]](#)
4. Amin K, Mileto A, Kolokythas O. MRI for liver iron quantification: concepts and current methods. *Semin Ultrasound CT MR*. 2022;43(4):364-370. [\[Crossref\]](#)
5. Clark PR, St Pierre TG. Quantitative mapping of transverse relaxivity (1/T(2)) in hepatic iron overload: a single spin-echo imaging methodology. *Magn Reson Imaging*. 2000;18(4): 431-438. [\[Crossref\]](#)
6. Henninger B, Alustiza J, Garbowski M, Gandon Y. Practical guide to quantification of hepatic iron with MRI. *Eur Radiol*. 2020;30(1):383-393. [\[Crossref\]](#)
7. Git KA, Fioravante LAB, Fernandes JL. An online open-source tool for automated quantification of liver and myocardial iron concentrations by T2* magnetic resonance imaging. *Br J Radiol*. 2015;88(1053):20150269. [\[Crossref\]](#)
8. Zhang X, Peng J, Wang C, Feng Y, Feng Q, Li X, et al. Improved liver R2* mapping by averaging decay curves. *Sci Rep*. 2017;7(1):6158. [\[Crossref\]](#)
9. Tanner MA, He T, Westwood MA, Firmin DN, Pennell DJ, on behalf of the Thalassemia International Federation Heart T2*

- Investigators*. Multi-center validation of the transferability of the magnetic resonance T2* technique for the quantification of tissue iron. *Haematologica*. 2006;91(10):1388-91. [\[Crossref\]](#)
10. Wood JC, Enriquez C, Ghugre N, Tyzka JM, Carson S, Nelson MD, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood*. 2005;106(4):1460-1465. [\[Crossref\]](#)
11. Pepe A, Positano V, Santarelli MF, Sorrentino F, Cracolici E, De Marchi D, et al. Multislice multiecho T2* cardiovascular magnetic resonance for detection of the heterogeneous distribution of myocardial iron overload. *J Magn Reson Imaging*. 2006;23(5):662-668. [\[Crossref\]](#)
12. Reeder SB, Yokoo T, França M, et al. Quantification of liver iron overload with MRI: review and guidelines from the ESGAR and SAR. *Radiology*. 2023;307(1):221856. [\[Crossref\]](#)
13. El Sherif AM, Ibrahim AS, Elsayed MA, et al. The impact of magnetic resonance imaging in the assessment of iron overload in heart and liver in transfusion-dependent thalassemic children: minia experience. *Egypt J Radiol Nucl Med*. 2021;52. [\[Crossref\]](#)
14. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. *Radiographics*. 2009;29(5):1433-1449. [\[Crossref\]](#)
15. Li J, Lin H, Liu T, et al. Quantitative susceptibility mapping (QSM) minimizes interference from cellular pathology in R2* estimation of liver iron concentration. *J Magn Reson Imaging*. 2018;48(4):1069-1079. [\[Crossref\]](#)
16. Meloni A, Rienhoff HY, Jones A, Pepe A, Lombardi M, Wood JC. The use of appropriate calibration curves corrects for systematic differences in liver R2* values measured using different software packages. *Br J Haematol*. 2013;161(6):888-891. [\[Crossref\]](#)
17. Abou Zahr R, Burkhardt BEU, Ehsan L, et al. Real-world experience measurement of liver iron concentration by R2 vs. R2 Star MRI in hemoglobinopathies. *Diagnostics (Basel)*. 2020;10(10):768. [\[Crossref\]](#)
18. Hankins JS, McCarville MB, Loeffler RB, et al. R2* magnetic resonance imaging of the liver in patients with iron overload. *Blood*. 2009;113(20):4853-4855. [\[Crossref\]](#)
19. Garbowski MW, Carpenter JP, Smith G, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. *J Cardiovasc Magn Reson*. 2014;16(1):40. [\[Crossref\]](#)
20. McCarville MB, Hillenbrand CM, Loeffler RB, et al. Comparison of whole liver and small region-of-interest measurements of MRI liver R2* in children with iron overload. *Pediatr Radiol*. 2010;40(8):1360-1367. [\[Crossref\]](#)

21. Kirk P, He T, Anderson LJ, et al. International reproducibility of single breathhold T2* MR for cardiac and liver iron assessment among five thalassemia centers. *J Magn Reson Imaging*. 2010;32(2): 315-319. [Crossref]
22. Ternifi R, Pouletaut P, Sasso M, Miette V, Charleux F, Bensamoun S. Improvements of liver MR imaging clinical protocols to simultaneously quantify steatosis and iron overload. *IRBM*. 2018;39(3):219-225. [Crossref]



Supplementary Figure 1. The location of region of interest as placed on (a) the right lobe of the liver, avoiding blood vessels, diaphragm or biliary tree and (b) on the cardiac septum.



Supplementary Figure 2. A 15-year-old female patient with B-thalassemia major who had multiple transfusions, and chelation therapy: (a) FerriScan® report showing an average liver iron concentration (LIC) of 1.8 mg/g dry tissue; (b) R2* analysis calibration curve indicating an LIC of 2.2 mg/g dry tissue.