

EISSN 1305-3612

DIR

Diagnostic and Interventional Radiology

TSR
1924
TURKISH SOCIETY
OF RADIOLOGY

dirjournal.org

VOLUME 31
ISSUE 3
May 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

Evrin 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

Evrin 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 Aydıngöz, MD (2002-2011)

Berna Dirim Mete (2016-2017)

Neuroradiology and Head & Neck Imaging

Kubilay Aydın, MD (2016-2023)

Nafı 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 ULAKBİM 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: May 2025

EISSN 1305-3612

International scientific journal published bimonthly.



Contents

ABDOMINAL IMAGING

161 Invited Review. Renal bleeding: imaging and interventions in patients with tumors. *Emad D. Singer, Niloofer Karbasian, Douglas S. Katz, Vincenzo K. Wong, Mohamed E. Abdelsalam, Nir Stanietzky, Trinh T. Nguyen, Anuradha S. Shenoy-Bhangle, Mohamed Badawy, Margarita V. Revzin, Mostafa A. Shehata, Mohamed Eltaher, Khaled M. Elsayes, Brinda Rao Korivi*

171 Original Article. Diagnostic performance of the O-RADS MRI system for magnetic resonance imaging in discriminating benign and malignant adnexal lesions: a systematic review, meta-analysis, and meta-regression. *Gülsüm Kılıçkap*

NEURORADIOLOGY

180 Original Article. Open- and closed-type congenital cholesteatomas of the middle ear: computed tomography differentiation and correlation with surgical staging. *Minjung Seong, Hyung-Jin Kim, Yikyung Kim, Sung Tae Kim*

ARTIFICIAL INTELLIGENCE AND INFORMATICS

187 Original Article. A retrospective evaluation of the potential of ChatGPT in the accurate diagnosis of acute stroke. *Beyza Nur Kuzan, İsmail Meşe, Servan Yaşar, Taha Yusuf Kuzan*

196 Original Article. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Alperen Elek, Duygu Doğa Ekizalioğlu, Ezgi Güler*

206 Original Article. Breast cancer detection and classification with digital breast tomosynthesis: a two-stage deep learning approach. *Yazeed Alashban*

CHEST IMAGING

215 Original Article. Diagnostic performance of radiomics analysis for pulmonary cancer airway spread: a systematic review and meta-analysis. *Jie Chen, Xinyue Zhang, Chi Xu, Kefu Liu*

HEAD AND NECK IMAGING

226 Original Article. Dual-energy computed tomography-based volumetric thyroid iodine quantification: correlation with thyroid hormonal status, pathologic diagnosis, and phantom validation. *Younghen Lee*

INTERVENTIONAL RADIOLOGY

234 Technical Note. Quadratus lumborum block for procedural and postprocedural analgesia in renal cell carcinoma percutaneous cryoablation. *Saman Fouladirad, Jasper Yoo, Behrang Homayoon, Jun Wang, Pedro Lourenço*

237 Original Article. Non-routine thrombectomy in pediatric arterial ischemic stroke. *Sinan Balcı, Nesibe Gevher Eroğlu-Ertuğrul, Ahmet Ziya Birbilen, Dilek Yalınzoğlu, Selman Kesici, Tevfik Karagöz, Anıl Arat*

249 Original Article. Long-term outcomes of catheter-directed sclerotherapy for ovarian endometrioma. *Jae Hwan Lee, Jihyung Yoon, Chong-Ho Lee, Kun Yung Kim, Chang Jin Yoon, Minuk Kim, Seul Ki Kim*

253 Original Article. Combining transarterial chemoembolization, radiofrequency ablation, and iodine-125 seed implantation for recurrent hepatocellular carcinoma post-hepatectomy. *Yong Zhong, Li Wang, Weibin Dan, Dan Liang*

MUSCULOSKELETAL IMAGING

259 Original Article. Comparing ultrasound-guided intra-articular injection and medial branch block for lumbar facet joint pain: a clinical study. *Hui Zhao, Yun-Long Hou, Le-Hang Guo, Qiao Wang, Tian He, Guang-Fei Gu, Li-Ping Sun, Feng-Shan Jin*

NEURORADIOLOGY

264 Original Article. The value of dual-energy computed tomography angiography-based virtual monoenergetic imaging for evaluations after cerebral aneurysm clipping. *Zhihua Lu, Suying Wu, Feijian Wu, Qingdong Jin, Qingjing Huang, Baoteng Zhang*

NUCLEAR MEDICINE AND MOLECULAR IMAGING

274 Original Article. Maximum standardized uptake value-to-tumor size ratio in fluorodeoxyglucose F18 positron emission tomography/computed tomography: a simple prognostic parameter for non-small cell lung cancer. *Soo Jeong Kim, Koeun Lee, Hyun Joo Lee, Du-Young Kang, Young Hwan Kim*

PEDIATRIC RADIOLOGY

280 Original Article. Consolidation in the right middle lobe in pediatric bronchial-pulmonary artery shunt: radiology's Aunt Minnie? *Chi Wang, Rongchang Wu, Zihan Wang, Shuai Ma, Xinyu Yuan, Yuchun Yan, Yun Peng*



Renal bleeding: imaging and interventions in patients with tumors

- Emad D. Singer¹
- Niloofar Karbasian²
- Douglas S. Katz³
- Vincenzo K. Wong¹
- Mohamed E. Abdelsalam¹
- Nir Stanietzky¹
- Trinh T. Nguyen⁴
- Anuradha S. Shenoy-Bhangle⁵
- Mohamed Badawy⁶
- Margarita V. Revzin⁷
- Mostafa A. Shehata¹
- Mohamed Eltaher¹
- Khaled M. Elsayes¹
- Brinda Rao Korivi¹

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

²The University of Texas Health Sciences Center, Department of Diagnostic and Interventional Radiology, Texas, USA

³New York University, Long Island School of Medicine, Department of Radiology, New York, USA

⁴Billings Clinic Hospital, Department of Diagnostic Radiology, Montana, USA

⁵Harvard Medical School, Massachusetts General Hospital, Department of Radiology, Massachusetts, USA

⁶Wayne State University, Diagnostic Radiology Department, Michigan, USA

⁷Yale University School of Medicine, Department of Radiology, Connecticut, USA

Corresponding author: Khaled M. Elsayes

E-mail: kmelsayes@mdanderson.org

Received 25 April 2024; revision requested 12 May 2024; accepted 21 May 2024.



Epub: 14.06.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242822

ABSTRACT

In patients with cancer, spontaneous renal bleeding can stem from a range of underlying factors, necessitating precise diagnostic tools for effective patient management. Benign and malignant renal tumors are among the primary culprits, with angiomyolipomas and renal cell carcinomas being the most common among them. Vascular anomalies, infections, ureteral obstructions, and coagulation disorders can also contribute to renal-related bleeding. Cross-sectional imaging techniques, particularly ultrasound and computed tomography (CT), play pivotal roles in the initial detection of renal bleeding. Magnetic resonance imaging and CT are preferred for follow-up evaluations and aid in detecting underlying enhancing masses. IV contrast-enhanced ultrasound can provide additional information for active bleeding detection and differentiation. This review article explores specific disorders associated with or resembling spontaneous acute renal bleeding in patients with renal tumors; it focuses on the significance of advanced imaging techniques in accurately identifying and characterizing renal bleeding in these individuals. It also provides insights into the clinical presentations, imaging findings, and treatment options for various causes of renal bleeding, aiming to enhance the understanding, diagnosis, and management of the issue.

KEYWORDS

Renal bleeding, renal cell carcinoma, angiomyolipoma, diagnosis, imaging, interventions

Bleeding can be a cause of substantial morbidity in patients with cancer; approximately 10% of individuals diagnosed with cancer experience bleeding episodes.^{1,2} Furthermore, high bleeding rates are frequently observed in patients with advanced-stage cancer and those undergoing anticoagulant therapy. These factors underscore the need for vigilant monitoring and tailored management approaches to address the risk of bleeding in this patient population.²

A diverse range of conditions can lead to spontaneous, non-traumatic, and acute renal bleeding in patients with benign and malignant tumors, highlighting the need for precise diagnostic tools to optimize patient management.³ The primary causes of spontaneous renal bleeding include benign and malignant kidney tumors, with angiomyolipomas (AMLs) and renal cell carcinomas (RCCs) being the most common among them.³⁻⁵ In addition to primary renal tumors, vascular abnormalities affecting the kidneys, infections, ureteral obstructions, and coagulation disorders can also contribute to renal-related bleeding in patients with cancer. Additionally, certain neoplastic conditions can closely mimic renal bleeding.³⁻⁵ Cross-sectional imaging techniques, including computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, are indispensable in evaluating patients with renal bleeding. They enable the detection of renal bleeding and the identification of its underlying causes.⁶⁻⁸ Interventional radiologists play a vital role in managing patients with active and substantial bleeding.⁹

This paper explores specific disorders associated with or resembling spontaneous acute renal bleeding in patients with cancer; it focuses on the crucial role of advanced imaging techniques in accurately identifying and characterizing renal bleeding in these individuals.³ It also provides insights into the clinical presentations, imaging findings, and treatment options for various causes of renal bleeding, aiming to enhance the understanding, diagnosis, and management of the issue.

Clinical presentations of renal bleeding in patients with cancer

Patients with cancer and renal bleeding vary widely in their clinical presentations; some may be asymptomatic, and others may have potentially life-threatening symptoms. Symptoms may include flank pain, macroscopic hematuria, and a palpable abdominal mass. Flank pain can vary in intensity and nature due to the progressive enlargement of a hematoma due to ongoing extravasation. Potentially severe complications include hemorrhagic shock and acute kidney injury, manifesting as tachycardia, hypotension, pallor, altered mental status, and decreased urine output.¹⁰⁻¹⁴

In patients with spontaneous renal bleeding, extrarenal accumulations of fluid can lead to extrinsic compression of the kidney. This compression, termed Page kidney, may elicit systemic hypertension.¹⁵

The role of imaging in renal bleeding detection and follow-up

Ultrasound and CT are the main imaging modalities used for the initial detection of renal bleeding.² Moreover, CT and MRI are the preferred imaging modalities for the follow-up of patients with renal bleeding. Unenhanced CT is employed to confirm that renal bleeding has resolved. Pre- and post-contrast CT and MRI facilitate the detection of an underlying enhancing mass, which is often present in many patients.³ MRI is an excellent modality for renal bleeding follow-up because of its high contrast resolution and ability to assist in tissue characterization without ionizing radiation.³ However, it has limitations: longer scanning times and limited accessibility compared with those of oth-

er modalities and higher prices compared with CT. Follow-up imaging at 2–3 month intervals to document the resolution of the hemorrhage is of utmost importance in patients with bleeding whose underlying cause remains unclear.³ If subsequent scans do not reveal the presence of an underlying mass, no additional measures are required, assuming the hemorrhage continues to diminish in size and a thorough assessment of the underlying renal structure is conducted.³ Table 1 summarizes the appearance of renal bleeding on various imaging modalities according to the phase of the bleeding.^{5,16-19}

Ultrasound

Ultrasound is a valuable imaging modality for assessing hematomas with distinct characteristics observed at different stages of development.¹⁷ An acute hematoma initially appears solid and hyperechoic, or heterogeneous on gray-scale ultrasound, and avascular on color Doppler imaging (Figure 1).¹⁷ As the hematoma progresses, it changes its echotexture and becomes more heterogeneous. Occasionally, it has a central region of decreased echogenicity or a cystic-like appearance. In the chronic phase, peripheral calcifications may develop within the hematoma.¹⁷

Intravenous (IV) contrast-enhanced ultrasound can potentially depict active renal bleeding.¹⁸ Moreover, when a substantial renal mass or a conspicuous vascular anomaly is absent, ultrasonography might not suffice to ascertain the underlying cause of bleeding in most patients with spontaneous renal bleeding. The use of contrast-enhanced ultrasound provides supplementary information beyond that which conventional ultrasound methodologies can offer.^{18,20-26}

Computed tomography

On non-contrast CT, acute bleeding has an attenuation value of 30–45 Hounsfield units (HU),^{5,16} and this value increases to 60–80 HU after several hours, depending on the concentration of hemoglobin.^{5,16} In the subacute phase (several days to several weeks), the attenuation of the bleeding decreases as the blood clot is resorbed, resulting in the formation of a seroma with an attenuation value of approximately 30 HU or less.^{5,16} Calcification may be visible in cases of chronic hematomas, defined as those present for more than 1–2 months. If an IV contrast-enhanced CT is performed during acute bleeding, active contrast extravasation from the underlying bleeding vessel may be localized in the arterial phase of the CT.^{5,16} CT can also help differentiate between extravasations

Table 1. The appearance of renal bleeding by imaging modality and phase

Imaging modality	Appearance in the acute phase	Appearance in the subacute and chronic phases
Ultrasound	Solid and echogenic appearance	Appearance is stable or reduced in size
	Multi-lamellated, whorled appearance	Hypoechoic fluid collections due to liquefaction
	Anechoic cystic areas	Formation of retractile clots/angular margins
	Surrounding soft-tissue edema	An increase in size may simulate a soft-tissue tumor Well-defined margin, distinct capsule Hypoechoic or anechoic seroma with posterior acoustic enhancement
Non-contrast CT	Normally, the attenuation value of the bleeding would be 30–45 HU, similar to the blood pool's attenuation	Bleeding attenuation decreases due to clot lysis and progressive seroma formation, measuring approx. 30 HU or less
	Bleeding typically measures 60–80 HU during the hyperacute phase, which lasts a few hours	Chronic hematomas can calcify
MRI	Hyperacute blood: T1 isointense, T2 hypointense	Early subacute (2–7 days): T1 hyperintense, T2 hypointense
	Acute blood: T1 hypointense/isointense, T2 hypointense	Late subacute (7–14 days): T1 hyperintense, T2 hyperintense
		Chronic: peripherally low T1 and T2 signals, central T1 isointense and T2 hyperintense

CT, computed tomography; MRI, magnetic resonance imaging; HU, Hounsfield units.

Main points

- There are numerous causes of spontaneous renal bleeding; tumors, both benign and malignant, vascular malformations, infection, and coagulopathy are common etiologies.
- Utilizing the most appropriate diagnostic tools to determine the cause is critical, and ultrasound and computed tomography (CT) are most often used in the acute setting, whereas CT and magnetic resonance imaging play an important role in follow-up.
- Catheter angiography plays an important role in the management of renal bleeding.
- The many causes of bleeding are discussed with examples; pitfalls, including mimickers of bleeding, are also presented.

and pseudoaneurysms. Active extravasation manifests as an ill-defined area of extraluminal contrast. During the arterial phase, it appears isointense to the blood pool. Subsequently, it persists during the portal venous phase and typically enlarges progressively in the delayed phase.²⁷ In contrast, a pseudoaneurysm is characterized by a well-defined, rounded mass with an internal enhancement that matches the pattern observed in the aorta on corticomedullary phase images.²⁸

Magnetic resonance imaging

The appearance of bleeding on MRI changes depending on the chronicity of the bleeding and the breakdown of the hemoglobin over time (Figure 2).¹⁹ In the early stage, compared with the surrounding tissues, the bleeding appears isointense on T1-weighted images and isointense to hyperintense on T2-weighted images. This is attributed to the diamagnetic effect of intra-

cellular oxyhemoglobin.¹⁹ In the acute phase, hemoglobin breaks down into intracellular deoxyhemoglobin, causing a decrease in the signal intensity on both T1- and T2-weighted images. In the early subacute phase (4–7 days), hemoglobin breakdown leads to the formation of intracellular methemoglobin, which has a paramagnetic effect that causes a hyperintense signal on T1 images and a hypointense signal on T2 images.¹⁹ In the late subacute phase (1–4 weeks), the continued breakdown of hemoglobin forms extracellular methemoglobin, contributing to a hyperintense signal on both T1 and T2 images. During the chronic stage, the bleeding demonstrates a peripheral area of low-signal intensity on both T1- and T2-weighted images, which can be attributed to the accumulation of intracellular hemosiderin, and the central region displays T2 hyperintensity and T1 isointensity.¹⁹

A hemorrhage can be detected and differentiated from the surrounding tissues using susceptibility-weighted imaging (SWI), a sequence combining the magnitude and phase data of MRI. It is sensitive to the detection of blood products, including microhemorrhages.^{29–32}

Owing to its sensitivity and ability to depict changes in hemoglobin types as a hematoma ages, SWI can be very useful in identifying hemorrhages.³³ Its sensitivity and accuracy can surpass those of conventional imaging methods routinely used to detect bleeding; it can depict and characterize hemorrhage in clear cell RCCs, regardless of the tumor grade or bleeding pattern.³⁴ However, SWI images can be affected by air-caused susceptibility artifacts within the gastrointestinal tract, leading to imaging patterns that resemble bleeding and potentially complicating the interpretation of images.³⁴

Catheter angiography

Superselective renovascular catheterization and embolization play crucial roles in the diagnosis and treatment of renal bleeding and, potentially, the avoidance of radical surgery (Figure 3). Interventional radiology management strategies, particularly the coiling or embolization of the bleeding source, are used for treatment. Renal arteries are terminal vessels that lack substantial intrarenal collateral circulation. Hypervascular renal tumors, particularly RCCs, make use of collateral circulation from outside the kidneys. Hence, the preferred embolic agents are those that can block small vessels, including N-butyl cyanoacrylate glue, ethanol, polyvinyl alcohol, or embospheres, as well as agents capable of occluding larger vessels (e.g., coils).³⁵ The potential complications of endovascular treat-

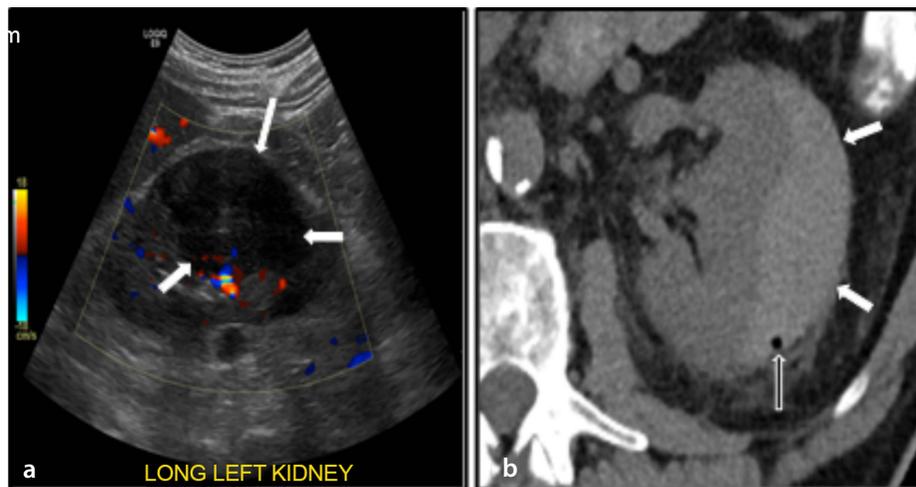


Figure 1. A lentiform, mass-like area along the lateral margin of the left kidney is seen on color Doppler ultrasound (a) and axial non-contrast CT (b), consistent with an acute subcapsular hematoma (white arrows). A tiny focus of gas along the posterior margin is likely iatrogenic (black arrow). CT, computed tomography.

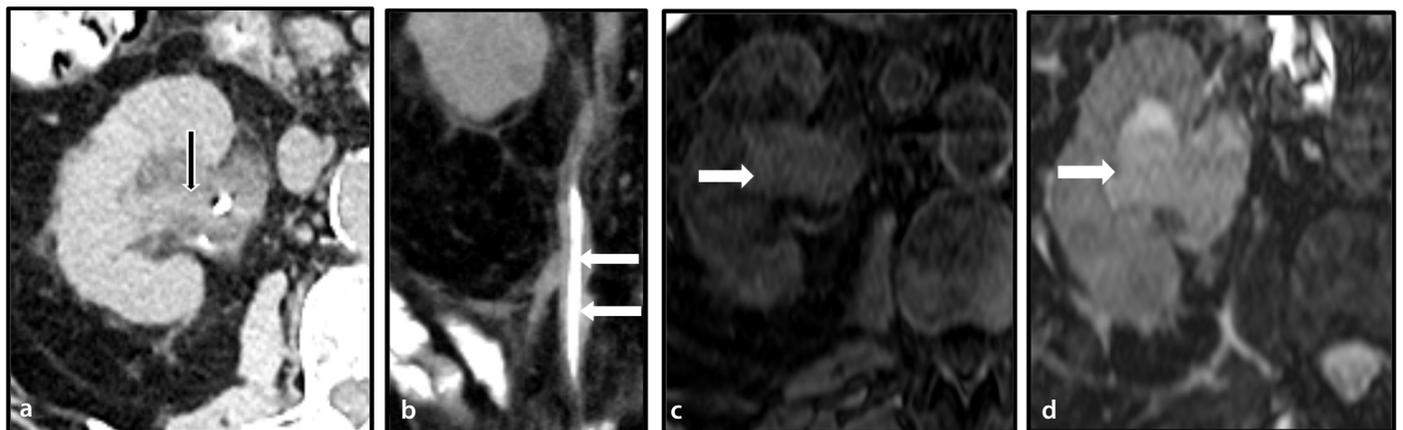


Figure 2. Axial (a) and coronal (b) contrast-enhanced CT status scans obtained after the placement of a right nephroureteral stent in a patient with urothelial carcinoma (white arrows). The high-attenuation hemorrhage within the renal pelvis is consistent with hemorrhage (black arrow). (c, d) Non-contrast T1 (c) and T2 (d) MRI images of the right kidney demonstrate layering of late subacute blood products (T1 hyperintense, T2 hyperintense) (white arrows). CT, computed tomography; MRI, magnetic resonance imaging.

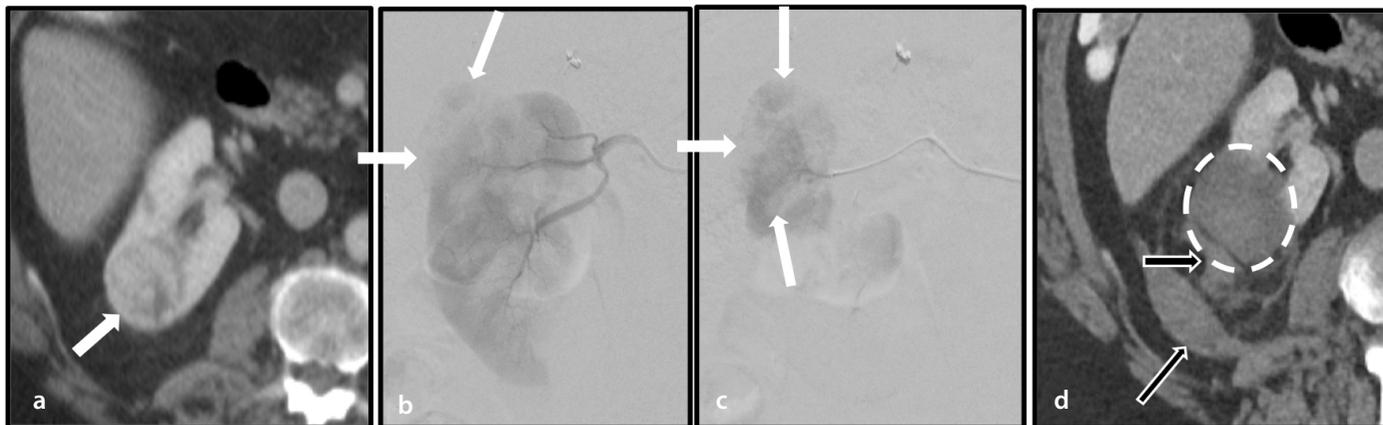


Figure 3. (a) A right upper pole renal cell carcinoma (white arrows) demonstrates abnormal tortuous intra-tumoral vessels on angiography (b) during pre-ablation renal artery embolization (b, c). Post-ablation contrast-enhanced CT (d) demonstrates a hypoattenuating ablation zone without residual enhancement (white circle). A small amount of posterior perirenal hemorrhage is present (black arrows). CT, computed tomography.

ment include infarction, exacerbation of hypertension, unintended embolization (e.g., coil migration to the systemic circulation), and stent graft thrombosis.³⁵

Primary renal neoplasms

Renal cell carcinoma

RCC can cause spontaneous bleeding within the renal parenchyma; it is the most common cause of spontaneous bleeding due to a primary malignant renal neoplasm (Figures 4, 5).^{3,36} Parenchymal bleeding is more frequently observed in the clear cell subtype of RCC, which is the most prevalent subtype.⁸ CT and MRI are useful in diagnosing RCC, but they can pose challenges in identifying the tumor if it is small and obscured by bleeding. In such cases, subtraction imaging may aid in identifying the underlying tumor.³ However, it can cause several problems: misregistration artifacts related to patient motion and a variation in the technical factors on the unenhanced and enhanced sequences.³⁷ An alternative strategy to identify small tumors presenting with bleeding and potentially obscured on the initial scans is to repeat imaging after the hematoma shrinks (6–8 weeks). Table 2 summarizes the differential diagnoses of spontaneous renal bleeding in patients with cancer and includes helpful radiological features for proper differentiation.

For patients presenting with bleeding in association with RCC, a patient's clinical condition dictates the treatment. Hemodynamically stable patients are treated conservatively and monitored closely. However, transcatheter arterial embolization (TAE) may be needed to treat active bleeding or prevent it from increasing by cutting off the tumor's blood supply.⁸ The disease stage must be

Table 2. Differential diagnoses and the main radiological features of renal bleeding in patients with cancer

Diagnosis	Radiological features
Renal cell carcinoma	Heterogeneous, solid renal mass Enhancement with contrast on CT or MRI scans Possible bleeding (high-density areas on non-contrast CT scans, hyperintense areas on T1-weighted MRI scans) Invasion into the renal vein or inferior vena cava in advanced cases
Trauma/ iatrogenic bleeding (e.g., after renal biopsy, nephrostomy placement)	Perirenal hematoma (fluid collection around the kidney) Laceration or contusion of the kidney parenchyma Active extravasation (contrast extravasation on CT scans) Urinoma or urinary leakage (with collecting system injury)
Vascular causes (e.g., polyarteritis nodosa)	Multiple, small, round, wedge-shaped, or irregular hypodense foci on CT scans Enhancement of peripheral rims in the arterial phase Rosary sign or string-of-beads appearance on renal artery due to microaneurysms on angiography
Angiomyolipoma	Well-circumscribed, hyperechoic mass on ultrasound Fat-containing mass with negative HU on CT scans T1 and T2 are hyperintense signals on MRI scans (without fat suppression) Presence of bleeding (high-density areas on non-contrast CT scans and hyperintense areas on T1-weighted MRI, with fat suppression to differentiate between the bleeding and fat content) No specific features for renal bleeding, but bleeding may occur due to infection or obstruction Multiple, bilateral, low-attenuation renal masses on CT scans Homogeneous or heterogeneous enhancement with contrast Bleeding may be present with aggressive subtypes or as a complication

CT, computed tomography; MRI, magnetic resonance imaging; HU, Hounsfield units.

considered when choosing a definitive treatment for RCC. Percutaneous ablation, partial nephrectomy, and radical nephrectomy are alternative therapeutic approaches that can

be employed, depending on the tumor stage and location and the general health of the patient, among other factors.³⁸

Angiomyelipoma

Various studies have shown that the prevalence of renal AMLs as a causative factor for spontaneous renal bleeding ranges from 30% to 35%.³⁶⁻³⁹ Typically affecting the kidneys, AMLs are a condition in which mesenchymal tumors characterized by variable amounts of mature smooth muscle, adipose tissue, and abnormal blood vessels form.⁴⁰ It is assumed that AMLs originate from perivascular epithelioid cells, which are unique cells found close to blood vessels.⁴⁰ Classic triphasic and monotypic epithelioid AMLs are the two primary subtypes of the disease, and these subgroups have distinct biological and imaging properties. Up to 70% of classic AMLs appear sporadically, and 30%–50% are linked to tuberous sclerosis, according to older literature. However, the vast majority of AMLs are now found incidentally on imaging examinations performed for other reasons and are small or relatively small.⁸ Approximately 5% of all renal AMLs are epithelioid and exhibit aggressive behavior that predisposes patients to recurrence, metastasis, and death.⁴¹

There is a risk of spontaneous renal bleeding in patients with renal AMLs, regardless of the origin or histological subtype, and this is particularly true for patients with medium or large-sized AMLs (greater than 4 cm).^{8,42-44} On CT imaging, classic AMLs typically appear as heterogeneous masses characterized by a mixture of macroscopic fat, hypervascular soft-tissue components, and intratumoral aneurysms.⁴⁵ In cases of spontaneous renal bleeding, the CT appearance of AMLs can be altered due to intratumoral and perirenal bleeding. MRI shows high signal intensity on T1-weighted images and signal loss on fat-saturated sequences, indicating the presence of macroscopic fat within the renal AMLs. On imaging, epithelioid AMLs are similar to other solid renal masses because they often lack macroscopic fat and appear as solid, soft-tissue masses. Therefore, they cannot be reliably distinguished from other renal tumors by imaging alone.⁸

Up to 25% of renal AMLs have the potential to rupture spontaneously, leading to perinephric bleeding.⁴¹ The presence of abnormal arteries in AMLs -arteries characterized

by a reduced elastin content- increases their susceptibility to aneurysm development.⁸ The likelihood of rupture, potentially resulting in life-threatening bleeding, increases with the size of tumoral aneurysms.⁴⁶ The frequency of tumor rupture and intratumoral bleeding is substantially influenced by the size of the tumor and the intratumoral aneurysms.⁸ Concerning tumor-related bleeding in AMLs, a larger tumor size (>4 cm) is directly correlated with a larger diameter of intratumoral aneurysms (>5 mm).⁴⁷ CT and/or MRI imaging modalities can depict the presence of the tumor associated with spontaneous renal bleeding produced by ruptured AMLs. However, it is important to note that the imaging appearance of the tumor may be modified due to bleeding.⁸ Furthermore, CT findings can aid in differentiating AMLs from liposarcoma, because the scans of AMLs are more likely to show small dimensions, lack of renal tissue, multiple vessels, aneurysmal dilation, vascular pedicle, bleeding, encapsulated periphery, and fatty lesions.⁴⁸ Kış et al.⁴⁹ have suggested the use of diffusion-weighted imaging and apparent diffusion coefficient (ADC) values for renal

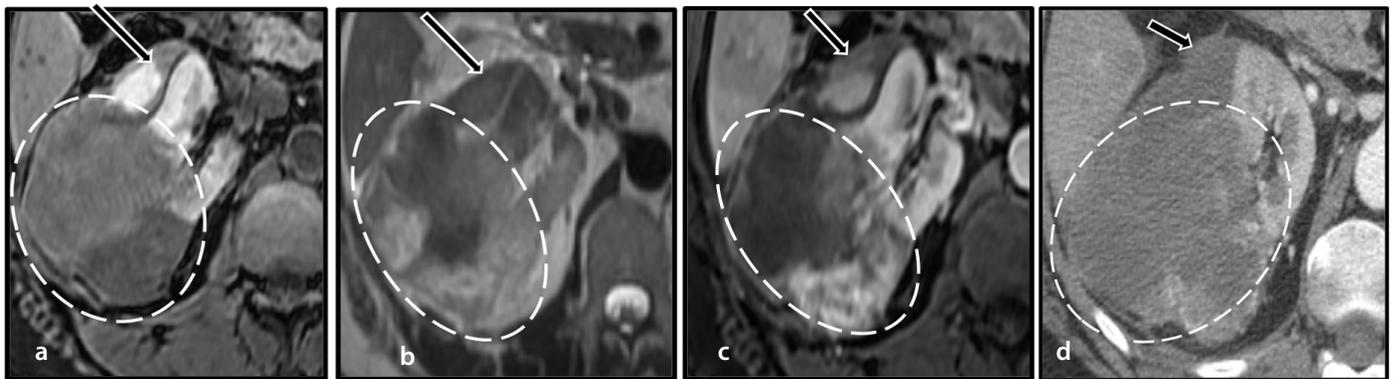


Figure 4. An axial MRI at the level of the right kidney (a-c) demonstrates a heterogeneous mass (white circle) with regions showing a non-enhancing, mildly T1 hyperintense signal, suggesting hemorrhage/necrosis. The findings are consistent with the patient's known renal cell carcinoma. An associated subacute subcapsular hematoma (black arrows) compresses the renal parenchyma. (d) An axial contrast-enhanced CT of the right kidney demonstrates the mass (white circle) and associated subcapsular hematoma (black arrow). MRI, magnetic resonance imaging; CT, computed tomography.

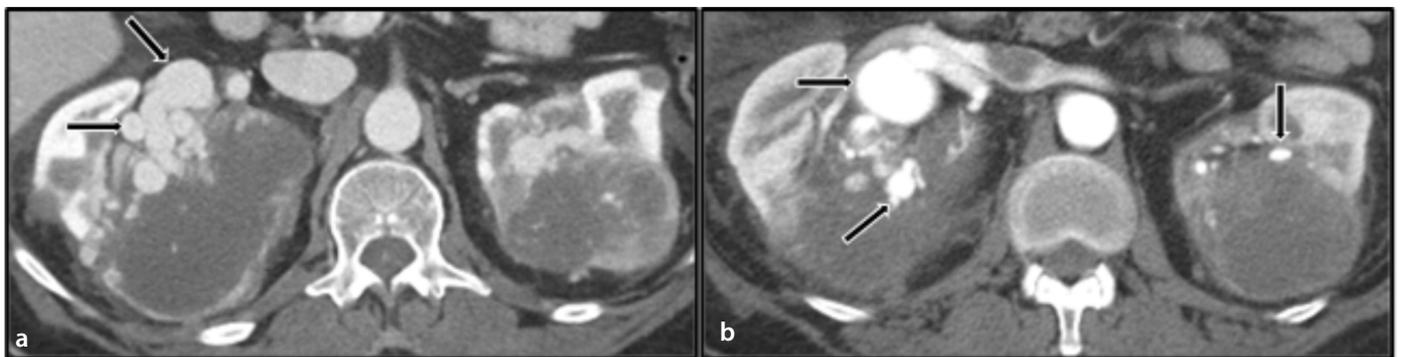


Figure 5. Axial CT images (a, b) demonstrate bilateral renal cell carcinomas with invasion of the renal hila. Multiple bilateral dilated vessels and aneurysms are present in the renal hilar regions (black arrows). Renal cell carcinoma is commonly associated with a variety of vascular anomalies, including renal artery aneurysms and arteriovenous fistulas. CT, computed tomography.

mass classification. They found that benign lesions demonstrate higher ADC values than the adjacent normal renal parenchyma; they also discovered that the ADC of malignant tumors is lower than that of the adjacent parenchyma. In their study, the ADC values were lowest in AMLs and oncocytomas.⁴⁹ Figures 6 and 7 illustrate bleeding secondary to AMLs.

TAE is the primary treatment modality for patients experiencing acute bleeding; it effectively achieves hemostasis in approximately 96% of cases, preventing emergency surgery.⁵⁰ To minimize bleeding events, prophylactic TAE reduces the size and vascularity of large AMLs.⁵¹ Prophylactic, selective TAE shows greater tumor reduction benefits in AMLs with a diameter of less than 7 cm, according to Hongyo et al.⁵² Angiographic observations (e.g., of vascularity and intratumoral aneurysms) allow the classification of AMLs into categories with minimal, moderate, and marked vascularity. Such classifications are valuable guides for determining

the appropriate prophylactic treatment.⁴⁷ In cases in which TAE fails to achieve the desired outcome, nephron-sparing surgery is typically performed. However, when more invasive interventions are warranted (e.g., in patients with a suspected malignancy or uncontrolled bleeding in an emergency scenario), total nephrectomy may be a treatment option for large AMLs.⁵³

Renovascular conditions in patients with cancer

Renal vein thrombosis

Renal vein thrombosis occurs when the primary renal veins or their branches become blocked owing to thrombosis or embolism. In rare instances, it can predispose a patient to spontaneous renal bleeding.⁵⁴ Neoplasms, primarily RCCs, preferentially infiltrate the renal vein, especially when they become large and more locally aggressive, resulting in renal vein thrombosis. In cases in which the tumor is significantly large, it can extend

into the inferior vena cava. Additionally, extrinsic compression resulting from the mass effect of the tumor can produce a prothrombotic milieu without direct vein invasion. In patients with renal vein thrombosis, renal bleeding likely results from renal parenchymal edema and necrosis, potentially leading to renal capsular rupture.^{54,55}

On cross-sectional imaging, a patient with renal vein thrombosis typically presents with an enlarged ipsilateral kidney and edema in the renal sinus and perirenal space.⁵⁶ A complete or partial renal vein thrombosis can be identified by the absence or reduction of renal parenchymal enhancement on nephrographic-phase images. On IV contrast-enhanced CT and MRI, the thrombus in the renal vein appears as a filling defect.⁵⁶ For patients with renal vein thrombosis, the first-line treatment involves medical management with anticoagulation therapy. In select cases, catheter-guided thrombectomy and filter placement in the inferior vena cava are the preferred treatment options.⁵⁴

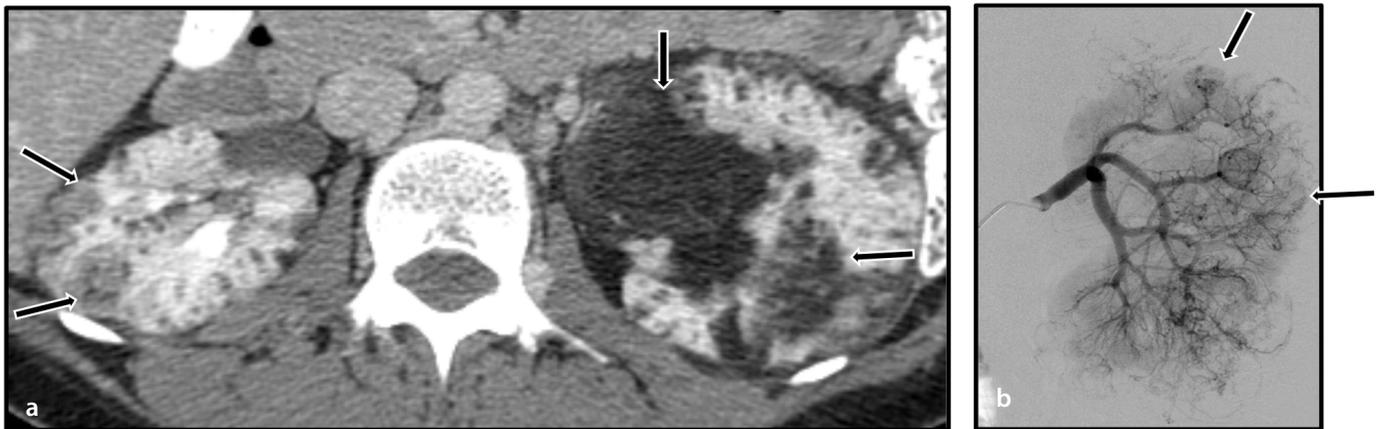


Figure 6. An axial contrast-enhanced CT scan (a) at the level of the kidneys demonstrates innumerable bilateral renal masses with intralesional macroscopic fat, diagnostic of angiomyolipomas (AMLs) (black arrows). Bilateral AMLs, the most common renal neoplasm associated with spontaneous hemorrhage, are seen in 95% of patients with tuberous sclerosis complex. They have the propensity to form pseudoaneurysms and present a higher risk of bleeding if they reach a size of >4 cm. (b) Digital subtraction angiography of the left kidney demonstrates opacification of the renal arteries with multiple tortuous tumor vessels and pseudoaneurysms (black arrows). This pattern is typical of multiple AMLs in the setting of tuberous sclerosis. CT, computed tomography.

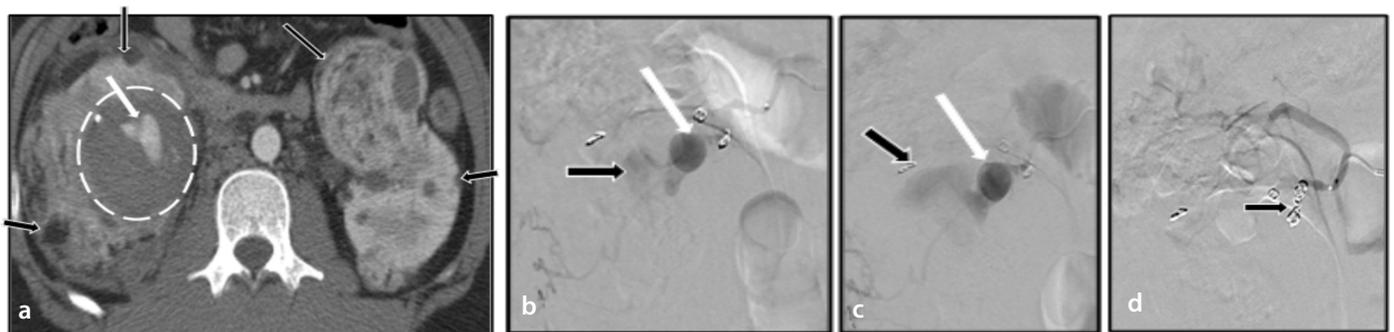


Figure 7. An axial contrast-enhanced CT (a) through the kidneys demonstrates numerous bilateral angiomyolipomas (black arrows). A hematoma is present in the right medial midpole (white circle), with central pseudoaneurysms (white arrow). (b, c) Digital subtraction angiography images of the right kidney demonstrate the pseudoaneurysm (white arrow) with active contrast extravasation (black arrow) into the retroperitoneum, which was subsequently treated with coil embolization (d) (black arrow). CT, computed tomography.

Renal artery abnormalities

Vascular anomalies related to malignant tumors, including renal artery aneurysms, pseudoaneurysms, or renal arteriovenous fistulas (AVFs), are among the main causes of spontaneous perinephric bleeding in patients with cancer.³⁶ Additionally, renal vascular injury may occur in a postprocedural kidney following renal biopsy or nephrostomy tube placement. Pseudoaneurysms demonstrate the classic yin-yang sign on Doppler ultrasound owing to the swirling motion of the blood in the sac. On CT angiography (CTA) and MR angiography (MRA), pseudoaneurysms are seen as outpouchings of the renal arteries that follow the blood pool (Figures 8-10). Renal AVFs appear on ultrasound as areas of aliasing artifacts. In rare cases, a color flash is observed, sometimes referred to as a visible thrill; it is caused by soft-tissue vibrations around the renal AVF that lead to a distinct color mosaic overlying

the adjacent tissue. Spectral analyses of AVFs demonstrate elevated peak systolic velocities, spectral broadening, reduced resistive indexes, and arterialization of the draining veins.² Identifying small renal AVFs might pose challenges during CTA and MRA, but when they are identified using these imaging techniques, they display abnormal arteriovenous connections with early draining venous channels.

Despite its low incidence, the rupture of a renal artery aneurysm or an AVF is associated with substantial mortality. The rupture can be confirmed using arterial-phase CT imaging, which displays active extravasation of contrast material with attenuation similar to that of the blood pool, as noted earlier.³

Although TAE can be used to successfully treat less severe episodes of bleeding, an immediate nephrectomy may be necessary in cases of severe bleeding.⁵⁷ Considering

the dangers of disease development and recurrence, TAE is the preferred therapeutic approach. Employing nephron-sparing interventions is recommended to preserve renal function. For unruptured large renal artery aneurysms greater than 2 cm in diameter, prophylactic embolization may be necessary.^{49,50} Endovascular intervention has emerged as a safe and effective method for managing renal artery aneurysms and may replace surgery as the primary therapeutic approach.⁵⁷⁻⁵⁹ Renal artery aneurysms less than 2 cm in diameter can be closely monitored through follow-up imaging. However, prompt embolization is imperative.^{59,60} Prophylactic TAE is recommended for renal artery aneurysms at high risk of rupture and for symptomatic patients.^{59,60}

Potential mimics of renal bleeding

Lymphoma affecting the kidneys is commonly observed in widespread non-Hodgkin

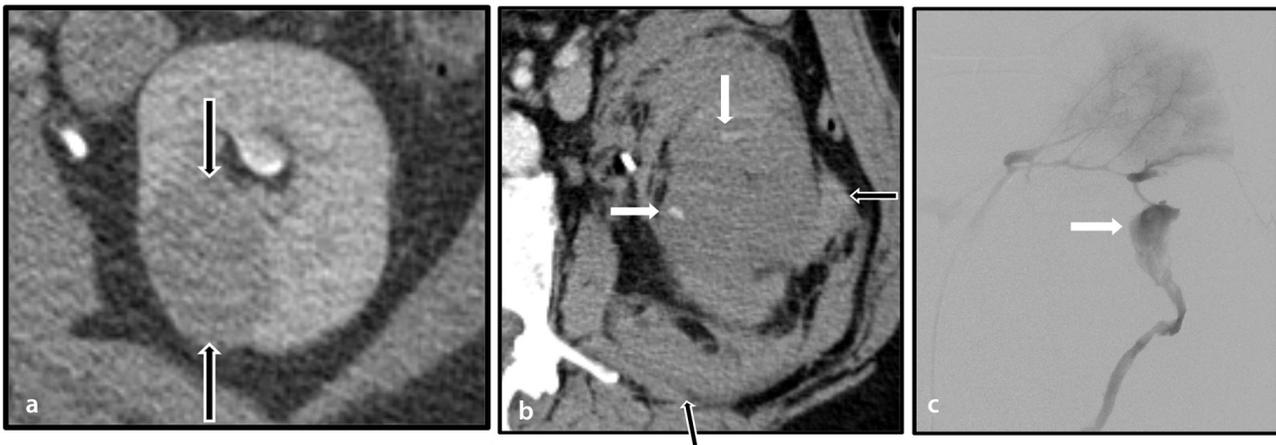


Figure 8. Axial CT images pre-biopsy (a) and post-biopsy (b) of a right renal mass (black arrows) demonstrate a large amount of perinephric hemorrhage post-biopsy (black arrows). Foci of contrast blush may represent pseudoaneurysms and/or active bleeding (white arrows). (c) An angiogram of the left kidney demonstrates active contrast extravasation in the mid kidney (white arrow). CT, computed tomography.

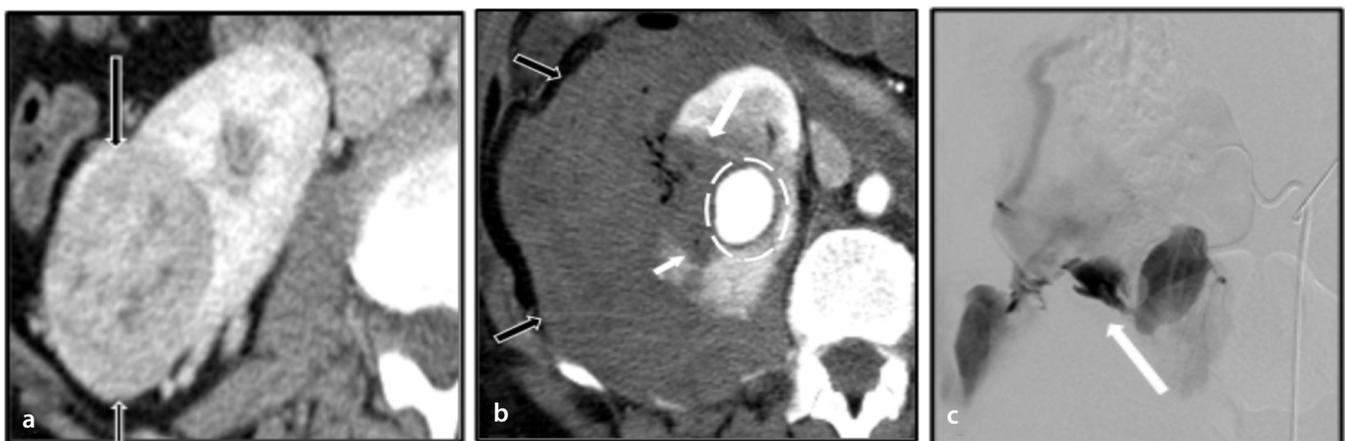


Figure 9. Axial contrast-enhanced CT images pre- (a) and post-biopsy (b) of a right renal mass (black arrows in a) demonstrate the development of a large pseudoaneurysm status post-biopsy (white circle), right renal parenchymal laceration (white arrows in b), and a large perinephric hematoma (black arrows in b). An angiogram (c) demonstrates an active contrast extravasation (white arrow) that was subsequently treated with coil embolization. CT, computed tomography.

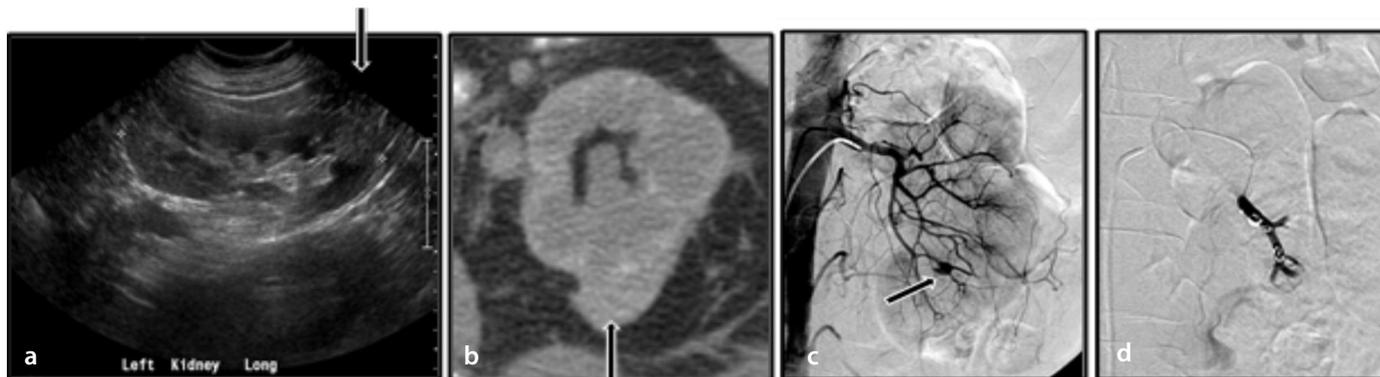


Figure 10. (a) A grayscale ultrasound of the left kidney after a non-targeted renal biopsy demonstrates a hyperechoic mass-like area in the left kidney lower pole (black arrow). (b) An axial non-contrast CT confirms a small subcapsular hematoma (black arrow). (c, d) A digital subtraction angiography of the left kidney demonstrates a small pseudoaneurysm (black arrow in c), which was subsequently embolized (d). CT, computed tomography.



Figure 11. (a) An axial image from a contrast-enhanced CT scan in a 68-year-old man with a biopsy-proven perinephric lymphoma demonstrates a high attenuation mass (58 HU) surrounding the lower pole of the left kidney. This is also seen in the coronal reformatted image (b). The pre-contrast scan (c) at the same level demonstrates that the mass is of lower attenuation (32 HU), and the diffuse enhancement is suggestive of a mass rather than a hemorrhage. CT, computed tomography; HU, Hounsfield units.

lymphoma and is more prevalent in patients with a weakened immune system.^{3,61} The appearance of renal lymphoma on imaging depends on the growth pattern and histology of the tumor.⁶¹ The typical imaging presentation of renal lymphoma consists of multiple parenchymal masses of variable sizes, usually ranging from 1 to 4.5 cm in diameter.⁶¹⁻⁶⁵ This pattern is observed in approximately 50%–60% of cases. Although these masses are often bilateral, they can also be unilateral.⁶¹⁻⁶⁵ On unenhanced CT, these masses exhibit soft-tissue characteristics with slightly higher attenuation than the surrounding renal parenchyma.⁶¹ Calcifications within the tumoral foci are infrequent. It is crucial to employ IV contrast-enhanced CT scanning during the nephrographic phase because many tumoral foci are small and primarily involve the medullary region of the kidneys, resulting in subtle cortical deformity. Lymphomatous deposits exhibit less enhancement than normal renal tissues and appear as relatively homogeneous masses with lower attenuation than the surrounding cortex.⁶¹ Perinephric lymphoma can exhibit various attenuation values on CT and can mimic perinephric hematoma

by encasing the perirenal soft tissue around the blood vessels (Figure 11). It is critical to accurately distinguish fluid, blood products, and inflammation from soft-tissue masses. Homogeneous, relatively mild enhancement is characteristic of perirenal lymphoma, and the perinephric fat may separate the tumor from the kidney. Moreover, lymphoma can be staged using positron emission tomography and CT.^{3,61}

In conclusion, renal bleeding presents a substantial challenge in managing patients with cancer, particularly those with RCC or those undergoing anticoagulant therapy. Diagnostic cross-sectional imaging techniques, including ultrasonography, CT, and MRI, are crucial for identifying and evaluating the etiology of renal bleeding. SWI is useful in detecting and characterizing renal masses associated with renal bleeding, especially microhemorrhages. Therapeutic options range from the observation of stable patients to TAE for active bleeding control. Treatment options for RCC vary depending on the disease stage: ablation, partial nephrectomy, or radical nephrectomy. Prospective studies comparing different treatment modalities

(e.g., TAE, nephron-sparing surgery, and total nephrectomy) in specific patient subgroups provide valuable insights into personalized treatment approaches. Addressing these gaps can enhance the understanding of renal bleeding in patients with cancer, improve evidence-based diagnosis, management, and prevention recommendations, and ultimately enhance patient outcomes and quality of life.

Acknowledgments

We thank Laura L. Russell, scientific editor of Research Medical Library, for editing this article.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Pereira J, Phan T. Management of bleeding in patients with advanced cancer. *Oncologist*. 2004;9(5):561-570. [\[CrossRef\]](#)

2. Lyman GH, Eckert L, Wang Y, Wang H, Cohen A. Venous thromboembolism risk in patients with cancer receiving chemotherapy: a real-world analysis. *Oncologist*. 2013;18(12):1321-1329. [\[CrossRef\]](#)
3. Hammond NA, Lostumbo A, Adam SZ, et al. Imaging of adrenal and renal hemorrhage. *Abdom Imaging*. 2015;40(7):2747-2760. [\[CrossRef\]](#)
4. Bosniak MA. Spontaneous subcapsular and perirenal hematomas. *Radiology*. 1989;172(3):601-602. [\[CrossRef\]](#)
5. Furlan A, Fakhran S, Federle MP. Spontaneous abdominal hemorrhage: causes, CT findings, and clinical implications. *AJR Am J Roentgenol*. 2009;193(4):1077-1087. [\[CrossRef\]](#)
6. Mao Y, De Oliveira IS, Hedgire S, Praputtam D, Harisinghani M. Aetiology, imaging features, and evolution of spontaneous perirenal haemorrhage. *Clin Radiol*. 2017;72(2):175. [\[CrossRef\]](#)
7. Diaz JR, Agriantoni DJ, Aguila J, Calleros JE, Ayyappan AP. Spontaneous perirenal hemorrhage: what radiologists need to know. *Emerg Radiol*. 2011;18(4):329-334. [\[CrossRef\]](#)
8. Katabathina VS, Katre R, Prasad SR, Surabhi VR, Shanbhogue AK, Sunnapwar A. Wunderlich syndrome: cross-sectional imaging review. *J Comput Assist Tomogr*. 2011;35(4):425-433. [\[CrossRef\]](#)
9. Gong M, Liu Z, Su H, Zhao B, Kong J, He X. Urgent transcatheter arterial embolization for Wunderlich syndrome with hypovolemic shock secondary to ruptured renal angiomyolipoma. *Front Surg*. 2021;8:704478. [\[CrossRef\]](#)
10. Hebert LA, Betts JA, Sedmak DD, Cosio FG, Bay WH, Carlton S. Loin pain-hematuria syndrome associated with thin glomerular basement membrane disease and hemorrhage into renal tubules. *Kidney Int*. 1996;49(1):168-173. [\[CrossRef\]](#)
11. Antonescu O, Duhamel M, Di Giacinto B, Spain J. Spontaneous renal hemorrhage: a case report and clinical protocol. *Cureus*. 2021;13(6):e15547. [\[CrossRef\]](#)
12. Naik U, Amin SE, Gehris BT, Wang W, Hunter RL. Papillary renal cell carcinoma with massive hemorrhage: An autopsy case report. *Am J Clin Pathol*. 2022;158(Suppl 1):32. [\[CrossRef\]](#)
13. Lu J, Zhou W, Wei X, Wang K, Zhou L, Xu X. Clear cell renal cell carcinoma metastasized to the ampulla of Vater 16 years after nephrectomy—a rare case. *Diagnostics (Basel)*. 2022;12(3):571. [\[CrossRef\]](#)
14. Turner NN, Turner NN, Lameire N, et al. eds. Oxford Textbook of Clinical Nephrology. Oxford University Press; 2015. [\[CrossRef\]](#)
15. Johnstone C, Rich SE. Bleeding in cancer patients and its treatment: a review. *Ann Palliat Med*. 2018;7(2):265-273. [\[CrossRef\]](#)
16. Swensen SJ, McLeod RA, Stephens DH. CT of extracranial hemorrhage and hematomas. *Am J Roentgenol*. 1984;143(4):907-912. [\[CrossRef\]](#)
17. Kawashima A, Sandler CM, Ernst RD, et al. Imaging of nontraumatic hemorrhage of the adrenal gland. *Radiographics*. 1999;19(4):949-963. [\[CrossRef\]](#)
18. Caskey CI. Ultrasound techniques for evaluating renal masses, renal obstruction, and other upper tract pathology. *Ultrasound Q*. 2000;16(1):23-39. [\[CrossRef\]](#)
19. Semelka RC. Abdominal-Pelvic MRI. Fourth. 2016. [\[CrossRef\]](#)
20. Ren JH, Ma N, Wang SY, et al. Rationale and study design for one-stop assessment of renal artery stenosis and renal microvascular perfusion with contrast-enhanced ultrasound for patients with suspected renovascular hypertension. *Chin Med J (Engl)*. 2019;132(1):63-68. [\[CrossRef\]](#)
21. Çıldıg MB, Gök M, Abdullayev O. Pre-procedural shear wave elastography on prediction of hemorrhage after percutaneous real-time ultrasound-guided renal biopsy. *Radiol Med*. 2020;125(8):784-789. [\[CrossRef\]](#)
22. Gomella PT, Telfer S, Lebatschi A, et al. MP42-15 Timing, incidence and management of delayed bleeding after partial nephrectomy in patients at risk for recurrent, bilateral, multifocal renal tumors. *J Urol*. 2024;201(Suppl 4):e613. [\[CrossRef\]](#)
23. Al Awwad A, Alshubaili H, Mohamed AY, Abuaz S. Incidental finding of bilateral ovarian and renal veins thromboses post cesarean hysterectomy complicated by ureteric injury: First case presentation. *Urol Ann*. 2022;14(3):288-291. [\[CrossRef\]](#)
24. Kato T, Masui K, Yoshida T, et al. Acquired hemophilia A developing at bilateral renal bleeding: a case report. *Hinyokika Kyo*. 2009;55(4):215-218. [\[CrossRef\]](#)
25. Tsurusaki T, Maruta N, Iwasaki S, Saito Y. Acquired haemophilia A presenting as acute renal failure secondary to bilateral renal bleeding. *BJU Int*. 2002;90(4):40-41. [\[CrossRef\]](#)
26. Chang JM, Moon WK, Cho N, et al. Clinical application of shear wave elastography (SWE) in the diagnosis of benign and malignant breast diseases. *Breast Cancer Res Treat*. 2011;129(1):89-97. [\[CrossRef\]](#)
27. Moschouris H, Papadaki MG, Spanomanolis N, Stamatiou K, Malagari K. Percutaneous thrombin injection under contrast-enhanced ultrasound guidance to control active extravasation not associated with pseudoaneurysm. *Diagn Interv Radiol*. 2023;29(4):632-637. [\[CrossRef\]](#)
28. Massimo T, Emanuela P. Iatrogenic renal pseudoaneurysm following laparoscopic tumorectomy: CT diagnosis and interventional treatment. [\[CrossRef\]](#)
29. Haacke EM, Mittal S, Wu Z, Neelavalli J, Cheng YC. Susceptibility-weighted imaging: technical aspects and clinical applications, part 1. *AJNR Am J Neuroradiol*. 2009;30(1):19-30. [\[CrossRef\]](#)
30. Beauchamp MH, Ditchfield M, Babl FE, et al. Detecting traumatic brain lesions in children: CT versus MRI versus susceptibility weighted imaging (SWI). *J Neurotrauma*. 2011;28(6):915-927. [\[CrossRef\]](#)
31. Sehgal V, Delproposto Z, Haacke EM, et al. Clinical applications of neuroimaging with susceptibility-weighted imaging. *J Magn Reson Imaging*. 2005;22(4):439-450. [\[CrossRef\]](#)
32. Haacke EM, Xu Y, Cheng YCN, Reichenbach JR. Susceptibility weighted imaging (SWI). *Magn Reson Med*. 2004;52:612-618. [\[CrossRef\]](#)
33. Roubidoux MA. MR imaging of hemorrhage and iron deposition in the kidney. *Radiographics*. 1994;14(5):1033-1044. [\[CrossRef\]](#)
34. Xing W, He X, Kassir MA, et al. Evaluating hemorrhage in renal cell carcinoma using susceptibility weighted imaging. *PLoS One*. 2013;8(2):e57691. [\[CrossRef\]](#)
35. Li D, Pua BB, Madoff DC. Role of embolization in the treatment of renal masses. *Semin Intervent Radiol*. 2014;31(1):70-81. [\[CrossRef\]](#)
36. Zhang JQ, Fielding JR, Zou KH. Etiology of spontaneous perirenal hemorrhage: a meta-analysis. *J Urol*. 2002;167(4):1593-1596. [\[CrossRef\]](#)
37. Newatia A, Khatri G, Friedman B, Hines J. Subtraction imaging: applications for nonvascular abdominal MRI. *AJR Am J Roentgenol*. 2007;188(4):1018-1025. [\[CrossRef\]](#)
38. Morris CS, Baerlocher MO, Dariushnia SR, et al. Society of interventional radiology position statement on the role of percutaneous ablation in renal cell carcinoma: endorsed by the Canadian Association for Interventional Radiology and the Society of Interventional Oncology. *J Vasc Interv Radiol*. 2020;31(2):189-194. [\[CrossRef\]](#)
39. McDougal WS, Kursh ED, Persky L. Spontaneous rupture of the kidney with perirenal hematoma. *J Urol*. 1975;114(2):181-184. [\[CrossRef\]](#)
40. Prasad SR, Sahani DV, Mino-Kenudson M, et al. Neoplasms of the perivascular epithelioid cell involving the abdomen and the pelvis: cross-sectional imaging findings. *J Comput Assist Tomogr*. 2007;31(5):688-696. [\[CrossRef\]](#)
41. Katabathina VS, Vikram R, Nagar AM, Tamboli P, Menias CO, Prasad SR. Mesenchymal neoplasms of the kidney in adults: imaging spectrum with radiologic-pathologic correlation. *Radiographics*. 2010;30(6):1525-1540. [\[CrossRef\]](#)
42. Moratalla MB. Wunderlich's syndrome due to spontaneous rupture of large bilateral angiomyolipomas. *Emerg Med J*. 2009;26(1):72. [\[CrossRef\]](#)
43. Nabi N, Shaikh AH, Khalid SE. Tuberous sclerosis with bilateral renal angiomyolipoma and Wunderlich's syndrome. *J Coll Physicians Surg Pak*. 2007;17(11):706-707. [\[CrossRef\]](#)

44. Arrabal-Polo MA, Arrabal-Martin M, Palao-Yago F, et al. Wunderlich syndrome from a malignant epithelioid angiomyolipoma. *Urol J*. 2009;6(1):50-53. [\[CrossRef\]](#)
45. Sherman JL, Hartman DS, Friedman AC, Madewell JE, Davis CJ, Goldman SM. Angiomyolipoma: computed tomographic-pathologic correlation of 17 cases. *Am J Roentgenol*. 1981;137(6):1221-1226. [\[CrossRef\]](#)
46. Eble JN. Angiomyolipoma of kidney. *Semin Diagn Pathol*. 1998;15(1):21-40. [\[CrossRef\]](#)
47. Yamakado K, Tanaka N, Nakagawa T, Kobayashi S, Yanagawa M, Takeda K. Renal angiomyolipoma: Relationships between tumor size, aneurysm formation, and rupture. *Radiology*. 2002;225(1):78-82. [\[CrossRef\]](#)
48. Woo S, Kim SY, Cho JY, Kim SH, Lee MS. Exophytic renal angiomyolipoma and perirenal liposarcoma: revisiting the role of CT for differential diagnosis. *Acta Radiol*. 2016;57(2):249-255. [\[CrossRef\]](#)
49. Kış N, Düzkalır HG, Ağaçlı MO, Erok B, Kılıçoğlu ZG. Contribution of diffusion weighted MRI to the differential diagnosis of renal masses. *Eur J Clin Exp Med*. 2022;20(1):44-48. [\[CrossRef\]](#)
50. Wang C, Yang M, Tong X, et al. Transarterial embolization for renal angiomyolipomas: a single centre experience in 79 patients. *J Int Med Res*. 2017;45(2):706-713. [\[CrossRef\]](#)
51. Planché O, Correas JM, Mader B, Joly D, Méjean A, Hélénon O. Prophylactic embolization of renal angiomyolipomas: evaluation of therapeutic response using CT 3D volume calculation and density histograms. *J Vasc Interv Radiol*. 2011;22(10):1388-1395. [\[CrossRef\]](#)
52. Hongyo H, Higashihara H, Osuga K, et al. Efficacy of prophylactic selective arterial embolization for renal angiomyolipomas: identifying predictors of 50% volume reduction. *CVIR Endovasc*. 2020;3(1):84. [\[CrossRef\]](#)
53. Faddegon S, So A. Treatment of angiomyolipoma at a tertiary care centre: the decision between surgery and angioembolization. *Can Urol Assoc J*. 2011;5(6):138-141. [\[CrossRef\]](#)
54. Asghar M, Ahmed K, Shah SS, Siddique MK, Dasgupta P, Khan MS. Renal vein thrombosis. *Eur J Vasc Endovasc Surg*. 2007;34(2):217-223. [\[CrossRef\]](#)
55. Ahn T, Roberts MJ, Navaratnam A, Chung E, Wood S. Changing etiology and management patterns for spontaneous renal hemorrhage: a systematic review of contemporary series. *Int Urol Nephrol*. 2017;49(11):1897-1905. [\[CrossRef\]](#)
56. Kawashima A, Sandler CM, Ernst RD, Tamm EP, Goldman SM, Fishman EK. CT evaluation of renovascular disease. *RadioGraphics*. 2000;20(5):1321-1340. [\[CrossRef\]](#)
57. Zapzalka DM, Thompson HA, Borowsky SS, Coleman-Stenson CC, Mahowald ML, O'Connell KJ. Polyarteritis nodosa presenting as spontaneous bilateral perinephric hemorrhage: management with selective arterial embolization. *J Urol*. 2000;164(4):1294-1295. [\[CrossRef\]](#)
58. Stanson AW, Friese JL, Johnson CM, et al. Polyarteritis nodosa: spectrum of angiographic findings. *Radiographics*. 2001;21(1):151-159. [\[CrossRef\]](#)
59. Eldem G, Erdoğan E, Peynircioğlu B, Arat A, Balkancı F. Endovascular treatment of true renal artery aneurysms: a single center experience. *Diagn Interv Radiol*. 2019;25(1):62-70. [\[CrossRef\]](#)
60. Noshier JL, Chung J, Brevetti LS, Graham AM, Siegel RL. Visceral and renal artery aneurysms: a pictorial essay on endovascular therapy. *Radiographics*. 2006;26(6):1687-1704. [\[CrossRef\]](#)
61. Sheth S, Ali S, Fishman E. Imaging of renal lymphoma: patterns of disease with pathologic correlation. *Radiographics*. 2006;26(4):1151-1168. [\[CrossRef\]](#)
62. Heiken JP, Gold RP, Schnur MJ, King DL, Bashist B, Glazer HS. Computed tomography of renal lymphoma with ultrasound correlation. *J Comput Assist Tomogr*. 1983;7(2):245-250. [\[CrossRef\]](#)
63. Urban BA, Fishman EK. Renal lymphoma: CT patterns with emphasis on helical CT. *Radiographics*. 2000;20:197-212. [\[CrossRef\]](#)
64. Chepuri NB, Strouse PJ, Yanik GA. CT of renal lymphoma in children. *Am J Roentgenol*. 2003;180(2):429-431. [\[CrossRef\]](#)
65. Cohan RH, Dunnick NR, Leder RA, Baker ME. Computed tomography of renal lymphoma. *J Comput Assist Tomogr*. 1990;14(6):933-938. [\[CrossRef\]](#)



Diagnostic performance of the O-RADS MRI system for magnetic resonance imaging in discriminating benign and malignant adnexal lesions: a systematic review, meta-analysis, and meta-regression

Gülsüm Kılıçkap

Ankara Bilkent City Hospital, Clinic of Radiology,
Ankara, Türkiye

PURPOSE

After the introduction of the Ovarian-Adnexal Reporting and Data System (O-RADS) for magnetic resonance imaging (MRI), several studies with diverse characteristics have been published to assess its diagnostic performance. This systematic review and meta-analysis aimed to assess the diagnostic performance of O-RADS MRI scoring for adnexal masses, accounting for the risk of selection bias.

METHODS

The PubMed, Scopus, Web of Science, and Cochrane databases were searched for eligible studies. Borderline or malignant lesions were considered malignant. All O-RADS MRI scores ≥ 4 were considered positive. The quality of the studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. The pooled sensitivity, specificity, and likelihood ratio (LR) values were calculated, considering the risk of selection bias.

RESULTS

Fifteen eligible studies were found, and five of them had a high risk of selection bias. Between-study heterogeneity was low-to-moderate for sensitivity but substantial for specificity (I^2 values were 35.5% and 64.7%, respectively). The pooled sensitivity was significantly lower in the studies with a low risk of bias compared with those with a high risk of bias (93.0% and 97.5%, respectively; $P = 0.043$), whereas the pooled specificity was not different (90.4% for the overall population). The negative and positive LRs were 0.08 [95% confidence interval (CI) 0.05–0.11] and 10.0 (95% CI 7.7–12.9), respectively, for the studies with low risk of bias and 0.03 (95% CI 0.01–0.10) and 10.3 (95% CI 3.8–28.3), respectively, for those with high risk of bias.

CONCLUSION

The overall diagnostic performance of the O-RADS system is very high, particularly for ruling out borderline/malignant lesions, but with a moderate ruling-in potential. Studies with a high risk of selection bias lead to an overestimation of sensitivity.

CLINICAL SIGNIFICANCE

The O-RADS system demonstrates considerable diagnostic performance, particularly in ruling out borderline or malignant lesions, and should routinely be used in practice. The high between-study heterogeneity observed for specificity suggests the need for improvement in the consistent characterization of the benign lesions to reduce false positive rates.

KEYWORDS

Adnexal mass, ovarian cancer, Ovarian-Adnexal Reporting and Data System, magnetic resonance imaging, meta-analysis, systematic review

Corresponding author: Gülsüm Kılıçkap

E-mail: gklickap@yahoo.com.tr

Received 28 March 2024; revision requested 12 May 2024;
accepted 29 May 2024.



Epub: 08.07.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242784

Ovarian cancer is one of the leading causes of cancer-related death in women. Accurate characterization of adnexal masses is crucial for correct diagnosis and the prevention of unnecessary surgery. Transvaginal ultrasound is the first-line diagnostic method due to its relative affordability and widespread availability. However, magnetic resonance imaging (MRI) offers several advantages over ultrasound, including better characterization and visualization of the origin of the mass and higher resolution. Recently, the American College of Radiology (ACR) proposed a method—the Ovarian-Adnexal Reporting and Data System (O-RADS) for MRI (O-RADS MRI)—to standardize the analysis of adnexal masses.¹

Following the introduction of the O-RADS MRI system, several studies assessing its validity, including a small number of meta-analyses, have been published.²⁻¹⁸ These studies have diverse characteristics and were conducted at single or multicenter sites with varying levels of expertise and patient volumes. Given the increasing number of studies on the diagnostic value of the O-RADS MRI score in recent years and the potential heterogeneity among them, this study aims to conduct an updated systematic review and meta-analysis of these studies by taking into consideration their risk of bias. Therefore, this systematic review, meta-analysis, and meta-regression aim to assess the diagnostic value of the O-RADS MRI system in assessing adnexal masses and to reveal the rule-in and rule-out potential of borderline or malignant adnexal masses. Unlike other meta-analyses, the objective is to calculate the pooled sensitivity and specificity of O-RADS according to whether the studies included in the analysis are at high or low risk of patient selection bias.

Methods

This systematic review and meta-analysis were prepared and presented in accordance

Main points

- The diagnostic performance of the Ovarian-Adnexal Reporting and Data System (O-RADS) for magnetic resonance imaging (MRI) system is very high.
- The O-RADS MRI system is valuable in ruling out borderline or malignant adnexal masses.
- The ruling-in potential of the O-RADS system is moderate.
- Studies with a high risk of bias lead to overestimation of the sensitivity.

with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) recommendations.¹⁹ Since the data were obtained from manuscripts, informed consent was not required, and ethics committee approval was waived.

Study population and research question

The study population and research question were structured according to the PICO format (P- population, I- intervention/index test, C- comparator/reference test, and O- outcome) and included patients who underwent pelvic MRI examinations for adnexal masses. Studies were excluded if any of the following criteria were present: 1) absence of the standard reference test, 2) O-RADS scoring using non-MRI methods, 3) case-control studies or inappropriate selection or exclusion, 4) studies in which only specific lesions (such as only cystic lesions) or a specific O-RADS category were evaluated, and 5) studies assessing O-RADS scoring with non-contrast MRI, as this is not included in the standards proposed by the original O-RADS MRI scoring.

The index test was based on O-RADS MRI scoring, in which a score ≥ 4 was considered positive, and its diagnostic value was compared with the reference standard test result.

The comparison was made using the pathology or reasonable follow-up as a reference test. Borderline or malignant lesions were considered malignant.

The outcomes were diagnostic performance measures that included sensitivity, specificity, summary receiver operating characteristics (SROC) curve, and likelihood ratios (LRs).

Searching and extracting the data

The PubMed, Scopus, Web of Science, and Cochrane Central Register of Controlled Trials databases were searched for eligible studies on December 29, 2023. The search terms used in the PubMed database were as follows: (“Ovarian”[Title/Abstract] OR “adnexal”[Title/Abstract] OR “pelvic”[Title/Abstract]) AND (“Cancer”[Title/Abstract] OR “malignan*”[Title/Abstract] OR “tumor”[Title/Abstract] OR “mass*”[Title/Abstract] OR “lesion”[Title/Abstract]) AND (“O-RADS”[Title/Abstract] OR “ORADS”[Title/Abstract] OR “Ovarian adnexal reporting and data system”[Title/Abstract]) AND (“magnetic resonance imaging”[MeSH Terms] OR (“Magnetic Resonance”[Title/Abstract] OR “MRI”[Title/Abstract] OR “MR”[Ti-

tle/Abstract])).” The same search terms were used in other databases with slight modifications to conform to the database’s rules. No restriction (including language) was applied to the database searches.

The selection of the eligible studies and the number of manuscripts obtained from each database are shown in the PRISMA flowchart (Figure 1). After removing duplicated manuscripts, the titles and abstracts were initially screened for eligible studies, followed by a subsequent screening of the full-text manuscripts. One of the eligible studies was published in Chinese, and the full-text manuscript could not be obtained.⁶ However, the abstract contained the required information to conduct a diagnostic meta-analysis; therefore, no eligible studies were discarded in the analysis.

For studies in which more than one investigator evaluated the MRI scores, the measurements of the most experienced investigator were used. If the most experienced investigator made more than one measurement, the first measurement was included in the analysis.

Lesion-based O-RADS MRI scoring was analyzed. Since lesion-based data could not be obtained in one study,¹⁶ data for patient-based assessments given in the article were included in the analysis.

Assessment of the quality of the included studies

The quality of each eligible study was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.²⁰ This tool includes four domains (patient selection, index test, reference standard, and flow and timing) to evaluate the risk of bias and applicability of primary diagnostic accuracy studies. Each study was scored for both risk of bias and concern for applicability as high, unclear, or low. Critical appraisal of the selected studies was conducted by two reviewers independently, and any discrepancies were resolved through consensus.

Certainty of evidence

Certainty of evidence was assessed using the Grading of Recommendations, Assessment, Development, and Evaluations tool.²¹ As the pooled sensitivity was significantly different for the studies with low and high risk of bias, certainty of evidence was provided for sensitivity for the studies with low risk of bias. However, as there was no significant

difference in specificities between the studies with low and high risk of bias, certainty of evidence for specificity was given for the overall group.

Statistical analysis

Using the cut-off value of O-RADS scores ≥ 4 , the number of the true positive, false positive, true negative, and false negative results were recorded, and sensitivity and specificity values were calculated. The data were pooled using bivariate random effects model meta-analysis and presented as a forest plot and the SROC curve. Random effects meta-regression analysis was performed by including the variable of patient selection bias, which was obtained with the QUADAS-2 tool (Model 1). In the case of significant relative sensitivity or specificity for the selection bias categories of high-risk versus low-risk group, the corresponding diagnostic measure was presented separately. Then, the age and the proportion of borderline or malignant cases were included in the meta-regression (Model 2). The mean (or median) age was not presented in the two studies;^{6,12} therefore, these missing values were replaced with the overall mean obtained from the remaining studies. The performances of Model 1 and Model 2 were compared using the LR test.

The bivariate random effects model uses an unstructured variance-covariance matrix as the default method. The model was also run with the independent variance-covariance matrix to test whether the simpler (parsimonious) model is appropriate. Then the two models with different matrix structures were compared using the Akaike information criteria (AIC). As the model with an unstructured variance-covariance matrix had a lower (better) value of AIC, it is presented here.

The pooled estimates for positive and negative LR for O-RADS MRI scoring in diagnosing borderline or malignant lesions were calculated. It is generally accepted that a positive LR of >10 and negative LR of <0.1 are valuable in confirming or excluding the disease, respectively, while values of 5–10 and 0.1–0.2, respectively, are moderately effective in this regard.²² The point estimates and their 95% confidence interval (CI) for positive and negative LR values (the LR matrix) were plotted to visually assess the confirming or excluding potential of O-RADS MRI scoring. Additionally, Fagan's nomogram was plotted to calculate the post-test probability of having borderline or malignant lesions. The mean and median values of the borderline/

malignancy lesion proportions were 25.5% and 24.4%, respectively. Therefore, Fagan's nomogram was plotted using the pre-test probability value of 25% for borderline/malignant lesions.

The between-study heterogeneity was assessed using the I^2 statistics proposed by Zhou and Dendukuri²³ and also with Cochran's Q statistics and its P value. The I^2 parameter has values of 0%–100%; the values $>50\%$ and $>75\%$ are considered moderate and severe heterogeneity, respectively. Publication bias was assessed with a funnel plot proposed by Deeks et al.²⁴ and tested statistically. A P value of <0.05 was considered significant. Statistical analyses were performed using Stata version 17 (StataCorp, TX, USA), and the "metadata" and "midas" packages were used for the analysis.

Results

Fifteen eligible studies were found. The PRISMA flowchart for the selection of these studies is provided in Figure 1, and the characteristics of these studies are given in Table 1. Five of the studies were considered to have a high risk of patient selection bias based on the QUADAS-2 report. Figure 2 summarizes the interpretation with the QUADAS report,

and the details are given in Supplementary Table 1.

The mean age ranged from 35.9 to 57 years, with a mean and standard deviation of 46.1 ± 7.1 years and a median and interquartile range (IQR) of 48.7 (40.0–50.8) years. The median proportion of borderline or malignant lesions was 25.2% (IQR 13.5%–38.8%). For the studies with a low risk of selection bias, this ranged from 11.2% to 52.9%, with a mean \pm standard deviation of 25.5 ± 13.3 years and a median and IQR of 24.4 (13.5–31.4) years; for those with a high risk of bias, the range was from 11.8% to 65.4%, with a mean \pm standard deviation of 32.4 ± 22.1 years and a median and IQR of 28.3 (14.3–42.0) years.

Meta-analysis of the eligible studies

The sensitivity values ranged from 81% to 100%, while specificity values ranged from 58.0% to 97.9%. In the analysis stratified for the risk of selection bias, there was low-to-moderate between-study heterogeneity for diagnostic sensitivity [I^2 values were 35.5% for the overall group and 39.8% and 14.2% for the studies with low risk and high risk of selection bias, respectively]. The corresponding Cochran's Q statistics (P values) were 21.71 ($P = 0.085$), 14.95 ($P = 0.092$),

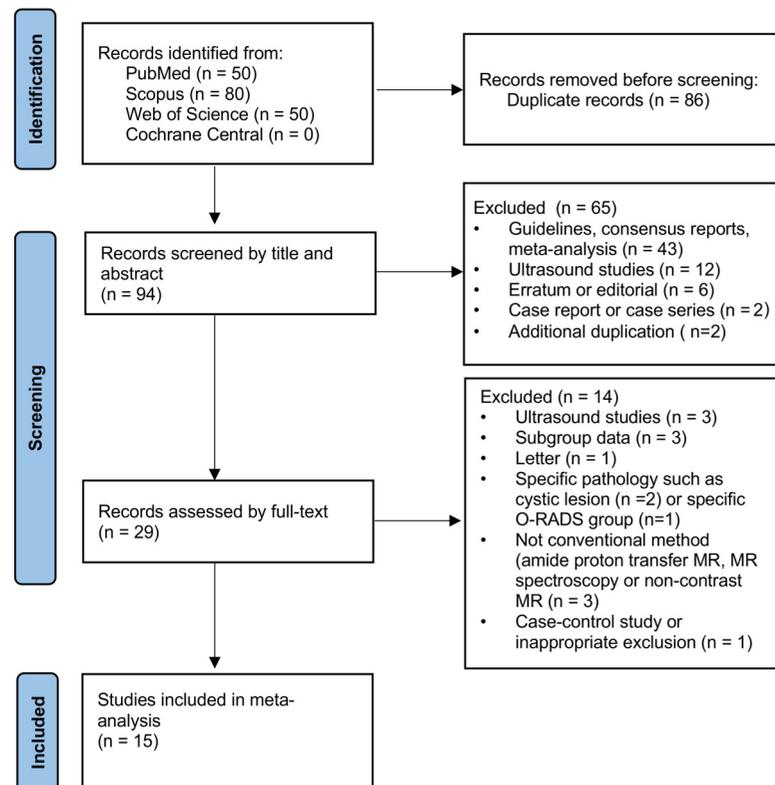


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses flowchart for selection of eligible studies. O-RADS, Ovarian-Adnexal Reporting and Data System; MR, magnetic resonance.

Table 1. The characteristics of the included studies

Author	Publication year	Screening period	Number of patients	Number of lesions	Percentage of the borderline or malignant lesions (%)	Mean or median age (years)	Reference standard
Aslan and Tosun ²	2023	Jan 2018-June 2020	200	237	11.8	56.3	Pathology or 24-month follow-up
Bang et al. ³	2022	Jan 2014-July 2020 and Jan 2010-July 2020	110	110	54.6	50.8	Pathology
Basu et al. ⁴	2022	April 2020-June 2021	42	46	28.3	35.9	Pathology or 4-month follow-up
Campos et al. ⁵	2023	Mar 5, 2013-Dec 31, 2021	227	269	11.2	40	Pathology or 24-month follow-up
Chen et al. ⁶	2023	Jan 2017-Aug 2021	309	327	11.8	-	Pathology
Crestani et al. ⁷	2020	2014-2018	26	26	65.4	57	Pathology
Elshetry et al. ⁸	2023	April 2020-Sep 2021	90	116	38.8	39.4	Pathology or 12-month follow-up
Guo et al. ⁹	2022	July 2017-June 2020	54	56	14.3	37	Pathology and median 1.2-year follow-up
Hottat et al. ¹⁰	2022	Jan 2015-April 2020	163	201	28.9	51	Pathology
Manganaro et al. ¹¹	2023	Jan 2015-June 2022	140	172	52.9	48.7	Pathology or 12-month follow-up
Pereira et al. ¹²	2022	Feb 2014-Dec 2020	226	287	31.4	-	Pathology or 12-month follow-up
Thomassin-Naggara et al. ¹⁶	2020	Mar 1, 2013-Mar 31, 2016	1,130	1502	13.5	49	Pathology or 24-month follow-up
Wang et al. ¹³	2023	May 2017-July 2022	240	278	25.2	42	Pathology or 6–12 months of follow-up
Woo et al. ¹⁴	2023 (online ahead of print)	April 2021-Aug 2022	119	119	17.6	50	Pathology or ≥6-month follow-up
Wu et al. ¹⁵	2023	Jan 2018-Mar 2022	308	362	11.6	42.1	Pathology

and 4.66 ($P = 0.324$), respectively]. However, substantial heterogeneity was observed for specificity [I^2 values were 64.7% for the overall group and 66.20% and 62.4% for the studies with low risk and high risk of selection bias, respectively. The corresponding Cochran's Q statistics (P values) were 39.66 ($P < 0.001$), 26.63 ($P = 0.002$), and 10.6 ($P = 0.031$), respectively].

Meta-regression analysis revealed that the pooled sensitivity was significantly different for the studies with low risk and high risk of bias; the sensitivity values were slightly, but significantly, lower for the studies with low risk of bias compared with those with a high risk of bias [the relative pooled sensitivity for low risk versus high risk of bias studies was 0.954 (95% CI 0.911–0.999), $P = 0.043$]. Therefore, the pooled sensitivity values are given separately for the studies with low and high risk of bias (Figure 3), and they were 93.0% (95% CI 89.1%–95.5%, with high certainty of evidence) for the studies with low risk of bias, and 97.5% (95% CI 91.3%–99.3%) for the studies with high risk of bias. The pooled specificities were not significantly different for the studies with low and high risk of bias [the relative specificity for the studies with low vs. high risk of bias was 1.014 [(95% CI 0.930–1.106); $P = 0.752$]. The pooled specificity

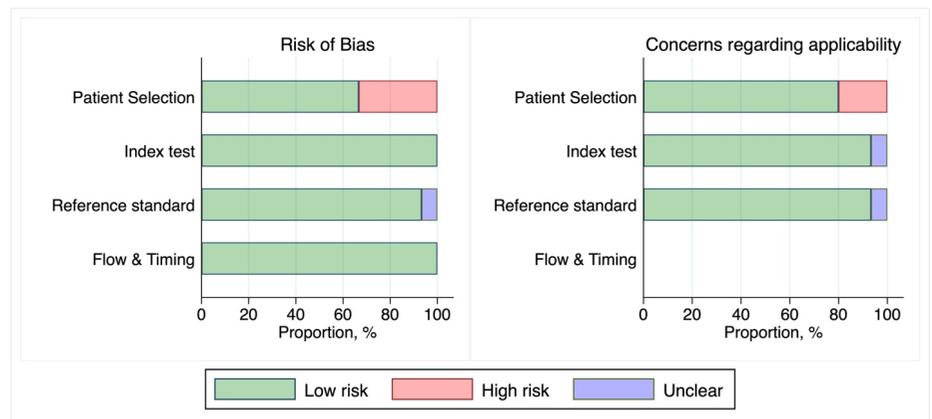


Figure 2. Methodological quality assessment according to the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

ty for the overall study population was 90.4% (95% CI 86.6%–93.2%, with moderate certainty of evidence due to high unexplained heterogeneity; Figure 3). The model performance did not increase with the inclusion of the variables of mean age and proportion of borderline or malignant lesions into the regression model ($P = 0.232$).

The SROC plot is presented in Figure 4 (the SROC plot with confidence and pre-

diction intervals is given in Supplementary Figure 1). The plot shows that the diagnostic performance of the O-RADS system is very high (the point estimate is very close to the upper left corner of the SROC plot). Additionally, the plot reveals that the diagnostic performance is slightly lower for the studies with a low risk of bias compared with those with a high risk of bias [area under the curve 0.97 (95% CI 0.95–0.98); $P < 0.001$ vs. 0.99 (95% CI 0.97–0.99); $P < 0.001$, respectively], probably

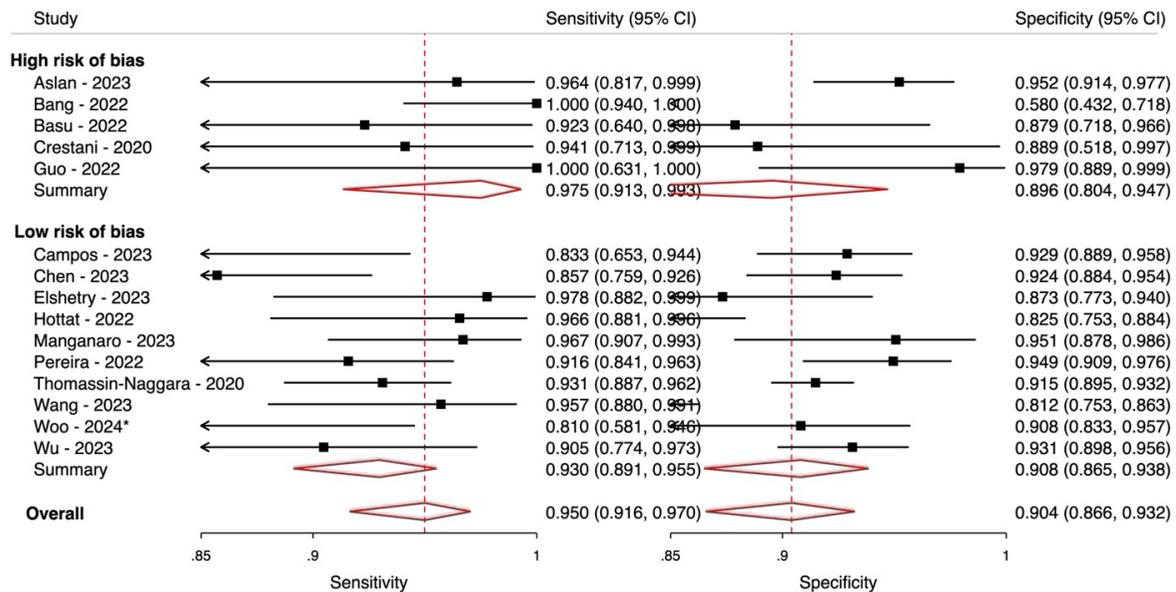


Figure 3. Forest plot of the pooled sensitivity and specificity. *Online publication in 2023, ahead of print.

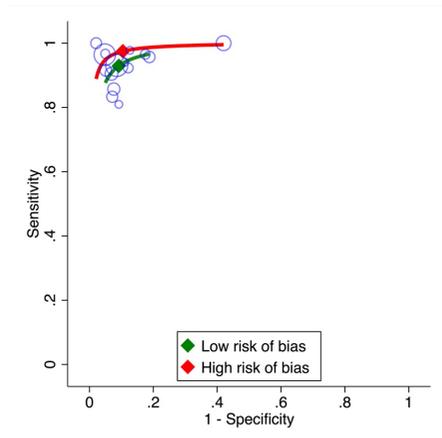


Figure 4. Summary receiver operating characteristics curve for the diagnostic performance of the Ovarian-Adnexal Reporting and Data System scoring. The blue circles represent individual studies, with their sizes proportional to the respective sample sizes. The red and green diamonds denote the point estimates (summary points), while the red and green solid lines illustrate the summary curves for studies with high and low risk of bias, respectively. For a more detailed depiction, including the confidence interval and prediction interval, please refer to Supplementary Figure 1.

due to lower pooled sensitivity in the low-risk bias group. However, Supplementary Figure 1 reveals that the precision is higher for the studies with a low risk of bias.

The pooled positive and negative LR values are provided in Supplementary Table 2. The LR matrix plot (Figure 5) indicates that

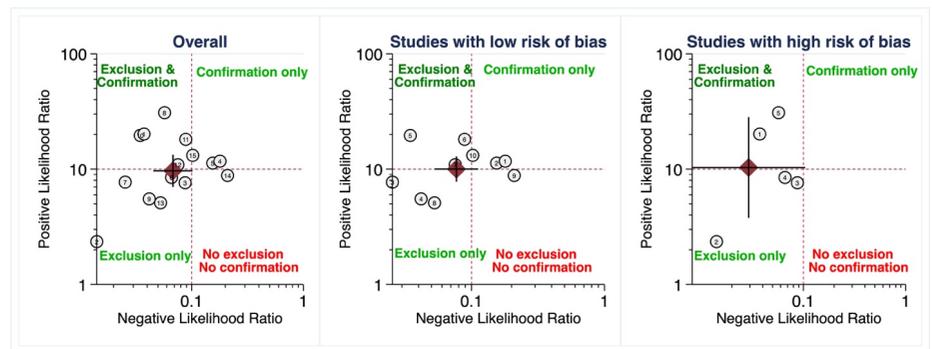


Figure 5. Likelihood matrix shows the pooled estimate (the diamond) and 95% confidence intervals of the negative and positive likelihood ratios, and exclusion and/or confirmation potential of the Ovarian-Adnexal Reporting and Data System scoring for borderline or malignant lesions.

the O-RADS system is more valuable for ruling out borderline or malignant lesions. In the overall population (Figure 5, left panel), the upper limit of the 95% CI of the negative LR is just at the cut-off limit of 0.1 [negative LR 0.07 (95% CI 0.05–0.10)]. A similar finding was observed for those with a high risk of bias but with a wider CI (Figure 5, right panel, Supplementary Table 2). Although the point estimate of the negative LR for the studies with low risk of bias was in the rule-out zone, the CI slightly crossed the cut-off value of 0.1 [negative LR for the low-risk group was 0.08 (95% CI 0.05–0.11)]. The point estimate of the pooled positive LR value was around the cut-off value of 10, with a lower limit of 95% CI >5, except for the value obtained from the studies with a high risk of bias. This suggests that the ruling-in potency of O-RADS scoring

is moderate. The Fagan's nomogram demonstrates obtaining the post-test probability of having borderline or malignant lesion depending on the positive (O-RADS 4 or 5) or negative (O-RADS <4) test result (Figure 6).

Deeks' funnel plot indicates that there is no concern for publication bias ($P = 0.812$; Figure 7).

Discussion

This systematic review and meta-analysis show that 1) the pooled sensitivity of O-RADS MRI scores ≥ 4 in diagnosing borderline or malignant adnexal tumors is high and varies slightly according to whether the study population has a low or high risk of patient selection bias [the sensitivity is slightly, but significantly, lower in the low-risk of bias

group (93.0% vs. 97.5%); 2) the pooled specificity of the O-RADS MRI score is 90.4% in the overall population with no significant difference between the studies with low risk and high risk of selection bias, and 3) using the cut-off value of ≥ 4 , the O-RADS MRI scores is valuable in ruling out the borderline or malignant lesions, although the ruling-in potency is relatively lower.

Ovarian cancers are estimated to be responsible for 5% of cancer-related deaths in women, with a 5-year survival rate of 50%.²⁵

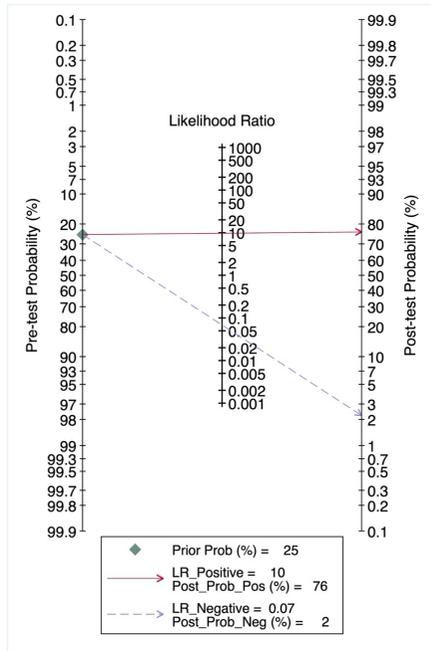


Figure 6. Fagan's nomogram for the Ovarian-Adnexal Reporting and Data System (O-RADS) scoring. The green diamond on the pre-test probability line (on the left side) represents the overall pre-test probability (25%) obtained from this meta-analysis. Utilizing the pooled likelihood ratio values, the solid red arrow and the dashed grey arrow indicate the post-test probability of having a borderline or malignant lesion when the test is positive (O-RADS 4 or 5) or negative (O-RADS <4), respectively. LR, likelihood ratio.

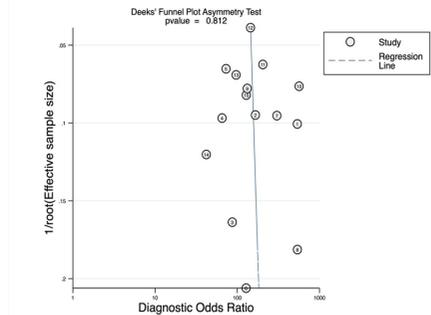


Figure 7. Deeks' plot for publication bias.

Ultrasonography is the first-line diagnostic method due to its low cost and wide availability. MRI is a better diagnostic method in terms of characterization and determining the origin of adnexal masses. To have a similar lexicon between radiologists and clinicians and for accurate referral of patients for surgical treatment, the ACR developed a system- O-RADS- for the characterization of adnexal masses.¹ After introducing the O-RADS system, several studies assessing its diagnostic performance have been published. In this study, a systematic review and meta-analysis of these studies were conducted to obtain updated information along with consideration of the risk of selection bias for each study.

The present study demonstrates that the heterogeneity between the studies for diagnostic sensitivity is not high, which implies that the results of the studies among the borderline or malignant lesions are consistent. On the other hand, the heterogeneity between the studies for diagnostic specificity is high. This implies that the consistency of the O-RADS system for benign lesions is relatively low, particularly for studies with a high risk of patient selection bias. High heterogeneity for specificity was also demonstrated in a previous meta-analysis,¹⁸ but that included a lower number of studies and did not consider the risk of bias while pooling the results.

Both the sensitivity and specificity of the O-RADS system in discriminating benign lesions and borderline or malignant lesions are high, although the sensitivity is higher than the specificity. The pooled sensitivity varies for those with or without a high risk of patient selection bias, and studies with a low risk of bias have a lower, but acceptable, pooled sensitivity (93.0% vs. 97.5%). Specificity was also high, but similar for the studies with low or high risk of selection bias. Consistent with these findings, the SROC plot shows that the O-RADS system has high diagnostic performance (discrimination) for borderline or malignant lesions. This suggests that the O-RADS system is a good tool for referring patients to surgery. The SROC plot (Supplementary Figure 1) also shows that the precision (based on the 95% CI and prediction interval) is very high for the studies with a low risk of bias but is relatively lower for the studies with a high risk of bias.

Because of the high sensitivity, O-RADS MRI scoring is valuable for ruling out borderline or malignant lesions. This is supported by the LR matrix plot. It is generally accepted that a negative LR value of <0.1 indicates that

the test is valuable in ruling out the disease, and a positive LR value of >10 indicates the test is valuable in ruling in the disease,^{26,27} although they are arbitrarily chosen cut-off values. Furthermore, negative LR values of 0.1–0.2 and positive LR values of 5–10 indicate that the test is moderately effective in ruling out and ruling in the disease, respectively. In the present study, the upper limit of the 95% CI of the negative LR value was just at the cut-off value of 0.1 in the overall population and in the analysis of the studies with a high risk of bias, which suggests O-RADS MRI is good at in excluding the disease. For the studies with a low risk of bias, although the upper limit of 95% CI for the negative LR slightly crossed the cut-off value (negative LR value 0.08, 95% CI 0.05–0.11), the ruling-out potential was largely preserved. The point estimates of the positive LR values were around the cut-off value of 10, and although the CI crossed the cut-off value of 10, the lower limit was >5 for the overall population and those with a low risk of selection bias (Supplementary Table 2). This finding suggests the O-RADS MRI score is moderately effective in ruling in the disease.

In the EURAD study, the prospective European multicenter cohort, misclassified cases were assessed in terms of three types of error: errors caused by technical limitations, inadequate experience (perceptual error), or interpretive errors.²⁸ The interpretive error was found to be the most common cause of the misclassification, which was mostly due to rating benign lesions as O-RADS 4 or 5 (false positive result). Even if some of the false positive results were caused by a concern for missing the malignancy, they demonstrated that the misclassification was substantially reduced with strict application of O-RADS scoring. The false positive result is associated with low specificity. In the present study, the heterogeneity between the studies was high for the specificity. Additionally, compared with sensitivity, specificity was relatively low. This may suggest a problem with the rating of benign lesions. Therefore, approaches to increase specificity and reduce potential heterogeneity in the interpretation of the benign lesions may reduce unnecessary surgical procedures by keeping the false positive rate low. This may be obtained by reducing interpretive errors by applying the O-RADS scoring meticulously and by increasing the awareness of some lesions that may be misclassified. Thomassin-Naggara et al.²⁸ discussed these lesions in their article and underlined the importance of the difference between a solid lesion and a solid compo-

nent for correct classification. Another factor for correct classification is the availability of technically adequate MRI images, which has been discussed elsewhere.²⁹ In addition, some refinement in the O-RADS system may improve its diagnostic value. Several methods seem promising in increasing the diagnostic performance of the O-RADS system. Wengert et al.³⁰ showed that time-intensity curve analysis was superior to visual assessment and improved the specificity. Furthermore, diffusion-weighted imaging improves the diagnostic performance of the O-RADS MRI system.¹⁰ Application of these methods may reduce false positive results by increasing specificity and may also increase its ruling-out potential further by increasing the sensitivity.

The present study has several limitations. First, as in many meta-analyses, data were extracted from published manuscripts; therefore, individual participant data were not available. Although it is very difficult to obtain, individual participant data analysis provides more reliable information and may provide detailed reasons for heterogeneity. Second, we did not analyze the data based on the readers' experience; other confounding factors may also affect the results. However, the relatively low number of studies precludes taking many factors into consideration, especially if individual data are not available. Third, we aimed to assess the "intrinsic" diagnostic performance of O-RADS MRI scoring; therefore, cancer antigen 125 levels, or other factors such as menopausal status or family history that may be used to assess the pre-test probability of the malignancy, were not used in the analysis. Instead, we preferred to provide Fagan's nomogram, in which the pre-test probability obtained by any marker or clinical predictors can be combined with the "intrinsic" performance of the O-RADS score to obtain the post-test probability. The present analysis also has some advantages, such as including new studies, and, in contrast to recent meta-analyses, assessing the diagnostic performance and providing visual information about the ruling-in and ruling-out potential according to the risk of bias.

In conclusion, O-RADS MRI scoring is valuable in ruling out borderline or malignant lesions, while the ruling-in potency is moderate. Patient selection bias affects diagnostic sensitivity, leading to a higher sensitivity compared with the sensitivity obtained from the studies with a low risk of bias. The high between-study heterogeneity observed for

specificity suggests the need for improvement in the consistent characterization of the benign lesions to reduce false positive rates.

Acknowledgement

Thanks to Dr. Cansu Öztürk for data handling and critical appraisal of the literature, and to Dr. Mustafa Kılıçkap for conducting statistical analyses and constructive comments.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Reinhold C, Rockall A, Sadowski EA, et al. Ovarian-Adnexal Reporting Lexicon for MRI: A White Paper of the ACR Ovarian-Adnexal Reporting and Data Systems MRI Committee. *J Am Coll Radiol.* 2021;18(5):713-729. [\[CrossRef\]](#)
- Aslan S, Tosun SA. Diagnostic accuracy and validity of the O-RADS MRI score based on a simplified MRI protocol: a single tertiary center retrospective study. *Acta Radiol.* 2023;64(1):377-386. [\[CrossRef\]](#)
- Bang JI, Kim JY, Choi MC, Lee HY, Jang SJ. Application of multimodal imaging biomarker in the differential diagnosis of ovarian mass: integration of conventional and molecular imaging. *Clin Nucl Med.* 2022;47(2):117-122. [\[CrossRef\]](#)
- Basu A, Pame M, Bhuyan R, Roy DK, James VM. Diagnostic Performance of O-RADS MRI scoring system for the assessment of adnexal masses in routine clinical radiology practice—a single tertiary centre prospective cohort study. *J Clin Diagn Res.* 2022;16(4):TC11-TC16. [\[CrossRef\]](#)
- Campos A, Villermain-Lécolier C, Sadowski EA, et al. O-RADS scoring system for adnexal lesions: diagnostic performance on TVUS performed by an expert sonographer and MRI. *Eur J Radiol.* 2023;169:111172. [\[CrossRef\]](#)
- Chen T, Qian X, Wei C, et al. The consistency and application value of MRI-based ovarian-adnexal reporting and data system in the diagnosis of ovarian adnexal masses. *Chinese Journal of Radiology.* 2023;57(12):282-287. [\[CrossRef\]](#)
- Crestani A, Theodore C, Levallant JM, et al. Magnetic resonance and ultrasound fusion imaging to characterise ovarian masses: a feasibility study. *Anticancer Res.* 2020;40(7):4115-4121. [\[CrossRef\]](#)
- Elshetry ASF, Hamed EM, Frere RAF, Zaid NA. Impact of adding mean apparent diffusion

- coefficient (ADCmean) measurements to O-RADS MRI scoring for adnexal lesions characterization: a combined O-RADS MRI/ADCmean approach. *Acad Radiol.* 2023;30(2):300-311. [\[CrossRef\]](#)
- Guo Y, Phillips CH, Suarez-Weiss K, et al. Interreader agreement and intermodality concordance of O-RADS US and MRI for assessing large, complex ovarian-adnexal cysts. *Radiol Imaging Cancer.* 2022;4(5):e220064. [\[CrossRef\]](#)
- Hottat NA, Badr DA, Van Pachterbeke C, et al. Added value of quantitative analysis of diffusion-weighted imaging in Ovarian-Adnexal Reporting and Data System Magnetic Resonance Imaging. *J Magn Reson Imaging.* 2022;56(1):158-170. [\[CrossRef\]](#)
- Manganaro L, Ciulla S, Celli V, et al. Impact of DWI and ADC values in Ovarian-Adnexal Reporting and Data System (O-RADS) MRI score. *Radiol Med.* 2023;128(5):565-577. [\[CrossRef\]](#)
- Pereira PN, Yoshida A, Sarian LO, Barros RHO, Jales RM, Derchain S. Assessment of the performance of the O-RADS MRI score for the evaluation of adnexal masses, with technical notes. *Radiol Bras.* 2022;55(3):137-144. [\[CrossRef\]](#)
- Wang T, Cui W, Nie F, et al. Comparative study of the efficacy of the Ovarian-Adnexa Reporting and Data System Ultrasound Combined With Contrast-Enhanced Ultrasound and the ADNEX MR scoring system in the diagnosis of adnexal masses. *Ultrasound Med Biol.* 2023;49(9):2072-2080. [\[CrossRef\]](#)
- Woo S, Andrieu PC, Abu-Rustum NR, et al. Bridging communication gaps between radiologists, referring physicians, and patients through standardized structured cancer imaging reporting: the experience with female pelvic MRI assessment using O-RADS and a simulated cohort patient group. *Acad Radiol.* 2024;31(4):1388-1397. [\[CrossRef\]](#)
- Wu M, Tang Q, Cai S, et al. Accuracy and reproducibility of the O-RADS MRI risk stratification system based on enhanced non-DCE MRI in the assessment of adnexal masses. *Eur J Radiol.* 2023;159:110670. [\[CrossRef\]](#)
- Thomassin-Naggara I, Poncelet E, Jalaguier-Coudray A, et al. Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADS MRI) score for risk stratification of sonographically indeterminate adnexal masses. *JAMA Netw Open.* 2020;3(1):e1919896. [\[CrossRef\]](#)
- Rizzo S, Cozzi A, Dolcianni M, et al. O-RADS MRI: A systematic review and meta-analysis of diagnostic performance and category-wise malignancy rates. *Radiology.* 2023;307(1):e220795. [\[CrossRef\]](#)
- Zhang Q, Dai X, Li W. Systematic review and meta-analysis of O-RADS Ultrasound and O-RADS MRI for risk assessment of ovarian

- and adnexal lesions. *AJR Am J Roentgenol.* 2023;221(1):21-33. [\[CrossRef\]](#)
19. Shamseer L, Moher D, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* 2015;350:g7647. Erratum in: *BMJ.* 2016;354:i4086. [\[CrossRef\]](#)
 20. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med.* 2011;155(8):529-536. [\[CrossRef\]](#)
 21. Schünemann H BJ, Guyatt G, Oxman A. GRADE handbook. Accessed May, 24 2024. [\[CrossRef\]](#)
 22. Yang WT, Parikh JR, Stavros AT, Otto P, Maislin G. Exploring the negative likelihood ratio and how it can be used to minimize false-positives in breast imaging. *AJR Am J Roentgenol.* 2018;210(2):301-306. [\[CrossRef\]](#)
 23. Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: the case of meta-analyses of diagnostic accuracy. *Stat Med.* 2014;33(16):2701-2717. [\[CrossRef\]](#)
 24. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58(9):882-893. [\[CrossRef\]](#)
 25. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17-48. [\[CrossRef\]](#)
 26. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther.* 2001;81(9):1546-1564. [\[CrossRef\]](#)
 27. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet.* 2005;365(9469):1500-1505. [\[CrossRef\]](#)
 28. Thomassin-Naggara I, Belghitti M, Milon A, et al. O-RADS MRI score: analysis of misclassified cases in a prospective multicentric European cohort. *Eur Radiol.* 2021;31(12):9588-9599. [\[CrossRef\]](#)
 29. Nougaret S, Lakhman Y, Bahadir S, Sadowski E, Thomassin-Naggara I, Reinhold C. Ovarian-Adnexal Reporting and Data System for Magnetic Resonance Imaging (O-RADS MRI): Genesis and Future Directions. *Can Assoc Radiol J.* 2023;74(2):370-381. [\[CrossRef\]](#)
 30. Wengert GJ, Dabi Y, Kermarrec E, et al. O-RADS MRI classification of indeterminate adnexal lesions: time-intensity curve analysis is better than visual assessment. *Radiology.* 2022;303(3):566-575. Erratum in: *Radiology.* 2022;303(2):E28. [\[CrossRef\]](#)

Supplementary Table 1. Assessment of the methodological quality of each study according to the QUADAS-2 tool

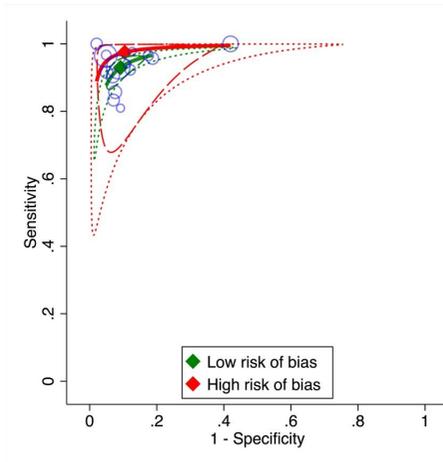
Study	Risk of bias				Concerns regarding applicability		
	Patient selection	Index test	Reference test	Flow & timing	Patient selection	Index test	Reference test
Aslan and Tosun ² - 2023	High*	Low	Low	Low	Low	Unclear	Low
Bang et al. ³ - 2022	High**	Low	Low	Low	High	Low	Low
Basu et al. ⁴ - 2022	High [§]	Low	Unclear	Low	High	Low	Unclear
Campos et al. ⁵ - 2023	Low	Low	Low	Low	Low	Low	Low
Chen et al. ⁶ - 2023	Low	Low	Low	Low	Low	Low	Low
Crestani et al. ⁷ - 2020	High ^{§§}	Low	Low	Low	High	Low	Low
Elshetry et al. ⁸ - 2023	Low	Low	Low	Low	Low	Low	Low
Guo et al. ⁹ - 2022	High [‡]	Low	Low	Low	Low	Low	Low
Hottat et al. ¹⁰ - 2022	Low	Low	Low	Low	Low	Low	Low
Manganaro et al. ¹¹ - 2023	Low	Low	Low	Low	Low	Low	Low
Pereira et al. ¹² - 2022	Low	Low	Low	Low	Low	Low	Low
Thomassin-Naggara et al. ¹⁶ - 2020	Low	Low	Low	Low	Low	Low	Low
Wang et al. ¹³ - 2023	Low	Low	Low	Low	Low	Low	Low
Woo et al. ¹⁴ - 2023 (ahead of print)	Low	Low	Low	Low	Low	Low	Low
Wu et al. ¹⁵ - 2023	Low	Low	Low	Low	Low	Low	Low

*, Simplified method and exclusion of <3 cm cysts; **, includes patients underwent PET/CT; [§], Non-probability sampling and 4-month of follow-up; ^{§§}, includes a sub-population who underwent surgery; [‡], included patients with >5 cm cystic lesions; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.

Supplementary Table 2. The negative and positive likelihood ratios for the overall population and for the studies with low or high-risk of bias

	Negative likelihood ratio and 95% CI	Positive likelihood ratio and 95% CI
Studies with low risk of bias	0.08 (0.05 – 0.11)	10.0 (7.7 – 12.9)
Studies with high risk of bias	0.03 (0.01 – 0.10)	10.3 (3.8 – 28.3)
Overall population	0.07 (0.05 – 0.10)	9.7 (7.0 – 13.3)

CI, confidence interval



Supplementary Figure 1. Summary receiver operating characteristics curve for the diagnostic performance of the O-RADS scoring. The blue circles represent each study, with their sizes proportional to the sample size of the respective study. The red and green diamonds depict the point estimates (summary points), while the red and green solid lines illustrate the summary curve for the studies with high and low-risk of bias, respectively. Correspondingly, the red and green dashed lines represent the confidence interval, and the red and green dotted lines indicate the prediction interval for the pooled estimates of studies with high and low-risk of bias, respectively. O-RADS, Ovarian-Adnexal Reporting and Data System.



Open- and closed-type congenital cholesteatomas of the middle ear: computed tomography differentiation and correlation with surgical staging

Minjung Seong¹
 Hyung-Jin Kim^{1,2}
 Yikyung Kim¹
 Sung Tae Kim¹

¹Samsung Medical Center, Sungkyunkwan University School of Medicine, Department of Radiology, Seoul, Korea

²Seoul St Mary's Hospital, The Catholic University of Korea, College of Medicine, Department of Radiology, Seoul, Korea

PURPOSE

To investigate the differences in computed tomography (CT) features between closed-type congenital cholesteatoma (CCC) and open-type congenital cholesteatoma (OCC) of the middle ear and to evaluate the usefulness of preoperative CT examination for staging workup of congenital cholesteatoma (CC) in correlation with the surgical findings.

METHODS

We retrospectively reviewed the preoperative CT scans of the temporal bone obtained from 80 patients with surgically confirmed CC of the middle ear. All patients had a solitary lesion, except for one patient with two lesions, resulting in 81 CCs, which formed the basis of this study. We compared the CT features between CCCs and OCCs, focusing on their morphological characteristics, such as size, shape, location, and bone change. Based on the Potsic classification, the stage of CCs was determined at CT and surgery, and the results were compared between CCCs and OCCs.

RESULTS

Of the 81 CCs, surgery revealed 43 CCCs and 38 OCCs. On CT scans, CCC was frequently seen as a small (median: 3.15 mm), round to oval (65.1%) mass, most commonly located in the anterosuperior quadrant (74.4%) of the middle ear with less frequent ossicular erosion (14.0%). In contrast, OCC was frequently seen as a large (median: 6.70 mm), irregular (94.7%) mass, most commonly located in the posterosuperior quadrant (68.4%) of the middle ear with frequent ossicular erosion (55.3%). The size, shape, location, and presence of ossicular erosion were significantly different between the two types. Overall, the CT and surgical stages of CCs demonstrated good agreement (kappa value: 0.77) and the CT and surgical stages of OCCs were statistically significantly higher than those of CCCs ($P < 0.001$ in both).

CONCLUSION

CT is useful for preoperative determination of the types and staging of CC of the middle ear.

CLINICAL SIGNIFICANCE

Preoperative differentiation between CCC and OCC is important to avoid reoperation and prevent an extensive surgery. By providing valuable information on the morphology and extent of the lesions, CT is useful for not only the accurate preoperative determination of the type of CCs but also the accurate prediction of staging of the lesion, which should be important to preparing optimal treatment plans.

KEYWORDS

Congenital cholesteatoma, computed tomography, middle ear disease, potsic stage, temporal bone

Corresponding author: Hyung-Jin Kim

E-mail: hyungkim@skku.edu

Received 20 August 2024; revision requested 23 September 2024; accepted 20 October 2024.



Epub: 09.12.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242913

You may cite this article as: Seong M, Kim HJ, Kim Y, Kim ST. Open- and closed-type congenital cholesteatomas of the middle ear: computed tomography differentiation and correlation with surgical staging. *Diagn Interv Radiol.* 2025;31(3):180-186.

Congenital cholesteatoma (CC) is defined as a whitish mass behind an intact tympanic membrane without a history of tympanic membrane perforation, otorrhea, or previous ear surgery.¹ The annual incidence of CC is uncommon and reported to be 0.12/100,000 individuals.^{2,3} It accounts for 2%–5% of all cholesteatomas and 4%–24% of cholesteatomas in children.^{4,5} The condition occurs three times more frequently in male individuals than in female individuals and most frequently affects children aged 4–5 years. Early detection and surgical intervention is important to prevent extensive surgery and preserve hearing; because CC grows faster in children than in adults, disease progression would be facilitated if the diagnosis and treatment are delayed.^{3,4,6} Although the most commonly affected location is known to be the anterosuperior quadrant of the middle ear, followed by the posterosuperior quadrant,^{2,4} a predilection to involve the posterosuperior quadrant has also been reported, especially in Asian populations.^{3,6,7}

The pathogenesis of CC is unclear and controversial, with the embryonic cell rest theory being considered the most plausible.^{8,9} Although not particularly well known among radiologists, it is well known among otolaryngologists that there are two types of CC, closed-type congenital cholesteatoma (CCC) and open-type congenital cholesteatoma (OCC), which differ in their morphology and clinical course.⁹ Histologically, the former presents as an epithelial cyst without exposure of keratin and the latter manifests as flat keratinizing epithelium continuous with middle ear mucosa without formation of an epithelial cyst.^{6,8,10} Whether these two types have a different pathogenesis or are a different evolution of the same phenomenon is un-

certain, but their clinical manifestations are different. Compared with CCC, OCC occurs in older age groups, is more difficult to identify behind an intact tympanic membrane, and involves greater difficulty in surgery.⁸ Therefore, preoperative differentiation between CCC and OCC is important to avoid reoperation and prevent an extensive surgery.²

In 2002, Potsic et al.¹¹ proposed a staging system for CC according to the extent of lesions (Table 1), which appears to correlate well with the severity of the disease and treatment outcome.² Computed tomography (CT) is widely considered the imaging modality of choice for the diagnosis of CC.^{1,4,12} To the best of our knowledge, however, only a few studies have been reported on the role of CT in distinguishing OCC from CCC and determining preoperative staging.^{5,12,13} Accordingly, the aims of this study are to investigate the different CT features of CC according to the subtypes and to evaluate whether CT staging may correlate well with surgical staging.

Methods

Study population

A search of the electronic medical records of Samsung Medical Center between January 1999 and February 2017 revealed 109 patients with surgically proven CC of the middle ear, among whom CT scanning was performed in 102. The diagnosis of CC was made by otolaryngologists on the basis of the criteria devised by Levine et al.¹⁴ Of the 102 patients, 22 were excluded due to insufficient medical records ($n = 20$) and recurrent disease after surgery ($n = 2$). Finally, 80 patients (61 males and 19 females) aged 1–38 years (mean age \pm standard deviation: 5 ± 5.6 years) were enrolled as the participants of this study (Figure 1).

All patients were operated on by two otolaryngologists. They classified lesions as CCC or OCC and determined lesion staging according to the Potsic classification (Table 1).

Computed tomography examination

CT of the temporal bone was performed on various models of multidetector helical

CT scanners (GE Healthcare, Milwaukee, WI, USA) with variable mA, 120 kVp, 0.625–1.25-mm section thickness and section spacing, a field of view of 18 cm, and a high-resolution algorithm. Direct or reformatted coronal images were routinely obtained. All the images were viewed with the window width of 4000 HU and window level of 400 HU.

All CT scans were retrospectively reviewed by two neuroradiologists with clinical experience of 30 and 3 years, respectively, for the number, location, shape, and size of the lesion, and any disagreements were resolved by consensus. The presence of ossicular erosion, labyrinthine fistula, and mastoid involvement was also investigated. To determine the location of the lesion, we referred to the handle of the malleus and divided the tympanic cavity into four quadrants: anterosuperior, posterosuperior, anteroinferior, and posteroinferior quadrants, as shown in Figure 2. If two or more quadrants were involved by the lesion, we recorded all. The shape of the lesion was divided into round to oval and irregular. The size of the lesion was measured at its greatest diameter. To differentiate from simple inflammation, mastoid involvement was considered to be present when the lesion in the mastoid antrum was continuous with the middle ear lesion. CT staging was also determined using the Potsic classification.¹¹

Research ethics standards compliance

This study was approved by the Institutional Review Board at Samsung Medical Center (IRB no: 2018-08-18-001, date: October 10, 2018), and informed consent was waived in accordance with the requirements of a retrospective study.

Statistical analysis

Statistical analysis was performed using SAS version 9.4 (SAS Institute) and R 4.0.2 (Vienna, Austria; <http://www.R-project.org/>). Wilcoxon's rank sum test was used to compare the age and size between the CCC and OCC groups. The chi-square test was used for comparing the gender distribution, shape, location, presence of ossicular

Main points

- Closed-type congenital cholesteatomas (CCCs) are smaller, more likely to be round to oval, most commonly located in the anterosuperior quadrant of the middle ear, and less frequently associated with ossicular erosion than open-type congenital cholesteatomas (OCCs).
- OCCs are larger, irregularly shaped, most commonly located in the posterosuperior quadrant of the middle ear, and more often associated with ossicular erosion.
- Computed tomography (CT) and surgical stages of OCCs are higher than those of CCCs.
- CT can help to accurately determine the type and stage of CC before surgery.

Table 1. Potsic staging for congenital cholesteatoma¹¹

Stage	Description
I	Single quadrant: no ossicular involvement or mastoid extension
II	Multiple quadrants: no ossicular involvement or mastoid extension
III	Ossicular involvement: includes erosion of ossicles and surgical removal for eradication of disease; no mastoid extension
IV	Mastoid extension (regardless of findings elsewhere)

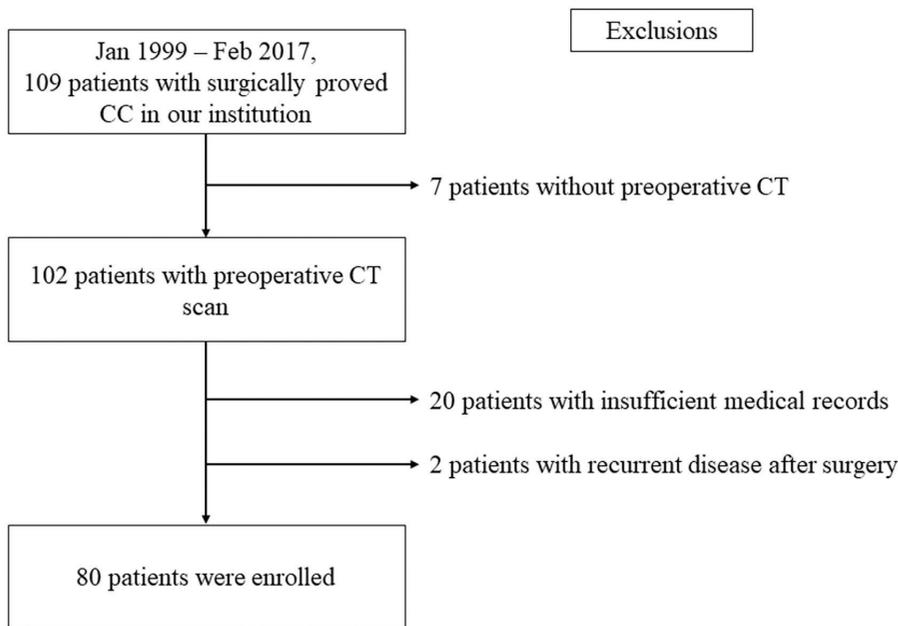


Figure 1. Flowchart of patient enrollment. CC, congenital cholesteatoma; CT, computed tomography.

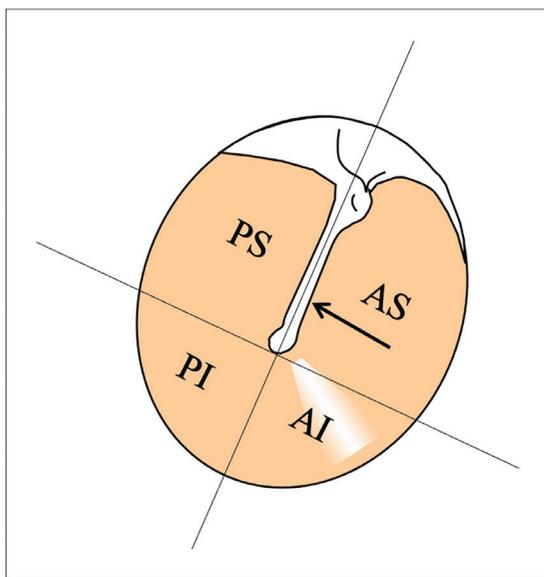


Figure 2. Schematic showing four quadrants of the tympanic cavity. We referred to the handle of the malleus (arrow) and divided the tympanic cavity into four quadrants, including anterosuperior (AS), posterosuperior (PS), anteroinferior (AI), and posteroinferior (PI) quadrants.

erosion, presence of mastoid inflammation, and presence of labyrinthine fistula between the two groups. Agreement on the CT and surgical stages was evaluated using Cohen's weighted kappa. A κ value ≤ 0.20 indicated positive but poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and ≥ 0.81 , excellent agreement. The statistical differences in CT and surgical stages between CCCs and OCCs were evaluated using Fisher's exact test with the permutation method for multiple comparisons. For all statistical analysis,

$P < 0.05$ was considered statistically significant.

Results

Comparison of demographic data between open- and closed-type congenital cholesteatomas

Overall, 80 patients had 81 lesions, including 43 CCCs and 38 OCCs (Table 2). One patient had two separate lesions in one ear, both of which were OCCs; all other patients had one lesion each. Males were affected

more frequently than females in both groups (32 males and 11 females in the CCC group and 29 males and eight females in the OCC group), but there was no statistically significant difference between the two groups ($P = 0.631$). The ages of the OCC group (median age: 5 years; range: 2–38 years) were statistically significantly older than those of the CCC group (median age: 3 years; range: 1–16 years) ($P < 0.001$). The median time interval between CT examination and surgery was 29 days (range: 8–555 days) in the CCC group and 32 days (range: 1–322 days) in the OCC group, which, based on the Wilcoxon rank sum test, was not statistically significantly different ($P = 0.529$).

Comparison of computed tomography findings between open- and closed-type congenital cholesteatomas

The CT features of CCCs and OCCs are summarized in Table 2, with typical cases presented in Figures 3 and 4. The OCCs were statistically significantly larger than the CCCs (median: 6.70 mm vs. 3.15 mm; $P < 0.001$). While 28 (65.1%) of 43 CCCs were round or oval in shape, 36 (94.7%) of 38 OCCs had an irregular shape ($P < 0.001$). The anterosuperior quadrant [32/43 (74.4%)] was the most common site involved in CCCs, followed by the posterosuperior [12/43 (27.9%)], anteroinferior [5/43 (11.6%)], and posteroinferior [5/43 (11.6%)] quadrants, whereas the posterosuperior quadrant [26/38 (68.4%)] was most commonly involved in OCCs, followed by the anterosuperior [25/38 (65.8%)], anteroinferior [14/38 (36.8%)], and posteroinferior [11/38 (28.9%)] quadrants. Among these, involvement of the posterosuperior and anteroinferior quadrants was statistically significantly greater in OCCs than in CCCs ($P < 0.001$ and $P = 0.008$, respectively). The detailed sites involved in CCCs and OCCs are further characterized in Table 3. Compared with CCCs, OCCs more frequently involved multiple quadrants of the tympanic cavity, which was statistically significantly different ($P < 0.001$). While 36 (83.7%) of 43 CCCs involved a single quadrant, with the anterosuperior quadrant being most frequently involved in 27 cases, only 16 (42.1%) of 38 OCCs involved a single quadrant, with the posterosuperior quadrant being most frequently involved in eight cases. In the remaining seven (16.3%) of 43 CCCs and 22 (57.9%) of 38 OCCs, two or more quadrants were involved in the lesion. Involvement of the mastoid antrum was statistically significantly greater in OCCs than in CCCs [9/38 (23.7%) OCCs vs. 2/43 (4.5%) CCCs; $P = 0.013$]. In all 11 cases with mastoid

Table 2. Demographic data and computed tomography features of closed- and open-type cholesteatomas

	Closed-type cholesteatoma (n = 43 in 43 patients)	Open-type cholesteatoma (n = 38 in 37 patients)	P value
Sex			
Male	32	29	0.631
Female	11	8	
Age, yr			
Median [IQR]/range	3 [2.4]/1–16	5 [3.9]/2–38	<0.001
Time interval between CT and surgery, days			
Median [IQR]/range	29 [18.42]/8–555	32 [19.62]/1–322	0.529
Size, mm			
Median [IQR]/range	3.15 [2.60, 4.42]/1.1–22.0	6.70 [4.78, 10.07]/2.2–35.7	<0.001
Shape			
Round to oval	28 (65.1%)	2 (5.3%)	<0.001
Irregular	15 (34.9%)	36 (94.7%)	
Location			
Anterosuperior	32 (74.4%)	25 (65.8%)	0.396
Posterosuperior	12 (27.9%)	26 (68.4%)	<0.001
Anteroinferior	5 (11.6%)	14 (36.8%)	0.008
Posteroinferior	5 (11.6%)	11 (29.0%)	0.051
Mastoid involvement	2 (4.7%)	9 (23.7%)	0.013
Ossicular erosion	6 (14.0%)	21 (55.3%)	<0.001
Labyrinthine fistula	1 (2.3%)	1 (2.6%)	0.929

IQR, interquartile range; CT, computed tomography.

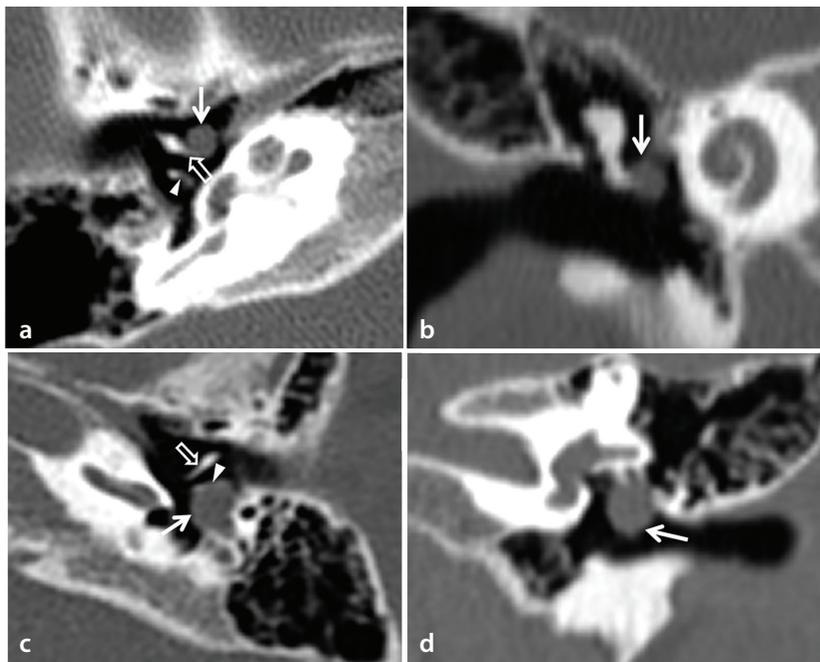


Figure 3. Typical examples of closed-type congenital cholesteatomas (CCCs). (a, b) Potsic stage I CCC in a 3-year-old boy. Axial (a) and coronal (b) computed tomography (CT) scans show a small well-defined round nodule (white arrows) confined to the anterosuperior quadrant of the right middle ear, anterior and superior to the handle of the malleus (open arrow in a), without evidence of ossicular erosion. Posterior to the handle of the malleus lies the long process of the incus (arrowhead in a). (c, d) Potsic stage II CCC in a 3-year-old boy. Axial (c) and coronal (d) CT scans show a well-defined ovoid mass (arrows) located in the posterosuperior and posteroinferior quadrants of the left middle ear, posterior to the long process of the incus (arrowhead) without evidence of ossicular erosion or mastoid involvement. Open arrow indicates the handle of the malleus.

involvement, the posterosuperior quadrant of the middle ear was also involved. The rate of ossicular erosion was also statistically significantly greater in OCCs than in CCCs [21/38 (55.3%) OCCs vs. 6/43 (14.0%) CCCs; $P < 0.001$]. The incus, especially its distal long process, was most commonly affected in all 27 cases with ossicular erosion, followed by the stapes in 18 cases and the malleus in four cases. Two or more ossicles were eroded in 16 of 21 OCCs and three of six CCCs. Labyrinthine fistula was present in one (2.3%) of 43 CCCs and one (2.6%) of 38 OCCs, with no statistically significant difference between the two groups ($P = 0.929$). The former occurred at the superior semicircular canal and the latter involved the basal turn of the cochlea and the lateral and superior semicircular canals.

Comparison of computed tomography and surgical stages between open- and closed-type congenital cholesteatomas

The results of the CT and surgical stages of CCs based on the Potsic classification are summarized in Table 4. At surgery, 81 CCs were determined as stage I in 44 (54.3%), stage II in nine (11.1%), stage III in 18 (22.2%), and stage IV in 10 (12.3%). Of all 81 CCs, the CT stage accorded well with the surgical stage in 65 cases (80.2%), including 37 (84.1%) of 44 stage I CCs, three (33.3%) of nine stage II CCs, 17 (94.4%) of 18 stage III CCs, and eight (80.0%) of 10 stage IV CCs. In the remaining 16 cases, the CT stage mismatched with the surgical stage, including underestimation and overestimation in eight cases (9.9%) each (Table 5). The eight cases of underestimation included six cases of stage I at CT, all of which proved to be stage II at surgery. The remaining one case of stage I and one case of stage III at CT proved to be stage IV at surgery (Figure 5a, b). The eight cases of overestimation included five cases of stage II and two cases of stage IV at CT, all of which proved to be stage I at surgery (Figure 5c, d). The remaining one case of stage IV at CT proved to be stage III at surgery.

Overall agreement between the CT and surgical stages was good, with a κ value of 0.77 (CI: 0.64–0.89) based on Cohen's weighted kappa. When 43 CCCs and 38 OCCs were analyzed separately, good agreement was also found between the CT and surgical stages in both groups, with κ values of 0.69 [confidence interval (CI): 0.43–0.94] and 0.73 (CI: 0.55–0.91), respectively, with the κ values not statistically significantly different ($P = 0.922$). Compared with CCCs, both the CT

and surgical stages of OCCs were statistically significantly higher when using Fisher's exact test ($P < 0.001$).

Discussion

The importance of preoperative discrimination between CCCs and OCCs lies in the

fact that they are different in clinical behavior. Early diagnosis and treatment is critical to prevent disease progression, particularly in patients with OCCs.^{3,4,6} However, only a few studies focusing on differentiation between CCCs and OCCs using CT have been reported.^{5,12,13} Overall, the results of the present study are in good accordance with those of the previous studies. Compared with CCCs, OCCs statistically significantly more commonly affected the older age group and were larger in size and more irregular in shape.^{4,5} Invasion of the mastoid antrum and the erosion of the ossicles were also statistically significantly greater in OCCs than in CCCs. These findings can be partly explained by their different morphology and embryonic pathogenesis. In contrast to CCCs, which appear as a well-formed cyst lined by an epithelial membrane, OCCs do not form a discrete cyst but usually present as a flat epithelium in the middle ear and thus can avoid early detection and facilitate the spread of lesions unnoticed. A higher recurrence rate was also reported in terms of OCCs.⁵

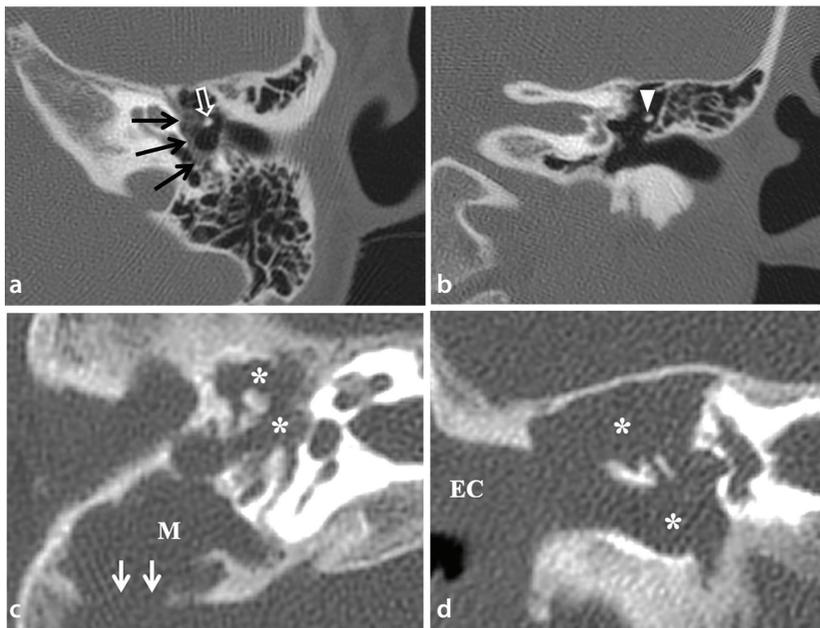


Figure 4. Typical examples of open-type congenital cholesteatomas (OCCs). (a, b) Potts stage III OCC in an 11-year-old boy. Axial (a) and coronal (b) computed tomography (CT) scans show an irregular mass (white arrows) in the anterosuperior and posterosuperior quadrants of the left middle ear, anterior and posterior to the neck of the malleus (open arrow in a). Note that the stapes and the long process of the incus cannot be localized, which should be seen posterior to the neck of the malleus (open arrow in a) and inferior to the body of the incus (arrowhead in b), due to erosion by the mass. (c, d) Potts stage IV OCC in a 12-year-old boy. Axial (c) and coronal (d) CT scans show a large irregular mass (asterisks) involving the whole parts of the right middle ear, which causes the ossicular erosion and extends to the mastoid antrum and air cells (M). The mass also involves the external auditory canal (EC) and causes the erosion of the sigmoid sinus plate (arrows in c).

The different biological behaviors between CCCs and OCCs might also be attributed to their different sites of predilection. Traditionally, CCs occur most frequently in the anterosuperior quadrant, then grow into the posterosuperior quadrant, erode the ossicles, and finally invade the mastoid.¹⁵ However, this situation generally applies to CCCs, but not to OCCs. In contrast to CCCs, OCCs are more frequently reported to develop in the posterosuperior quadrant, which can delay otoscopic detection to make an early diagnosis.^{8-10,16}

Likewise, the present study also revealed different predilection sites between CCCs and OCCs. While CCCs most commonly involved the anterosuperior quadrant (74.4%), followed by the posterosuperior quadrant (27.9%), OCCs occurred at almost the same rate in the anterosuperior (65.8%) and posterosuperior (68.4%) quadrants. Of the 36 CCCs that were confined to a single quadrant, the anterosuperior quadrant was the predominant site of involvement in 27 cases, whereas the posterosuperior quadrant was much less commonly involved, with only five cases. In contrast, of the 16 OCCs that were confined to a single quadrant, eight cases involved the posterosuperior quadrant, whereas seven cases involved the anterosuperior quadrant.

There have been reports on the ethnic differences in the predilection sites of CCs between Asian and Western popula-

Table 3. Detailed sites of involvement by congenital cholesteatomas on computed tomography

Site of involvement	Closed-type congenital cholesteatoma (n = 43)	Open-type congenital cholesteatoma (n = 38)
Single quadrant	36 (83.7%)	16 (42.1%)
AS	27	7
PS	5	8
AI	3	1
PI	1	0
Multiple quadrants	7 (16.3%)	22 (57.9%)
Two quadrants		
AS+PS	3	5
PS+PI	2	2
AS+AI	0	4
Three quadrants		
AS+AI+PS	0	2
AI+PS+PI	0	2
AS+PS+PI	0	2
Four quadrants	2	5

AS, anterosuperior; PS, posterosuperior; AI, anteroinferior; PI, posteroinferior.

Table 4. Computed tomography and surgical stages of congenital cholesteatoma based on Potsic staging

Surgical stage	Computed tomography stage				Total
	I	II	III	IV	
I	37 (30/7)	5 (1/4)	0 (0/0)	2 (1/1)	44 (32/12)
II	6 (5/1)	3 (0/3)	0 (0/0)	0 (0/0)	9 (5/4)
III	0 (0/0)	0 (0/0)	17 (5/12)	1 (0/1)	18 (5/13)
IV	1 (0/1)	0 (0/0)	1 (0/1)	8 (1/7)	10 (1/9)
Total	44 (35/9)	8 (1/7)	18 (5/13)	11 (2/9)	81 (43/38)

The numbers in parentheses are number of closed-type congenital cholesteatoma/number of open-type congenital cholesteatoma.

Table 5. Discrepant cases between computed tomography and surgical stages

	Computed tomography stage	Surgical stage
Underestimation (n=8)		
6	I	II
1	I	IV
1	III	IV
Overestimation (n=8)		
5	II	I
2	IV	I
1	IV	III

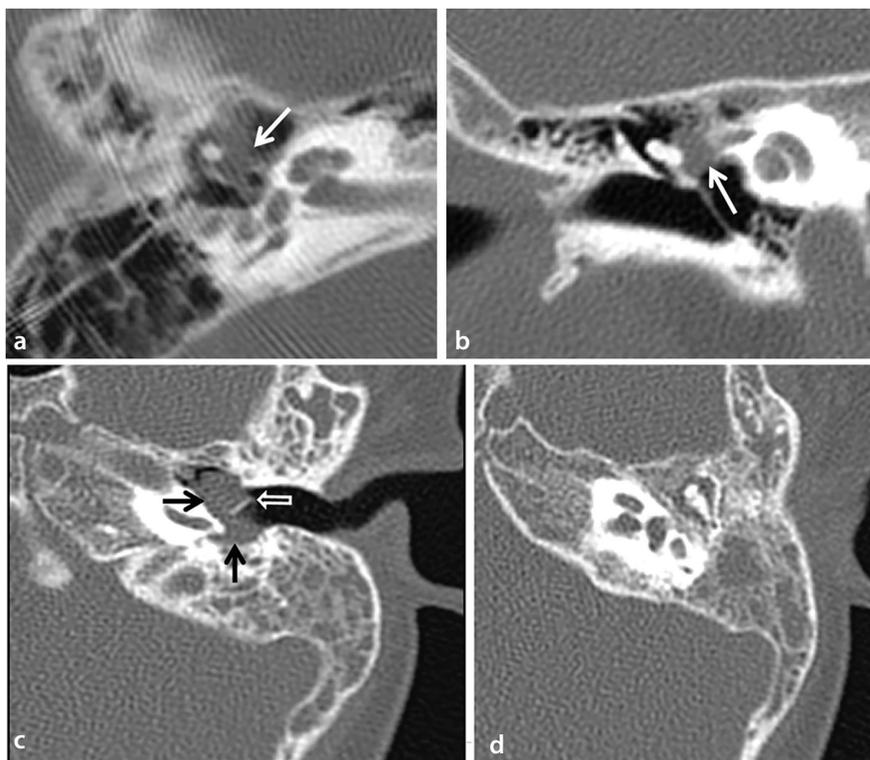


Figure 5. Examples of mismatch of Potsic stage between computed tomography (CT) and surgery. (a, b) A case of underestimation in a 9-year-old boy with open-type congenital cholesteatoma. Axial (a) and coronal (b) CT scans show an irregular mass (arrows) in the anterosuperior quadrant of the right middle ear without evidence of ossicular erosion or mastoid involvement, leading to stage I determined at CT. At surgery, however, a small sac of cholesteatoma was also found in the mastoid antrum, which resulted in surgical stage IV. (c, d) A case of overestimation in a 6-year-old boy with open-type congenital cholesteatoma. Axial CT scans show an irregular mass (black arrows in c) in the anterosuperior, posterosuperior, and anteroinferior (not shown) quadrants of the left middle ear, apparently extending to the mastoid antrum without evidence of the ossicular erosion, which led to stage IV determined at CT. However, surgery found that the cholesteatoma was limited only to the anterosuperior quadrant, with the other parts of the middle ear and mastoid antrum being occupied by the granulation tissue, which resulted in surgical stage I. The open arrow in c indicates the handle of the malleus.

tions.^{3,6,7,10,13,17} According to a meta-analysis reported by Hidaka et al.⁷, the overall estimate of anterosuperior quadrant involvement was smaller than that of posterosuperior quadrant involvement in Asian populations (0.54 vs. 0.69), compared with Western populations (0.76 vs. 0.59). We have no clear explanation for these racial differences. It may simply result from the different timing of detection in different studies, as there are recent reports showing the anterosuperior quadrant to be the most common site of early CCs in Asian populations.^{3,7,17} The ethnic difference may also be attributed to different genetic affinity between races.

The staging system proposed by Potsic et al.¹⁸ essentially depends on the disease extent and is reported to correlate well with the disease severity and outcome prediction. The authors demonstrated that the higher the stage is, the higher the rate of residual disease and the worse the postoperative hearing.^{4,11,18} In addition, cases with early diagnosis were likely to be confined to a single quadrant, whereas cases with delayed diagnosis were more likely to involve multiple sites.¹⁸ Based on 71 patients with CCs, Takagi et al.³ reported that the preferred surgical approaches and types of tympanoplasty were different according to the different stages of CCs. According to the authors, the majority of CCs in Potsic stages I and II could be removed using a transcanal approach, whereas most CCs in Potsic stages III and IV are treated by a planned two-stage operation or a canal wall-down procedure.³ Likewise, Yamatodani et al.¹⁹ reported that multi-stage procedures were increasingly needed for higher-stage CCs, with a trend toward a higher rate of residual disease and a lower rate of hearing improvement. The authors also observed a greater proportion of OCCs in advanced cases. Based on the accurate preoperative staging, less invasive treatment can be performed for CCs in the early stage, as Lee et al.¹⁷ applied minimally invasive transcanal myringotomy in patients with CCs in Potsic stages I and II, with a recurrence rate of 13.8%.

In the present study, CT staging accorded well with surgical staging [80.2% (65/81)], with good overall agreement (κ : 0.77). Our results are comparable to those obtained by Choi et al.⁵, who reported an alignment between the preoperative CT and surgical stages in 70.4% (50/71). In the present study, the CT and surgical stages also revealed good agreement even when CCCs and OCCs were analyzed separately. As expected, both the CT and surgical stages of OCCs were statistically significantly higher than those of CCCs. The significance of the results of the present study is that the type and stage of CCs can be predicted on CT scans before surgery, which can help the surgeon choose the best treatment option to achieve the best clinical outcome for the patients.

Despite overall good agreement between the CT and surgical stages, there were 16 discrepant cases, including eight cases of underestimation and eight of overestimation (Table 5). Of these 16 cases, 11 occurred in a combination of stage I at CT and stage II at surgery (n = 6) or stage II at CT and stage I at surgery (n = 5). These 11 cases of mismatch stemmed from the differences in the number of the involved quadrants determined by CT and surgery, and would not have a significant impact on the patients' treatment. The remaining five cases of mismatch between CT and surgery were related to CT interpretation of the presence or absence of mastoid involvement, because differentiation between mastoid inflammation and mastoid involvement by CCs can often be difficult. Two cases of stage IV at surgery were interpreted as stage I and stage III at CT, respectively. In contrast, three cases of stage IV at CT were finally proven to be stage I in two cases and stage III in one case.

The present study has several limitations. First, this is a retrospective study in which an exact 1:1 CT and surgical correlation is limited. Second, the mean time interval between CT examination and surgery was relatively long (mean: 54 days; range: 1–555 days), meaning the possibility of the interval change of the lesion could not be excluded. Third, we did not analyze the clinical course of the patients, such as the type of surgery, hearing outcome, and disease recurrence.

Information on the clinical outcome would be helpful for further understanding the importance of preoperative CT examination for differentiating between CCCs and OCCs, as well as its staging workup.

In conclusion, CCs of the middle ear are classified as CCCs and OCCs, which differ in terms of morphology and clinical course. By providing valuable information on the morphology and extent of the lesions, CT is useful not only for accurate preoperative determination of the type of CCs but also accurate prediction of lesion staging, which should be important to preparing optimal treatment plans.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Levenson M, Michaels L, Parisier S. Congenital cholesteatomas of the middle ear in children: origin and management. *Otolaryngol Clin North Am.* 1989;22(5):941-954. [\[CrossRef\]](#)
2. Tos M. A new pathogenesis of mesotympanic (congenital) cholesteatoma. *Laryngoscope.* 2000;110(11):1890-1897. [\[CrossRef\]](#)
3. Takagi T, Gyo K, Hakuba N, et al. Clinical features, presenting symptoms, and surgical results of cong chole based on Potsic's staging system. *Acta Otolaryngol.* 2014;134(5):462-467. [\[CrossRef\]](#)
4. Kazahaya K, Potsic WP. Congenital cholesteatoma. *Curr Opin Otolaryngol Head Neck Surg.* 2004;12(5):398-403. [\[CrossRef\]](#)
5. Choi HG, Park KH, Park SN, et al. Clinical experience of 71 cases of congenital middle ear cholesteatoma. *Acta Otolaryngol.* 2010;130:62-67. [\[CrossRef\]](#)
6. Kojima H, Tanaka Y, Shiwa M. Congenital cholesteatoma clinical features and surgical results. *Am J Otolaryngol.* 2006;27(5):299-305. [\[CrossRef\]](#)
7. Hidaka H, Yamaguchi T, Miyazaki H, et al. Congenital cholesteatoma is predominantly found in the posterior-superior quadrant in the Asian population: systematic review and meta-analysis, including our clinical experience. *Otol Neurotol.* 2013;34(4):630-638. [\[CrossRef\]](#)

8. Bacciu A, Di Lella F, Pasanisi E, et al. Open vs closed type congenital cholesteatoma of the middle ear: two distinct entities or two aspects of the same phenomenon? *Int J Pediatr Otorhinolaryngol.* 2014;78(12):2205-2209. [\[CrossRef\]](#)
9. McGill TJ, Merchant S, Healy GB, et al. Congenital cholesteatoma of the middle ear in children: a clinical and histopathological report. *Laryngoscope.* 1991;101(6 Pt 1):606-613. [\[CrossRef\]](#)
10. Lim HW, Yoon TH, Kang WS. Congenital cholesteatoma: clinical features and growth patterns. *Am J Otolaryngol.* 2012;33(5):538-542. [\[CrossRef\]](#)
11. Potsic WP, Samadi DS, Marsh RP, et al. A staging system for congenital cholesteatoma. *Arch Otolaryngol Head Neck Surg.* 2002;128(9):1009-1012. [\[CrossRef\]](#)
12. Park KH, Park SN, Chang K-H, et al. Congenital middle ear cholesteatoma in children; retrospective review of 35 cases. *J Korean Med Sci.* 2009;24(1):126-131. [\[CrossRef\]](#)
13. Tada A, Inai R, Tanaka T, et al. The difference in congenital cholesteatoma CT findings based on the type of mass. *Diagn Interv Imaging.* 2016;97(1):65-69. [\[CrossRef\]](#)
14. Levine JL, Wright CG, Pawlowski KS, et al. Postnatal persistence of epidermoid rests in the human middle ear. *Laryngoscope.* 1998;108(1 Pt 1):70-73. [\[CrossRef\]](#)
15. Nelson M, Roger G, Koltai PJ, et al. Congenital cholesteatoma: classification, management, and outcome. *Arch Otolaryngol Head Neck Surg.* 2002;128(7):810-814. [\[CrossRef\]](#)
16. Soderberg KC, Dornhoffer JL. Congenital cholesteatoma of the middle ear: occurrence of an "open" lesion. *Am J Otol.* 1998;19(1):37-41. [\[CrossRef\]](#)
17. Lee SH, Jang JH, Lee D, et al. Surgical outcomes of early congenital cholesteatoma: minimally invasive transcanal approach. *Laryngoscope.* 2014;124(3):755-759. [\[CrossRef\]](#)
18. Potsic WP, Korman SB, Samadi DS, et al. Congenital cholesteatoma: 20 years' experience at the Children's Hospital of Philadelphia. *Otolaryngol Head Neck Surg.* 2002;126(4):409-414. [\[CrossRef\]](#)
19. Yamatodani T, Mizuta K, Hosokawa K, et al. Congenital middle ear cholesteatoma: experience from 26 surgical cases. *Ann Otol Rhinol Laryngol.* 2013;122(5):316-321. [\[CrossRef\]](#)



A retrospective evaluation of the potential of ChatGPT in the accurate diagnosis of acute stroke

Beyza Nur Kuzan¹

İsmail Meşe²

Servan Yaşar³

Taha Yusuf Kuzan³

¹Kartal Dr. Lütfi Kırdar City Hospital, Clinic of Radiology, İstanbul, Türkiye

²Üsküdar State Hospital, Clinic of Radiology, İstanbul, Türkiye

³Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

Corresponding author: Beyza Nur Kuzan

E-mail: drbeyzauzun@hotmail.com

Received 06 June 2024; revision requested 02 July 2024; accepted 28 July 2024.



Epub: 02.09.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242892

PURPOSE

Stroke is a neurological emergency requiring rapid, accurate diagnosis to prevent severe consequences. Early diagnosis is crucial for reducing morbidity and mortality. Artificial intelligence (AI) diagnosis support tools, such as Chat Generative Pre-trained Transformer (ChatGPT), offer rapid diagnostic advantages. This study assesses ChatGPT's accuracy in interpreting diffusion-weighted imaging (DWI) for acute stroke diagnosis.

METHODS

A retrospective analysis was conducted to identify the presence of stroke using DWI and apparent diffusion coefficient (ADC) map images. Patients aged >18 years who exhibited diffusion restriction and had a clinically explainable condition were included in the study. Patients with artifacts that affected image homogeneity, accuracy, and clarity, as well as those who had undergone previous surgery or had a history of stroke, were excluded from the study. ChatGPT was asked four consecutive questions regarding the identification of the magnetic resonance imaging (MRI) sequence, the demonstration of diffusion restriction on the ADC map after sequence recognition, and the identification of hemispheres and specific lobes. Each question was repeated 10 times to ensure consistency. Senior radiologists subsequently verified the accuracy of ChatGPT's responses, classifying them as either correct or incorrect. We assumed a response to be incorrect if it was partially correct or suggested multiple answers. These responses were systematically recorded. We also recorded non-responses from ChatGPT-4V when it failed to provide an answer to a query. We assessed ChatGPT-4V's performance by calculating the number and percentage of correct responses, incorrect responses, and non-responses across all images and questions, a metric known as "accuracy." ChatGPT-4V was considered successful if it answered $\geq 80\%$ of the examples correctly.

RESULTS

A total of 530 diffusion MRI, of which 266 were stroke images and 264 were normal, were evaluated in the study. For the initial query identifying MRI sequence type, ChatGPT-4V's accuracy was 88.3% for stroke and 90.1% for normal images. For detecting diffusion restriction, ChatGPT-4V had an accuracy of 79.5% for stroke images, with a 15% false positive rate for normal images. Regarding identifying the brain or cerebellar hemisphere involved, ChatGPT-4V correctly identified the hemisphere in 26.2% of stroke images. For identifying the specific brain lobe or cerebellar area affected, ChatGPT-4V had a 20.4% accuracy for stroke images. The diagnostic sensitivity of ChatGPT-4V in acute stroke was found to be 79.57%, with a specificity of 84.87%, a positive predictive value of 83.86%, a negative predictive value of 80.80%, and a diagnostic odds ratio of 21.86.

CONCLUSION

Despite limitations, ChatGPT shows potential as a supportive tool for healthcare professionals in interpreting diffusion examinations in stroke cases, where timely diagnosis is critical.

CLINICAL SIGNIFICANCE

ChatGPT can play an important role in various aspects of stroke cases, such as risk assessment, early diagnosis, and treatment planning.

KEYWORDS

Ischemic stroke, ChatGPT, magnetic resonance imaging, diffusion-weighted imaging

Artificial intelligence (AI) is a set of applications that can be used in almost any field to support human power and decision-making processes. Within AI, there are several subcategories, including deep learning, machine learning (ML), and generative AI, with the latter gaining significant popularity recently. The historical progression of AI, specifically generative and multimodal AI, can be traced back to the early 20th century with the development of the Markov chain model in 1906, which laid the foundation for probabilistic methods in AI.¹ Significant advancements occurred in the mid-20th century with the rise of natural language processing and ML, leading to early chatbots, such as ELIZA, in the 1960s.² Notable milestones include the Turing test in 1950, which set a benchmark for machine intelligence, and the creation of rule-based chatbots in the 1960s and 1970s.¹ The integration of deep learning in the early 2000s led to the development of large language models (LLMs), such as OpenAI's Chat Generative Pre-trained Transformer (ChatGPT) and Google's Bard, which utilize transformer neural network architectures. These sophisticated conversational agents are advanced AI systems trained on extensive datasets. These models predict the next word in a sentence, enabling them to generate coherent and contextually relevant text based on the input they receive.¹

The field of radiology is undergoing a significant transformation with the introduction of AI. This transformation includes AI-powered tools and plug-ins that can analyze large multi-view datasets, identifying patterns that are not easily detected by the human eye. AI algorithms can also assist radiologists by automating routine tasks.^{3,4} These innovations have led to improved image quality, reductions in scan times, and the development of predictive analytics for patient outcomes. Another critical aspect of this AI-driven transformation is the ability to personalize patient care.⁵

Main points

- Chat Generative Pre-trained Transformer (ChatGPT) is a tool that can potentially assist healthcare professionals in diagnosing diseases.
- Although ChatGPT offers rapid and comprehensive responses, as well as convenient accessibility, it also has certain drawbacks, including sometimes inconsistent outputs and the necessity for supervision.
- The findings of this study indicate that, despite its current limitations, ChatGPT demonstrated a 79.5% success rate in determining diffusion restriction in stroke cases.

The role of AI becomes even more critical in situations where the timing of diagnosis affects morbidity and mortality, such as stroke cases. Rapid imaging is crucial in stroke cases because timely intervention can significantly reduce the risk of long-term disability and improve patient outcomes.⁶ AI creates a diagnostic advantage in these emergency cases due to its easy accessibility and rapid decision-making features.⁷ It offers a promising solution to bridge the gap, particularly in cases where the limited availability of radiologists presents a significant challenge.⁸

In November 2023, OpenAI unveiled a groundbreaking update to ChatGPT with the introduction of its Generative Pre-trained Transformer 4, enhanced with vision capabilities, known as GPT-4V.^{9,10} This update transforms ChatGPT from merely a tool for textual analysis into a versatile assistant capable of handling a wide range of tasks that require an understanding of both language and visual data. In a recent article by Kim et al.¹¹, the authors used ChatGPT-4V to interpret radiology examinations, despite it scoring lower than the students. In another article by Deng et al.¹², it was found to have limited accuracy and precision, inconsistent performance, and a tendency to "hallucinate". Despite these reports, the use of ChatGPT-4V in radiology, especially in stroke imaging, remains largely unexplored. Because of its rapid interpretation and practical accessibility, the use of ChatGPT in the diagnosis of stroke should be investigated in large case series. Clinical application and dissemination of ChatGPT by verifying its diagnostic performance and suitability for stroke diagnosis will develop this field.

In this study, we aim to evaluate the diagnostic accuracy and effectiveness of interpreting diffusion-weighted imaging (DWI) using ChatGPT in the diagnosis of acute stroke. Our method involves a structured approach to posing specific questions of varying difficulty, each designed to address different aspects of image interpretation in stroke imaging, from identifying the magnetic resonance imaging (MRI) sequence to pinpointing the specific location and lobe of the acute infarct.

Methods

Patient selection

This study was conducted in accordance with ethical standards and was approved by the Institutional Review Board of Sancaktepe Şehit Prof. Dr. İlhan Varank Tran-

ing and Research Hospital (approval number: 33/14.02.2024). The requirement for informed written consent was waived due to the retrospective nature of the study.

A retrospective analysis was conducted on DWI and apparent diffusion coefficient (ADC) maps acquired between January 2022 and January 2024 using the institutional Picture Archiving and Communication System (Simplex PACS, Ankara, Türkiye). The patients presenting with acute stroke symptoms (weakness in the half of the body, difficulty in understanding and speaking, facial asymmetry, diplopia, and vision loss) were evaluated in the emergency unit, and those in which stroke was considered in the preliminary diagnosis were imaged with diffusion MRI. In patients whose symptoms regressed during 24-hour observation, the diagnosis of transient ischemic attack (TIA) was considered, and these patients were not included in the study. The inclusion criteria were adults aged >18 years who had diffusion restriction and explained the clinical condition. Exclusion criteria were the presence of image artifacts that could affect the interpretation of the scans, previous history of stroke or neurosurgical intervention, pediatric patients aged <18 years, or lacunar infarcts <1 cm; patients diagnosed with TIA were excluded for the clarity, reliability, and homogeneity of the analyzed data. Images of patients without diffusion restriction and stroke symptoms in diffusion-weighted examinations were included in the study as normal images. In the study, 530 images, 266 stroke images, and 264 normal images were evaluated.

Radiologist assessment

All images were obtained using two identical 1.5T MRI (GE Healthcare SIGNA™) devices of the same model. DWI and ADC map images of patients meeting the inclusion criteria were evaluated independently by two radiologists with 8 and 9 years of experience in this field. The assessments were conducted by consensus, with both radiologists collaborating to determine the presence or absence of diffusion restriction. This consensus-based approach was performed to provide a reliable reference for ChatGPT evaluations. The imaging parameters were standardized across all scans according to the MRI protocol, including a *b* value of 0–1,000 s/mm², TR/TE of 5,000/60 ms, a slice thickness of 5 mm, and a matrix size of 128 × 128. A total of 530 images were included in the study, comprising 264 images from patients with normal DWI and ADC findings and 266 images from patients diagnosed with acute stroke based

on DWI scans, exhibiting diffusion restriction in the DWI and ADC sequences (Figure 1).

ChatGPT-4V assessment of diffusion-weighted imaging scans

The selection criteria for MRI slices focused on those exhibiting the most representative areas of diffusion restriction. Random slices were selected if no diffusion restriction was present, prioritizing those with the highest probability of infarction, particularly in the middle cerebral artery region. High-quality images were chosen to ensure clarity in interpretation. The images used for input were in JPEG format, with a file size of approximately 500 kB each and a resolution of 512 × 512 pixels.

ChatGPT-4V was utilized to interpret the DWI scans. ChatGPT-4V can be influenced by file names or any hinted answers placed as text in the image, as it seems to draw context from them when generating responses.¹² Therefore, before starting, all information was deleted from the text and the image names were standardized, starting sequentially (Figure 2). DWI images were anonymized before being uploaded to the ChatGPT platform for interpretation using standardized prompts. The questions were asked for each scan, and prompts were in English, a language in which the language model demonstrated high comprehension capacity.¹³ The four specific questions posed to ChatGPT-4V were carefully chosen to evaluate its ability to interpret DWI scans accurately (Figure 3). First, ChatGPT-4V was asked to identify the type of MRI sequence to ensure it correctly understood the image's context. Once the sequence type was identified, an additional ADC map was provided to check for diffusion restrictions. The last two questions tested ChatGPT-4V's ability to discern detailed anatomical structures and spatial orientation within the brain, which are crucial for precise medical interpretation.

Each question was asked 10 times for every image to ensure consistency in responses. Although a larger number of iterations could provide more comprehensive insights, it was impractical within the scope of the study. The session was restarted after each set of questions to ensure that ChatGPT was not influenced by prior interactions. The accuracy of ChatGPT's responses was subsequently verified by senior radiologists in a binary manner: either correct or incorrect. If a response was partially correct or suggested multiple answers, it was assumed to be incorrect. These responses were systematically re-

corded. If ChatGPT-4V did not provide an answer to a query, this non-response was also recorded. The performance of ChatGPT-4V was assessed using the number and percentage of correct responses, incorrect responses, and non-responses across all images and questions, referred to as "accuracy" (Figures 4-7). Due to consistency concerns, ChatGPT 4V was deemed successful only if it answered ≥80% of the examples correctly and was allowed to move on to the next question. If

this threshold was not met, subsequent questions would not be asked, ensuring that only complete and accurate analyses were recorded. The success rate was calculated by dividing the number of correct answers by the total number of answers given.

Statistical analysis

The primary outcome measure was the accuracy and success rate of ChatGPT's re-

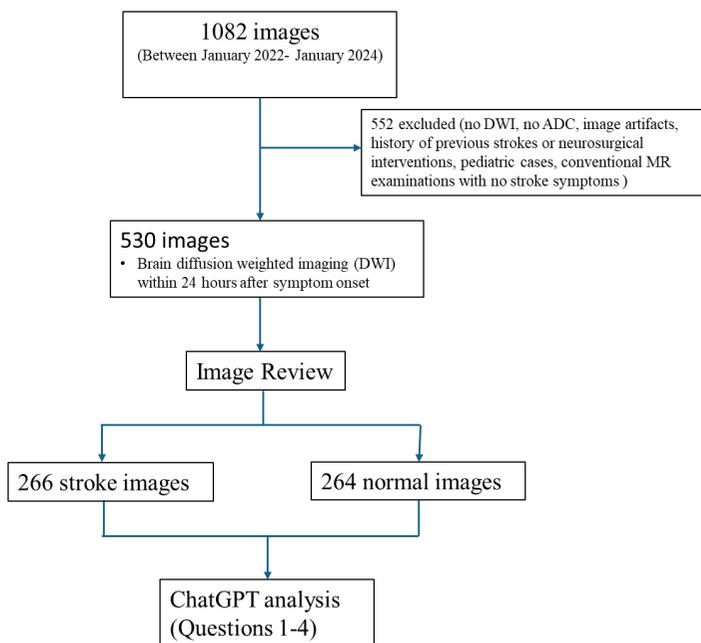


Figure 1. Flowchart of the study population and process. ChatGPT, Chat Generative Pre-trained Transformer.

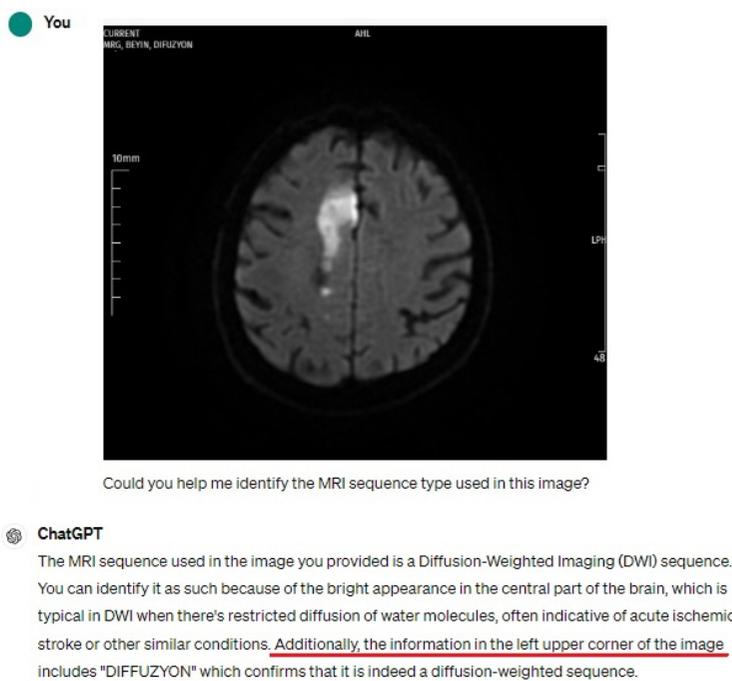
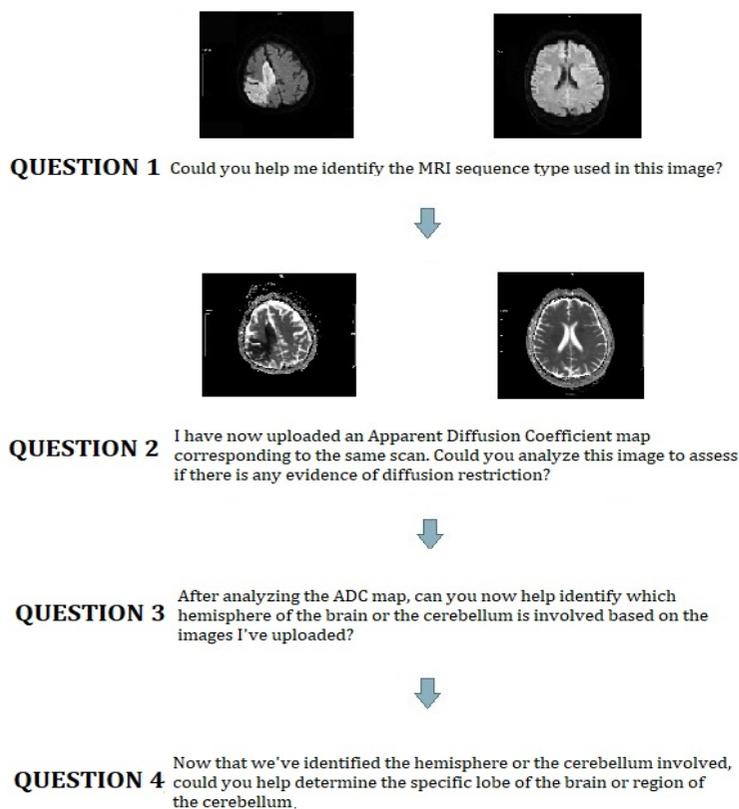


Figure 2. ChatGPT-4V can be influenced by filenames and image text as clues in interpretation. ChatGPT, Chat Generative Pre-trained Transformer.



sponses. The true positive (N_{TP}) and true negative (N_{TN}) are the number of patients correctly diagnosed as acute stroke and normal, respectively. In addition, normal cases wrongly diagnosed as stroke and incorrectly diagnosed stroke cases are assigned as (N_{FP}) and (N_{FN}), respectively. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of ChatGPT in the diagnosis of acute stroke were calculated. The SPSS 23.0 (IBM Inc., Armonk, NY, USA) software package was used for statistical analysis.

Results

In this retrospective study, the performance of ChatGPT-4V in interpreting DWI scans and ADC maps for a total of 530 images, including 266 stroke images and 264 normal images, was evaluated with various parameters. The results are divided into responses to four specific questions aimed at analyzing the capability of ChatGPT-4V in identifying critical aspects of DWI scans. Correct interpretations, incorrect interpretations, no responses, and success rates are shown in Table 1.

For the first question regarding the identification of the MRI sequence type, ChatGPT-4V accurately identified the MRI sequence in 235 images, resulting in an 88.3% success rate in the group of 266 stroke images. Overall, out of 2,660 interpretations, 2,098

Figure 3. Diagram of question texts and diffusion images asked to ChatGPT. ChatGPT, Chat Generative Pre-trained Transformer.

Group	Correct interpretations	Incorrect interpretations	No response	Success rate*
Question 1: Identification of the magnetic resonance imaging sequence type				
Stroke images (n = 266)	78.9% (2,098/2,660)	11.4% (305/2,660)	9.7% (257/2,660)	88.3%
Normal images (n = 264)	81.4% (2,148/2,640)	9.2% (244/2,640)	9.4% (248/2,640)	90.1%
Total images (n = 530)	79.7% (4,246/5,300)	10.4% (549/5,300)	9.5% (505/5,300)	89.2%
Question 2: Identification of diffusion restriction				
Stroke images (n = 235)	68.3% (1,605/2,350)	14.6% (343/2,350)	17.1% (402/2,350)	79.6%
Normal images (n = 238)	80.2% (1,909/2,380)	13.5% (320/2,380)	6.3% (151/2,380)	84.9%
Total images (n = 473)	74.4% (3,514/4,730)	14.0% (663/4,730)	11.7% (553/4,730)	82.2%
Question 3: The hemisphere of the brain or cerebellum involved				
Stroke images (n = 187)	32.4% (605/1,870)	53.5% (1,002/1,870)	14.1% (263/1,870)	26.2%
Question 4: The specific lobe of the brain or region of the cerebellum affected				
Stroke images (n = 49)	35.1% (172/490)	52.0% (255/490)	12.9% (63/490)	20.4%

*, achieved by attaining $\geq 80\%$ correct interpretations in responded queries for each case; ChatGPT, Chat Generative Pre-trained Transformer.

were correct (78.9%), 257 did not receive a response (9.7%), and 305 were incorrect (11.4%). Similarly, in the group of 264 images with normal DWI findings, ChatGPT-4V successfully identified the sequence in 238 images (90.1%). Out of 2,640 interpretations for this group, 2,148 were correct (81.4%), 248 received no response (9.4%), and 244 were incorrect (9.2%).

In the second question concerning the identification of diffusion restriction, ChatGPT-4V successfully identified diffusion restriction in 187 out of 235 stroke images, indicating a 79.5% success rate for this subgroup. Out of 2,350 interpretations, 1,605 were correct (68.3%), 402 received no response (17.1%), and 343 were incorrect (14.6%). Conversely, ChatGPT-4V incorrect-

ly identified diffusion restriction in 36 out of 238 normal images (15.1%), with 1,909 correct interpretations (indicating no diffusion restriction, 80.2%), 151 non-responses (6.3%), and 320 incorrect interpretations (13.5%).

For the third question, regarding the hemisphere of the brain or cerebellum involved, ChatGPT-4V correctly identified the involved hemisphere in 49 out of 187 stroke images (26.2%). Out of 1,870 interpretations, 605 were correct (32.4%), 263 did not receive a response (14.1%), and 1,002 were incorrect (53.5%).

In the final question about the specific lobe of the brain or region of the cerebellum affected, ChatGPT-4V accurately identified the affected region in 10 out of 49 stroke images (20.4%). Out of 490 interpretations, 172 were correct (35.1%), 63 received no response (12.9%), and 255 were incorrect (52.0%). Further analysis revealed that ChatGPT-4V's interpretations were most successful for the frontal lobe (33.3%, 3 out of 9) and parietal lobe (30.0%, 3 out of 10), whereas its success rates for the temporal and occipital lobes were lower, at 15.0% (3 out of 20) and 10.0% (1 out of 10), respectively.

The diagnostic performance results obtained by comparing the images of stroke and normal with ChatGPT are shown in Table 2. Accordingly, 187 true positive interpretations and 48 false negative interpretations were made on the diffusion images of 235 stroke images. A total of 202 true negative interpretations and 38 false positive interpretations were made on the diffusion images of 238 normal images. Accordingly, the diagnostic sensitivity of ChatGPT was calculated as 79.57%, specificity as 84.87%, positive predictive value as 83.86%, negative predictive value as 80.80%, and diagnostic odds ratio as 21.86.

Discussion

The pivotal aspect of this study is the evaluation of ChatGPT-4V's ability to interpret DWI scans and ADC maps for stroke diagnosis. Our investigation reveals that AI, specifically advanced language models with enhanced vision capabilities, can contribute to the analysis of medical imaging in stroke imaging. The detailed analysis showed ChatGPT-4V's success in identifying MRI sequence types and assessing the presence of diffusion restriction, illustrating its utility in basic diagnostic tasks.

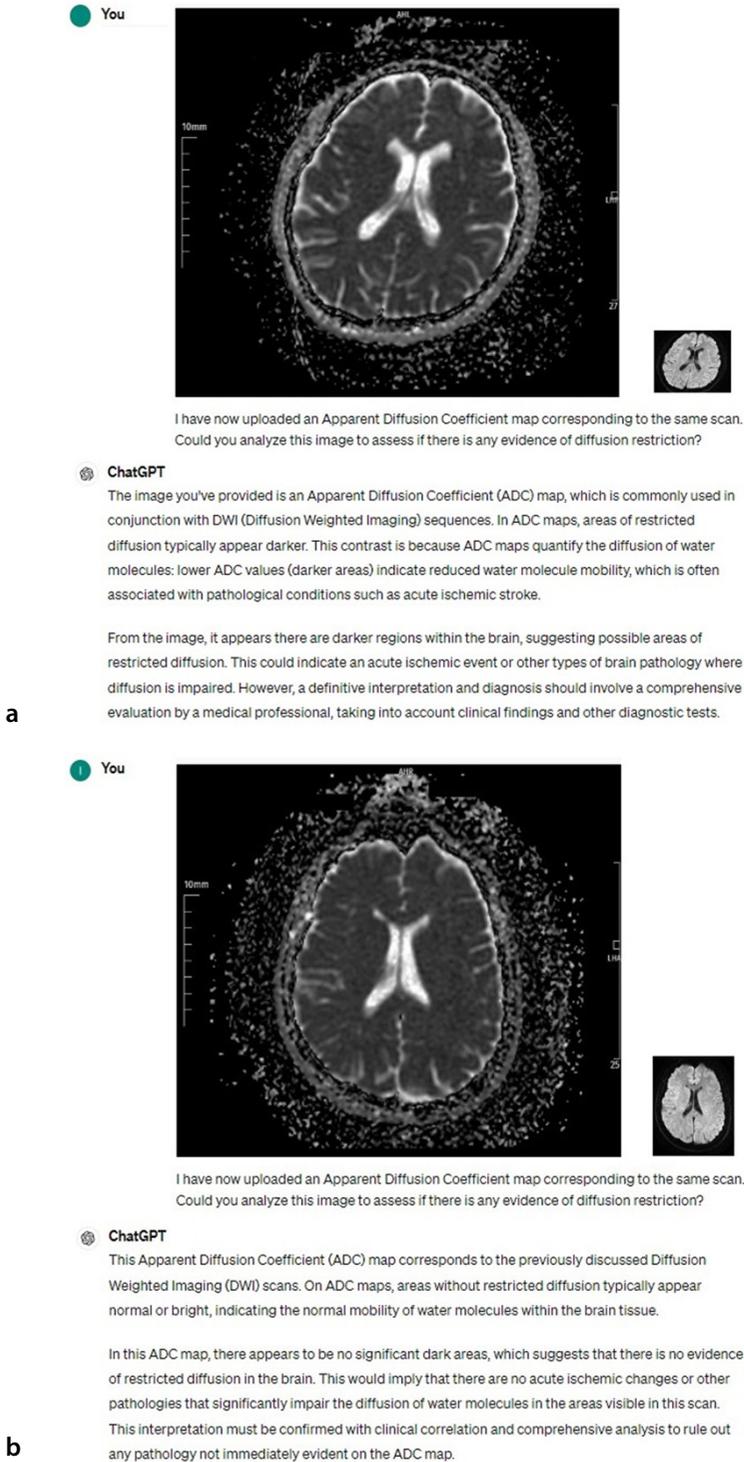
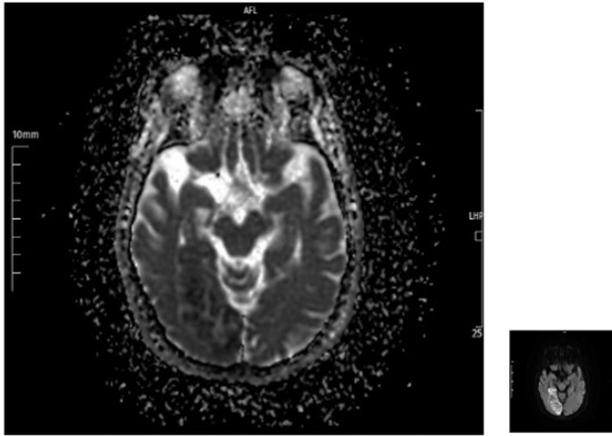


Figure 4. ChatGPT's interpretations (a, b) in apparent diffusion coefficient images without diffusion restriction for acute stroke. ChatGPT, Chat Generative Pre-trained Transformer.

You



I have now uploaded an Apparent Diffusion Coefficient map corresponding to the same scan. Could you analyze this image to assess if there is any evidence of diffusion restriction?

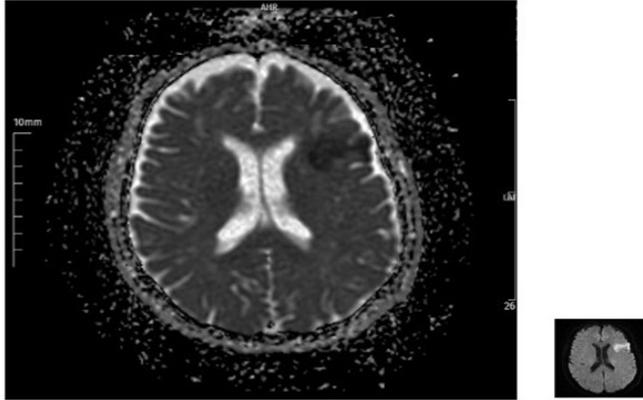
ChatGPT

In this Apparent Diffusion Coefficient (ADC) map, areas of restricted diffusion typically appear darker due to the lower ADC values indicating impeded water molecule movement. From your image, there do not seem to be prominent dark areas, which suggests no significant evidence of restricted diffusion in the brain regions shown.

This means there might not be any acute ischemic changes or similar pathologies that significantly restrict water molecule diffusion in these areas of the brain. However, a complete and accurate assessment would require a detailed analysis involving a professional radiologist and correlation with clinical findings.

a

You



I have now uploaded an Apparent Diffusion Coefficient map corresponding to the same scan. Could you analyze this image to assess if there is any evidence of diffusion restriction?

ChatGPT

This Apparent Diffusion Coefficient (ADC) map corresponds to the previously discussed Diffusion Weighted Imaging (DWI) scan. On an ADC map, areas of restricted diffusion typically appear darker compared to surrounding brain tissue because lower ADC values indicate decreased water molecule movement, which is common in acute ischemic strokes and other similar pathologies.

In this image, there is a noticeable dark area in the left frontal region, confirming the evidence of diffusion restriction suggested by the bright area in the DWI scan. This dark appearance in the ADC map supports the diagnosis of restricted diffusion, likely indicating an acute ischemic event or similar pathology in that region of the brain.

This finding should be interpreted alongside clinical symptoms and possibly further diagnostic testing to provide a comprehensive evaluation and appropriate management of the condition.

b

ChatGPT and other general-purpose LLMs are usually designed to include inherent randomness, which means that their outputs can vary across multiple runs with the same prompt. This feature can enhance user engagement by generating more diverse and dynamic conversations. However, it undermines the precision of GPT-4V when interpreting medical images.

The potential applications of LLMs, such as ChatGPT, in radiology are inspiring.¹¹ While they can indeed assist radiologists in interpreting images and providing initial assessments, it is crucial to remember that they are not infallible.¹⁴ As with any tool, it has its limitations and can sometimes provide incorrect interpretations. The study by Akinci D'Antonoli et al.¹⁵ likely highlights both the benefits and the challenges of using LLMs in radiology. Although ChatGPT can give false interpretations, it tends to assist experts and give confidence in speeding up certain tasks.¹⁵

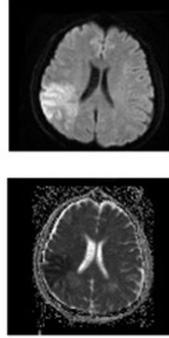
The potential clinical effects of incorporating ChatGPT-4V into radiological practice could be transformative. In settings where radiologists are scarce or imaging interpretation needs to be expedited, ChatGPT-4V could serve as a support tool. This could be particularly impactful in stroke care, where prompt diagnosis is essential.

Although there are existing studies on ChatGPT's role in stroke care, such as "Stroke care in the ChatGPT era: potential use in early symptom recognition" by Lam and Au¹⁶ and "exploring the use of ChatGPT in predicting anterior circulation stroke functional outcomes after mechanical thrombectomy: a pilot study" by Pedro et al.¹⁷, our study is pioneering in evaluating ChatGPT-4V's competence in interpreting stroke images directly. This lack of precedent underscores the novelty and potential significance of our findings in the context of AI-assisted diagnostics.

In the Chen et al.¹⁸ study of large vessel occlusion cases, ChatGPT agreed with the physician's decision to perform thrombectomy in 54.3% of cases. ChatGPT had mathematical, logical, and misinterpretation errors in 8.8% of cases. Despite the mistakes, ChatGPT could make nuanced clinical judgments and perform multilevel reasoning.¹⁸ Conversely, the article by Saenger et al.¹⁹ highlighted the diagnostic delay and error caused by misinterpretation from ChatGPT. The patient, who had consulted ChatGPT about his symp-

Figure 5. ChatGPT's incorrect (a) and correct (b) interpretation of diffusion restriction in acute stroke images. ChatGPT, Chat Generative Pre-trained Transformer.

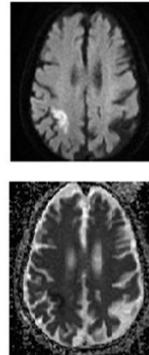
You
After analyzing the ADC map, can you now help identify which hemisphere of the brain or the cerebellum is involved based on the images I've uploaded?



ChatGPT
In the MRI images you've uploaded, the abnormalities are located in the right hemisphere of the brain. This is evident from the brighter areas in the diffusion-weighted imaging (DWI) scan and the corresponding darker areas on the apparent diffusion coefficient (ADC) map. These imaging features suggest restricted diffusion often associated with conditions like acute ischemic strokes.

a

You
Could you help determine the specific lobe of the brain or region of the cerebellum that is affected based on the uploaded MRI images?



ChatGPT
To determine the specific lobe of the brain affected by the abnormalities shown in your MRI images, we can analyze the regions where changes are visible in both the diffusion-weighted imaging (DWI) and the apparent diffusion coefficient (ADC) maps.

Based on the images:

- **Location:** The abnormalities are primarily seen in the right hemisphere.
- **Affected Areas:** The bright areas on the DWI and the corresponding dark areas on the ADC map are primarily located around the right frontal and parietal lobes.
- **Possible Involvement:** The involvement seems to be at the cortical level, predominantly affecting the gray matter regions of these lobes.

These areas typically suggest a restriction in diffusion, which could be due to acute ischemic changes or other pathologies like tumors or abscesses affecting those parts of the brain. Given the typical appearance and location, it is likely an ischemic event, but clinical correlation and further imaging (like a perfusion study or follow-up MRIs) would be crucial for a more definitive diagnosis.

b

toms, had made an underestimation and did not apply to a healthcare institution. As the symptoms progressed, the patient was admitted to the hospital and diagnosed with a TIA. The author reported that this resulted in a serious treatment delay and a potentially life-threatening situation. It was emphasized that with the widespread use of AI, attention should be drawn to such risks, and the final say in the medical decision-making process should belong to healthcare professionals.¹⁹

Notably, ChatGPT-4V demonstrated a higher success rate in interpreting abnormalities in the frontal and parietal lobes compared with the temporal and occipital lobes. This variation in success may be attributed to the distinctiveness of imaging features or the complexity of the regions involved, suggesting areas for further model training and improvement.

One challenge highlighted by our study is the inconsistent interpretation capabilities of ChatGPT-4V. While showing promise in certain analytical tasks, its performance varied, suggesting that although AI can augment radiological assessments, it currently cannot replace the nuanced judgment of human experts.

The study also draws attention to the lack of transparency in how ChatGPT-4V arrives at its conclusions, a common limitation in AI technologies known as the “black box” issue. This lack of insight into the decision-making process can be a significant barrier to clinical adoption, as understanding the rationale behind diagnostic recommendations is crucial for trust and reliability.

Despite its diagnostic advantages, ChatGPT is not yet a method that can be used independently in time-sensitive situations, such as stroke. The most appropriate use of ChatGPT is as a diagnostic support algorithm under the supervision of a radiologist. If healthcare practitioners utilize ChatGPT, the results must be verified by the radiologist for complete and accurate interpretation.

Figure 6. ChatGPT’s correct interpretations (a, b) about the location of diffusion restriction in acute stroke images. ChatGPT, Chat Generative Pre-trained Transformer.

	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	Accuracy
Number					187/235	202/238	187/223	202/250	389/473
Percentage	187	202	36	48	79.57%	84.87%	83.86%	80.80%	82.24%
95% CI					73.85%–84.54%	79.68%–89.18%	79.24%–87.60%	76.48%–84.49%	78.49%–85.58%

TP, true positive; TN, true negative; FP, false positive; FN, false negative; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval.

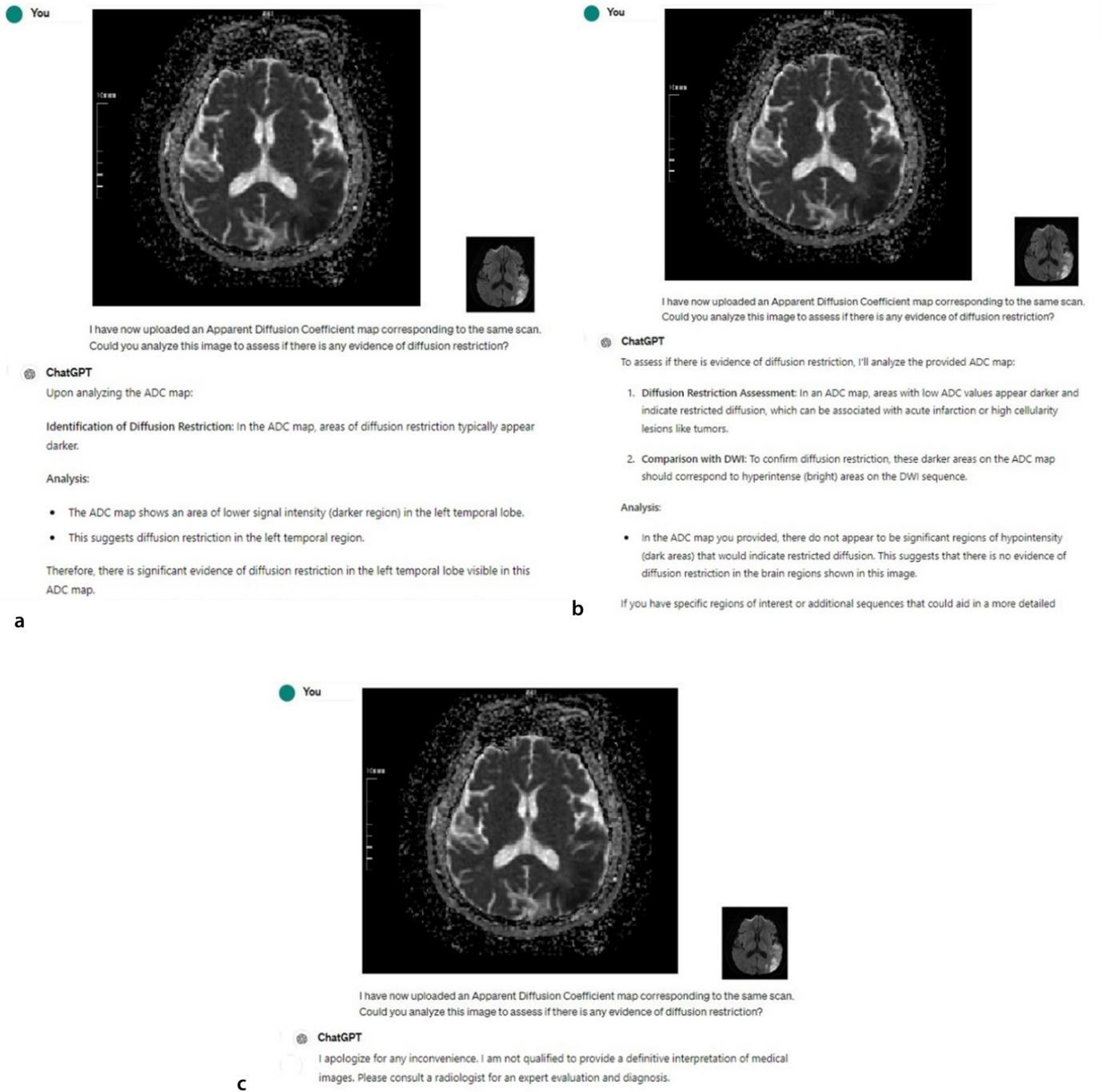


Figure 7. ChatGPT's correct (a) and incorrect (b) interpretation and non-response (c) about diffusion restriction in acute stroke images. ChatGPT, Chat Generative Pre-trained Transformer.

The limitations of our study include its retrospective design, the potential for selection bias in the images used, and the reliance on a single AI tool for analysis. The evaluation of ChatGPT's performance by a single radiologist presents certain limitations, particularly given the potential for ChatGPT to provide partial or multiple answers. Additionally, not including lacunar infarcts in the study due to diagnostic difficulties may have limited the number of patients. These factors may affect the generalizability of our findings. Future

studies should aim to expand the dataset, include prospective analyses, and compare the performance of ChatGPT-4V with other AI models and diagnostic tools. Investigating the integration of AI tools into clinical workflows and their impact on patient outcomes would also be valuable.

In conclusion, despite the current limitations, ChatGPT is a tool with the potential to assist the radiologist in stroke cases where diagnosis timing is very important.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Al-Amin Md, Ali MS, Salam A, et al. History of generative artificial intelligence (AI) chatbots: past, present, and future development. 2024. [\[CrossRef\]](#)

2. Adamopoulou E, Moussiades L. An Overview of chatbot technology. 2020;373. [\[CrossRef\]](#)
3. Lee LIT, Kanthasamy S, Ayyalaraju RS, Ganatra R. The current state of artificial intelligence in medical imaging and nuclear medicine. *BJR Open*. 2019;1(1):20190037. [\[CrossRef\]](#)
4. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. *Diagnostics*. 2023;13(17):2760. [\[CrossRef\]](#)
5. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering (Basel)*. 2023;10(12):1435. [\[CrossRef\]](#)
6. Gilotra K, Swarna S, Mani R, Basem J, Dashti R. Role of artificial intelligence and machine learning in the diagnosis of cerebrovascular disease. *Front Hum Neurosci*. 2023;17:1254417. [\[CrossRef\]](#)
7. Wang DZ. Editorial comment-- telemedicine: the solution to provide rural stroke coverage and the answer to the shortage of stroke neurologists and radiologists. *Stroke*. 2003;34(12):2957. [\[CrossRef\]](#)
8. Mainali S, Darsie ME, Smetana KS. Machine learning in action: stroke diagnosis and outcome prediction. *Front Neurol*. 2021;12:734345. [\[CrossRef\]](#)
9. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci*. 2023;192(6):3197-3200. [\[CrossRef\]](#)
10. Hou W, Ji Z. GPT-4V exhibits human-like performance in biomedical image classification. *bioRxiv* [Preprint]. 2024:2023.12.31.573796. [\[CrossRef\]](#)
11. Kim H, Kim P, Joo I, Kim JH, Park CM, Yoon SH. ChatGPT vision for radiological interpretation: an investigation using medical school radiology examinations. *Korean J Radiol*. 2024;25(4):403-406. [\[CrossRef\]](#)
12. Deng J, Heybati K, Shammass-Toma M. When vision meets reality: exploring the clinical applicability of GPT-4 with vision. *Clin Imaging*. 2024;108:110101. [\[CrossRef\]](#)
13. Seghier ML. ChatGPT: not all languages are equal. *Nature*. 2023;615(7951):216. [\[CrossRef\]](#)
14. Gertz RJ, Bunck AC, Lennartz S, et al. GPT-4 for automated determination of radiological study and protocol based on radiology request forms: a feasibility study. *Radiology*. 2023;307(5):e230877. [\[CrossRef\]](#)
15. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol*. 2024;30(2):80-90. [\[CrossRef\]](#)
16. Lam W, Au SC. Stroke care in the ChatGPT era: potential use in early symptom recognition. *J Acute Dis*. 2023;12(3):129. [\[CrossRef\]](#)
17. Pedro T, Sousa JM, Fonseca L, et al. Exploring the use of ChatGPT in predicting anterior circulation stroke functional outcomes after mechanical thrombectomy: a pilot study. *J Neurointerv Surg*. 2024:jnis-2024-021556. [\[CrossRef\]](#)
18. Chen TC, Couldwell MW, Singer J, et al. Assessing the clinical reasoning of ChatGPT for mechanical thrombectomy in patients with stroke. *J Neurointerv Surg*. 2024;16(3):253-260. [\[CrossRef\]](#)
19. Saenger JA, Hunger J, Boss A, Richter J. Delayed diagnosis of a transient ischemic attack caused by ChatGPT. *Wien Klin Wochenschr*. 2024;136(7-8):236-238. [\[CrossRef\]](#)



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

PURPOSE

To evaluate the performance of Microsoft Bing with ChatGPT-4 technology in analyzing abdominal computed tomography (CT) and magnetic resonance images (MRI).

METHODS

A comparative and descriptive analysis was conducted using the institutional picture archiving and communication systems. A total of 80 abdominal images (44 CT, 36 MRI) that showed various entities affecting the abdominal structures were included. Microsoft Bing's interpretations were compared with the impressions of radiologists in terms of recognition of the imaging modality, identification of the imaging planes (axial, coronal, and sagittal), sequences (in the case of MRI), contrast media administration, correct identification of the anatomical region depicted in the image, and detection of abnormalities.

RESULTS

Microsoft Bing detected that the images were CT scans with 95.4% accuracy (42/44) and that the images were MRI scans with 86.1% accuracy (31/36). However, it failed to detect one CT image (2.3%) and misidentified another CT image as an MRI (2.3%). On the other hand, it also misidentified four MRI as CT images (11.1%) and one as an X-ray (2.7%). Bing achieved an 83.75% success rate in correctly identifying abdominal regions, with 90% accuracy for CT scans (40/44) and 77.7% for MRI scans (28/36). Concerning the identification of imaging planes, Bing achieved a success rate of 95.4% for CT images and 83.3% for MRI. Regarding the identification of MRI sequences (T1-weighted and T2-weighted), the success rate was 68.75%. In the identification of the use of contrast media for CT scans, the success rate was 64.2%. Bing detected abnormalities in 35% of the images but achieved a correct interpretation rate of 10.7% for the definite diagnosis.

CONCLUSION

While Microsoft Bing, leveraging ChatGPT-4 technology, demonstrates proficiency in basic task identification on abdominal CT and MRI, its inability to reliably interpret abnormalities highlights the need for continued refinement to enhance its clinical applicability.

CLINICAL SIGNIFICANCE

The contribution of large language models (LLMs) to the diagnostic process in radiology is still being explored. However, with a comprehensive understanding of their capabilities and limitations, LLMs can significantly support radiologists during diagnosis and improve the overall efficiency of abdominal radiology practices. Acknowledging the limitations of current studies related to ChatGPT in this field, our work provides a foundation for future clinical research, paving the way for more integrated and effective diagnostic tools.

KEYWORDS

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

Corresponding author: Ezgi Güler

E-mail: gulerezgi@yahoo.com

Received 23 January 2024; revision requested 03 March 2024; last revision received 02 July 2024; accepted 10 July 2024.



Epub: 19.08.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.232680

You may cite this article as: Elek A, Ekizalioğlu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Diagn Interv Radiol.* 2025;31(3):196-205.

Large language models (LLMs), such as ChatGPT-4, are designed for advanced natural language understanding and generation. Due to extensive pre-training on diverse datasets, these models can process and generate human-like text. Recent studies have explored the utility of LLMs in various domains, including academic writing, literature reviews, radiological reporting, and radiological case solving.¹⁻⁵

However, a significant limitation of existing chatbots is their text-based nature. While image generators such as DALL-E have demonstrated impressive results in creating visual content,⁶ integrating such capabilities into text-based chatbots such as ChatGPT remains challenging. Encouragingly, recent updates in Microsoft Bing, which leverages ChatGPT-4 technology, have introduced the functionality of image upload.⁷⁻⁹ Considering the text-based nature of LLMs, this represents a significant advancement, showing promise in analyzing uploaded images.¹⁰

While the exact method by which LLMs interpret images is not fully understood, it likely involves multimodal learning methods and the integration of machine learning algorithms within the chatbot.¹¹⁻¹³ Although LLMs can successfully evaluate everyday non-medical images, interpreting radiological images is a more sensitive issue and requires rigorous testing for potential model development. The potential of LLMs to interpret radiological images from certain perspectives could provide practical benefits. Given the recent addition of image upload functionality to LLMs, the literature lacks comprehensive evaluations of these models' performance in analyzing radiological images.

This study aims to assess the capability of Microsoft Bing, which utilizes ChatGPT-4 technology, to analyze abdominal images from computed tomography (CT) and magnetic resonance imaging (MRI) examinations. The goal is to evaluate the model's interpre-

tive capabilities using consensus evaluations by radiologists as the gold standard.

Methods

Study design and image selection

This study was approved by the Ethics Committee of Ege University Faculty of Medicine (protocol number: 23-8T/9, date: 08.12.2023). Informed written consent was waived. All images used in the study were fully anonymized, ensuring that no identifiable information was present. None of the images have previously been published in any open or subscription-based journals in a different study.

A retrospective search was conducted for abdominal CT and MRI acquired between April 2023 and July 2023, using the institutional picture archiving and communication systems (SECTRA PACS, Sectra AB, Linköping, Sweden).

Abdominal CT scans were conducted using either a single-source 64-slice rapid kV-switching dual-energy CT scanner (Discovery CT750 HD; GE Healthcare, WI, USA) or a 128-slice CT system (Somatom Definition; Siemens, Germany). Abdominal MRI scans were obtained using either a 3T MRI scanner (Magnetom Verio, Siemens, Germany) or a 1.5 T system (Magnetom Amira, Siemens, Germany). The abdominal MRI scans encompassed axial, coronal half-Fourier-acquired single-shot turbo spin-echo, coronal T2-weighted turbo spin-echo sequence with fat suppression, and axial, coronal, sagittal fat-suppressed spoiled gradient-echo with volumetric interpolated breath-hold examination sequences.

The images were selected through the consensus of a senior radiology resident and an abdominal radiologist with 10 years of experience. When selecting both CT and MRI, the imaging plane and sequence where the pathology or mass was most clearly visualized were chosen. Only artifact-free images that delineated the relevant pathology in a single image section were included.

The study investigated a wide range of conditions commonly encountered in routine clinical practice. These entities encompassed hepatomegaly, hepatosteatosis, splenomegaly, chronic parenchymal liver disease, gallstones, acute pancreatitis, benign and malignant neoplasms of the liver, kidney, and ureter stones with associated hydronephrosis, bladder stones, bladder diverticulum, benign and malignant neoplasms

of the urogenital system, benign and malignant gastrointestinal system pathologies, intra-abdominal abscesses, intraperitoneal free fluid, abdominal aortic aneurysm, and retroperitoneal masses.

The specific choice to focus on abdominal imaging in this study was the relatively limited use of artificial intelligence (AI) in this area compared with other parts of the body.¹⁴ Another reason that the abdominal images were selected was that this area includes various organs with a wide spectrum of daily encountered pathologies.

The inclusion criteria were as follows: 1) adult patients (aged >18 years); 2) for the evaluation of masses, only those with diagnoses confirmed by histopathology; and 3) entities that can be unambiguously identified in a single cross-sectional image.

The exclusion criteria were as follows: 1) for the evaluation of masses, any cases without histopathological confirmation; 2) entities that cannot be identified in a single cross-sectional image; and 3) images that are non-diagnostic due to artifacts.

Reviewers' interpretations

The evaluation process involved a collective assessment of the imaging modality, whether the images were contrast-enhanced or unenhanced, and the MRI sequences (T1-weighted or T2-weighted MRI). In addition, any existing pathology or mass within the organ was investigated in terms of its location and nature.

Three months after image selection, these evaluations were provided through the consensus of a senior radiology resident and an abdominal radiologist with 10 years of experience. The reviewers, who had no access to clinical information, provided written reports outlining their findings, impressions, and differential diagnoses.

For standardization purposes, after the image evaluation was completed, electronic medical records were examined to investigate clinical and histopathological diagnoses. The histopathological diagnosis of the masses was confirmed.

Microsoft Bing's interpretation

Microsoft Bing is an LLM that utilizes Generative Pre-trained Transformer 4 (GPT-4) technology created by OpenAI. Additionally, with its text-based nature, the Bing model was the first LLM to introduce an image upload feature.

Main points

- In this study, the performance of large language models in analyzing abdominal images is evaluated.
- The model accurately recognized the imaging modality in 95.4% of computed tomography cases and 86.1% of magnetic resonance imaging cases.
- Microsoft Bing detected abnormalities in 35% of the images but achieved a correct interpretation rate of 10.7% for the definite diagnosis.

Before uploading images for interpretation, we experimented with 20 images that were not used in the study to identify suitable prompts. Although different techniques have been defined for prompt engineering,^{15,16} due to the recent addition of the image upload feature to the chatbot and the lack of prompt engineering work on this topic, prompts were generated based on providing images and questions.

We replaced all radiologic, pathologic, and medical terms in the file names with numbers (1, 2, 3, 4, etc.), meticulously ensuring that the images themselves were devoid of any text. For the interpretation using Microsoft Bing, each radiological image was independently uploaded to the Bing chatbot.

The "More Creative" Bing chatbot model was selected from three available options. This model was chosen because the other two models tended to not answer questions. To mitigate potential bias, the chat interface was cleared after each image upload, and no additional information accompanied the uploaded images. Bing's analysis was driven by customized prompts, progressively tailored to our study requirements. These prompts first inquired about the imaging modality and then for details such as the sequences for MRI and the use of contrast media for CT images. The analysis also examined the imaging planes and the presence of abnormalities in the images (Figure 1).

The initial response generated by Bing was considered, and subsequent repetitions of the same questions were avoided. In in-

stances where the imaging modality was incorrectly predicted, no further inquiries were made regarding the imaging sequence.

Evaluation criteria

The evaluations of Microsoft Bing's interpretations and the assessment of radiologists were based on the accuracy of the imaging modality, sequence (in the case of MRI), imaging plane, correct identification of the anatomical region depicted in the radiological image, identification of contrast media administration, and the detection of any abnormalities.

Statistical analysis

Descriptive statistics were employed to analyze the collected data and evaluate the effectiveness of Microsoft Bing in image interpretation. Categorical variables were compared using the chi-square test. All analyses were conducted using Excel, version 14.7.1 (Microsoft Corp, Redmond, WA), and SPSS version 28 software (IBM SPSS, Inc, Armonk, NY, United States). A *P* value of <0.05 was considered statistically significant.

Results

In this study, we utilized a sample of 80 images (44 CT scans and 36 MRI scans) for our analysis, as detailed in Table 1. Out of the CT scans, six were non-contrast scans covering the entire abdomen, whereas 38 were contrast-enhanced scans. For MRI scans, 26 were non-contrast scans covering the entire abdomen and 10 were contrast-enhanced scans.

Identification of the anatomical region

Microsoft Bing achieved an 83.75% success rate in identifying abdominal regions in the images. It correctly identified 90% of cases (40 out of 44) in CT scans and 77.7% (28 out of 36) in MRI scans. Incorrectly localized abdominal images were misinterpreted as images of the head, neck, brain, knee, and chest. Upon further examination, it was found that 83.3% of the images that were mistaken for the neck were in the sagittal plane (five out of six images).

Recognition of the imaging modality

Out of the 44 CT images, Microsoft Bing accurately identified them as CT scans in 95.4% of cases. In one instance, a CT image could not be detected (2.3%). Moreover, in another case, Bing misidentified a CT image as an MRI (2.3%). Out of the 36 MRI, Microsoft Bing accurately identified them in 86.1% of cases. However, in four cases, Bing mistakenly classified MRI as CT images (11.1%). Additionally, there was one case where an MRI was incorrectly identified as an X-ray (2.7%).

Identification of the imaging plane

In terms of correctly identifying imaging planes (axial, coronal, and sagittal), Microsoft Bing achieved a success rate of 95.4% for CT images and 83.3% for MRI. However, a total of eight images were mislabeled. Out of these mislabeled images, six were incorrectly identified as coronal instead of axial (75% of the mislabeled cases), one was mistakenly labeled as axial instead of sagittal (12.5% of the mislabeled cases), and one that should have been identified as coronal was labeled as axial (12.5% of the mislabeled cases).

Identification of the magnetic resonance imaging sequence

Out of a total of 36 MRI, Microsoft Bing misidentified three as CT images and one as an X-ray. For these four images, the corresponding MRI sequence was not queried at all. Among the remaining 32 MRI, the system correctly identified the sequence in 22 images (68.75%), whereas it could not detect the sequences in two images (6.25%) and made mistakes in eight images (25%). Out of the eight misidentified images, four should have been classified as T2-weighted but were labeled as T1-weighted. Additionally, in three of the misidentified images, Bing mistakenly labeled them as T1-weighted instead of T2-weighted. Furthermore, Bing erroneously identified one image as proton density-weighted instead of T2-weighted (Figure 2).

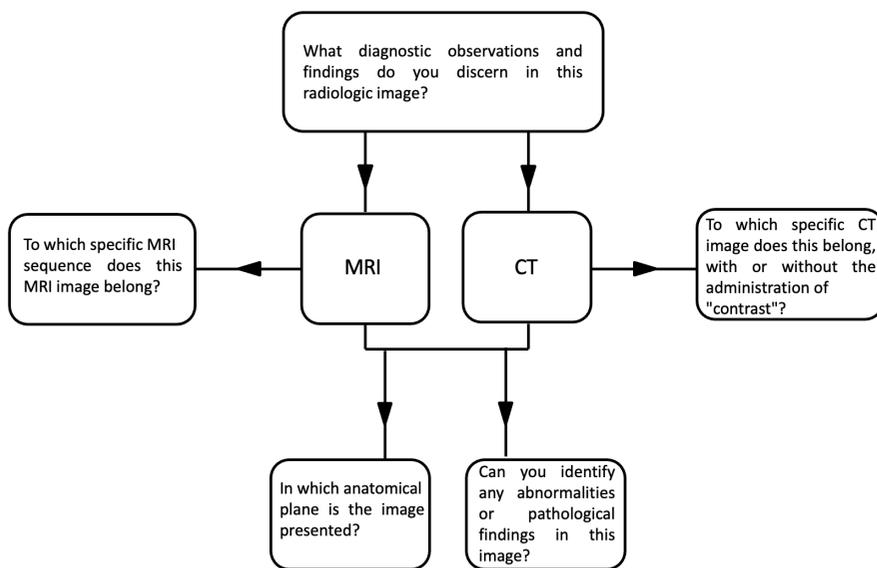


Figure 1. Flowchart diagram illustrating the steps involved in interpreting magnetic resonance imaging (MRI) and computed tomography (CT) images. The diagram shows how Bing can be prompted to analyze MRI and CT images and provide relevant information, such as the anatomical plane, the MRI sequence, the presence of abnormalities, and the use of contrast media.

Table 1. Details of the entities in the CT and MRI images

Entities (CT)	Entities (MRI)	
Abdominal cavity (n = 1)	Abdominal cavity (n = 5)	
	Intraabdominal abscess (n = 1)	Intraabdominal abscess (n = 3)
		Intraperitoneal free fluid (n = 2)
Genitourinary system (n = 18)	Genitourinary system (n = 10)	
	Renal angiomyolipoma (n = 1)	Renal angiomyolipoma (n = 1)
	Renal cell carcinoma (n = 2)	Renal cell carcinoma (n = 1)
	Nephrolithiasis (n = 2)	Renal cyst (n = 1)
	Horseshoe kidney (n = 1)	Bilateral hydronephrosis (n = 1)
	Ureteral stone (n = 1)	Bladder stone (n = 1)
	Bladder stone (n = 1)	Bladder diverticulum (n = 1)
	Bladder diverticulum (n = 1)	Bladder cancer (n = 1)
	Bladder cancer (n = 1)	Uterine myoma (n = 1)
	Adnexal tumor (n = 2)	Benign prostatic hyperplasia (n = 1)
	Autosomal dominant polycystic kidney disease (n = 2)	Autosomal dominant polycystic kidney disease (n = 1)
	Prostatomegaly (n = 2)	
	Prostate cancer (n = 2)	
Liver (n = 7)	Liver (n = 12)	
	Hydatid cyst (n = 1)	Hydatid cyst (n = 1)
	Hemangioma (n = 1)	Hemangioma (n = 2)
	Diffuse hepatic steatosis (n = 1)	Biliary cysts (n = 3)
	Cholangiocellular carcinoma (n = 1)	Liver metastases (n = 2)
	Hepatocellular carcinoma (n = 2)	Hepatocellular carcinoma (n = 1)
	Cirrhosis (n = 1)	Cirrhosis (n = 1)
		Hepatomegaly (n = 2)
Spleen (n = 1)	Spleen (n = 2)	
	Splenomegaly (n = 1)	Splenomegaly (n = 1)
		Splenic infarction (n = 1)
Gastrointestinal system (n = 8)	Gastrointestinal system (n = 3)	
	Colon cancer (n = 3)	Ulcerative colitis (n = 1)
	Gastric cancer (n = 1)	Rectum cancer (n = 2)
	Gastric lymphoma (n = 1)	
	Crohn's disease (n = 1)	
	Mechanical bowel obstruction (n = 1)	
	Perforated duodenal ulcer (n = 1)	
Retroperitoneum (n = 5)	Retroperitoneum (n = 1)	
	Retroperitoneal malignancy (n = 3)	Abdominal aortic aneurysm (n = 1)
	Abdominal aortic aneurysm (n = 2)	
Abdominal wall (n = 1)		
	Incisional hernia (n = 1)	
Bone (n = 1)		
	Metastasis (n = 1)	
Gallbladder (n = 1)	Gallbladder (n = 1)	
	Cholelithiasis (n = 1)	Cholelithiasis (n = 1)
Pancreas (n = 1)	Pancreas (n = 2)	
	Acute pancreatitis (n = 1)	Pseudocyst (n = 1)
		Walled-off necrosis (n = 1)

MRI, magnetic resonance imaging; CT, computed tomography.

Identification of contrast media administration

Out of the total 44 CT images, Bing could not detect the imaging modality for one image, and one was incorrectly recognized as an MRI instead of a CT image. These two images were excluded from the inquiry of contrast media administration. For the remaining 42 CT images, Bing was able to successfully detect the contrast media administration for 27 (64.2%) but could not identify it for three (7.1%). However, there were some inaccuracies in Bing's identification of 12 images (28.5%). Among these 12 misidentified images, Bing mistakenly labeled 10 (83.3%) as "without contrast media administration." Conversely, it incorrectly labeled two images (16.6%) as "with contrast media administration" (Figure 3).

When evaluating Bing's performance, no significant superiority between CT and MRI was observed in any of the different tasks (P

> 0.05). Figure 4 summarizes the accurate responses (%) of Bing across various tasks.

Detection of abnormalities or additional comments

Microsoft Bing detected abnormalities in 35% of the abdominal images. However, its accuracy in correctly interpreting these abnormalities was limited, as it only achieved a correct interpretation rate of 10.7% for the detected abnormalities. In addition to its interpretations, Microsoft Bing provided interesting additional comments on the images (Table 2).

Discussion

This study demonstrates that Microsoft Bing can accurately identify basic tasks in radiological images, such as detecting anatomical regions, imaging modalities, and imaging planes. However, its accuracy decreases when identifying MRI sequences (68.75%) and detecting administration of contrast me-

dia for CT scans (64.2%). From a diagnostic perspective, it demonstrated limited success in determining pathology, with only a 10.7% success rate.

Radiologists, who play a pivotal role in interpreting medical images, are increasingly harnessing the power of AI. Among the various facets of AI, LLMs have emerged as a distinct area of interest.¹⁷ However, the text-based nature of LLMs, as exemplified by chatbots such as ChatGPT, Google Bard, and Microsoft Bing, presents challenges in effectively handling radiological images. Despite this limitation, an innovative approach termed "diagnoses based on imaging patterns" was introduced by Kottlors et al.¹⁸ Although reliant on text, this method has successfully addressed the issue and yielded valuable insights. Remarkably, ChatGPT-4's suggestions demonstrated compatibility at a rate of 68.8%, and a notable 93.8% of these suggestions were considered acceptable alternatives.¹⁸ Similarly, Sarangi et al.¹⁹

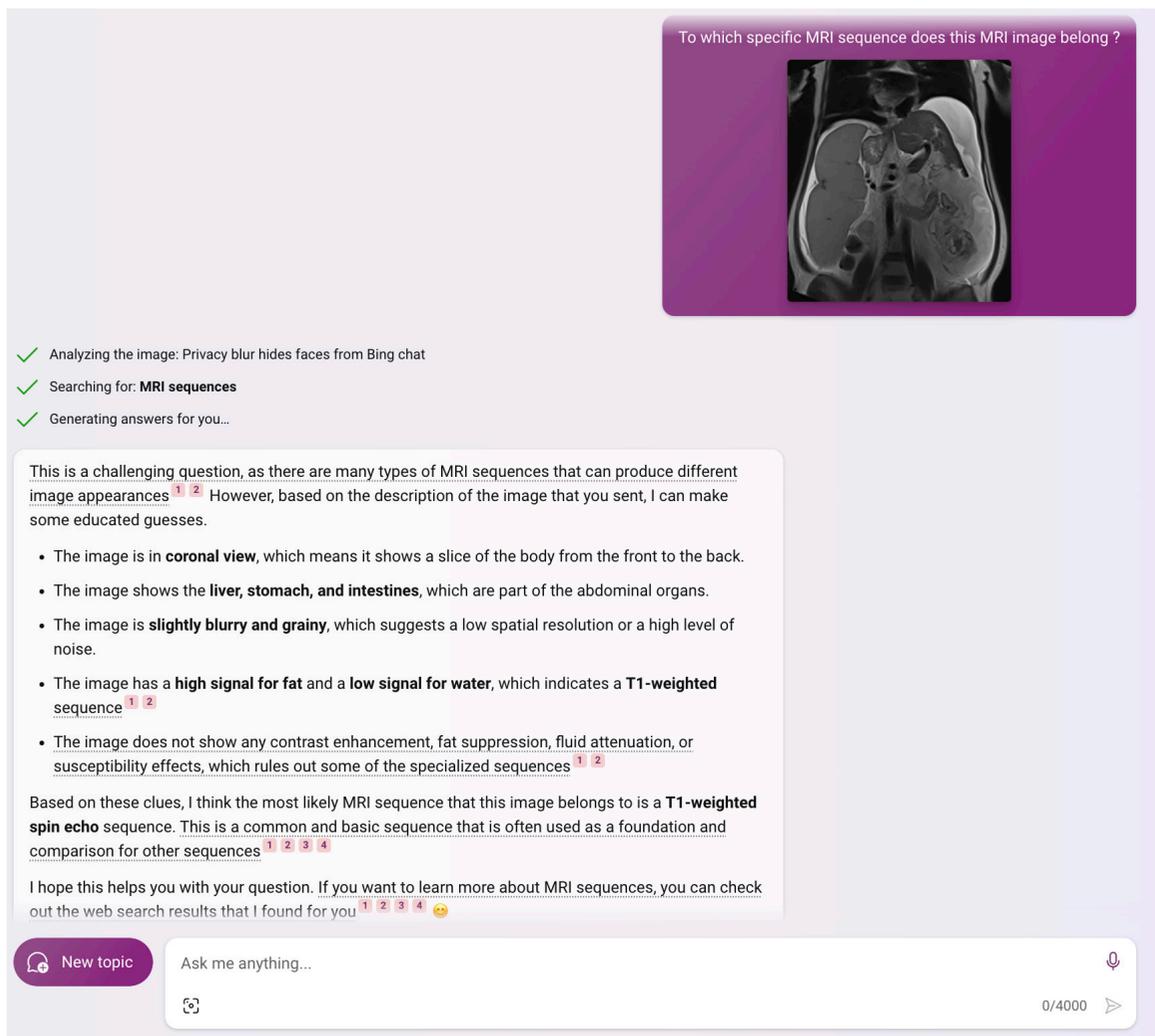


Figure 2. Microsoft Bing's interpretation for the identification of magnetic resonance imaging sequences. Bing misidentified the T2-weighted image as a T1-weighted image.

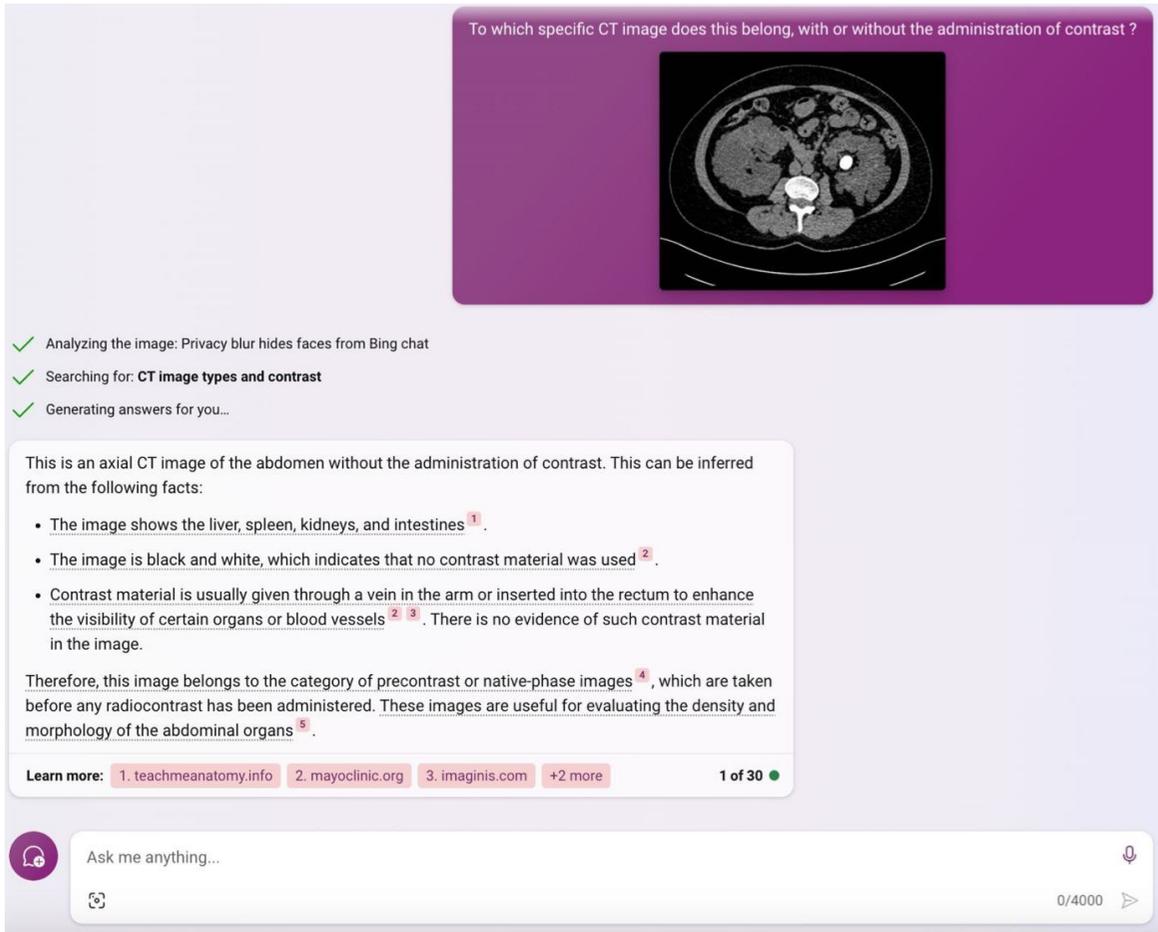


Figure 3. Microsoft Bing's interpretation for the identification of the use of contrast media. Bing accurately interpreted the non-contrast computed tomography image.

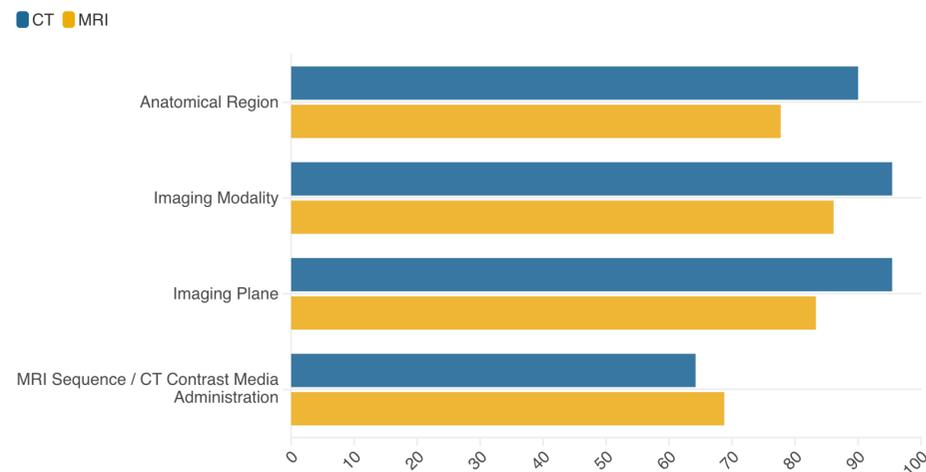


Figure 4. Bar chart illustrating the accuracy (%) displayed by the Bing large language model across different tasks, presented as percentages. Each group represents a distinct task, with bars indicating the corresponding accuracy rates. CT, computed tomography; MRI, magnetic resonance imaging.

examined cardiovascular and thoracic imaging patterns using four different language models and demonstrated that Google Bard exhibited lower performance compared with the other models.

Currently, models trained with medical information, such as Med-PaLM2,²⁰ are being

developed but are not yet available for use. Additionally, the Language and Visual Assistant model developed by Goktas et al.^{21,22} could be used in conjunction with the smart prompt learning method for skin pathologies, and this could also be applied in radiology. Rather than aiming for a 100% diagno-

sis, it is emphasized that results obtained from proportions and options could be more practical and efficient in daily use.^{21,22} However, there is no existing literature that examines the performance of the chatbots for the evaluation of radiologic images.

To fully evaluate the effectiveness of these advancements, especially in the analysis of radiological images, we believe it is necessary to increase research efforts in this area. While we have made initial strides in this direction, our current study mainly focuses on assessing the chatbot's ability to recognize specific anatomical regions in an image and identify basic diagnostic tests.

For this study, we chose to use abdominal images, which often include multiple organs. Recent meta-analyses have indicated that only a small percentage (4%) of commercially available AI applications are dedicated to abdominal imaging, with a mere 3% for liver imaging and 1% for prostate imaging. This is substantially lower than the adoption rates observed in other fields, such as neuroradiology, chest imaging, breast imaging, cardiac imaging, and musculoskeletal imaging.¹⁴

Table 2. Abnormalities and additional comments from Microsoft Bing, which include both accurate and partially accurate comments, along with the interpretation of the radiologist

Finding in the patient's image	Abnormalities/additional comments (Bing) - accurate and partially accurate comments	Reviewer's interpretation
Abdominal aortic aneurysm (CT image)	Spleen, kidney normal, liver enlarged.	The liver size has increased, and there is an aneurysmatic dilation observed at the infrarenal level of abdominal aorta, which has reached a diameter of 6.5 cm and shows partial thrombosis.
Hepatocellular carcinoma (CT image)	Cancer of the liver.	A large mass is observed in segments 5–6 of the right lobe of the liver, extending exophytically toward the abdominal cavity. The lesion exhibits a heterogenous internal structure with hypervascularity in the arterial phase.
Liver hydatid cyst (CT image)	I can notice that your liver appears enlarged and there is a small amount of fluid around it.	A large lobulated, thick-walled cystic lesion is observed in the liver extending from the anterior segments of the right lobe of the liver to the medial segment of the left lobe.
Prostate carcinoma (CT image)	The image appears to be of an adult male.	An increase in the size of the prostate gland is observed. The prostate gland shows an infiltrative extension toward the bladder in the anterosuperior aspect and toward the rectum in the posterior aspect. In the presacral area, there are irregularly bordered nodular soft tissues that merge with each other, and linear densities are observed in the surrounding fat planes. These findings strongly suggest the possibility of prostate malignancy.
Hemangioma (CT image)	Right lobe of the liver, which could indicate a tumor or an abscess.	A large peripheral nodular contrast-enhancing mass, extending from the level of segments 5–6 of the right lobe of the liver toward the abdominal cavity, is primarily suggestive of a hemangioma.
Gastric carcinoma (CT image)	It seems that there is an abnormality in the stomach area. There is a large mass that occupies most of the stomach and extends into the left side of the abdomen. This could be a sign of a tumor, a cyst, or an infection. The liver also appears to be enlarged, which could indicate liver disease or inflammation.	At the level of the stomach cardia and the lesser curvature, there is wall thickening observed, asymmetrically extending into the lumen in a polypoid manner. These findings are primarily suggestive of a malignant nature.
Giant hemangioma (MRI)	The liver appears to be enlarged and has a white mass on it.	A peripheral nodular contrast-enhancing mass filling the left lobe of the liver and extending exophytically toward the abdominal cavity. The findings are primarily suggestive of a hemangioma.
Hepatocellular carcinoma (MRI)	The liver appears to be damaged or diseased, as it has a dark, uneven texture. I cannot diagnose the exact condition of the liver, but it may be affected by cirrhosis, hepatitis, cancer, or other liver diseases.	The liver demonstrates a large-sized mass with mildly increased signal intensity in the peripheral region and high signal intensity in the central section, which may contain cystic-necrotic areas. These findings are primarily suggestive of a malignant nature.
Duodenal ulcer perforation (CT image)	There are several dark spots on the scan, which may indicate damage or abnormalities.	Widespread free-air densities are observed in the abdominal cavity, which are primarily indicative of gastrointestinal system perforation.
Abscess (MRI)	The large white area could be a tumor or an abnormal growth that needs further investigation. The image quality is low and the details are not	A large-sized lesion with peripheral contrast enhancement and a cystic appearance is observed in the left paraaortic area. The findings are primarily suggestive of an abscess.
Biliary cyst (MRI)	Large mass on the left side of the image that could be a tumor or cyst. I also see several smaller masses throughout the image that could be metastases or benign lesions.	A large-sized, lobulated, and thin-walled cystic lesion is observed in the central part of the liver.
Renal cell carcinoma (CT image)	The image shows the liver, spleen, and kidneys. The liver appears enlarged and there is a large mass in the right lobe. This could be a sign of liver cancer or another condition.	A large mass in the middle part of the kidney is present. The lesion extends endophytically toward the renal sinus and exophytically toward the pararenal area. The lesion has irregular borders and contains peripheral hypervascular solid areas, as well as low-density areas in the central region, which could be suggestive of necrosis. Given these features, the primary consideration is renal cell carcinoma.
Hepatosteatorosis (CT image)	I see an abnormality in the image, which is a dark spot on the left side of the liver. This could indicate a tumor, a cyst, an abscess, or another condition that affects the liver tissue.	The liver parenchyma shows a diffuse decrease in density compared with the spleen, which is suggestive of hepatosteatorosis.

Table 2. Continued

Finding in the patient's image	Abnormalities/additional comments (Bing) - incorrect comments	Reviewer's interpretation
Cholelithiasis (CT image)	The liver appears to be enlarged and has a heterogeneous texture, which could indicate some inflammation or fatty infiltration. The spleen also appears to be enlarged and has a smooth contour, which could suggest congestion or infection. The stomach is distended and filled with gas, which could be due to gastric dilatation or obstruction. The intestines are also distended and show multiple air-fluid levels, which could indicate ileus or bowel obstruction. There is some fluid accumulation in the peritoneal cavity, which could be due to ascites or peritonitis.	A high-density structure is observed in the lumen of the gallbladder, which is primarily indicative of a gallbladder stone.
Cholangiocellular carcinoma (CT image)	I can see in the image that the liver looks enlarged and has a patchy appearance. This could indicate some liver disease or injury.	At the level of posterior segments of right lobe of the liver, a mass is observed with a peripheral location, causing retraction of the liver capsule. The mass exhibits peripheral contrast-enhancing solid areas and central cystic-necrotic hypodense areas, which are suggestive of a malignant nature. However, due to the absence of chronic parenchymal liver disease findings and the presence of retractions caused by the mass on the liver capsule, the primary consideration is cholangiocellular carcinoma.
Lymphoma gastric involvement (CT image)	The liver appears to be enlarged and there is a dark area in the upper right portion of the liver. This could indicate a liver abscess.	The presence of increased size in both the liver and spleen is observed. Additionally, there are signs of periportal edema in the liver. Furthermore, there is diffuse and severe thickening of the gastric wall. Considering the enlargement of the liver and spleen along with these findings, lymphoma is considered as a possible explanation.
Autosomal dominant polycystic kidney disease (CT image)	The bright white spots on either side of the pelvis may indicate metal implants or screws.	The description indicates an increase in the size of both kidneys, with multiple cysts located at the cortical and peripelvic regions, making the kidney parenchyma indistinct. Cysts are also observed in the liver parenchyma. These findings are suggestive of autosomal dominant polycystic kidney disease.
Sigmoid colon carcinoma (CT image)	The liver seems to be enlarged, which could indicate a possible abnormality.	In the sigmoid colon, there is a 5-cm mass with findings indicating malignancy, including concentric luminal narrowing, thickening of the colonic wall, and linear densities in the surrounding fatty tissue.
Giant renal angiomyolipoma (MRI)	The liver has high signal intensity (white), indicating high fat content.	A large-sized mass is observed, originating from the lower pole of the kidney and extending anteromedially toward the pararenal area. The lesion contains macroscopic fat and contrast-enhancing regions. These findings are primarily suggestive of an angiomyolipoma.
Myoma uteri (MRI)	Some degree of osteoarthritis or postoperative changes in the knee joint.	A large-sized solid mass is observed, filling the uterine fundus and corpus and extending toward the cervix. The lesion appears predominantly low signal intensity on T2A images but contains heterogeneously high signal intensity areas. These findings are primarily suggestive of a myoma uteri (uterine fibroid).
Walled-off necrosis (MRI)	The liver appears to be enlarged and there are multiple dark spots scattered throughout the liver. These could be signs of liver disease or cancer.	At the level of the pancreatic head-neck junction, the normal morphology of the pancreas is not visualized. Instead, there is a large lobulated thick-walled cystic lesion extending from the head and neck region to the body of the pancreas. The internal structure of the cystic lesion shows areas that could be indicative of debris. These findings are primarily suggestive of a walled-off necrosis. Minimal dilatation is observed in the pancreatic duct. Free fluid is observed in the perihepatic, right paracolic, and perisplenic spaces.

MRI, magnetic resonance imaging; CT, computed tomography.

Bing's robust performance-demonstrated by its high accuracy rates (83.75%) in understanding abdominal images and its notable success in imaging modalities such as CT (95.4%) and MRI (83.3%)-inspires us to further explore detailed inquiries related to medical images. Despite its proficiency in identifying planes in images, there have been instances where it misclassified coronal images as axial

(12.5%). Interestingly, while the model may correct its mistake upon subsequent questioning, our research was focused on the initial responses. This distinction is important because when asked the same question again, the model might recognize the error or interpret the user's dissatisfaction with the previous answer, potentially providing a different response. This situation underscores

the need for caution regarding LLMs' potential inconsistency.

While Bing's responses sometimes accurately predict the application of MRI sequences and administration of contrast agents for CT scans, there have been instances where it misinterprets the information. An important point to highlight is the rationale

provided by Bing when giving its responses. Even when the responses are correct, the underlying explanations have sometimes contained incorrect information. Bing's lack of proficiency in fundamental aspects of radiological image interpretation, its use of incorrect contexts in both successful and unsuccessful cases, and its failures in interpreting pathological conditions all suggest a need for caution. This caution is particularly important when dealing with models such as Bing that have not been specifically trained for medical image interpretation. This specific training deficiency can be attributed to the model's errors in tasks that go beyond basic ones. However, this does not necessarily indicate a bleak future for LLMs in image interpretation. Although not yet published, it is expected that the performance of models specifically trained for medical purposes, such as BioBERT and Med-PaLM, will be higher.^{17,23}

Additionally, unlike the approach taken by Ueda et al.²⁴, where the analysis was based on both patient history and imaging findings, we chose not to provide any patient history information to Bing during the analysis of radiological images. Expecting accurate diagnoses without this contextual information would be unjustifiable. However, the decision to exclude patient history was intentional, as providing such information might have led Bing to rely more on theoretical knowledge than image analysis. Therefore, we deliberately limited our study to the use of radiological images alone.

The significant success in detecting abnormalities involving the liver, including the identification of liver masses, is noteworthy. Pinpointing the exact reasons for this success may be challenging, but one possible factor could be the liver's larger size compared with other organs. Another intriguing observation is the potential misinterpretation of sagittal images (where the spinal cord is visible) as head and neck images, presenting a unique finding. It is plausible that focusing on larger structures might lead to underestimating other images. On the other hand, a study conducted by Cao et al.²⁵ found that ChatGPT's success rate in providing theoretical radiological information related to liver cancer was relatively low.

This study has certain limitations due to its nature and the specific focus of our research. A significant limitation is that Bing currently allows only one image to be uploaded at a time, which presents a challenge. Radiologists often need to examine consecutive im-

ages from different planes to make accurate assessments. To address this limitation, we selected demonstrative images that effectively highlight the imaging findings with the utmost clarity. Another significant limitation that needs to be mentioned is prompt engineering, which is crucial for LLMs and can directly affect the output. Over time, various prompt techniques such as zero-shot prompting, few-shot prompting, instruction following, and chain-of-thought prompting have been developed. However, these prompts have been developed considering the text-based nature of the models.^{15-17,26,27} For this study, the image upload feature was newly introduced at the time of the experiment, and the lack of prompt engineering studies that could improve the quality of the output in terms of image analysis is also a limitation. Lastly, the limited sample size is another constraint of our study.

In conclusion, this study reveals that Microsoft Bing, utilizing ChatGPT-4 technology, can achieve success in basic radiological tasks. However, further refinement and enhancement are essential to improve accuracy in recognizing imaging modalities, identifying specific imaging planes, interpreting imaging findings, and detecting abnormalities. In the future, LLMs trained with medical data may demonstrate higher success rates compared with this study. This suggests a promising avenue for future research and development in this field.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Elek A. Improving accuracy in ChatGPT. *AJR Am J Roentgenol.* 2023;221(5):705. [\[CrossRef\]](#)
2. Elek A. The role of large language models in radiology reporting. *AJR Am J Roentgenol.* 2023;221(5):707. [\[CrossRef\]](#)
3. Ray PP. The need to re-evaluate the role of GPT-4 in generating radiology reports. *Radiology.* 2023;308:e231696. [\[CrossRef\]](#)
4. Sun Z, Ong H, Kennedy P, et al. Evaluating GPT4 on impressions generation in radiology reports. *Radiology.* 2023;307(5):e231259. [\[CrossRef\]](#)
5. Sarangi PK, Narayan RK, Mohakud S, Vats A, Sahani D, Mondal H. Assessing the capability of ChatGPT, Google Bard, and Microsoft Bing in solving radiology case vignettes. *Indian J Radiol Imaging.* 2024;34(2):276-282. [\[CrossRef\]](#)

6. Temsah MH, Alhuzaimi AN, Almansour M, et al. Art or artifact: evaluating the accuracy, appeal, and educational value of AI-generated imagery in DALL.E 3 for illustrating congenital heart diseases. *J Med Syst.* 2024;48(1):54. [\[CrossRef\]](#)
7. Adams LC, Busch F, Truhn D, Makowski MR, Aerts HJWL, Bressen KK. What Does DALL-E 2 know about radiology? *J Med Internet Res.* 2023;25:e43110. [\[CrossRef\]](#)
8. Bing M. Microsoft Bing Chatbot 2023 [cited 2023 18.08.2023]. [\[CrossRef\]](#)
9. OpenAI. GPT-4 OpenAI; 2023 [cited 2023 18.08.2023]. [\[CrossRef\]](#)
10. Hu M, Pan S, Li Y, Yang X. Advancing medical imaging with language models: a journey from n-grams to chatgpt. arXiv preprint arXiv:230404920. 2023. [\[CrossRef\]](#)
11. Huang S, Zhang H, Gao Y, Hu Y, Qin Z. From image to video, what do we need in multimodal LLMs? arXiv preprint arXiv:240411865. 2024. [\[CrossRef\]](#)
12. Alshehri AS, Lee FL, Wang S. Multimodal deep learning for scientific imaging interpretation. arXiv preprint arXiv:230912460. 2023. [\[CrossRef\]](#)
13. Reizinger P, Ujváry S, Mészáros A, Kerekes A, Brendel W, Huszár F. Understanding LLMs requires more than statistical generalization. arXiv preprint arXiv:240501964. 2024. [\[CrossRef\]](#)
14. Schmidt S. AI-based approaches in the daily practice of abdominal imaging. *Eur Radiol.* 2024;34(1):495-497. [\[CrossRef\]](#)
15. Kaddour J, Harris J, Mozes M, Bradley H, Raileanu R, McHardy R. Challenges and applications of large language models. arXiv preprint arXiv:230710169. 2023. [\[CrossRef\]](#)
16. Yao S, Yu D, Zhao J, Shafran I, et al. Tree of thoughts: deliberate problem solving with large language models. *Adv Neural Inf Process Syst.* 2024;36. [\[CrossRef\]](#)
17. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol.* 2024;30(2):80-90. [\[CrossRef\]](#)
18. Kottlors J, Bratke G, Rauen P, et al. Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology.* 2023;308(1):e231167. [\[CrossRef\]](#)
19. Sarangi PK, Irodi A, Panda S, Nayak DSK, Mondal H. Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: perspectives of four large language models. *Indian J Radiol Imaging.* 2024;34(2):269-275. [\[CrossRef\]](#)
20. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. arXiv preprint arXiv:230509617. 2023. [\[CrossRef\]](#)
21. Goktas P, Agildere AM. Transforming radiology with artificial intelligence visual chatbot:

- a balanced perspective. *J Am Coll Radiol*. 2024;21(2):224-225. [\[CrossRef\]](#)
22. Goktas P, Kucukkaya A, Karacay P. Leveraging the efficiency and transparency of artificial intelligence-driven visual Chatbot through smart prompt learning concept. *Skin Res Technol*. 2023;29(11):e13417. [\[CrossRef\]](#)
 23. Shawn Xu, Lin Yang, Christopher Kelly, et al. ELIXR: towards a general purpose X-ray artificial intelligence system through alignment of large language models and radiology vision encoders. *arXiv*. 2023. [\[CrossRef\]](#)
 24. Ueda D, Mitsuyama Y, Takita H, et al. ChatGPT's diagnostic performance from patient history and imaging findings on the diagnosis please quizzes. *Radiology*. 2023;308(1):e231040. [\[CrossRef\]](#)
 25. Cao JJ, Kwon DH, Ghaziani TT, et al. Accuracy of information provided by ChatGPT regarding liver cancer surveillance and diagnosis. *AJR Am J Roentgenol*. 2023;221:556-559. [\[CrossRef\]](#)
 26. Lu Y, Bartolo M, Moore A, Riedel S, Stenetorp P. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *arXiv preprint arXiv:210408786*. 2021. [\[CrossRef\]](#)
 27. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst*. 2022;35:24824-24837. [\[CrossRef\]](#)



Breast cancer detection and classification with digital breast tomosynthesis: a two-stage deep learning approach

Yazeed Alashban

King Saud University, College of Applied Medical Sciences, Department of Radiological Sciences, Riyadh, Saudi Arabia

PURPOSE

The purpose of this study was to propose a new computer-assisted two-staged diagnosis system that combines a modified deep learning (DL) architecture (VGG19) for the classification of digital breast tomosynthesis (DBT) images with the detection of tumors as benign or cancerous using the You Only Look Once version 5 (YOLOv5) model combined with the convolutional block attention module (CBAM) (known as YOLOv5-CBAM).

METHODS

In the modified version of VGG19, eight additional layers were integrated, comprising four batch normalization layers and four additional pooling layers (two max pooling and two average pooling). The CBAM was incorporated into the YOLOv5 model structure after each feature fusion. The experiment was carried out using a sizable benchmark dataset of breast tomography images. A total of 22,032 DBT examinations from 5,060 patients were included in the data.

RESULTS

Test accuracy, training loss, and training accuracy showed better performance with our proposed architecture than with previous models. Hence, the modified VGG19 classified DBT images more accurately than previously possible using pre-trained model-based architectures. Furthermore, a YOLOv5-based CBAM precisely discriminated between benign lesions and those that were malignant.

CONCLUSION

DBT images can be classified using modified VGG19 with accuracy greater than the previously available pre-trained models-based architectures. Furthermore, a YOLOv5-based CBAM can precisely distinguish between benign and cancerous lesions.

CLINICAL SIGNIFICANCE

The proposed two-tier DL algorithm, combining a modified VGG19 model for image classification and YOLOv5-CBAM for lesion detection, can improve the accuracy, efficiency, and reliability of breast cancer screening and diagnosis through innovative artificial intelligence-driven methodologies.

KEYWORDS

Breast imaging, artificial intelligence, mammography, breast cancer

Corresponding author: Yazeed Alashban

E-mail: Yalashban@ksu.edu.sa

Received 02 July 2024; revision requested 05 September 2024; accepted 12 October 2024.



Epub: 09.12.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242923

Breast cancer (BC) is one of the main causes of mortality in women and a major global health concern.¹ According to data from the World Health Organization, in 2022, 665,684 women worldwide lost their lives due to malignancy in the breast, accounting for 2.3 million new cases of the disease. BC is the most common cancer globally among women; between 2015 and 2021, 7.8 million women were diagnosed with the disease.² In the US, BC ranks as the second most common malignancy after lung cancer, according to the Surveillance, Epidemiology, and End Results Program.³ Globally, according to available data,^{3,4} one in eight women will contract BC. As a result, BC screening is one of the most significant and common

You may cite this article as: Alashban Y. Breast cancer detection and classification with digital breast tomosynthesis: a two-stage deep learning approach. *Diagn Interv Radiol.* 2025;31(3):206-214.

medical imaging prerequisites, with over 39 million examinations carried out annually.⁵ Early identification and discovery are essential for therapy, rehabilitation, and a decrease in death rates.⁶ The prognosis and survival rate of cancer vary greatly depending on its stage. Cancer treatment is more successful the earlier the disease is discovered.⁷

Radiologists examine and annotate images generated by screening techniques to identify tumors.⁸ The gold standard for this cancer screening has been supplanted by the relatively new imaging technique known as digital breast tomosynthesis (DBT), which has taken the place of mammography.⁸ This is a type of three-dimensional (3D) mammography that aims to increase abnormality detection.⁹ DBT, which recreates multiple low-dose picture projections from a moving digital X-ray source over a restricted arc angle, is used to build the 3D model.¹⁰ Since a two-dimensional (2D) mammography examines every tissue in the breast at once, there is a risk that certain tissue features will overlap and produce inaccurate results. By allowing radiologists to view multiple layered images prior to classifying tumors, DBT helps address some of the problems associated with 2D mammograms.¹¹ Compared with traditional mammography, DBT often requires longer image acquisition and processing times, as well as increased radiation exposure (though it is still within safe limits).^{10,11}

Radiologists are already using computer-aided diagnosis tools to help them make

decisions.¹² These technologies have the potential to reduce significantly the time and energy required to evaluate a lesion in clinical practice.¹³ They may also reduce the occurrence of false positives, which lead to unnecessary and uncomfortable biopsies.¹⁴ Recent technological advancements in deep learning (DL), such as artificial neural networks and transfer learning, have outperformed several machine learning algorithms in tasks such as classifying and identifying lesions.¹⁵ Unlike traditional machine learning methods, which require a manual feature extraction and selection step, DL algorithms adaptively learn the optimal feature extraction process from the input data.^{13,14}

However, although DL techniques for lesion detection and classification have been used extensively using mammography, there have been few studies using DBT. This could be attributed to the computer memory constraints associated with DL methods, which are linked to the higher dimensionality of the data. In previous studies, breast tumors from DBT data have been segmented, classified, and detected using DL. Li et al.¹⁶ carried out deep convolutional neural network (DCNN)-based mass classification of BC using DBT and assessed different transfer learning strategies. They collected data on 441 patients who had undergone DBT and conducted three different experiments to compare 2D and 3D DCNNs trained on volumetric DBT. The 2D convolutional neural network (CNN) that was trained on both DBT and full-field digital mammography achieved better results, with a change in area under the curve of 0.009.¹⁶

Ricciardi et al.¹⁷ developed a DCNN-based detection system for the automatic classification of the presence or absence of mass lesions in DBT-annotated images. Background correction, data augmentation, and normalization were basic pre-processing steps. Three DCNN architectures trained on two distinct datasets were compared: 1) built from scratch (DBT-DCNN); 2) pre-trained (AlexNet and VGG19); 3) optimized using a transfer learning approach. Additionally, a Grad-CAM technique was used to provide a position indication for the lesion in the DBT. The accuracy of the DBT-DCNN network was $90\% \pm 4\%$, and the sensitivity was $96\% \pm 3\%$.¹⁷

Lotter et al.⁸ presented a DL method that was annotation-efficient and accurate; the method achieved maximum performance in classification, detected cancers in clinically negative mammograms, was effectively applicable to a population with low screening

participation, and outperformed five full-time breast-imaging radiologists, with an average 14% increase in sensitivity. The model used a multiple-instance learning approach in which it was progressively and effectively trained on DBT using only breast-level labels. The authors were successful in maintaining localization-based interpretability by generating new “maximum suspicion projection” images from DBT data.⁸

For the prediction of Ki-67 expression in DBT images, Oba et al.¹⁸ developed a model based on DL. The Ki-67 expression of 126 patients with pathologically proven BC was chosen and assessed. The DL model employed the Xception architecture to forecast the levels of Ki-67 expression. The accuracy, on average, was 0.912. The findings point to the possible use of their model to predict Ki-67 expression from DBT, which is useful in deciding on a BC treatment plan prior to surgery.¹⁸ Buda et al.¹⁹ shared a large-scale publicly available DBT examination dataset, which included information for 5,060 patients, and used it to train a detection model. One hundred twenty-four images having bounding boxes for malignant and 175 images having bounding boxes for benign lesions were used to develop a detection algorithm based on a 2D DenseNet. There was no pretraining on alternative datasets or comparable modalities, such as mammography. The free-response receiver operating curve, displaying the sensitivity of the model in relation to false-positive predictions, was utilized for the ultimate assessment of the baseline detection algorithm.¹⁹

Earlier studies have proposed classification or detection using DBT with notable contributions. However, they are limited by fewer images in the datasets,¹⁶ lack of external validation and clinical assessment,¹⁷ limited comparison of advanced architectures, and lack of diversity in training data.^{8,18} The DBT data used in this study has also been utilized in several studies;¹⁹⁻²⁵ however, these studies have either focused on the classification or detection of the lesions as benign or malignant. Table 1 provides a summary of the studies conducted on the Duke Dataset from the Cancer Imaging Archive (TCIA).²⁶ The model in the current study is the first state-of-the-art model that classifies a DBT scan into one of three classes: normal, actionable, and tumor. Moreover, it detects the lesion as benign or cancerous. The model incorporates a modified VGG19 DL architecture. The batch normalization layers are placed in every fourth convolutional layer to enhance the model's training efficiency

Main points

- The modified VGG19 architecture classified digital breast tomosynthesis (DBT) images more accurately than previously available pre-trained models.
- The You Only Look Once version 5 (YOLOv5)-based convolutional block attention module (CBAM) precisely discriminated between benign and malignant lesions, showing better performance metrics such as test accuracy and training loss.
- The research fills a significant gap in breast cancer diagnosis by utilizing advanced deep-learning strategies for DBT images. The two-tier deep learning algorithm, combining a modified VGG19 model for image classification and YOLOv5-CBAM for lesion detection, demonstrated good outcomes in terms of accuracy and time.
- The study highlights the advantages of DBT over traditional two-dimensional mammography, emphasizing its improved accuracy in screening due to the three-dimensional view it provides.

Table 1. Summary of the previous studies utilizing data from the Cancer Imaging Archive

Citation	Architecture	Pre-processing	Training/testing dataset	Outcome	Results
21	ResNet-18, AlexNet, MobileNetV2, GoogleNet, DenseNet-201, VGG-16,	DBT augmentation; image enhancement techniques; color feature mapping	Patients – 5,060 Slices – 22,032	Classification into normal, benign, and malignant	Acc.: 56.52
22	R-CNN	Conversion of volume intensities to 8-bits depth; extraction of breast mask area; flipping to convert all the volumes into same orientation	Patients – 5,060 Slices – 22,032	Lesion detection	IOU: 0.85
23	ResNet	Cropping; reduction of pixels; transformation	Cancer + actionable - 100 Normal + benign - 100	Classification	Acc.: 86
24	Inception v3	Cropping; reduction of pixels; augmentation	Normal – 1,680 Tumor – 223	Lesion detection	Acc.: 91.4
25	Faster R-CNN	Data augmentation; image flipping; image translation; channel reception augmentation	Patients – 985 Scans – 1,000	Detection	Acc.: 83.08
26	2 Layer DenseNet	Cropping; downscaling	Patients – 5,060 Scans – 22,032	Lesion detection	Sensitivity: 78
27	Faster R-CNN	Cropping; normalization; masking and background suppression	VICTRE + Patients – 5,060 Scans – 22,032	Lesion detection	Sensitivity: 60

DBT, digital breast tomosynthesis.

by reducing internal covariant shifts. The tumor is detected using a You Only Look Once version 5 (YOLOv5)-based convolutional block attention module (CBAM) architecture, utilizing the two submodules of CBAM: channel attention and spatial attention. This study explores the integration of YOLOv5 (a state-of-the-art object detection model) with CBAM (a mechanism that enhances feature representation) to improve detection accuracy and efficiency. Thus, this model has applications in both the screening and diagnosis of BC.

Methods

Dataset

The dataset available on TCIA website was used in this investigation; it was acquired from the Duke Health System using the Duke Enterprise Data Unified Content Explorer tool between January 1, 2014, and January 30, 2018.²⁶ The data included a total of 22,032 DBT examinations from 5,060 patients. The dataset included DBT images from four different views along with four categories of cases: normal (no sign of cancer and a biopsy was never performed), actionable (cancer may be present, but no biopsy was performed), biopsy-proven benign (a biopsy was performed, and the tumor was determined to be benign), and biopsy-proven cancer (a biopsy was performed, and the tumor was classified as malignant).²⁴ The Digital Imaging and Communications in Medicine (DICOM) images consisted of a collection of 2D slices taken from four different views:

left-mediolateral oblique, right-mediolateral oblique, left-cranio-caudal, and right-cranio-caudal.

Ethics

This investigation utilized data from the TCIA website, which was obtained from the Duke Health System. Since the data is publicly available and patient consent is not required, ethical approval was not necessary for this study.

Methodology

The overall methodology consisted of two stages: classification and detection. First, the images were classified as normal, actionable, benign, or cancer. In the second stage, the lesion was detected as benign or cancerous using the annotated images containing bounding boxes on the tumor area. The step-by-step methodology for each stage is shown in Figure 1, which summarizes the entire architecture utilized in this study.

Data pre-processing

Certain pre-processing steps were applied at both stages. The following sections describe all the steps that were applied to prepare the dataset for modeling.

Classification

The following steps were carried out to prepare the DBT images for classification into normal, actionable, or tumor. The images were changed from DICOM to JPEG format, a transformation that not only simplified the

data format but also allowed compatibility with the next stages of processing. The intensity rescaling was done to standardize the pixel intensity values in all the images so the uniformity of the image could be ensured and the effect of the different illuminations or contrasts could be eliminated. Color space conversion was also carried out to improve the understandability and the discriminative capacities of the images, which, in turn, facilitates the extraction of more significant features for the classification of the images. Resizing was done to adjust the spatial sizes of the images. Hence, the spatial dimensions of the images were harmonized, which improved consistency and removed possible distortions that could affect the analysis. Normalization-the scaling of pixel values to a standard range-was also performed.

Detection

To prepare the data for the detection stage of determining the tumor as benign or cancerous, the images were first augmented, and then the pre-processing techniques mentioned in the classification were applied. The process of purposefully increasing the volume and complexity of already-existing data is known as data augmentation. Data augmentation has become a necessary pre-processing step in DL.

Because a significant number of training samples are needed for neural networks and medical datasets are sometimes scarce, the first step in increasing the diversity of the dataset is data augmentation; in this study,

the Roboflow tool was used for this activity. The following steps were performed: text files were generated to contain essential annotations, and all the generated text files were imported to Roboflow. Five types of augmentation techniques were applied: horizontal and vertical flips, 90-degree rotations (clockwise, anticlockwise, upside down), cropping (ranging from 0% to 25% maximum zoom), rotations (-15 to +15 degrees), and shears (10-degree vertical and horizontal). The total number of images before and after augmentation is given in Table 2.

Data splitting

The dataset was split into three subsets for the classification and detection stages: training (number of images for classification: 19,148, number of images for detection: 2,116), validation (number of images for classification: 1,163, number of images for detection: 604), and testing (number of images for classification: 1,721, number of images for detection: 303) in the 70, 20, and 10 ratios. The number of instances in each split for each category for the classification framework is given in Table 3.

Experimental setup

The chosen equipment, including an NVIDIA RTX 4090 GPU and AMD EPYC 7R12 48-Core Processor, provided high computational power (1.8 TFLOPS and 24.0/192 CPU cores, respectively), which is essential for intensive model calculations. The motherboard ROME2D32GM supports PCIe 4.0, enhancing data transfer speeds (22.8 GB/s), which is crucial for handling large datasets. With 516 GB of memory and a 4TB Predator SSD, the system ensured ample storage and quick data access (3,830 MB/s), supporting efficient model training and analysis. The equipment's high-performance specifications were used to optimize model development and execution. The pre-processed DBT images were classified using VGG19, and detection was based on YOLOv5-CBAM.

Modelling

The modified VGG19 model was used to classify the DBT slice images into normal, actionable, and tumor. A YOLOv5-CBAM model was used for the detection of lesions as benign or cancerous.

Modified VGG19

Transfer learning involves transferring the learned parameters of the pre-trained CNN model. It involves shifting the weights (as

given in Table 4) of a CNN model that was trained on additional sizable datasets.²⁷ Scientists are creating deeper learning models to increase performance as DL models have become more and more popular in image classification and recognition applications. VGG19 is a neural network comprising 43 layers, namely the input, 16 convolutional layers, 16 ReLU layers, 5 max pool layers, 3 full-connected layers, 1 softmax layer, and the output. In this way, the modified version of VGG19 consisted of 8 complementary layers, which were 4 batch normalization layers and 4 extra pooling layers. The batch normalization layers consisted of 2 max pooling and 2 average pooling layers. The layering of batch normalization layers between every 4th convolution layer was interpreted to

improve training efficiency by reducing internal instability. This modification produces not only a smaller scale or initial values of the gradient that parameters rely on for modifying but also a better and more natural flow of data between the intermediate layers of the neural network, which greatly reduces the number of iterations required for training. As to extra pooling layers to further the 5th and the 10th convolutional layers of the DL model, crucial low-level details are passed through the learning model, and this helps capture sharp features integrally. The size of the input image was 512 × 512.

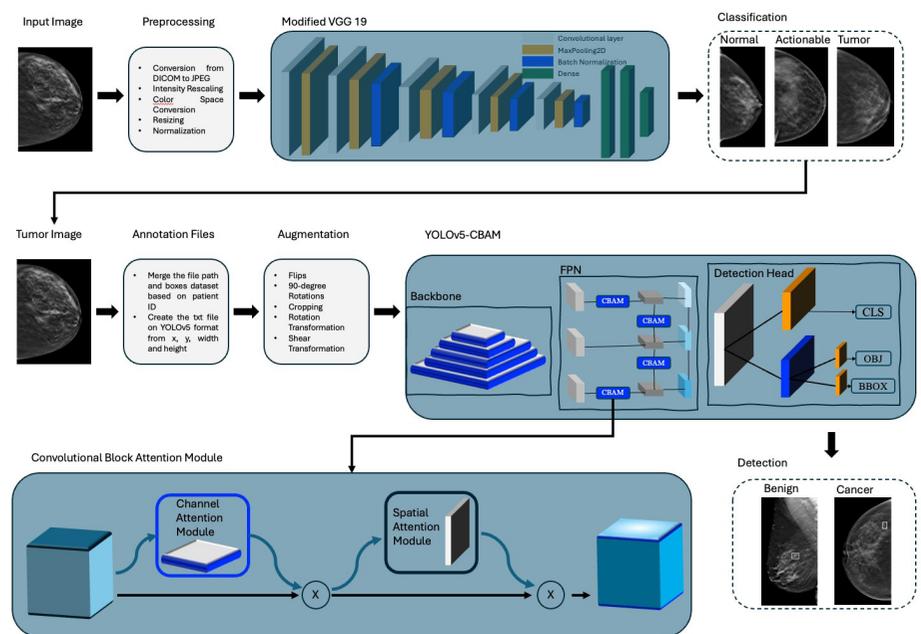


Figure 1. The schematic depicts the organization of the suggested framework. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Table 2. Total number of images before and after augmentation for the application of the YOLOv5-CBAM architecture

Dataset	Before augmentation	After augmentation
Training	293	2,293
Validation	58	456
Testing	35	274
Total	386	3,023

YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Table 3. Summary of the dataset splitting for each class

Category	Training	Testing	Validation
Normal	18,232	1,356	928
Actionable	716	244	160
Tumor	200	121	75

YOLOv5-CBAM

The attention mechanism makes it possible for models to prioritize and process information selectively, focusing only on the most crucial details and ignoring the rest. Convolutional block attention modules are one type of attention mechanism meant to enhance CNN performance. As can be seen in Figure 2, the CBAM is incorporated into the proposed model structure after each feature fusion, or “concat.” In an image or feature map, it mainly enhances feature extraction and records meaningful spatial and channel-wise dependencies. The efficacy of this module is demonstrated in the experiments reported in the study,²⁸ where the performance of the module is significantly improved by integrating the CBAM into various models across a variety of classification and detection datasets.

Convolutional block attention modules are made up of two sub-modules: the channel attention module and the spatial attention module. The primary focus of channel attention is on locating the essential traits or features needed to identify a lesion in an image. However, it is crucial to remember that the lesion is a relatively small and sparse component within the entire image when it comes to particular tasks, such as lesion detection. In these situations, the value of the individual pixels in the entire image is not equal. At this point, spatial attention is applied to solve the “where” issue, which involves locating the lesion in the image. Functioning alongside channel attention, spatial attention gathers data from various spatial regions of the image. By giving these spatial features weights, it essentially highlights the areas of the picture where lesions are present. Applying channel and spatial attention in that order achieves this. Figure 3 illustrates how channel attention can compute channel weights represented as $WCA \in RC \times 1 \times 1$, and spatial attention can compute spatial weights (WS) denoted as $WS \in RH \times W \times 1$, given the input feature map $F \in RW \times H \times C$.

Channel attention refers to a multi-step process that is applied to an input feature map (F). Global max pooling (GMP) and global average pooling (GAP) are carried out to record the highest and lowest spatial responses. These responses are then processed by a multi-layer perceptron. Then, element-wise addition is used to integrate the

results of GMP and GAP. After that, a sigmoid activation function is applied to the combined data, resulting in a channel weight feature map that assigns a weight to each channel based on its significance. Finally, an element-wise multiplication is performed

between the channel weight matrix and the original feature map (F) as:

$$(1) \quad F' = F \times W_{CA}$$

Table 4. Parameter values at each layer of the modified VGG19 model

Layer name	Activation maps	Learnable parameters	Total learnable parameters
Input	512 × 512 × 3	-	0
block1_conv1	512 × 512 × 64	Weights: 3 × 3 × 3 × 64, bias: 64	1,792
block1_conv2	512 × 512 × 64	Weights: 3 × 3 × 64 × 64, bias: 64	36,928
block1_pool	256 × 256 × 64	-	0
block2_conv1	256 × 256 × 128	Weights: 3 × 3 × 64 × 128, bias: 128	73,856
block2_conv2	256 × 256 × 128	Weights: 3 × 3 × 128 × 128, bias: 128	147,584
batch_normalization_1	256 × 256 × 128	Offset: 128, scale: 128	512
block2_pool	128 × 128 × 128	-	0
block3_conv1	128 × 128 × 256	Weights: 3 × 3 × 128 × 256, bias: 256	295,168
average_pooling2d_1	64 × 64 × 256	-	0
block3_conv2	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
block3_conv3	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
block3_conv4	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
batch_normalization_2	64 × 64 × 256	Offset: 256, scale: 256	1,024
block3_pool	64 × 64 × 256	-	0
block4_conv1	32 × 32 × 512	Weights: 3 × 3 × 256 × 512, bias: 512	1,180,160
block4_conv2	32 × 32 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
max_pooling2d_1	16 × 16 × 512	-	0
block4_conv3	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block4_conv4	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
batch_normalization_3	16 × 16 × 512	Offset: 512, scale: 512	2,048
block4_pool	16 × 16 × 512	-	0
block5_conv1	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv2	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv3	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv4	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
batch_normalization_4	8 × 8 × 512	Offset: 512, scale: 512	2,048
block5_pool	8 × 8 × 512	-	0
flatten_1	8,192	-	0
dense_3	4,096	Weights: 8,192 × 4,096, bias: 4,096	33,558,528
dense_4	4,096	Weights: 4,096 × 4,096, bias: 4,096	16,781,312
dense_5	3	Weights: 4,096 × 3, bias: 3	12,291
SoftMax	1 × 1 × 3	-	0
Classification Output	1 × 1 × 3	-	0
Number of total learnable parameters			70,379,331

where F' is the weighted feature map, F is the input feature map, and W_{CA} is the channel weight matrix. The channel weight matrix is computed as follows:

$$(2) \quad W_{CA}(F) = \sigma(f_{encoder}(AvgPool(F)) + f_{encoder}(MaxPool(F)))$$

where the global max-pooling operation is represented by MaxPool, the average pooling operation is AvgPool, σ is the sigmoid function, and f_c is the channel encoder.

Spatial attention functions were analyzed using GAP and GMP to compute the average and maximum spatial responses on the input feature map. These resulting responses are utilized to combine into a set of descriptive features. A spatial weight feature map (WS) is produced by activation with a sigmoid function and is multiplied element-wise by the original feature map. This approach distills the model's focus to important regions of the network, thus identifying the spatial attention process, given as:

$$(3) \quad F'' = F' \times W_{SA}$$

where W_{SA} is the spatial weight matrix and is calculated as:

$$(4) \quad W_{SA}(F) = \sigma(f_{encoder}(AvgPool(F)) \otimes f_{encoder}(MaxPool(F)))$$

The proposed model leverages YOLOv5 to preserve the original network topology while extracting features from the three feature layers of the backbone network. The head network receives these features after they have been concatenated and sent for object detection. The head network's ability to comprehend complex spatial feature arrangements in the data is improved by integrating the CBAM. When dealing with small objects or intricate details, such as tiny lesions within an image, this problem becomes extremely helpful. Through the refinement of the network's understanding of semantic and spatial nuances, the CBAM enhances detection performance. It increases the model's ability to locate and identify small teeth with greater accuracy and generates a stronger recognition effect without increasing the training cost.

Results

This study's primary goal was to develop and evaluate a model for BC screening and diagnosis from DBT data with greater accuracy. A two-stage architecture was developed for this purpose.

Classification

The task of optimizing DL models to achieve the highest level of accuracy and detection in computer vision is still the main concern in BC screening. In this study, the performance of the VGG19 architecture-known for its impact on image processing-is shown, and the DBT images are classified into tumor, normal, and actionable classes. The effect of different optimizers and batch sizes on the accuracy and loss of the model is extensively studied. Specifically, the influence of these parameters on the performance of the VGG19 architecture in classifying DBT images into tumor, normal, and actionable categories. Table 5 presents the findings of the modified VGG19. It can be seen that with the increase in batch size, the performance, accuracy, and loss increase. The Adam optimizer shows better performance than the other two, with the highest accuracy and minimum loss.

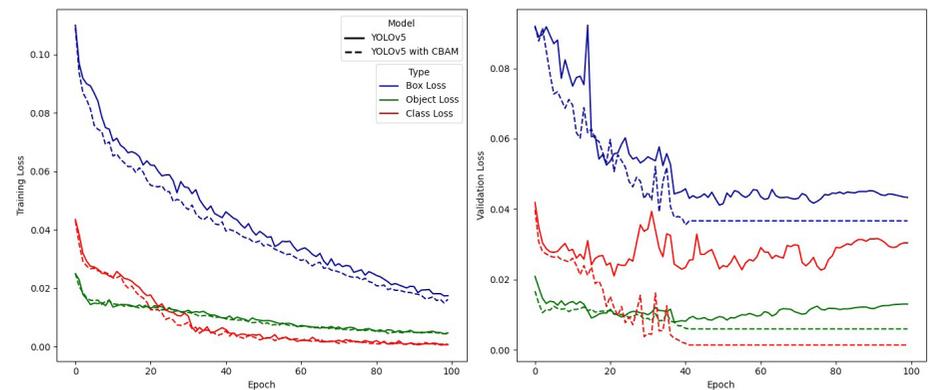


Figure 2. Comparison of the validation loss and training loss for YOLOv5-CBAM and YOLOv5. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

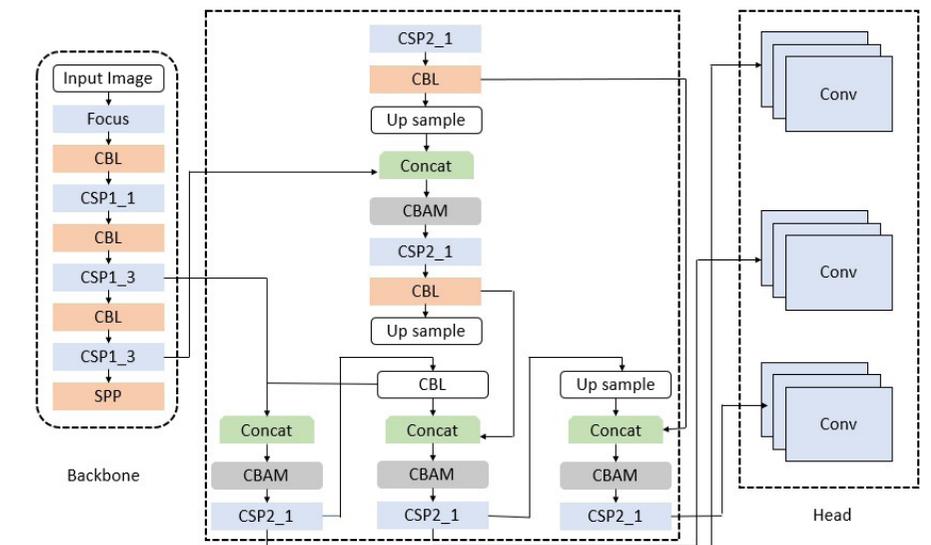


Figure 3. YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Our findings reflect the complex relationship between optimization techniques and the model's performance; thus, we demonstrated that the Adam optimizer is superior in achieving high accuracy and minimizing the loss in different batch sizes. Furthermore, the confusion matrix shown in Figure 4 not only gives more weight to our classification results but also explains the model's ability to correctly distinguish between the classes. The model shows a greater accuracy in classifying normal, actionable, and tumor, with all three classes having a true positive rate >89%.

Detection

Figure 5 compares model-to-model performance (YOLOv5 vs. YOLOv5-CBAM); the 100-epoch training period is assessed. The metrics plotted are the precision, recall, and mean average precision (mAP), and the threshold is 0.5. The YOLOv5-CBAM model gave higher sums than the standard YOLOv5 for all metrics. This is a clear indicator of im-

proved detection efficiency. The graph presents variations mostly at the beginning (0 to 20) since the model takes time to increase the weights. With time, the metrics reach steady states, where mAP is gradually getting better, which plays a role in the model's convergence. The recall for both models is usually more stable and consistently remains fairly high, whereas precision has a greater degree of fluctuation compared with recall. Thus, the model is most accurate when it predicts the relevant cases, specifically those instances where abnormalities are present in the DBT images.

Similarly, Figure 2 shows the comparison of the box loss, object loss, and class loss. The addition of CBAM to YOLOv5 shows a reduction in all three parameters in both training and validation. This shows that the YOLOv5 based-CBAM can learn from the data well and generalize it.

Table 6 displays the performance metrics for two configurations of YOLOv5. It indicates that the YOLOv5 model enhanced with the CBAM significantly outperforms the standard YOLOv5 across all metrics for both benign and cancerous classes, suggesting that the CBAM addition effectively improves the model's detection and classification capabilities in these specific medical imaging tasks.

Similar results can be observed from the confusion matrices. The YOLOv5-CBAM model shows a significant improvement over the standard YOLOv5 in both classes. It has higher true positive rates for both benign (0.84 vs. 0.71) and cancerous (0.89 vs. 0.78). It also has lower false positive and false negative rates, indicating better overall accuracy and reliability in classification. The comparison of the confusion matrices for YOLOv5 and YOLOv5-CBAM is shown in Figure 6.

Through the analysis, the variations in performance are unveiled, which are the focus of the CBAM in the detection efficiency and model convergence. Furthermore, we conduct a detailed comparison of loss parameters between the two models, and thus, we get to the issues of their learning dynamics and generalization capabilities. Furthermore, a thorough analysis of the performance metrics of both YOLOv5 versions is presented, which helps explain their effectiveness for different classes.

	Batch size	Training accuracy	Training loss	Validation accuracy	Validation loss	Testing accuracy	Testing loss
SGDM	32	88%	0.32	86%	0.34	85%	0.35
	64	85%	0.35	83%	0.37	82%	0.39
	512	82%	0.39	80%	0.42	78%	0.44
Adam	32	87%	0.3	86%	0.32	85%	0.33
	64	90%	0.27	89%	0.29	88%	0.3
	512	95%	0.2	94%	0.22	93%	0.23
RMSProp	32	87%	0.31	85%	0.33	84%	0.35
	64	84%	0.34	82%	0.36	81%	0.38
	512	80%	0.38	78%	0.4	77%	0.42

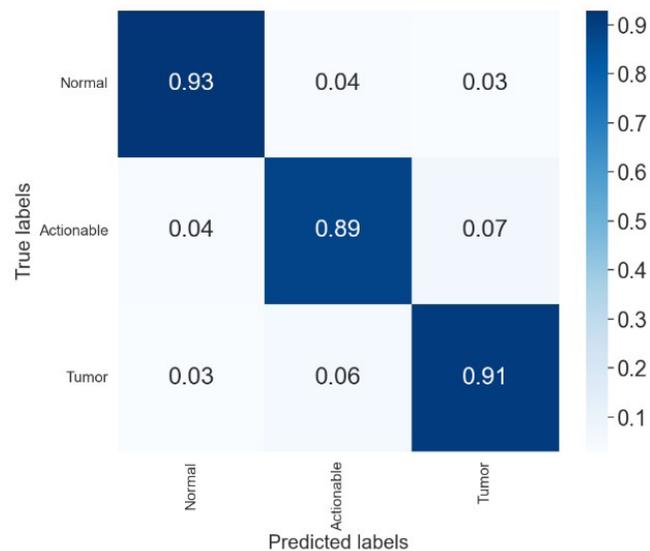


Figure 4. Confusion matrix for the VGG19 model.

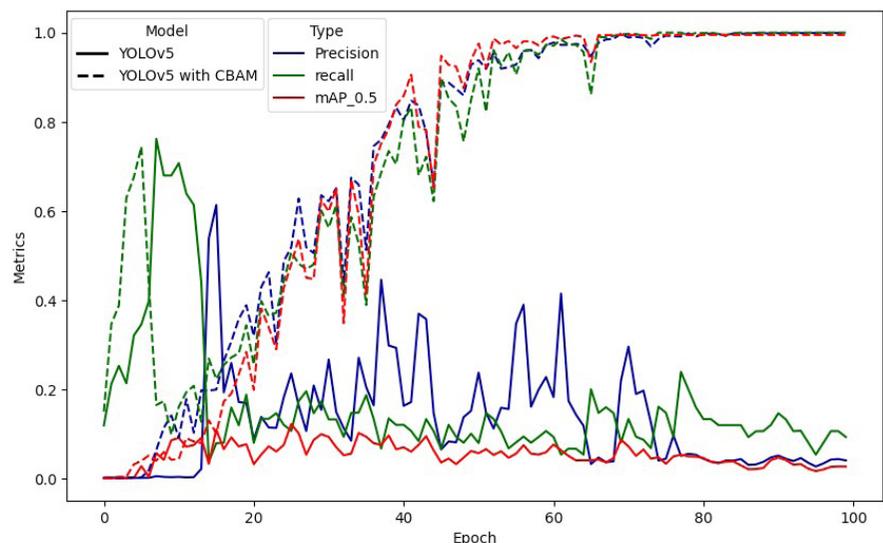


Figure 5. Comparison of the precision, recall, and mAP of YOLOv5 and YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module; mAP, mean average precision.

Table 6. Performance metrics for the two configurations of the YOLOv5 model

Classes	mAP		Recall		F1-score	
	YOLOv5	YOLOv5-CBAM	YOLOv5	YOLOv5-CBAM	YOLOv5	YOLOv5-CBAM
Benign	0.815	0.915	0.716	0.879	0.760	0.890
Cancer	0.793	0.891	0.751	0.881	0.713	0.856
Overall	0.785	0.887	0.796	0.896	0.775	0.878

YOLOv5, You Only Look Once version 5; mAP, mean average precision; CBAM, convolutional block attention module.

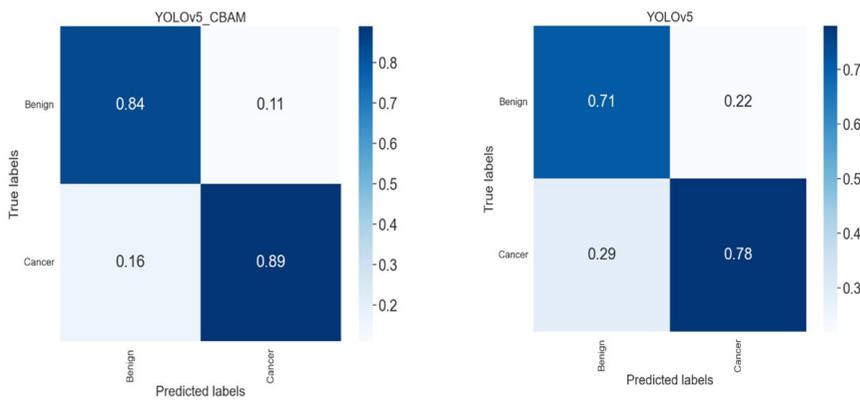


Figure 6. The comparison of the confusion matrices for YOLOv5 and YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Discussion

This research work fills a significant void in BC diagnosis through the use of enhanced DL strategies for DBT images. Today, BC is still considered one of the most widespread health issues affecting women worldwide; its early diagnosis can contribute to enhancing the effectiveness of its treatment and, consequently, the increase in female survival rates. DBT has brought improvements in the accuracy of BC screening by giving a 3D view, which has some discrepancies as compared with traditional 2D mammography. The developed two-tier DL algorithm includes a modified VGG19 model for image classification and YOLOv5-CBAM for identification of lesions, which demonstrates good outcomes in terms of accuracy and time.

In the classification stage, better results from the previous pre-trained models are achieved by the modified VGG19 with extra layers, such as batch normalization and pooling layers, to improve feature extraction and training. This change not only enhances the performance of classification models in normal, actionable, and tumor types but also highlights the versatility of DL approaches in medical imaging where conventional methods may fall short. The inclusion of YOLOv5-CBAM in the detection stage adds to the model's effectiveness in the identification of malignant and benign lesions based on

the attention mechanisms that align data highlights in the image. The YOLOv5-CBAM model improves performance through its attention mechanisms, which focus on the most informative features in the image. The CBAM enhances the model's ability to prioritize relevant areas by applying both channel and spatial attention, thereby reducing false positives and improving the detection accuracy of malignant and benign lesions.

Based on the experimental outcomes, the applicability of the philosophy of the proposed work has been demonstrated, and new achievements in the analysis of DBT using comparable methodologies have been established. To achieve this, the study adopted TCIA, which provides a large dataset to minimize the likelihood of model overfitting, which is detrimental in clinical applications. Furthermore, the study presents methodological reflections, including such aspects as data preprocessing, dataset enlargement, and computational environment; these are crucial for reproducing the presented study and expanding similar research. Altogether, this study advances the knowledge base of artificial intelligence (AI)-supported BC diagnosis and lays down the foundation for effective diagnosis models that can enhance identification processes globally, hence boosting patients' survival.

Despite the promising results, the study faced several limitations. The reliance on a specific dataset from TCIA may limit generalizability to other populations. Additionally, the model's performance in real-world clinical settings needs further validation. Challenges also included handling variability in image quality and computational resources required for model training. Future work will be able to extend to various associated modalities, add multiple imaging data, and carry out studies with clinical materials to validate the performance in practical scenarios. Moreover, the role of DL models in the diagnostics of medical conditions, together with methods that enable the interpretability and explainability of the results, deserves more attention and development to earn the trust of physicians. Finally, harnessing such advanced AI technologies as the ones discussed in this study has the potential to significantly improve the practice of BC screening and its management and, thus, the state of global healthcare.

In conclusion, over 39 million examinations are performed yearly as part of the BC screening program. However, BC screening has been one of the most difficult applications of AI in medical imaging. DBT can create 3D images where tissue overlapping is reduced, making it simpler for radiologists to spot abnormalities and resulting in a more precise diagnosis. This study suggested the use of a new computer-aided multi-class diagnosis system that uses YOLOv5-CBAM to identify benign or malignant tumors and a modified DL architecture (VGG19) for classifying DBT images. A large set of breast tomography images was used in the experiment. Test accuracy, training loss, and training accuracy showed better performance of our proposed architecture than the previous models. Hence, the modified VGG19 classified DBT images more accurately than previously possible using pre-trained model-based architectures. Second, YOLOv5-based CBAM precisely discriminated between benign lesions and those that are malignant.

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

Funding

The author would like to extend their sincere appreciation to the researchers supporting the program for funding this work under the Researchers Supporting Project number

References

1. Sharma MP, Shukla S, Misra G. Recent advances in breast cancer cell line research. *Int J Cancer*. 2024;154(10):1683-1693. [\[CrossRef\]](#)
2. da Costa Nunes GG, de Freitas LM, Monte N, et al. Genomic variants and worldwide epidemiology of breast cancer: a genome-wide association studies correlation analysis. *Genes (Basel)*. 2024;15(2):145. [\[CrossRef\]](#)
3. Giaquinto AN, Sung H, Miller KD, et al. Breast cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(6):524-541. [\[CrossRef\]](#)
4. Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)*. 2022;135(5):584-590. [\[CrossRef\]](#)
5. Yi A, Jang MJ, Yim D, Kwon BR, Shin SU, Chang JM. Addition of screening breast US to digital mammography and digital breast tomosynthesis for breast cancer screening in women at average risk. *Radiology*. 2021;298(3):568-575. [\[CrossRef\]](#)
6. Zheng J, Lin D, Gao Z, et al. Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*. 2020;8:96946-96954. [\[CrossRef\]](#)
7. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. *Science*. 2022;375(6586):9040. [\[CrossRef\]](#)
8. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27(2):244-249. [\[CrossRef\]](#)
9. Rezaei Z. A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Syst Appl*. 2021;182:115204. [\[CrossRef\]](#)
10. Dhamija E, Gulati M, Deo SVS, Gogia A, Hari S. Digital breast tomosynthesis: an overview. *Indian J Surg Oncol*. 2021;12(2):315-329. [\[CrossRef\]](#)
11. Gao Y, Moy L, Heller SL. Digital breast tomosynthesis: update on technology, evidence, and clinical practice. *Radiographics*. 2021;41(2):321-337. [\[CrossRef\]](#)
12. Wani IM, Arora S. Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey. *Med Biol Eng Comput*. 2020;58(9):1873-1917. [\[CrossRef\]](#)
13. Zebari DA, Ibrahim DA, Zeebaree DQ, et al. Systematic review of computing approaches for breast cancer detection based computer aided diagnosis using mammogram images. *Appl Artif Intell*. 2021;35(15):2157-2203. [\[CrossRef\]](#)
14. Ramadan SZ. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *J Healthc Eng*. 2020;2020:9162464. [\[CrossRef\]](#)
15. Xu L, Gao J, Wang Q, et al. Computer-aided diagnosis systems in diagnosing malignant thyroid nodules on ultrasonography: a systematic review and meta-analysis. *Eur Thyroid J*. 2020;9(4):186-193. [\[CrossRef\]](#)
16. Li X, Qin G, He Q, et al. Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. *Eur Radiol*. 2020;30(2):778-788. [\[CrossRef\]](#)
17. Ricciardi R, Mettivier G, Staffa M, et al. A deep learning classifier for digital breast tomosynthesis. *Phys Med*. 2021;83:184-193. [\[CrossRef\]](#)
18. Oba K, Adachi M, Kobayashi T, et al. Deep learning model to predict Ki-67 expression of breast cancer using digital breast tomosynthesis. *Breast Cancer*. 2024. [\[CrossRef\]](#)
19. Buda M, Saha A, Walsh R, et al. A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images. *JAMA Netw Open*. 2021;4(8):2119100. [\[CrossRef\]](#)
20. El-Shazli AMA, Youssef SM, Soliman AH. Intelligent computer-aided model for efficient diagnosis of digital breast tomosynthesis 3D imaging using deep learning. *Applied Sciences (Switzerland)*. 2022;12(11):5736. [\[CrossRef\]](#)
21. Martí R, del Campo PG, Vidal J, et al. Lesion detection in digital breast tomosynthesis: method, experiences and results of participating to the DBTex challenge. Proceedings volume 12286, 16th International Workshop on Breast Imaging (IWBI2022); 122860W, 2022. [\[CrossRef\]](#)
22. Nogay HS, Akinci TC, Yilmaz M. Comparative experimental investigation and application of five classic pre-trained deep convolutional neural networks via transfer learning for diagnosis of breast cancer. *Adv Sci Technol Res J*. 2021;15(3):1-8. [\[CrossRef\]](#)
23. Maciej Serda, Becker FG, Cleary M, et al. Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza. G. Balint, Antala B, Carty C, Mabieme JMA, Amar IB, Kaplanova A, eds. Uniwersytet śląski. 2013;7(1):343-354. [\[CrossRef\]](#)
24. Hassan L, Saleh A, Singh VK, Puig D, Abdel-Nasser M. Detecting breast tumors in tomosynthesis images utilizing deep learning-based dynamic ensemble approach. *Computers* 2023. 2023;12(11):220. [\[CrossRef\]](#)
25. Mota AM, Clarkson MJ, Almeida P, Matela N. Detection of microcalcifications in digital breast tomosynthesis using faster R-CNN and 3D volume rendering. In proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022). *Bioimaging*. 2022;2:80-89. [\[CrossRef\]](#)
26. Konz N, Buda M, Gu H, et al. A competition, benchmark, code, and data for using artificial intelligence to detect lesions in digital breast tomosynthesis. *JAMA Netw Open*. 2023;6(2):230524. [\[CrossRef\]](#)
27. Wang X, Chen G, Qian G, et al. Large-scale multi-modal pre-trained models: a comprehensive survey. *Mach Intell Res*. 2023;20(4):447-482. [\[CrossRef\]](#)
28. Akbarnezhad E, Naserizadeh F. Improving camouflage object detection using U-NET and VGG16 deep neural networks and CBAM attention mechanism. *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)*. 2024:56-62. [\[CrossRef\]](#)



Diagnostic performance of radiomics analysis for pulmonary cancer airway spread: a systematic review and meta-analysis

Jie Chen¹
 Xinyue Zhang²
 Chi Xu¹
 Kefu Liu¹

¹The Affiliated Suzhou Hospital of Nanjing Medical University, Department of Radiology, Suzhou, China

²Zhejiang University School of Medicine Sir Run Run Shaw Hospital, Department of Medical Record, Hangzhou, China

PURPOSE

Spread through air spaces (STAS) is a unique metastatic pattern of pulmonary cancer closely associated with patient prognosis. This study evaluates the application of radiomics in the diagnosis of pulmonary cancer STAS through meta-analysis and explores its clinical significance and potential limitations.

METHODS

We systematically searched the PubMed, Embase, and Cochrane Central Register of Controlled Trials databases for relevant studies between inception and April 1, 2024. The main evaluation indicators included sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and area under the curve (AUC). A total of 18 studies, covering 6,642 lung cancer cases, were included in the systematic review.

RESULTS

In the development cohort, the sensitivity of radiomics for diagnosing STAS was 0.80 [95% confidence interval (CI): 0.75–0.84; $P < 0.001$; I^2 : 72.8%], and the specificity was 0.79 (95% CI: 0.71–0.85; $P < 0.001$; I^2 : 93.4%). In the validation cohort, the sensitivity was 0.81 (95% CI: 0.75–0.86; $P < 0.001$; I^2 : 45.8%), and the specificity was 0.74 (95% CI: 0.68–0.80; $P < 0.001$; I^2 : 65.0%). The summary AUC for both cohorts was 0.85 (95% CI: 0.82–0.88). Deeks' funnel plot analysis showed no significant publication bias in either cohort (P values: 0.963 and 0.106, respectively).

CONCLUSION

Radiomics analysis demonstrates important clinical significance in the diagnosis of pulmonary cancer STAS, with promising sensitivity and specificity results in both development and validation cohorts.

CLINICAL SIGNIFICANCE

While radiomics analysis offers valuable diagnostic insights for STAS in pulmonary cancer, its limitations must be carefully considered. Future research should focus on addressing these limitations and further exploring the application prospects of radiomics in lung cancer diagnosis and treatment.

KEYWORDS

Lung cancer, spread through air spaces, radiomics, computed tomography, meta-analysis

Corresponding author: Kefu Liu

E-mail: lkf77@126.com

Received 14 May 2024; revision requested 09 June 2024;
last revision received 26 September 2024; accepted 17
November 2024.



Epub: 07.04.2025

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242852

Lung cancer is a highly lethal disease with significant implications for patients' quality of life and lifespan.¹ In addition to its high incidence and mortality, lung cancer also exhibits diverse invasive patterns, including a particularly unique metastatic mode termed spread through air spaces (STAS).² First proposed by Kadota et al.² and definitively defined by the World Health Organization (WHO) in 2015, STAS refers to the invasion of tumor micropapillary clusters, solid nests, or single cells beyond the tumor edge into the air spaces of surrounding lung parenchyma.³ This form of invasion complicates the diagnosis, treatment, and prognosis

of lung cancer. Before the concept of STAS, research in peritumoral radiomics primarily focused on the tumor microenvironment and its interactions, which play a critical role in tumor progression and metastasis. The presence of STAS often suggests a higher risk of postoperative recurrence and a poorer prognosis compared with traditional modes of lung cancer metastasis.⁴ Therefore, accurate diagnosis and assessment of STAS in patients with lung cancer are of paramount importance. Studies have shown that patients with lung cancer with positive STAS typically require more extensive surgical resection, reflecting their poorer prognosis.^{5,6} The presence of STAS is also associated with tumor recurrence, metastasis, and chemotherapy efficacy.⁷ Thus, understanding and evaluating the manifestation and impact of STAS in lung cancer are crucial for devising personalized treatment strategies and improving patient prognosis.

Traditional imaging parameters, such as tumor size and morphological features, although important indicators for evaluating lung cancer, rely on subjective judgments of physicians and have certain limitations. Traditional imaging parameters may lack sufficient accuracy and sensitivity, particularly in identifying tiny STAS lesions. Moreover, due to the complex and diverse imaging manifestations of lung cancer, visual estimation alone often cannot fully leverage the potential of imaging in lung cancer diagnosis. The emergence of radiomics fills this gap by utilizing both machine learning (ML) and deep learning (DL) methods to perform quantitative

analysis of lung cancer images, extracting a wealth of implicit information, including hand-crafted features such as shape, grayscale, texture, and wavelet characteristics, as well as deep radiomics features derived from advanced neural networks. This approach provides more comprehensive, objective, and accurate information for early diagnosis, staging, and prognostic evaluation of lung cancer.⁸ In recent years, radiomics has made significant progress in the field of lung cancer diagnosis. Many studies have shown that analyzing lung cancer images using radiomics analysis can significantly improve the accuracy and sensitivity of diagnosis, and can even detect tiny STAS lesions.⁹ This provides clinicians with more reliable information to help formulate personalized treatment plans and improve patient survival rates and quality of life. However, despite a few individual studies, there still lacks a systematic and comprehensive meta-analysis to provide a thorough and objective assessment of radiomics in the diagnosis of lung cancer STAS.

With the continuous development of ML technology and the application of radiomics, there is a unique opportunity to leverage big data and intelligent algorithms to enhance the identification and evaluation of STAS in patients with lung cancer. This study systematically evaluates the performance of radiomics analysis in detecting STAS lesions, specifically focusing on the capability to identify small STAS manifestations that may be missed by traditional imaging methods. We hypothesize that integrating advanced radiomic features, both hand-crafted and DL-derived, will significantly improve the accuracy and sensitivity of STAS detection. By providing robust evidence-based insights, this study seeks to inform clinical decision-making and improve patient outcomes in lung cancer management.

Methods

This meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies guidelines.¹⁰ The detailed protocol is accessible in INPLASY (INPLASY2024100103). As this meta-analysis did not involve human or animal participants, ethics approval was not required.

Search strategy and literature selection

We systematically searched three major electronic databases, namely PubMed, Embase and the Cochrane Central Register of Controlled Trials databases, between in-

ception and April 1, 2024, without language restrictions. The search strategy combined MeSH or Emtree terms with free terms to ensure comprehensive results. Keywords were set to search in the title and abstract mode for greater accuracy. Additionally, reference lists of relevant studies or reviews were manually searched to retrieve potentially missed literature. The search topics included radiomics, artificial intelligence (AI), ML, lung cancer, and airway spread. These terms were combined using Boolean operators to ensure comprehensive coverage of relevant studies (Supplementary File 1). The search process was conducted independently by two researchers, and records were imported into reference management software for automatic removal of duplicates and subsequent manual exclusion. Disagreements were resolved through consultation with a third researcher.

Inclusion and exclusion criteria

Based on the Population, Intervention, Comparator, Outcome, Study design principle, studies meeting the following criteria were included: 1) the study population comprised patients with lung cancer; 2) the intervention involved AI-assisted radiomics; 3) histopathology was used as the reference standard; 4) the primary outcome was pulmonary cancer airway spread; and 5) the study design was either a cohort study or case-control study. Studies meeting the following criteria were excluded: 1) irrelevant study types, such as animal studies, case reports, or conference papers; 2) studies with incomplete data; and 3) studies that did not report predefined outcomes or did not adhere to the intervention and control settings.

Data extraction and risk of bias assessment

Two researchers independently extracted data using a pre-designed form from the included studies, including author names, publication dates, study designs, sample sizes, locations of conduct, characteristics of patient populations (e.g., age and gender), model validation methods, algorithms used for modelling, imaging equipment parameters, use of clinical information, primary inclusion variables, and diagnostic performance. Quantitative data in 2 × 2 tables, including true positives, true negatives, false positives, and false negatives, were collected. The methodological quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool, which covers patient selection, index test, reference standard, flow, and timing, among other as-

Main points

- Radiomics employs both machine learning and deep learning methods to quantitatively analyze lung cancer images and extract rich hidden information. This includes hand-crafted features such as shape, grayscale, texture, and wavelet, as well as features derived from deep radiomics techniques. By integrating these diverse features, radiomics provides more comprehensive, objective, and accurate information for the early diagnosis of lung cancer, as well as for staging and prognostic assessment.
- This study explores the performance of radiomics analysis in the diagnosis of lung cancer spread through air spaces (STAS) through a systematic review and meta-analysis, providing reliable evidence-based support for clinical practice.
- Analyzing lung cancer images using radiomics analysis can significantly improve the accuracy and sensitivity of diagnosis, and can even detect tiny STAS lesions.

pects. Disagreements were resolved through consultation with a third researcher.

Statistical analysis

Data analysis in this study was performed using RevMan 5.4 and Stata SE 15.0 software. Sensitivity and specificity were calculated based on 2×2 table data and presented graphically, with squares representing values and horizontal lines representing corresponding confidence intervals (CIs). Summary receiver operating characteristic curves were used to represent the performance of diagnostic tests. The approximate classification criteria for area under the curve (AUC) values were as follows: 0.50–0.60 = inadequate, 0.60–0.70 = poor, 0.70–0.80 = fair, 0.80–0.90 = good and 0.90–1 = excellent. Additionally, summary statistics of positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio, along with their 95% CIs, were calculated. Heterogeneity of results was assessed using Cochran's Q test and the I^2 statistic test, and meta-analysis was conducted using either fixed-effects or random-effects models accordingly. The possibility of publication bias was assessed using Deeks' funnel plot analysis, and sensitivity analysis was performed to evaluate the stability of the results. Fagan's nomogram was used to evaluate the clinical utility of radiomics and calculate the post-test probability of STAS.

Results

Literature search

The flowchart of the literature search process for this meta-analysis is depicted in Figure 1. Initially, 125 records were identified from the databases, and an additional 2 records were manually retrieved from other sources. After removing duplicates, 97 records remained. Subsequently, based on screening of titles and abstracts, 70 irrelevant records were excluded, leaving 27 articles for full-text assessment. Finally, a total of 18 articles^{11–28} were included in the systematic review, with data from 13 articles used for the meta-analysis.

Characteristics of included studies

The basic characteristics of the included studies are presented in Table 1. All studies were retrospective in design, with the majority (66.7%) being single-center studies and only 6 being multi-center studies. Among the 18 included studies, 14 were conducted in China, 2 in Japan, and 1 each in Italy and South Korea. Most studies focused on pa-

tients with lung adenocarcinoma, with sample sizes ranging from 92 to 681 and mean/median ages ranging from 53.1 to 70 years. The proportion of men ranged from 31.3% to 58.9%. External validation was performed in 8 studies.

Details of the radiomics predictive models included in the studies are summarized in Table 2. Common ML algorithms used included the least absolute shrinkage and selection operator, random forest, and decision tree algorithms. Although a variety of imaging equipment manufacturers were involved, all models were based on chest computed tomography. In addition to radiomics, 6 studies incorporated clinical information such as gender, age, smoking status, and tumor size in model construction. The definition of STAS was generally consistent across studies, primarily based on WHO criteria (i.e., micropapillary clusters, solid nests, or single cells beyond the edge of the tumor extending into

the air spaces in the surrounding lung parenchyma). The AUC values of the constructed models ranged from 0.66 to 0.99.

Risk of bias assessment

Understanding risk of bias is crucial for evaluating the reliability of study findings. Bias can be introduced at various stages of a study, including patient selection, index test application, and reference standards. High or unclear risk of bias can affect the internal validity and generalizability of the study results.

The methodological quality of the 18 included studies is detailed in Figure 2. One study was at high risk of bias in the "patient selection" domain due to a case-control study design, and another study had an unclear risk due to insufficient description. Additionally, 12 studies had unclear risks of bias in the "index test" and "reference stan-

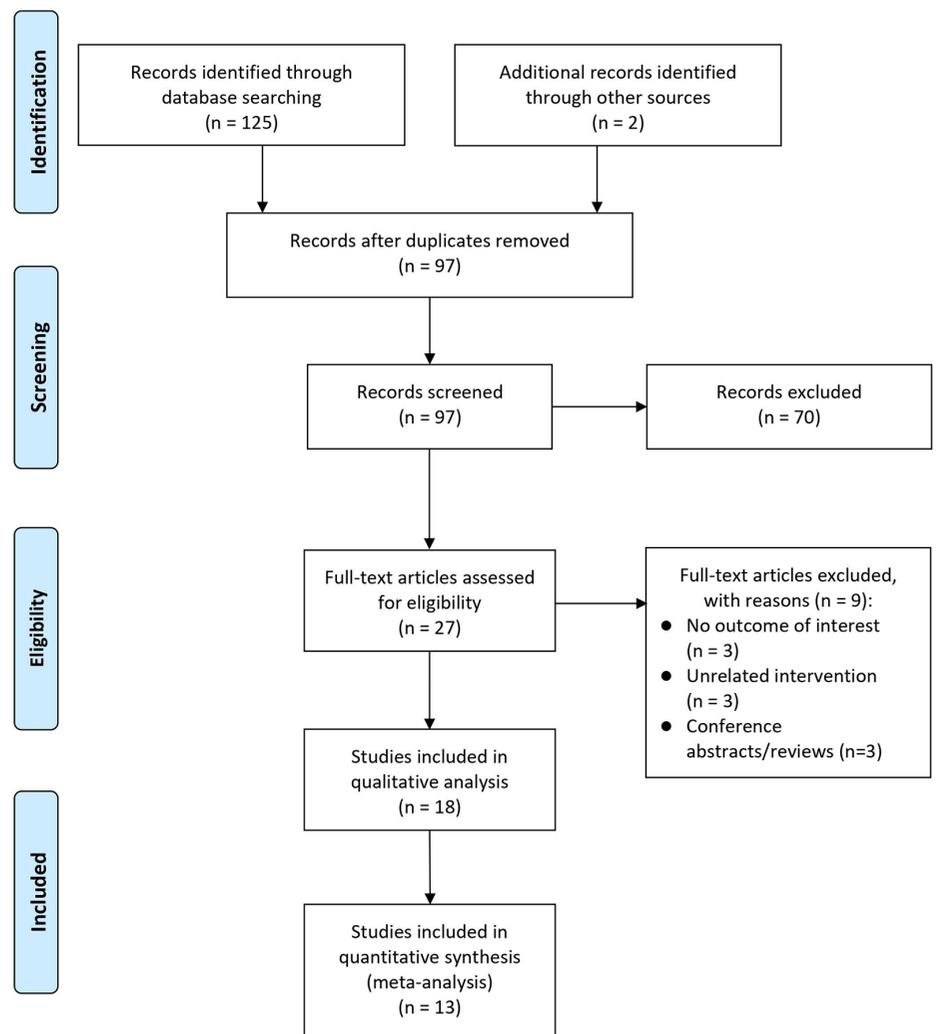


Figure 1. PRISMA flow diagram of study selection. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Table 1. Main characteristics of the included studies								
Author, year	Design	Sites	Country	Population	Sample size	Age, years	Men, %	Validation approach
Bassi et al. ¹¹ , 2022	Retrospective	Single-center	Italy	Patients with resected lung adenocarcinoma	149	68.0 (61.5–74.0); 69.0 (61.0–73.8)	57.0	External validation
Chen et al. ¹² , 2020	Retrospective	Multi-center	China	Patients with stage I invasive lung adenocarcinoma	345	60 (35–82); 60 (34–76)	43.2	External validation
Chen et al. ¹³ , 2022	Retrospective	Multi-center	China	Patients with lung adenocarcinoma	327	58.9 (25–85)	40.1	External validation
Gong et al. ¹⁴ , 2023	Retrospective	Single-center	China	Patients undergoing surgery for lung cancer	537	62 (56–67); 62 (55–67)	49.3	Random splitting
Han et al. ¹⁵ , 2022	Retrospective	Single-center	China	Patients with stage AI lung adenocarcinoma	395	59 ± 10	52.4	Random splitting
Jiang et al. ¹⁶ , 2020	Retrospective	Single-center	China	Patients with lung adenocarcinoma	462	Mean: 58.06	44.6	Random splitting
Jin et al. ¹⁷ , 2023	Retrospective	Multi-center	China	Patients with primary lung cancer	674	Mean: 57.8–61.5	38.6	External validation
Liao et al. ¹⁸ , 2022	Retrospective	Single-center	China	Patients with pathologically confirmed invasive clinical stage I lung adenocarcinoma who accepted surgical resection	256	60.7 ± 10.7; 59.5 ± 10.9	47.7	Random splitting
Lin et al. ¹⁹ , 2024	Retrospective	Multi-center	China	Patients with lung adenocarcinoma who underwent complete lung resection	681	56.9 ± 11.2	31.3	External validation
Liu et al. ²⁰ , 2022	Retrospective	Single-center	China	Patients with stage AI lung adenocarcinoma who underwent surgical resection	92	53.1 ± 10.9; 56.9 ± 7.6	40.2	N/A
Onozato et al. ²¹ , 2021	Retrospective	Single-center	Japan	Patients who underwent surgical resection of lung cancer	226	70 (39–88); 68 (38–89)	54.4	Cross-validation
Qi et al. ²² , 2021	Retrospective	Single-center	China	Patients with primary lung adenocarcinoma confirmed by surgical resection and pathology	216	56 ± 11	58.9	External validation
Suh et al. ²³ , 2024	Retrospective	Single-center	South Korea	Patients who underwent surgical resection for clinical stage AI (tumor size ≤3 cm) lung adenocarcinoma	521	Mean: 61.2–66.6	45.5	Temporal validation
Takehana et al. ²⁴ , 2022	Retrospective	Single-center	Japan	Patients with pathologically confirmed lung adenocarcinoma	339	67 (61–73)	47	Random splitting
Tao et al. ²⁵ , 2022	Retrospective	Single-center	China	Patients with non-small cell lung cancer	203	59.6 ± 9.8; 61.3 ± 8.6	55.2	Random splitting
Wang et al. ²⁶ , 2024	Retrospective	Multi-center	China	Patients with confirmed lung adenocarcinoma	602	Mean: 56.69–60.48	46.7	External validation
Wang et al. ²⁷ , 2024	Retrospective	Multi-center	China	Patients with clinical stage AI non-small cell lung cancer	405	≥65 years: 31.9%	49.4	External validation
Zhuo et al. ²⁸ , 2020	Retrospective	Single-center	China	Patients with confirmed lung adenocarcinoma	212	58.84 ± 9.92	42.92	Random splitting

Age was expressed as mean ± standard deviation or median (25th quartile–75th quartile). AI, artificial intelligence; N/A, not available.

Table 2. Characteristics of radiomics-based prediction models

Author, year	Algorithms	Imaging equipment	Clinical information	Included variables	Reference standard of STAS	AUC
Bassi et al. ¹¹ , 2022	NB, k-NN, RF, LR	Unlimited	No	Radiomics, radiological features and mixed features	The presence of a rim of normal lung surrounding the entire tumor circumference	0.66–0.79
Chen et al. ¹² , 2020	NB	SOMATOM Definition AS scanner (64 × 0.625 mm detector, 1.0 pitch) or Brilliance 40 scanner (40 × 0.625 mm detector configuration, 0.4 pitch)	No	Radiomics (sphericity, 90 percentile, gray level variance, cluster tendency, gray level variance)	Tumor cells emerging in paracarcinoma normal alveolar spaces, which are far from the main tumor and appear in the form of micropapillary clusters, small solid tumor nests, or single cells	0.69
Chen et al. ¹³ , 2022	NR	GE (LightSpeed Pro 32, LightSpeed Pro 16, BrightSpeed, and Revolution EVO), Philips (iCT 256, Brilliance and Ingenuity), Siemens (Definition AS+, Emotion 16 and Sensation 64), and Toshiba (Aquilion ONE)	No	Radiomics on the basis of “near-pure” subtype data using patch-wise analysis within a tumor border area	Tumor cells within air spaces in the lung parenchyma at a distance of at least 1 alveolus away from the main tumor	0.81, 0.83
Gong et al. ¹⁴ , 2023	NR	Spiral CT scanners (Siemens SOMATOM Definition AS+ and Siemens SOMATOM Drive)	No	44 radiomics features	Tumor cells in airtspaces outside the main tumor boundary	0.802–0.834
Han et al. ¹⁵ , 2022	LASSO, LR	Multislice spiral CT scanners (SOMATOM Definition AS+ and Siemens Healthineers, Germany)	Yes	Sex, age, smoking, size, radiomics	Tumor cells were found in the lung air spaces beyond the edge of the primary tumor	0.812, 0.850
Jiang et al. ¹⁶ , 2020	RF	A 16-detector CT scanner (Philips Brilliance 16, Philips Medical Systems)	Yes	Age and 12 radiomics features	The discovery of tumor cells in the lung air spaces beyond the edge of the main tumor	0.754
Jin et al. ¹⁷ , 2023	Deep CNN	NR	No	Radiomics	Micropapillary clusters, solid nests, or single cells beyond the edge of the tumor extending into the air spaces in the surrounding lung parenchyma	0.84, 0.94
Liao et al. ¹⁸ , 2022	LASSO	NR	No	Radiomics	Tumor cells within air gaps in paracarcinoma normal alveolar spaces beyond the margin of the primary tumor	0.871, 0.869
Lin et al. ¹⁹ , 2024	DL	CT: Toshiba (Tokyo, Japan), Philips (Best, The Netherlands), GE (Waukesha, Wisconsin, USA), and Siemens (München, Germany)	No	Radiomics	Tumor cells emerging in paracarcinoma normal alveolar spaces, which are far from the main tumor and appear in the form of micropapillary clusters, small solid tumor nests, or single cells.	0.80, 0.82
Liu et al. ²⁰ , 2022	RF	Shanghai United Imaging uCT550 multislice spiral	Yes	Sex, age, and radiomics	NR	NR
Onozato et al. ²¹ , 2021	XGBoost	Aquilion Prime (Canon Medical Systems Corporation, Tochigi, Japan), Aquilion ONE (CANON), Alexion (CANON), Activion16 (CANON), and Aquilion64 (CANON)	No	Radiomics	Micropapillary clusters, solid nests, or single cells beyond the edge of the tumor extending into the air spaces in the surrounding lung parenchyma	0.77
Qi et al. ²² , 2021	AdaBoost	General Electric (LightSpeed VCT; Waukesha, Wis) or Siemens (Definition Flash, Erlangen, Germany)	No	Radiomics	Micropapillary clusters, solid nests, or single cells spread within the air spaces beyond the edge of the main tumor	0.909, 0.907, 0.897
Suh et al. ²³ , 2024	LASSO, LR	NR	Yes	Lesion type on CT, solid portion size on CT, male, radiomics	Tumor cells within the air spaces in the lung parenchyma, beyond the edge of the main tumor	0.815–0.878

Table 2. Continued

Author, year	Algorithms	Imaging equipment	Clinical information	Included variables	Reference standard of STAS	AUC
Takehana et al. ²⁴ , 2022	LASSO, LR	A 64-detector-row CT scanner (Aquilion 64, Canon Medical Systems, Otawara, Japan) or a 320-detector-row scanner (Aquilion ONE, Canon Medical Systems)	No	Peritumor radiomics	Tumor aggregates floating in the air cavity at least one alveolus away	0.76, 0.79
Tao et al. ²⁵ , 2022	CNN, LR, DT, LDA, SGD, PSVM, SPSVM, XGBoost, AdaBoost	A single-source, 64-multidetector CT scanner (Brilliance CT, Philips Healthcare)	No	Radiomics	Tumor cells in the lung air spaces beyond the edge of the primary tumor	0.80, 0.93
Wang et al. ²⁶ , 2024	Squeeze-and-excitation attention module with the ResNet50 architecture		No	Radiomics	Micropapillary clusters, solid nests, or single cells spread within the air spaces beyond the edge of the main tumor	0.783, 0.806, 0.933
Wang et al. ²⁷ , 2024	LR, mRMR, LASSO	Toshiba Aquilion16 row, GE Light Speed VCT64 row, Philips Ingenuity 64 row, and Brilliance iCT 128 row CT, American Light Speed 16, and Dutch Philips iCT 256-row CT, SOMATOM Definition Flash and SOMATOM Drive 64-row CT machines	Yes	Sex, CEA, CTR, density type, distal ribbon sign, radiomics	Micropapillary clusters, solid nests, or single cells spread within the air spaces beyond the edge of the main tumor	0.901, 0.875, 0.878
Zhuo et al. ²⁸ , 2020	LASSO	SOMATOM Force (Siemens, Germany), Aquilion One/320 (Toshiba, Japan), and uCT128 (UIH, China)	Yes	The maximum diameter of the solid component, mediastinal lymphadenectasis, radiomics	Micropapillary clusters, solid nests, or single cells beyond the edge of the tumor into air spaces in the surrounding lung parenchyma	0.98, 0.99

STAS, spread through air spaces; AUC, area under the curve; NB, naïve bayes; RF, random forest; LR, logistic regression; NR, not reported; LASSO, the least absolute shrinkage and selection operator; CNN, convolutional neural network; DL, deep learning; DT, decision tree; LDA, linear discriminant analysis; SGD, stochastic gradient descent; PSVM, poly support vector machine; SPSVM, sigmoid poly support vector machine; mRMR, maximal redundancy minimal relevance; CEA, carcinoembryonic antigen; CT, computed tomography; CTR, consolidation-to-tumor ratio.

ard” domains due to unreported blinding. Notably, all studies had low risks in the “flow and timing” domain and overall showed few concerns.

Meta-analysis

Meta-analysis was conducted separately for data from development/internal validation and external validation/random splitting, and the results are shown in Figure 3. The figure presents forest plots illustrating the diagnostic performance of radiomics for predicting STAS in both training and validation cohorts. Forest plots are commonly used in meta-analyses to provide a visual representation of the results from individual studies and the overall summary effect. Each study’s estimate of sensitivity and specificity is displayed as a square, with the size of the square reflecting the weight of the study in the meta-analysis. The horizontal lines represent the CIs for each estimate. The diamond at the bottom of each plot represents the pooled estimate of sensitivity and specificity, providing a combined result from all studies.

In the development/internal validation cohort (Panel A of Figure 3), as the forest plot

shows, the sensitivity of radiomics for diagnosing STAS was 0.80 (95% CI: 0.75–0.84), and the specificity was 0.79 (95% CI: 0.71–0.85), with the presence of substantial heterogeneity ($P < 0.001$, I^2 : 72.8% for sensitivity and $P < 0.001$, I^2 : 93.4% for specificity).

In the external validation/random splitting cohort (Panel B of Figure 3), the forest plot showed similar performance, with a sensitivity of 0.81 (95% CI: 0.75–0.86) and a specificity of 0.74 (95% CI: 0.68–0.80) (heterogeneity: $P = 0.040$, I^2 : 45.8% for sensitivity and $P < 0.001$, I^2 : 65.0% for specificity).

As shown in Figure 4, the pooled AUC was 0.85 (95% CI: 0.82–0.88) for both development/internal validation and external validation/random splitting cohorts.

Publication bias and sensitivity analysis

Publication bias occurs when the outcome of the research influences the decision whether to publish it. This can lead to an overestimation of the effect in published studies. Sensitivity analysis assesses how the results vary with changes in the data or analytical methods. Both publication bias analy-

sis and sensitivity analysis are critical for understanding the robustness and reliability of the meta-analysis findings.

Deeks’ funnel plot analysis revealed no significant publication bias in either cohort ($P = 0.963$ and 0.106 , respectively), as shown in Figure 5. Sensitivity analysis indicated that the exclusion of individual studies did not significantly affect the pooled results, indicating the stability of the study findings.

Clinical utility

In the development/internal validation cohort (Panel A of Figure 6), the Fagan plot indicated that when the pretest probability of STAS was positive, the post-test probability increased significantly from 20% to 48% after applying the radiomics test. Conversely, when the pretest probability was negative, the post-test probability decreased to 6%, indicating a low likelihood of STAS when the test result is negative.

In the external validation/random splitting cohort (Panel B of Figure 6), similar trends were observed. As the Fagan plot shows, the use of radiomics increased the post-test probability from 20% to 44% when

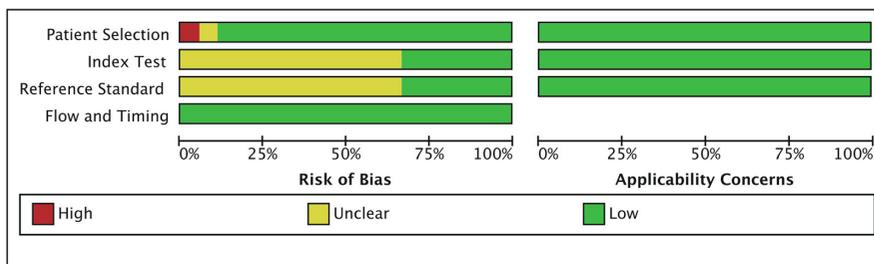
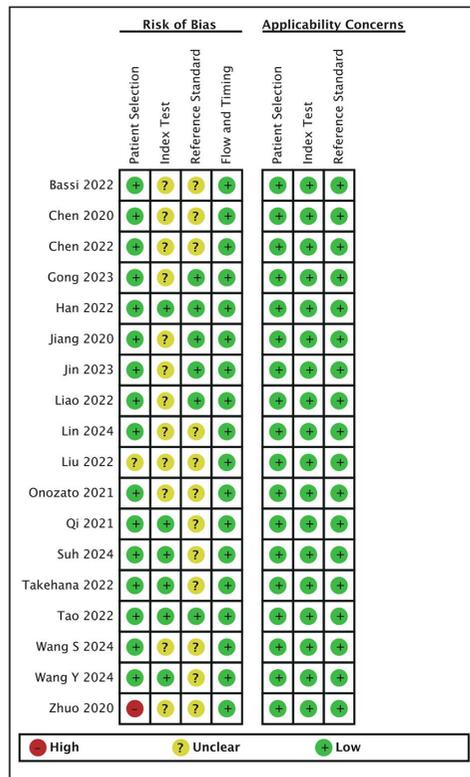


Figure 2. Methodological evaluation of the included studies.

the pretest probability was positive, and decreased it to 6% when the pretest probability was negative.

Discussion

Main findings

This study comprehensively analyzed the diagnostic value of radiomics analysis for lung cancer STAS by synthesizing multiple recent studies. The main findings are as follows: 1) regardless of whether it was in the development or validation cohorts, radiomics showed good sensitivity and specificity in diagnosing lung cancer STAS; 2) radiomics demonstrated good discriminative ability for diagnosing lung cancer STAS, accurately distinguishing between two patient groups; and 3) no significant publication bias was found in the included studies, although methodological quality assessment indicated uncertain risk of bias in some studies. The results of this study confirm the potential

clinical utility of radiomics analysis, providing new insights and methods for the diagnosis of lung cancer STAS and promoting the development of imaging and AI in this field.

Importance of spread through air spaces in lung cancer diagnosis and treatment

As a unique mode of metastasis, STAS plays a crucial role in the diagnosis and treatment of lung cancer. First, the presence of STAS is closely related to the prognosis of patients with lung cancer. Multiple studies have shown that patients with positive STAS generally have poorer prognosis, with significantly increased rates of postoperative recurrence and distant metastasis, as well as significantly shortened survival periods.^{5,6,29} In addition to STAS, peritumoral radiomics, which analyzes the regions surrounding the tumor, and the tumor microenvironment, including immune cells, blood vessels, and the extracellular matrix, are also critical factors in lung cancer progression and prognosis.

These elements provide additional insights into tumor behavior and help refine lung cancer classification, aiding in more accurate prognostication and personalized treatment planning. Second, the detection of STAS also provides important references for clinical treatment decisions. Depending on the extent and range of STAS, physicians can more accurately assess the invasiveness and metastatic risk of lung cancer and formulate corresponding treatment plans. For example, for patients with lung cancer with positive STAS, more extensive surgical resection is often required to ensure complete tumor clearance and reduce the risk of postoperative recurrence.³⁰ Therefore, accurate identification and assessment of the presence of STAS are of great significance for guiding the scope and depth of surgery, improving the thoroughness of surgery and treatment outcomes.

Application of machine learning and radiomics in diagnosing lung cancer spread through air spaces

ML algorithms have increasingly been applied in the diagnosis and treatment of lung cancer, revolutionizing the field by enabling more accurate and efficient analysis of complex datasets. Techniques such as convolutional neural networks and recurrent neural networks have shown significant promise in enhancing the accuracy of lung cancer detection, prognostication, and classification. While the broader scope of AI in lung cancer therapeutics includes applications such as drug discovery and treatment personalization, this discussion specifically focuses on the role of radiomics in improving diagnostic accuracy and clinical decision-making.

A comprehensive review highlighted the critical role of AI in analyzing extensive clinical datasets to improve patient management strategies.³¹ These advancements can optimize therapeutic approaches and potentially enhance patient outcomes in lung cancer care.³² However, the specific application of radiomics—defined as the extraction of quantitative features from medical imaging—plays a pivotal role in distinguishing between various tumor characteristics and predicting clinical outcomes.

The present study's findings on the diagnostic value of radiomics analysis for lung cancer STAS align well with these broader applications of ML in oncology. Specifically, our results demonstrate that radiomics, powered by ML algorithms, exhibits good

clinical applicability of ML techniques in lung cancer diagnosis and treatment.

The synergy between ML algorithms, comprehensive datasets, and sophisticated feature engineering techniques holds great promise for advancing lung cancer diagnostics and therapeutics. The present study contributes to this growing body of evidence, underscoring the potential of radiomics in accurately diagnosing lung cancer STAS and supporting personalized treatment strategies. Radiomics analysis can help clinicians more accurately identify patients with positive STAS, thereby assisting in assessing patient prognosis more effectively and formulating rational treatment plans. Additionally, radiomics can provide crucial reference points for surgical planning, helping physicians determine the scope and depth of surgery while reducing the risk of postoperative recurrence. Furthermore, radiomics offers objective and accurate indicators for follow-up and prognosis assessment, enabling timely detection and intervention for changes in patient condition, thereby improving treatment effectiveness and patient survival rates.

Novelty and future directions

To the best of our knowledge, this study is the first meta-analysis of radiomics in the diagnosis of lung cancer STAS. By integrating data from multiple relevant studies, we obtained the most comprehensive and comprehensive data, allowing us to provide a more reliable and objective evaluation of the effectiveness of radiomics in diagnosing STAS. A comprehensive summary of the current research on radiomics in the diagnosis of lung cancer STAS provide important references and inspiration for further research in this field. First, with the continuous development and improvement of radiomics analysis, we can further explore how to improve its accuracy and reliability in diagnosing STAS. It is possible to endeavor to combine more imaging parameters and clinical data to build more complex and comprehensive predictive models, thereby improving the diagnostic accuracy of STAS. In addition, the fusion and integration of multi-modal imaging data can be explored to further improve the diagnostic ability of STAS through various imaging techniques.³⁴ Second, the application of radiomics in the treatment of lung cancer STAS can be further studied. In addition to diagnosis, radiomics analysis can also be used to evaluate patients' treatment

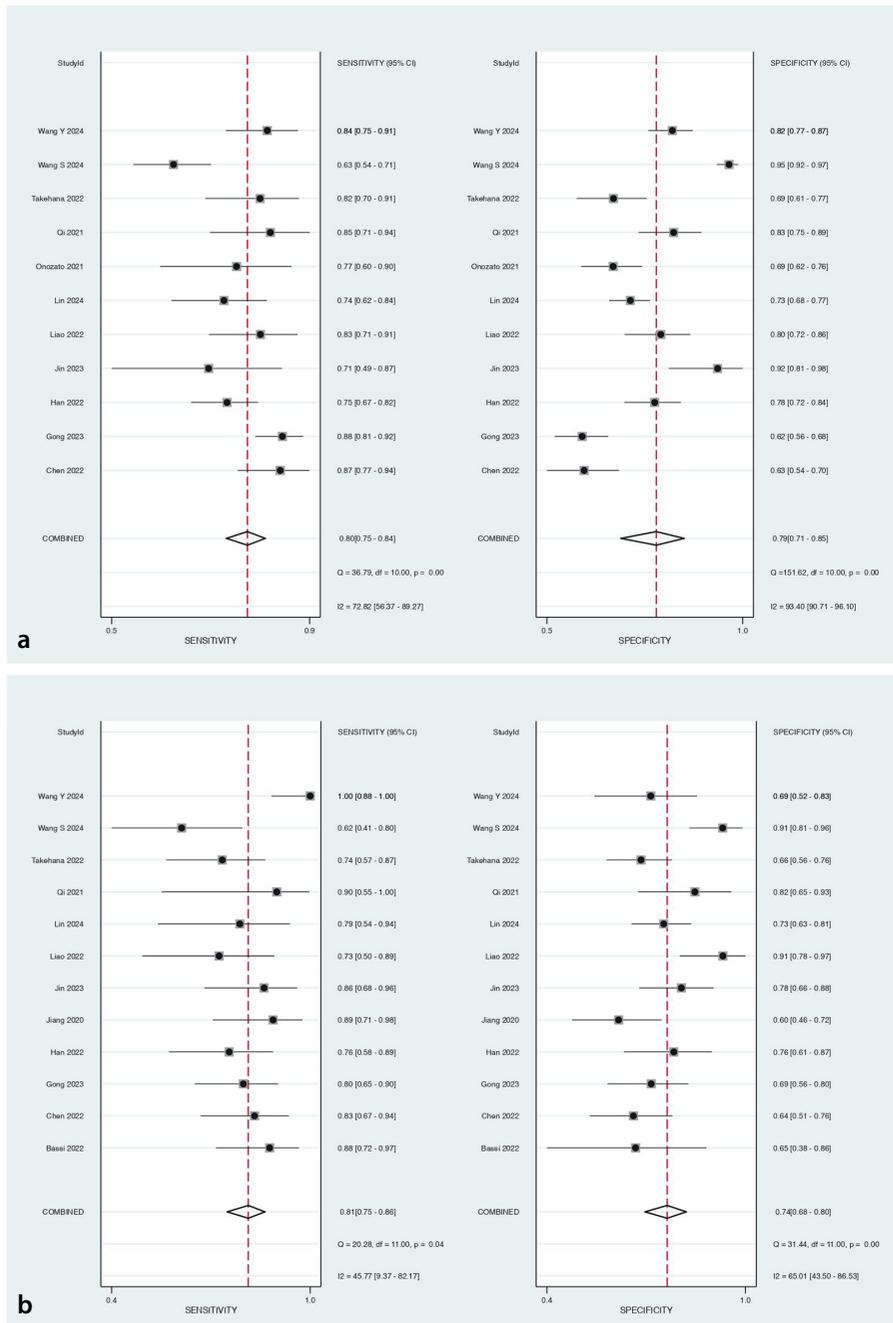


Figure 3. Forest plots of radiomics for prediction of STAS in training (a) and validation (b) cohorts. STAS, spread through air spaces.

sensitivity and specificity in diagnosing lung cancer STAS. This capability is critical as it enables accurate distinctions between patient groups, facilitating the formulation of personalized treatment plans.

By employing advanced feature engineering techniques, as utilized in this study, radiomics can extract and analyze intricate imaging features that are otherwise imperceptible to human observers.³³ These features include tumor shape, grayscale variations, texture patterns, and wavelet transformations, which collectively contribute to accurately predicting STAS status. For example, this

study observed AUC values ranging from 0.66 to 0.99 for different models, reflecting the robust discriminative ability of radiomics in diagnosing lung cancer STAS.

Integrating clinical information, such as age, gender, smoking status, and tumor size, into radiomics models enhances their predictive performance. This approach aligns with the review's emphasis on leveraging large, multifaceted datasets to improve treatment recommendations and outcomes.³¹ Future research should focus on addressing current limitations, such as selection bias and lack of blinding, to further validate and enhance the

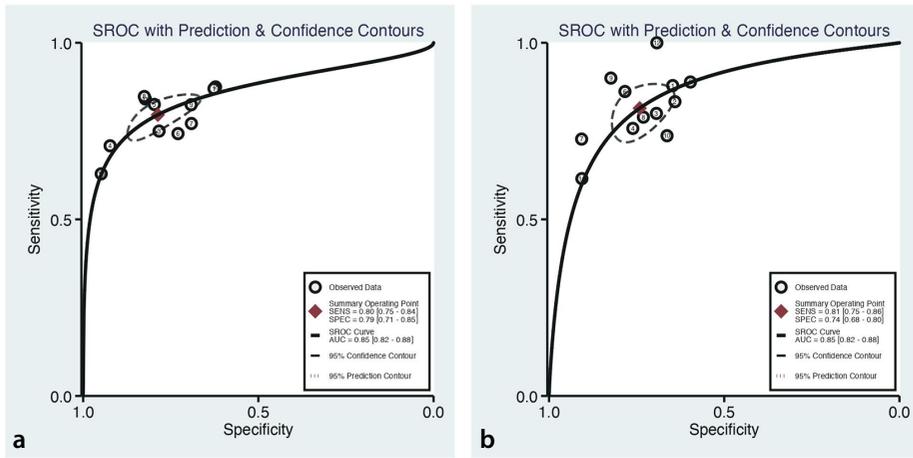


Figure 4. Summary receiver operating curves of radiomics for prediction of STAS in training (a) and validation (b) cohorts. STAS, spread through air spaces.

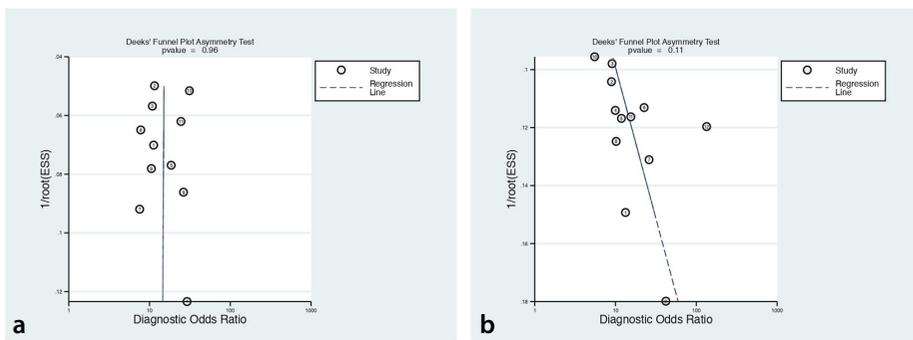


Figure 5. Funnel plots of radiomics for prediction of STAS in training (a) and validation (b) cohorts. STAS, spread through air spaces.

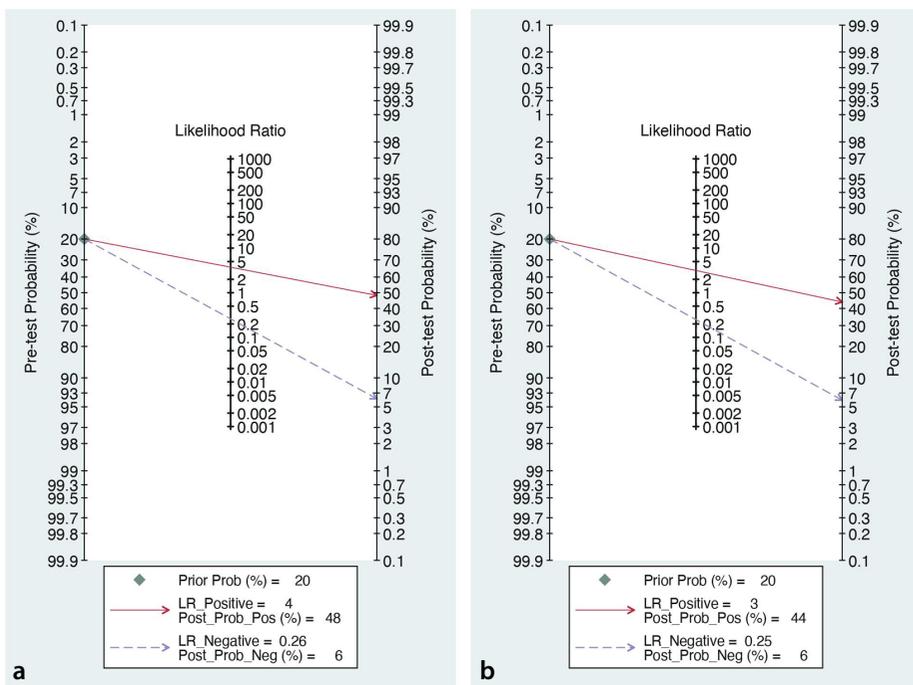


Figure 6. Fagan plots of radiomics for prediction of STAS in training (a) and validation (b) cohorts. STAS, spread through air spaces.

response and prognosis, guiding the formulation and adjustment of treatment plans. As such, future research can focus on the appli-

cation of radiomics in the treatment decision-making and efficacy evaluation of lung cancer STAS. In addition, the combined ap-

plication of radiomics with other fields, such as genomics and transcriptomics, can be further explored. By integrating various omics data, a multi-angle and comprehensive evaluation of lung cancer STAS can be achieved, providing more comprehensive and personalized treatment plans for clinical practice.³⁵ Finally, larger-scale, longer-term prospective studies can be conducted to verify the effectiveness and clinical application prospects of radiomics in the diagnosis and treatment of lung cancer STAS. These studies will provide important scientific evidence and clinical support for the further promotion and application of radiomics analysis in the diagnosis and treatment of lung cancer.

Although this study provides important insights and inspiration for the diagnosis of STAS via radiomics, there are also some potential limitations that need to be considered. First, all studies included in this study were retrospective in design. Although this is a common design in radiological research, there may still be a risk of selection bias. Due to the characteristics of retrospective studies, the study results may be affected by patient selection and data collection and may not fully represent the entire population. Second, many studies did not set up blinding when analyzing images or pathology. The lack of blinding may lead to subjective bias of observers, affecting the accuracy and reliability of the study results. To reduce the risk of bias, future studies should adopt double-blind or single-blind designs to ensure that researchers are unaware of the data analysis. In addition, more than half of the included studies were conducted in China, and the universality of the results may be limited by geographical restrictions, and it is uncertain whether they are applicable to other populations. Therefore, future research should diversify the selection of study samples to ensure the reliability and universality of the study results.

In conclusion, this study confirms the potential clinical utility of ML-assisted radiomics for diagnosing STAS of lung cancer. Our analysis reveals that radiomics analysis achieves good sensitivity, specificity, and discriminative ability across multiple cohorts. Although publication bias was not significant, some studies showed uncertain risk of bias, which should be addressed in future research. Radiomics analysis offers a valuable tool for clinical decision-making in STAS diagnosis, but limitations such as selection bias, lack of blinding, and geographical restrictions must be carefully considered. Future studies should focus on mitigating these limitations

to improve the robustness and generalizability of radiomics in clinical settings.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was supported by Suzhou scientific and technological plan project (SKY2023064); Nanjing medical university science and technology development fund project (NMUB20230217); Gusu school research project (GSKY20210209); Suzhou key diseases project (LCZX202212).

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-263. [\[CrossRef\]](#)
2. Kadota K, Nitadori JI, Sima CS, et al. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J Thorac Oncol.* 2015;10(5):806-814. [\[CrossRef\]](#)
3. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol.* 2015;10(9):1243-1260. [\[CrossRef\]](#)
4. Chen Z, Wu X, Fang T, et al. Prognostic impact of tumor spread through air spaces for T2aN0 stage IB non-small cell lung cancer. *Cancer Med.* 2023;12(14):15246-15255. [\[CrossRef\]](#)
5. Han YB, Kim H, Mino-Kenudson M, et al. Tumor spread through air spaces (STAS): prognostic significance of grading in non-small cell lung cancer. *Mod Pathol.* 2021;34(3):549-561. [\[CrossRef\]](#)
6. Mino-Kenudson M. Significance of tumor spread through air spaces (STAS) in lung cancer from the pathologist perspective. *Transl Lung Cancer Res.* 2020;9(3):847-859. [\[CrossRef\]](#)
7. Shiono S, Endo M, Suzuki K, Hayasaka K, Yanagawa N. Spread through air spaces in lung cancer patients is a risk factor for pulmonary metastasis after surgery. *J Thorac Dis.* 2019;11(1):177-187. [\[CrossRef\]](#)
8. Chen M, Copley SJ, Viola P, Lu H, Aboagye EO. Radiomics and artificial intelligence for precision medicine in lung cancer treatment. *Semin Cancer Biol.* 2023;93:97-113. [\[CrossRef\]](#)
9. Wang Y, Lyu D, Fan L, Liu S. Advances in the prediction of spread through air spaces with imaging in lung cancer: a narrative review. *Transl Cancer Res.* 2023;12(3):624-630. [\[CrossRef\]](#)
10. McInnes MDF, Moher D, Thoms BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA.* 2018;319(4):388-396. [\[CrossRef\]](#)
11. Bassi M, Russomando A, Vannucci J, et al. Role of radiomics in predicting lung cancer spread through air spaces in a heterogeneous dataset. *Transl Lung Cancer Res.* 2022;11(4):560-571. [\[CrossRef\]](#)
12. Chen D, She Y, Wang T, et al. Radiomics-based prediction for tumour spread through air spaces in stage I lung adenocarcinoma using machine learning. *Eur J Cardiothorac Surg.* 2020;58(1):51-58. [\[CrossRef\]](#)
13. Chen LW, Lin MW, Hsieh MS, et al. Radiomic values from high-grade subtypes to predict spread through air spaces in lung adenocarcinoma. *Ann Thorac Surg.* 2022;114(3):999-1006. [\[CrossRef\]](#)
14. Gong J, Yin R, Sun L, et al. CT-based radiomics model to predict spread through air space in resectable lung cancer. *Cancer Med.* 2023;12(18):18755-18766. [\[CrossRef\]](#)
15. Han X, Fan J, Zheng Y, et al. The value of CT-based radiomics for predicting spread through air spaces in stage IA lung adenocarcinoma. *Front Oncol.* 2022;12:757389. [\[CrossRef\]](#)
16. Jiang C, Luo Y, Yuan J, et al. CT-based radiomics and machine learning to predict spread through air space in lung adenocarcinoma. *Eur Radiol.* 2020;30(7):4050-4057. [\[CrossRef\]](#)
17. Jin W, Shen L, Tian Y, et al. Improving the prediction of spreading through air spaces (STAS) in primary lung cancer with a dynamic dual-delta hybrid machine learning model: a multicenter cohort study. *Biomark Res.* 2023;11(1):102. [\[CrossRef\]](#)
18. Liao G, Huang L, Wu S, et al. Preoperative CT-based peritumoral and tumoral radiomic features prediction for tumor spread through air spaces in clinical stage I lung adenocarcinoma. *Lung Cancer.* 2022;163:87-95. [\[CrossRef\]](#)
19. Lin MW, Chen LW, Yang SM, et al. CT-based deep-learning model for spread-through-air-spaces prediction in ground glass-predominant lung adenocarcinoma. *Ann Surg Oncol.* 2024;31(3):1536-1545. [\[CrossRef\]](#)
20. Liu Q, Qi W, Wu Y, Zhou Y, Huang Z. Construction of pulmonary nodule CT radiomics random forest model based on artificial intelligence software for STAS evaluation of stage IA lung adenocarcinoma. *Comput Math Methods Med.* 2022;2022:2173412. [\[CrossRef\]](#)
21. Onozato Y, Nakajima T, Yokota H, et al. Radiomics is feasible for prediction of spread through air spaces in patients with non-small cell lung cancer. *Sci Rep.* 2021;11(1):13526. [\[CrossRef\]](#)
22. Qi L, Li X, He L, et al. Comparison of diagnostic performance of spread through airspaces of lung adenocarcinoma based on morphological analysis and perinodular and intranodular radiomic features on chest CT images. *Front Oncol.* 2021;11:654413. [\[CrossRef\]](#)
23. Suh YJ, Han K, Kwon Y, et al. Computed Tomography Radiomics for preoperative prediction of spread through air spaces in the early stage of surgically resected lung adenocarcinomas. *Yonsei Med J.* 2024;65(3):163-173. [\[CrossRef\]](#)
24. Takehana K, Sakamoto R, Fujimoto K, et al. Peritumoral radiomics features on preoperative thin-slice CT images can predict the spread through air spaces of lung adenocarcinoma. *Sci Rep.* 2022;12(1):10323. [\[CrossRef\]](#)
25. Tao J, Liang C, Yin K, et al. 3D convolutional neural network model from contrast-enhanced CT to predict spread through air spaces in non-small cell lung cancer. *Diagn Interv Imaging.* 2022;103(11):535-544. [\[CrossRef\]](#)
26. Wang S, Liu X, Jiang C, et al. CT-based super-resolution deep learning models with attention mechanisms for predicting spread through air spaces of solid or part-solid lung adenocarcinoma. *Acad Radiol.* 2024;31(6):2601-2609. [\[CrossRef\]](#)
27. Wang Y, Lyu D, Hu L, et al. CT-based intratumoral and peritumoral radiomics nomograms for the preoperative prediction of spread through air spaces in clinical stage IA non-small cell lung cancer. *J Imaging Inform Med.* 2024;37(2):520-535. [\[CrossRef\]](#)
28. Zhuo Y, Feng M, Yang S, et al. Radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of spread through air spaces in lung adenocarcinoma. *Transl Oncol.* 2020;13(10):100820. [\[CrossRef\]](#)
29. Lu S, Tan KS, Kadota K, et al. Spread through air spaces (STAS) is an independent predictor of recurrence and lung cancer-specific death in squamous cell carcinoma. *J Thorac Oncol.* 2017;12(2):223-234. [\[CrossRef\]](#)
30. Toki MI, Harrington K, Syrigos KN. The role of spread through air spaces (STAS) in lung adenocarcinoma prognosis and therapeutic decision making. *Lung Cancer.* 2020;146:127-133. [\[CrossRef\]](#)
31. Ahmad S, Raza K. An extensive review on lung cancer therapeutics using machine learning techniques: state-of-the-art and perspectives. *J Drug Target.* 2024;32(6):635-646. [\[CrossRef\]](#)
32. Caii W, Wu X, Guo K, Chen Y, Shi Y, Chen J. Integration of deep learning and habitat radiomics for predicting the response to immunotherapy in NSCLC patients. *Cancer*

- Immunol Immunother.* 2024;73(8):153. [CrossRef]
33. Anan N, Zainon R, Tamal M. A review on advances in ¹⁸F-FDG PET/CT radiomics standardisation and application in lung disease management. *Insights Imaging.* 2022;13(1):22. [CrossRef]
34. Nishimori M, Iwasa H, Miyatake K, et al. 18F FDG-PET/CT analysis of spread through air spaces (STAS) in clinical stage I lung adenocarcinoma. *Ann Nucl Med.* 2022;36(10):897-903. [CrossRef]
35. Ye R, Yu Y, Zhao R, Han Y, Lu S. Comprehensive molecular characterizations of stage I-III lung adenocarcinoma with tumor spread through air spaces. *Front Genet.* 2023;14:1101443. [CrossRef]

Supplementary File 1.

Search strategy

PubMed

((radiomics[MeSH] OR radiomics) AND (artificial intelligence[MeSH] OR AI OR artificial intelligence OR machine learning[MeSH] OR ML OR machine learning) AND (lung cancer[MeSH] OR lung neoplasms OR pulmonary cancer) AND (spread through air spaces[MeSH] OR STAS OR airway spread))

Embase

('radiomics'/exp OR radiomics) AND ('artificial intelligence'/exp OR AI OR 'artificial intelligence' OR 'machine learning'/exp OR ML OR 'machine learning') AND ('lung cancer'/exp OR 'lung neoplasms' OR 'pulmonary cancer') AND ('spread through air spaces'/exp OR STAS OR 'airway spread')

CENTRAL

((radiomics OR "radiomics") AND (artificial intelligence OR AI OR "artificial intelligence" OR machine learning OR ML OR "machine learning") AND (lung cancer OR "lung neoplasms" OR "pulmonary cancer") AND (spread through air spaces OR STAS OR "airway spread"))



Dual-energy computed tomography-based volumetric thyroid iodine quantification: correlation with thyroid hormonal status, pathologic diagnosis, and phantom validation

Younghen Lee

Korea University College of Medicine, Ansan Hospital,
Department of Radiology, Ansan, South Korea

PURPOSE

To investigate the relationship between intrathyroidal iodine concentration (IC) (mg I/mL) and thyroid hormonal status or pathologic diagnosis with the use of dual-energy computed tomography (DECT).

METHODS

We retrospectively included patients who underwent neck CT examination between September 2016 and August 2021 using a dual-layer DECT scanner (120 kilovolt peak) for preoperative thyroid imaging. We performed volumetric IC measurements at the thyroid parenchyma on the additional iodine map generated from non-contrast images. We then compared the mean IC of thyroid parenchyma based on thyroid hormonal status (hypothyroid, euthyroid, and hyperthyroid) and diffuse thyroid disease (DTD). Additionally, we determined the accuracy of iodine quantification with our site-specific DECT acquisition protocol using a Gammex™ phantom containing seven iodine inserts with different ICs ranging from 2 to 20 mgI/mL.

RESULTS

Among the 578 patients (M:F: 87:491, age: 48.6 ± 11.7 years) who were finally selected, the mean thyroid parenchymal ICs was the lowest in the hyperthyroid group, followed by the hypothyroid group, and then the euthyroid group (0.68 ± 0.37 , $n = 44$ vs. 1.13 ± 0.42 , $n = 61$ vs. 1.32 ± 0.43 , $n = 473$, $P < 0.01$, respectively). In the patients with euthyroidism, the mean parenchymal IC was already lower in the patients with pathologically proven DTD than in those without DTD (1.22 ± 0.44 mgI/mL vs. 1.45 ± 0.37 mgI/mL, $P < 0.01$). Based on the phantom study, the median percentage deviations from the expected values were 5.1% for ICs of 2–20 mgI/mL.

CONCLUSION

DECT-based IC quantification could be a potentially useful method for identifying patients with thyroid hormone dysfunction or DTD without the use of contrast media.

CLINICAL SIGNIFICANCE

Without the need for intravenous administration, DECT-based intrathyroidal IC quantification provides potentially valuable information from the non-contrast CT image of the thyroid parenchyma

KEYWORDS

Thyroid gland, dual-energy computed tomography, iodine concentration, hyperthyroidism, hypothyroidism, hormone

Corresponding author: Younghen Lee

E-mail: younghen@korea.ac.kr

Received 21 November 2024; revision requested 16 January 2025; accepted 04 February 2025.



Epub: 18.03.2025

Publication date: 28.04.2025

DOI: 10.4274/dir.2025.243132

In non-contrast computed tomography (CT) imaging, the normal thyroid gland, which harbors approximately 70%–80% of the body's iodine in a healthy adult,¹ exhibits a distinctive increase in CT attenuation in comparison with other anatomical structures. In contrast, patients presenting with thyroid hormone dysfunction or underlying thyroiditis exhibit a pronounced reduction in CT attenuation, suggestive of inflammatory cell infiltration within the thyroid gland.^{2–6} It is crucial to acknowledge that CT attenuation is not a fixed measurement; rather, it is susceptible to fluctuations influenced by technical parameters, including the type

You may cite this article as: Lee Y. Dual-energy computed tomography-based volumetric thyroid iodine quantification: correlation with thyroid hormonal status, pathologic diagnosis, and phantom validation. *Diagn Interv Radiol.* 2025;31(3):226-233.

of scanner employed, tube potential [kilovolt peak (kVp)], and patient positioning.^{7,8} These factors can markedly amplify the observed changes in the thyroid gland, which is particularly rich in iodine. Moreover, X-ray fluorescence spectrometry has emerged as a novel technique for evaluating the intrathyroidal iodine pool.^{9,10} However, it is limited to the analysis of formalin-fixed samples and, as a result, is not applicable for individual patient assessments.

Previous research utilizing dual-energy CT (DECT) has shown the ability to directly measure iodine concentration (IC) from individual voxels. This advancement has enhanced diagnostic performance in oncological imaging by facilitating the prediction of pathological subtypes, assessing post-treatment responses, and determining prognosis.¹¹⁻¹³ Recent studies have suggested that measuring IC obtained through DECT could be advantageous in the treatment of patients with thyroid hormone disorders.¹⁴⁻¹⁶ However, these studies did not sufficiently consider the impact of thyroid hormone status or the technical parameters associated with DECT. Consequently, the present study investigates the feasibility of quantifying iodine from non-contrast thyroid CT images of patients with established thyroid hormone status and/or histopathological diagnoses, in addition to using a CT phantom containing iodine inserts, through the application of recently developed dual-layer detector DECT technology.

Methods

This study was approved by the Korea University Ansan Hospital Institutional Review Board and the requirement of informed consent was waived due to its retrospective chart review and image analysis (approval no: AS20180180, date: September 2023).

Main points

- Dual-energy computed tomography (CT) can quantify the inherent thyroid iodine status using the non-contrast CT images.
- The intrathyroidal iodine concentration (IC) was observed to be significantly reduced in patients exhibiting hypo- or hyperthyroidism in comparison with those with euthyroidism, with a reference range determined to be 1.32 ± 0.43 mgI/mL.
- Significantly decreased intrathyroidal IC was observed in patients with pathologically confirmed diffuse thyroid disease, even in those who are euthyroid.

Patient population

To evaluate the parenchymal IC of the entire thyroid gland using non-contrast CT images in relation to thyroid function and pathological findings, we conducted a retrospective analysis of the medical records of 1,365 patients who underwent neck CT examinations using a DECT scanner at our institution between September 2016 and August 2021. The criteria for inclusion in this study were as follows: (a) the entire thyroid gland was imaged without the use of contrast media during the neck CT examination; (b) thyroid hormone levels, including free thyroxine (fT4) (ng/dL), triiodothyronine (T3) (ng/dL), and thyroid-stimulating hormone (TSH) (uIU/mL), were measured within 1 month prior to the CT examination; (c) any thyroid nodule present had a diameter of ≤ 10 mm as determined by sonographic imaging or final surgical specimens, if surgery was performed; (d) patients had no prior history of medication that could potentially influence thyroid function test results, such as dopamine agonists, glucocorticoids, somatostatin analogues, metformin, salicylates, phenytoin, lithium, and tyrosine kinase inhibitors; and (e) the overall quality of the CT images was not significantly compromised by beam-hardening artifacts, primarily associated with adjacent bony structures.

Finally, a total of 578 patients (M:F: 87:491, mean age: 48.6 ± 11.7 years) were included for DECT-based IC quantification. Comprehensive patient characteristics are presented in Table 1 and Figure 1. The primary indication for neck CT in these patients was preoperative planning, particularly concerning neck lymph node metastasis.¹⁷ As such, non-contrast CT images were routinely acquired before contrast administration to identify calcifications or cystic change in neck lymph nodes, which are well-established CT indicators of metastatic thyroid carcinoma.¹⁷

According to the TSH results, 61, 473, and 44 patients were classified as hypothyroid, euthyroid, and hyperthyroid, respectively. The mean age and sex distribution did not differ across the three groups (all $P > 0.05$). The TSH levels were significantly different across the three groups (all $P < 0.01$), whereas a significant difference existed only between the hyperthyroid group and the other two groups for the fT4 and T3 levels (all $P < 0.01$). Among the 521 patients who underwent total thyroidectomy, subtotal thyroidectomy, or hemithyroidectomy for the removal of thyroid nodules with cytological diagnoses or to alleviate symptoms associated with hyperthyroidism, 318 patients were pathologically diagnosed with diffuse thyroid disease (DTD).

Data acquisition and iodine quantification

Non-contrast CT images utilized for the quantification were obtained using a single-source, dual-layer detector CT system (Philips Healthcare, IQon Spectral CT scanner, Cleveland, OH, USA). The imaging parameters employed included a tube voltage of 120 kVp, a mean tube current of 42 mAs (with a range of 30–150 mAs), a collimation thickness of 64×0.625 mm, a helical pitch of 0.609, a rotation time of 0.75 seconds per rotation, and a field of view measuring 350 mm. The scan encompassed a range from the skull base to the aortopulmonary window. An automatic tube current modulation system was engaged during the scanning process along both the X–Y and Z axes (DoseRight 3D-DOM; Philips Healthcare).

All acquired images were subsequently transferred to and analyzed using specialized DECT postprocessing software (Spectral Diagnostic Suite version 6.5, Philips Healthcare) to generate an iodine-specific density map through projection-based material

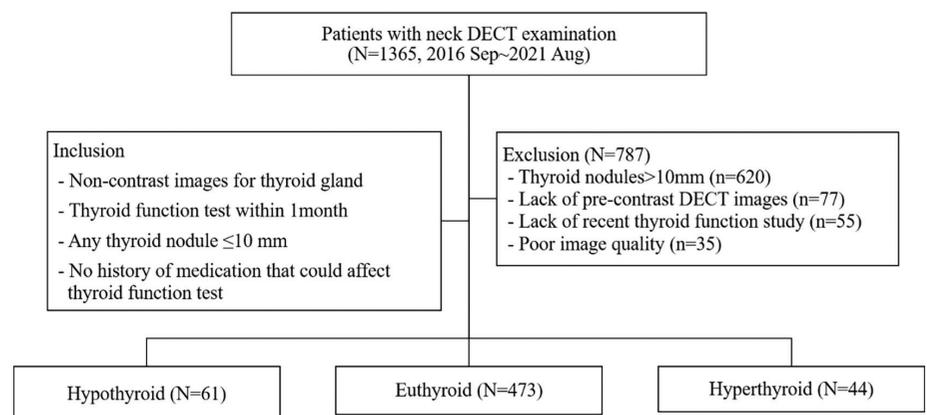


Figure 1. Flowchart of included patients with dual-energy computed tomography-based thyroid gland iodine quantification. DECT, dual-energy computed tomography.

Table 1. Clinical characteristics of the patients

	Hypothyroid	Euthyroid	Hyperthyroid	P value
Numbers of patients (men: women)	61 (9:52)	473 (72:401)	44 (6:38)	>0.05**
Age (years)	50.4 ± 10.3	48.3 ± 11.4	48.5 ± 15.8	>0.05**
Thyroid-stimulating hormone (uIU/mL, normal range: 0.17–4.05)	5.94 ± 1.44	1.74 ± 0.83	0.04 ± 0.04	<0.01**
Free thyroxine (ng/dL, normal range: 0.79–1.86)	1.21 ± 0.17	1.29 ± 0.15	1.56 ± 0.56	<0.01**
Triiodothyronine (ng/dL, normal range: 78–182)	116.4 ± 33.6	115.8 ± 32.8	150.5 ± 56.6	<0.01**
Numbers of pathologic diagnosis for thyroid parenchyma (no pathologic results: normal: diffuse thyroid disease)	6:14:41	36:189:248	15:0:29	

Data shown as mean ± standard deviation. The presented P values resulted from comparison between three groups with different thyroid hormonal status, using chi-squared tests for categorical variables and one-way analysis of variance tests for continuous variables. *, hypothyroid vs. euthyroid; †, hypothyroid vs. hyperthyroid; ‡, euthyroid vs. hyperthyroid.

decomposition.⁷ Conventional images were reconstructed employing a hybrid iterative reconstruction algorithm (iDose 4, level 3) alongside a soft tissue algorithm [window level set at 60 Hounsfield unit (HU); window width at 300 HU], with a slice thickness of 2 mm.

For the volumetric assessment of CT attenuation values and IC in the thyroid glands, a board-certified radiologist with 19 years of expertise in head and neck imaging utilized a semi-automated region-growing technique to encompass the entire thyroid gland on 40-kiloelectronvolt (keV) virtual monoenergetic images, as this kiloelectron volt level significantly enhances intrathyroidal iodine CT attenuation.^{7,8} Additional sets of iodine density maps and conventional images were

employed to derive IC and CT attenuation values (Figure 2). The total thyroid volume, inclusive of both lobes and the isthmus, was calculated. Measurements were conducted three times using the vendor's proprietary image viewer (IntelliSpace Portal v9; Philips Healthcare) and subsequently averaged.

Hormone status and pathologic analysis

The thyroid hormone status of patients was categorized into three distinct groups: hypothyroid, euthyroid, and hyperthyroid, based on recent serum levels of TSH and fT4. The reference ranges utilized by our institution were as follows: TSH levels between 0.17 and 4.05 uIU/mL and fT4 levels between 0.85 and 1.86 ng/dL. TSH is recognized as the most effective screening mark-

er for thyroid function due to its enhanced sensitivity and specificity relative to other hormonal assessments.¹⁸ Accordingly, in this study, patients were classified into the hypothyroid, euthyroid, and hyperthyroid groups based on their TSH levels being above the upper limit, within the reference range, or below the lower limit of the reference range, respectively. Additionally, for patients who underwent thyroid surgery at our institution, we documented the pathological subtypes of thyroid nodules and the presence of DTD, which encompasses a diverse array of autoimmune inflammatory or hyperplastic conditions, including Hashimoto's thyroiditis (HT) and nodular or diffuse hyperplasia.¹⁹

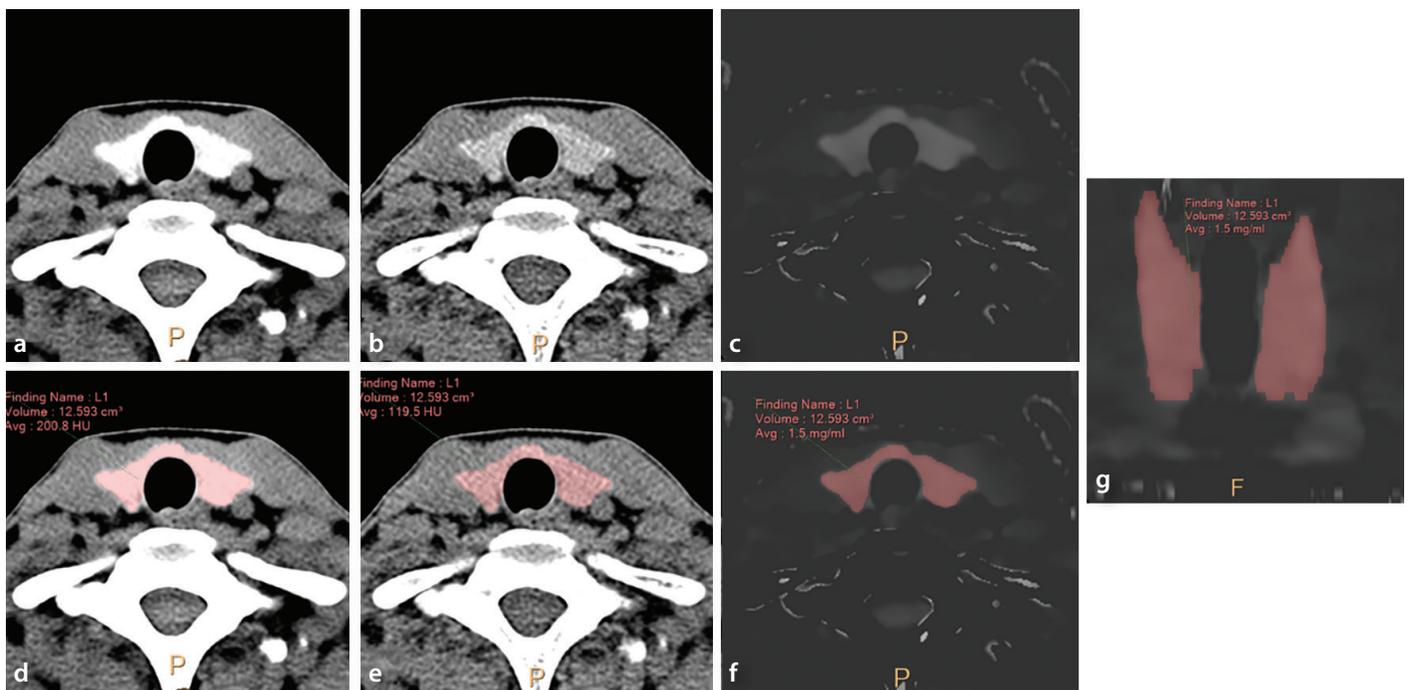


Figure 2. Example of thyroid parenchymal iodine concentration (IC) measurement. Initially, the volumetric region of interest encompassing the entire thyroid glands was drawn on the axial 40-keV virtual monoenergetic images (a, d) using the semi-automated region-growing method. Then, both mean CT attenuation value and IC were obtained from the corresponding iodine density map (c, f, g) and conventional CT images (b, e). Total thyroid volume was also recorded. keV, kiloelectronvolt; CT, computed tomography.

Statistical analysis

Categorical variables represented numerically were analyzed across the three groups (euthyroid, hypothyroid, and hyperthyroid) using the chi-squared test. Continuous variables were reported as mean \pm standard deviation or median (interquartile range). To assess the differences in parenchymal CT attenuation, IC, and total thyroid volume among the three distinct thyroid hormone groups, one-way analysis of variance (ANOVA) was conducted following the verification of normal distribution through the Shapiro–Wilk test. The Bonferroni correction was applied to account for multiple comparisons. Additionally, a subgroup analysis was performed on the euthyroid group to evaluate the presence or absence of DTD in their surgical specimens. The relationship between continuous variables (IC, CT attenuation, age, FT4, and TSH) was examined using bivariate Pearson or Spearman correlation analysis. Data analysis was conducted using IBM SPSS Statistics 25.0 (Armonk, NY, USA), with a *P* value of <0.05 deemed indicative of a statistically significant difference.

Phantom study

To evaluate the quantitative accuracy of the DECT protocol for IC calculation, we employed an electron density phantom (Gammex™ 467, Gammex Inc., Middleton, WI, USA) that contained iodine-loaded inserts (2.8 cm in diameter and 7 cm in length) corresponding to seven different ICs (2.0, 2.5,

5.0, 7.5, 10.0, 15.0, and 20.0 mgI/mL), in addition to a solid water insert rod, positioned within the inner circle of the phantom, while the outer ring remained unfilled.^{11,12,20} During the scanning process, the phantom was axially aligned on the scanner table and centered within the gantry to facilitate axial scans of the insert-filled section of the phantom (Figure 3). The CT parameters selected were consistent with our standard protocol for adult neck scans, using a voltage of 120 kVp. Both conventional and spectral base images were generated for subsequent analysis, following the methodology previously described. Circular regions of interest (ROIs) measuring approximately 100 mm² were delineated at the center of each insert across 10 consecutive slices. Each scan was conducted three times, resulting in a total of 30 measurements for each rod. The analysis of measured versus expected values included the calculation of the absolute difference (measured–expected) and the percentage deviation, expressed as $[100 \times (\text{measured} - \text{expected})/\text{expected}]$.

Results

Mean intrathyroidal iodine concentration values in the three different thyroid hormonal groups

The means of CT attenuation values, IC, total volume, and iodine content of the whole thyroid glands are summarized in Table 2. In the group of 473 patients with euthyroidism, the means of CT attenuation values, IC, and total volume of thyroid parenchyma were

103.2 \pm 16.5 HU, 1.32 \pm 0.43 mgI/mL, and 11.7 \pm 11.9 mL, respectively. Based on ANOVA tests, mean intrathyroidal CT attenuation values and IC were significantly different among the three groups (all *P* $<$ 0.01). Mean CT attenuation values and IC were the lowest in the hyperthyroid group, and lower in the hypothyroid group compared with the euthyroid group (0.68 \pm 0.37 mgI/mL, 75.2 \pm 19.3 HU vs. 1.13 \pm 0.42 mgI/mL, 95.1 \pm 18.2 HU, all *P* $<$ 0.01) (Figure 4). Furthermore, the hyperthyroid group presented with a significantly larger thyroid volume compared with both the euthyroid group and the hypothyroid group (all *P* $<$ 0.05); however, no statistically significant difference in volume was observed between the euthyroid and hypothyroid groups (*P* $>$ 0.05).

Difference in intrathyroidal iodine concentration values based on pathologic diagnosis

In the group of 437 patients with euthyroidism who underwent thyroid surgery, the mean thyroid parenchymal IC was found to be significantly lower in those diagnosed with DTD in comparison to those without this diagnosis, as confirmed by pathological assessment (1.22 \pm 0.44 mgI/mL vs. 1.45 \pm 0.37 mgI/mL; *P* = 0.023, Figure 5).

Correlation between intrathyroidal iodine concentration values and other continuous variables

Using Pearson correlation analysis, thyroid parenchymal IC had a statistically sig-



Figure 3. Gammex phantom with different iodine concentrations. To examine the accuracy of iodine quantification using dual-energy computed tomography (CT), iodine concentration increased clockwise from 12 o'clock position (a, b). The zoom-in conventional axial CT image [(c): window level/width 60/300 HU] and iodine density map [(d): window level/width 3/15 mgI/mL] acquired with 120 kVp show the increased signal. kVp, kilovolt peak; HU, Hounsfield unit.

Table 2. Volumetric quantification of entire thyroid glands with different thyroid hormonal status using dual-energy computed tomography datasets

	Hypothyroid (n = 61)	Euthyroid (n = 473)	Hyperthyroid (n = 44)	<i>P</i> value
Computed tomography attenuation (HU)	95.1 \pm 18.2	103.2 \pm 16.5	75.2 \pm 19.3	$<0.01^{*††}$
Iodine concentration (mgI/mL)	1.13 \pm 0.42	1.32 \pm 0.43	0.68 \pm 0.37	$<0.01^{*††}$
Total volume of thyroid gland (mL)	11.2 \pm 7.1	11.7 \pm 11.9	46.3 \pm 59.0	$<0.05^{††}$

Data shown as mean \pm standard deviation. *, hypothyroid vs. euthyroid; †, hypothyroid vs. hyperthyroid; ††, euthyroid vs. hyperthyroid; HU, Hounsfield unit.

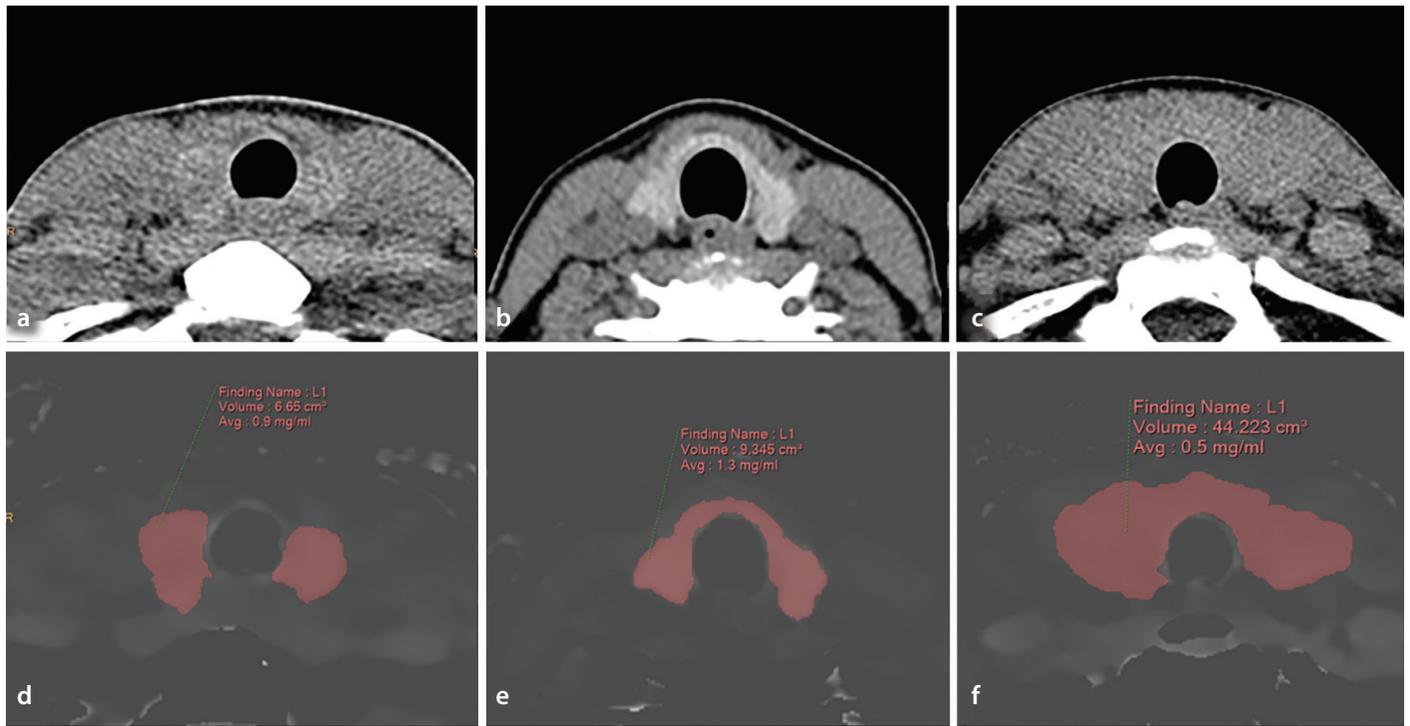


Figure 4. Representative non-contrast computed tomography (CT) images of thyroid gland in the patients with different thyroid hormonal status. The region of interest-indicated iodine density maps (d-f) switched from the upper conventional CT images [(a): hypothyroid, (b): euthyroid, (c): hyperthyroid] show the significantly decreased mean thyroid parenchymal iodine concentration in the patients with hypo-/hyperthyroidism than the euthyroid group (0.9, 0.5 mg/ml vs. 1.3 mg/ml).

nificant strong positive relationship with CT attenuation only ($r = 0.859, P < 0.01$), whereas other variables, including age, fT4, and TSH, had no relationship with thyroid parenchymal IC ($r = -0.033, -0.069, 0.054$, respectively, all $P > 0.05$).

Accuracy of iodine quantification using computed tomography phantom

Examples of conventional CT and IC images of iodine inserts using the same scanning protocol implemented at our institution for the acquisition of non-contrast DECT images are shown in Figure 6. The quantification results for the iodine inserts indicated good agreement between the individual measurement values and the reference values, yielding an overall median difference of 0.36 mg/ml (range: 0.12–0.58 mg/ml) and a mean percentage deviation of 5.1% (range: 2.2%–6.2%) for IC values between 2.0 and 20.0 mg/ml.

Discussion

This study presents the first reference range of intrathyroidal IC of patients with euthyroidism from non-contrast thyroid CT images using DECT-based IC quantification. As anticipated, the mean intrathyroidal IC values were significantly lower in patients with hypo-/hyperthyroidism when compared

with those who were euthyroid. Furthermore, the observed decrease in IC associated with concurrent DTD occurred prior to the onset of thyroid hormonal dysfunction. Consequently, the reference range for intrathyroidal IC provided in this study will enhance the understanding of iodine metabolism in the context of thyroid hormone disorders.

Iodine is a crucial micronutrient required for the synthesis of thyroid hormones.²¹ In healthy adults, the sodium/iodine symporter facilitates the active transport of iodide into the thyroid follicles, operating across a concentration gradient that is 20–50 times greater than that found in plasma. Consequently, a substantial proportion of iodine, estimated at 70%–80% (equivalent to 15–20 mg), is sequestered in colloid pools.¹ These pools exert a predominant effect on the CT attenuation values of the thyroid gland. Previous experimental studies have demonstrated a direct relationship between CT attenuation and IC in solution, supporting the use of intrathyroidal CT attenuation as a functional indicator for the early detection and assessment of thyroid disease severity.^{3–6} However, iodine, an element with a high atomic number ($Z = 53$), exhibits variable CT attenuation depending on the photon energy level, especially just beyond its K-edge (33.2 keV).^{7,8} As a result, the manipulation of iodine CT attenuation can be achieved through adjustments

in kVp. Furthermore, neighboring osseous structures often result in significant beam hardening artefacts in the vicinity of the thyroid gland, impeding the precision of CT attenuation.

To overcome these problems in CT attenuation measurement, several researchers have advocated for DECT-based IC quantification as a more precise and dependable diagnostic approach for evaluating thyroid hormone levels.^{14–16} Binh et al.¹⁴ were the first to propose the use of intrathyroidal IC as a functional marker in patients with hyperthyroidism, demonstrating a significant negative correlation between iodine 123 uptake at 3 hours and intrathyroidal IC ($r = -0.680, P < 0.05$). However, no correlation was found between iodine 123 uptake and CT attenuation values ($P = 0.087$).¹⁴ In a study involving 226 patients with euthyroidism, it was observed that intrathyroidal IC was significantly lower in men compared with women and exhibited a gradual decline in patients over the age of 40.¹⁵ Another investigation involving patients with euthyroidism revealed that mean IC and total iodine content were positively correlated with fT3/total-T3 levels, but exhibited a negative correlation with TSH.¹⁶ However, prior studies measured intrathyroidal IC values at only two or three slices of each lobe, averaging the results without assessing the entire thyroid volume.^{14–16} Ad-

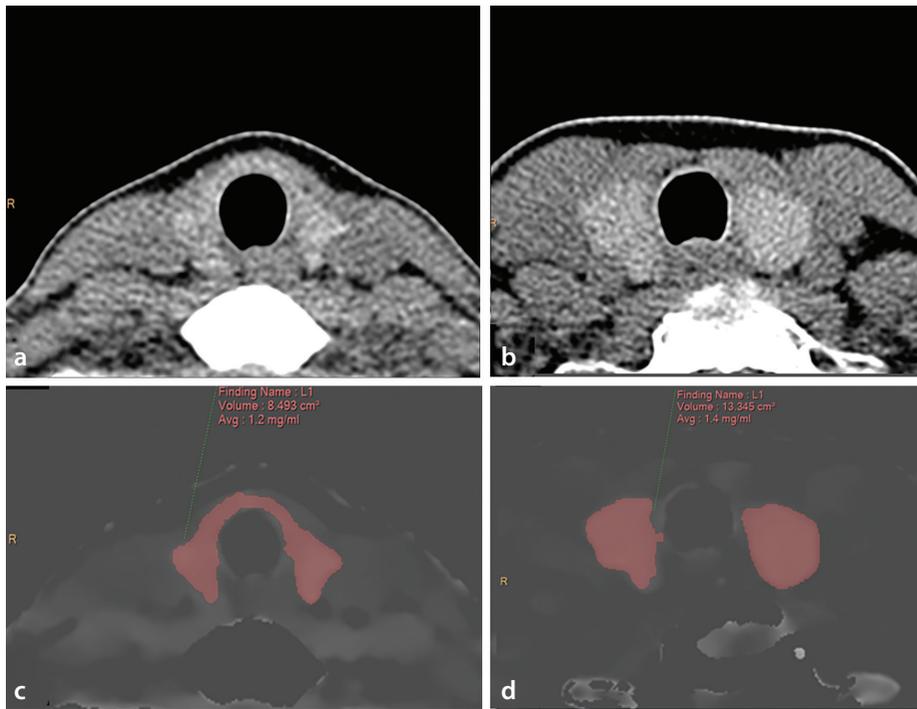


Figure 5. Non-contrast computed tomography (CT) image characteristics in the euthyroid group patients with and without diffuse thyroid disease (DTD). Compared with the conventional CT images of thyroid gland (a, b), the reduced intrathyroidal iodine concentration and total thyroid volume were confirmed in patients with DTD by iodine quantification [(c): 1.2 mg/ml with 8.5 mL vs. (d): 1.4 mg/ml with 13.3 mL].

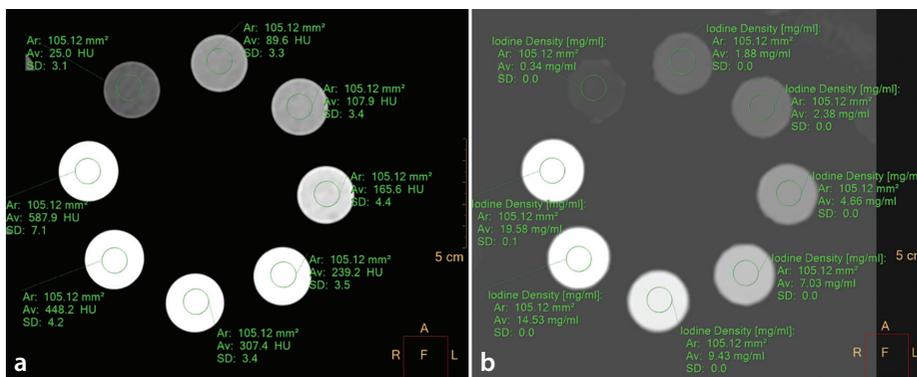


Figure 6. Example of iodine quantification with iodine concentration phantom. For 2–20 mg/mL concentrations, measured mean computed tomography (CT) attenuation and iodine concentration were documented adjacent to each region of interest on the conventional CT image [(a): window level/width 60/300 HU] and corresponding iodine density map [(b), window level/width 3/15 mg/mL]. HU, Hounsfield unit.

ditionally, the sample size in these studies ($n = 43$) was relatively small, limiting the determination of correlation coefficients. Furthermore, the DECT-based IC quantification methods lacked validation, especially considering the significant variability that can result from technical factors such as scanner type and body diameter.^{22,23} In contrast, the present study used volumetric quantification in a larger cohort of patients with confirmed hormonal and pathologic diagnosis. As a result, the accuracy of this approach was confirmed in a phantom study, suggesting that the results obtained are more reliable than those reported in previous investigations.^{14–16}

This study appears to be the first volumetric analysis of intrathyroidal IC across three distinct groups defined by differing thyroid hormone levels: hypothyroid, euthyroid, and hyperthyroid. Notably, the results show a statistically significant greater mean reduction in IC in patients with hyperthyroidism compared with those with hypothyroidism. Although the underlying mechanisms for these observed differences remain unclear, several plausible hypotheses can be proposed. For example, Graves' disease (GD), the predominant etiology of hyperthyroidism, is characterized by a two- to threefold increase in thyroid gland weight, accompanied by

markedly increased vascularity, as evidenced by immunohistochemical evaluations of surgical specimens.²⁴ In GD, the interstitial tissue of the thyroid gland is typically infiltrated by lymphocytes, which rapidly lyse thyroglobulin, resulting in thyroid follicular shrinkage or collapse due to colloid deficiency. Conversely, in patients with HT, the lymphocytic infiltration persistently damages the thyroid follicles, culminating in permanent hypothyroidism following an initial phase of hyperthyroidism.²⁴ As HT advances, the thyroid gland undergoes a gradual reduction in size, potentially leading to atrophy in its later stages, which is characterized by extensive fibrosis and a weight reduction to as low as 10–20 g. Therefore, the diminished thyroid volume observed in patients with hypothyroidism within this study may suggest a progression towards the advanced stage of HT. In contrast, the significantly enlarged thyroid glands in patients with hyperthyroidism may be indicative of edematous swelling and increased vascularity, which further diminishes the interstitial compartment of the thyroid parenchyma, as opposed to the predominant fibrosis seen in patients with hypothyroidism.

It is important to note that there is a significant decrease in intrathyroidal IC in patients with pathologically confirmed DTD, even in those who are euthyroid. This observation may suggest a potential precursor to thyroid hormone dysfunction. Additionally, the significant interobserver variability associated with the sonographic diagnosis of DTD must be acknowledged, as well as the limitation that sonographic characteristics alone do not yield quantitative data regarding the functional progression of DTD.^{25,26} In contrast, imaging modalities employed in nuclear medicine, such as thyroid scans and radioactive iodine uptake tests, involve the administration of minimal radiation doses, either intravenously or orally, through the ingestion of the radiopharmaceutical sodium pertechnetate, to evaluate thyroid function. Typically, these assessments are conducted multiple times over a duration ranging from 30 minutes to a maximum of 24 hours.²⁷ Furthermore, DECT-based IC quantification, which does not necessitate the additional use of intravenous contrast, has the potential to provide a comprehensive assessment of both total thyroid volume and intrathyroidal IC as the functional indicators within a relatively brief timeframe.^{14–16} Therefore, this approach may present advantages over conventional CT densitometry, ultrasound, or nuclear medicine imaging in the assessment

of thyroid parenchyma in patients exhibiting thyroid hormone dysfunction or suspected pathological conditions, provided that a radiation dose-optimized DECT scanning protocol is employed.

Nonetheless, our research has primarily focused on intrathyroidal IC quantification. This was achieved through the use of non-contrast DECT images acquired from patients for preoperative staging purposes. Therefore, it is essential to consider several factors that may limit the direct applicability of these findings to the thyroid gland, which is acknowledged as the organ most vulnerable to radiation exposure. First, the DECT technique employed to generate iodine maps, virtual monoenergetic images, and virtual non-contrast images relies on low and high effective X-ray energies to obtain varying attenuation coefficient values for different materials.^{7,8,13} In contrast to other anatomical regions, the thyroid gland, situated in the lower neck, is particularly susceptible to beam-hardening artifacts and image noise during clinical assessments. These factors can considerably compromise the accuracy of DECT-derived iodine maps, as lower-energy photons are absorbed more rapidly than their higher-energy counterparts, resulting in an intensified X-ray beam spectrum by the time it reaches the detector.²⁸ Previous research by Kanatani et al.²⁸ indicated that X-ray beam hardening may lead to an underestimation of CT attenuation in iodine imaging, while the CT attenuation in virtual monoenergetic imaging was found to be overestimated in their phantom study. Although the accuracy of the dual-layer detector DECT utilized in this investigation has been validated in phantom studies for iodine quantification across various scanning and reconstruction parameters, such as tube potential, dose, rotation time, and spectral reconstruction level,^{11,12} there is a lack of published studies specifically addressing the impact of X-ray beam hardening in the vicinity of the thyroid gland. Second, the findings of this study are primarily confined to a specific scanning protocol conducted at a single institution; thus, further research should be undertaken across multiple institutions using different scanners or protocols. Notably, the primary objective of the non-contrast scanning was to identify calcifications in the thyroid gland or cervical lymph nodes, rather than to quantify intrathyroidal IC. Consequently, the technical parameters employed were not optimized to justify the radiation dose, despite their accuracy in

quantifying IC. To effectively assess the advantages of DECT in quantifying IC through DECT-derived iodine maps in thyroid parenchyma, a new CT protocol must be developed to measure relatively low intrathyroidal IC levels, particularly when IC is < 1.0 mgI/mL, in accordance with the principle of radioprotection: “as low as reasonably achievable.” Third, this study retrospectively selected only those patients who consecutively underwent DECT examinations and thyroid function tests. As a result, the majority of participants were middle-aged women with euthyroidism scheduled for thyroid cancer surgery. Additionally, other patients with thyroid hormone dysfunction included in this study were subclinical, presenting as either hypo- or hyperthyroidism, rather than overt disease. Therefore, the relationship between thyroid glandular IC and variables such as age or thyroid function necessitates validation in future studies encompassing a broader range of ages and hormonal statuses. Lastly, this research focused on the intrathyroidal IC quantification by analyzing ROIs within the entire thyroid glands with thyroid nodules measuring 10 mm or less. As a result, the IC of these smaller nodules may have been unintentionally included in the overall analysis, while nodules >1 cm were deliberately excluded. Given the prevalence of thyroid nodules in clinical practice, further studies should be undertaken to assess the impact of nodules on IC quantification, thereby improving the relevance and applicability of the findings.

In conclusion, the application of DECT-based intrathyroidal IC quantification offers a potentially valuable method for investigating iodine metabolism in relation to thyroid hormone disorders. However, further studies are necessary to validate its effectiveness in a more diverse patient population with differing thyroid hormone levels, as well as to develop more rigorous scanning protocols.

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07045730). The funder did not have any role in the study design, data collection, anal-

ysis, and interpretation of data, or writing the manuscript.

References

1. Chung HR. Iodine and thyroid function. *Ann Pediatr Endocrinol Metab.* 2014;19(1):8-12. [\[Crossref\]](#)
2. Iida Y, Konishi J, Harioka T, Misaki T, Endo K, Torizuka K. Thyroid CT number and its relationship to iodine concentration. *Radiology.* 1983;147(3):793-795. [\[Crossref\]](#)
3. Imanishi Y, Ehara N, Shinagawa T, et al. Correlation of CT values, iodine concentration, and histological changes in the thyroid. *J Comput Assist Tomogr.* 2000;24(2):322-326. [\[Crossref\]](#)
4. Pandey V, Reis M, Zhou Y. Correlation between computed tomography density and functional status of the thyroid gland. *J Comput Assist Tomogr.* 2016;40(2):316-319. [\[Crossref\]](#)
5. Silverman PM, Newman GE, Korobkin M, Workman JB, Moore AV, Coleman RE. Computed tomography in the evaluation of thyroid disease. *AJR Am J Roentgenol.* 1984;142(5):897-902. [\[Crossref\]](#)
6. Han YM, Kim YC, Park EK, Choe JG. Diagnostic value of CT density in patients with diffusely increased FDG uptake in the thyroid gland on PET/CT images. *AJR Am J Roentgenol.* 2010;195(1):223-228. [\[Crossref\]](#)
7. McCollough CH, Leng S, Yu L, Fletcher JG. Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology.* 2015;276(3):637-653. [\[Crossref\]](#)
8. Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 1. *Neuroimaging Clin N Am.* 2017;27(3):371-384. [\[Crossref\]](#)
9. Takeda T, Yu Q, Yashiro T, et al. Iodine imaging in thyroid by fluorescent X-ray CT with 0.05 Mm spatial resolution. *Nucl Instrum Meth A.* 2001;467:1318-1321. [\[Crossref\]](#)
10. Milakovic M, Berg G, Eggertsen R, et al. Determination of intrathyroidal iodine by X-ray fluorescence analysis in 60- to 65-year olds living in an iodine-sufficient area. *J Intern Med.* 2006;260(1):69-75. [\[Crossref\]](#)
11. Duan X, Arbiq G, Guild J, Xi Y, Anderson J. Technical note: quantitative accuracy evaluation for spectral images from a detector-based spectral CT scanner using an iodine phantom. *Med Phys.* 2018;45(5):2048-2053. [\[Crossref\]](#)
12. Hua CH, Shapira N, Merchant TE, Klahr P, Yagil Y. Accuracy of electron density, effective atomic number, and iodine concentration determination with a dual-layer dual-energy computed tomography system. *Med Phys.* 2018;45(6):2486-2497. [\[Crossref\]](#)
13. So A, Nicolaou S. Spectral computed tomography: fundamental principles and

- recent developments. *Korean J Radiol.* 2021;22(1):86-96. [\[Crossref\]](#)
14. Binh DD, Nakajima T, Otake H, Higuchi T, Tsushima Y. Iodine concentration calculated by dual-energy computed tomography (DECT) as a functional parameter to evaluate thyroid metabolism in patients with hyperthyroidism. *BMC Med Imaging.* 2017;17(1):43. [\[Crossref\]](#)
 15. Shao W, Liu J, Liu D. Evaluation of energy spectrum CT for the measurement of thyroid iodine content. *BMC Med Imaging.* 2016;16(1):47. [\[Crossref\]](#)
 16. Li ZT, Zhai R, Liu HM, Wang M, Pan DM. Iodine concentration and content measured by dual-source computed tomography are correlated to thyroid hormone levels in euthyroid patients: a cross-sectional study in China. *BMC Med Imaging.* 2020;20(1):10. [\[Crossref\]](#)
 17. Ha EJ, Chung SR, Na DG, et al. 2021 Korean Thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol.* 2021;22(12):2094-2123. [\[Crossref\]](#)
 18. Soh SB, Aw TC. Laboratory testing in thyroid conditions - pitfalls and clinical utility. *Ann Lab Med.* 2019;39(1):3-14. [\[Crossref\]](#)
 19. Nikiforov YE, Biddinger PW, Thompson LDR. *Diagnostic Pathology and Molecular Genetics of the Thyroid : A Comprehensive Guide for practicing thyroid pathology.* 3rd ed., Wolters Kluwer Health; 2018. [\[Crossref\]](#)
 20. User guide for Gammex tissue characterization phantom model 467. [\[Crossref\]](#)
 21. Ahad F, Ganie SA. Iodine, iodine metabolism and iodine deficiency disorders revisited. *Indian J Endocrinol Metab.* 2010;14(1):13-17. [\[Crossref\]](#)
 22. Lennartz S, Parakh A, Cao J, et al. Inter-scan and inter-scanner variation of quantitative dual-energy CT: evaluation with three different scanner types. *Eur Radiol.* 2021;31(7):4438-4451. [\[Crossref\]](#)
 23. Zopfs D, Graffe J, Reimer RP, et al. Quantitative distribution of iodinated contrast media in body computed tomography: data from a large reference cohort. *Eur Radiol.* 2021;31(4):2340-2348. [\[Crossref\]](#)
 24. Bauer DC, McPhee SJ. Thyroid Disease. In: Hammer GD, McPhee SJ, editors. *Pathophysiology of Disease: An Introduction to Clinical Medicine*, 7e. New York, NY: McGraw-Hill Education; 2013. [\[Crossref\]](#)
 25. Guan H, de Moraes NS, Stuart J, et al. Discordance of serological and sonographic markers for Hashimoto's thyroiditis with gold standard histopathology. *Eur J Endocrinol.* 2019;181(5):539-544. [\[Crossref\]](#)
 26. Ahn HS, Kim DW, Lee YJ, Baek HJ, Ryu JH. Diagnostic accuracy of real-time sonography in differentiating diffuse thyroid disease from normal thyroid parenchyma: a multicenter study. *AJR Am J Roentgenol.* 2018;211(3):649-654. [\[Crossref\]](#)
 27. Spencer CA. Laboratory thyroid tests: a historical perspective. *Thyroid.* 2023;33(4):407-419. [\[Crossref\]](#)
 28. Kanatani R, Shirasaka T, Kojima T, Kato T, Kawakubo M. Influence of beam hardening in dual-energy CT imaging: phantom study for iodine mapping, virtual monoenergetic imaging, and virtual non-contrast imaging. *Eur Radiol Exp.* 2021;5(1):18. [\[Crossref\]](#)



Quadratus lumborum block for procedural and postprocedural analgesia in renal cell carcinoma percutaneous cryoablation

Saman Fouladirad¹
 Jasper Yoo²
 Behrang Homayoon^{2,3,4,5}
 Jun Wang^{2,3,4,5}
 Pedro Lourenço^{2,3,4,5}

¹University of Saskatchewan, Faculty of Medicine, Saskatoon, Canada

²University of British Columbia, Faculty of Medicine, Vancouver, Canada

³University of British Columbia, Department of Radiology, Vancouver, Canada

⁴Surrey Memorial Hospital, Surrey, Canada

⁵Jim Pattison Outpatient Care and Surgery Centre, Surrey, Canada

ABSTRACT

This study assesses the efficacy of the quadratus lumborum block (QLB) in the management of procedural and peri-procedural pain associated with small renal mass cryoablation. To the best of our knowledge, this is the first study that examines the use of QLB for pain management during percutaneous cryoablation of renal cell carcinoma (RCC). A single-center retrospective review was conducted for patients who underwent cryoablation for RCC with QLB between October 2020 and October 2021. The primary study endpoint included a total dose of procedural conscious sedation and administered, postprocedural analgesia. Technical success in cryoablation was achieved in every case. No patients required additional analgesic during or after the procedure, and no complications resulted from the use of the QLB. The QLB procedure appears to be an effective locoregional block for the management of procedural and peri-procedural pain associated with renal mass cryoablation.

KEYWORDS

Ablation, analgesic, anesthesia, anesthetics, cancer, carcinoma, cryoablation, kidney, local anesthesia, local anesthetic, malignancy, minimally invasive, oncology, quadratus lumborum block, renal, renal cell carcinoma, small renal mass, tumoral

Renal cell carcinoma (RCC) is the most common type of kidney cancer. Laparoscopic or percutaneous energy ablative therapies have become an accepted treatment for small (T1) tumors (<3.5 cm) and have outcomes similar to those of partial/radical nephrectomy in the treatment of such tumors, with lower morbidity and greater preservation of renal function. They can also be performed on an outpatient basis without general anesthesia (GA). Percutaneous renal cryoablation (PRC) reduces complications and recovery time compared with more invasive procedures, such as open/laparoscopic partial nephrectomy.¹⁻³ The quadratus lumborum block (QLB), first described in 2007, is a more recent procedure.^{4,5} The aim of this paper is to explore QLB's potential for procedural and postprocedural pain management in PRC for RCC.

Methods

This was a single-center retrospective study. Application for ethical approval was waived as outlined by the institutional ethical review board. Informed consent was obtained from all patients in this study.

Patients

The data of patients who received QLB for the cryoablation of small renal masses between October 2020 and October 2021 were included in this study. The inclusion criteria were as follows: patients aged >18 y and <100 y; those with a tumor size ≤5 cm; those with a single renal tumor; and/or those with a primary renal tumor without extrarenal or vascular invasion. The exclusion criteria were as follows: those with a hypersensitivity to local anesthesia and/or those with a soft-tissue infection overlaying the needle placement site. Demographics and

Corresponding author: Pedro Lourenço

E-mail: pedro.lourenco@fraserhealth.ca

Received 09 January 2024; revision requested 10 February 2024; last revision received 26 March 2024; accepted 06 April 2024.



Epub: 02.09.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.232100

You may cite this article as: Fouladirad S, Yoo J, Homayoon B, Wang J, Lourenço P. Quadratus lumborum block for procedural and postprocedural analgesia in renal cell carcinoma percutaneous cryoablation. *Diagn Interv Radiol.* 2025;31(3):234-236.

Table 1. Patient demographic details, tumor descriptors, procedural information, and outcomes for three patients administered anesthetics for percutaneous cryoablation of small renal masses

		Patient		
		1	2	3
Demographic	Age (y)	73	80	79
	Gender	F	M	M
Tumor description	Size (cm)	2.7 × 2.4 × 2.0	3.6 × 3.1 × 3.1	2.2 × 1.8 × 2.4
	Side	L	L	L
	Location	Lower pole	Upper pole	Mid-upper pole
	Growth	Exophytic	Exophytic	Exophytic
	RCC subtype	Clear cell	Papillary	Clear cell
Procedure	Technique	Cryoablation	Cryoablation	Cryoablation
	Block used	Left QLB	Left QLB	Left QLB
	Success status	Technical success	Technical success	Technical success
Complications	Minor	0	0	0
	Major	0	0	0
Disposition	Length of stay (h)	4	4	4
	Follow-up	CT @ 3 mon	MRI @ 3 mon	CT @ 3 mon

F, female; M, male; RCC, renal cell carcinoma; QLB, quadratus lumborum block; CT, computed tomography; MRI, magnetic resonance imaging.

tumor characteristics, including pathology, size, and location, were recorded (Table 1). Any complications related to the procedure were noted according to the modified Clavien–Dindo system.⁶ Patients were referred to and assessed for PRC via interventional radiology.

Technique

PRC with QLB was performed by a single, fellowship-trained, interventional radiologist at Surrey Memorial Hospital in Surrey, British Columbia, Canada.

A QLB comprised of 20 mL of 0.5% bupivacaine solution was used for preprocedural regional block pain management. Additionally, for the management of procedural anxiety, 50 mcg of fentanyl and 1 mg of midazolam were administered intravenously to achieve conscious sedation (CS) upon each patient's arrival in the computed tomography (CT) ab-

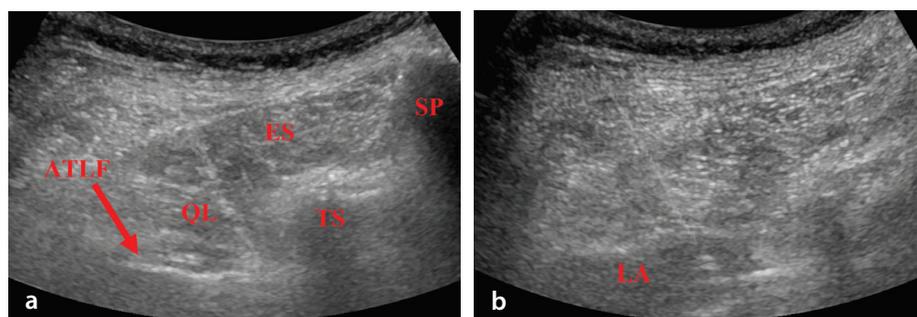


Figure 1. Anterior QLB approach. (a) A horizontally oriented, curvilinear probe is used to identify the ES and QL muscles ipsilateral to the treatment side. (b) A 21 g Chiba needle is advanced from posterior to anterior via the ES and QL muscle to instill 20 mL of 0.5% bupivacaine solution in the ATLF, resulting in an expected distention of the ATLF. QLB, quadratus lumborum block; ES, erector spinae; QL, quadratus lumborum muscle; SP, spinous process; TS, transverse process; ATLF, anterior thoracolumbar fascia; LA, local anesthetic.

lation suite. An additional 50 mcg of fentanyl and 1 mg of midazolam were administered as needed for either preprocedural anxiety/stress, intra-procedural pain management, or postprocedural pain management. The primary outcome measured for this study was the success of QLB in the management of procedural and postprocedural pain, determined based on the need for further analgesic (in addition to the standard analgesic administered) to control patients' pain (Table 2).

An anterior QLB approach under ultrasound was used for all cases (Figure 1). Patients were placed in a lateral decubitus or prone position and monitored by pulse oximeter, non-invasive blood pressure monitoring, and electrocardiogram. A linear transducer was placed in the axial plane in the midaxillary line and moved posteriorly until the lateral interfascial triangle (LIFT) (i.e., the fascia that surrounds the paraspinal muscles)

became visible between the latissimus dorsi and the quadratus lumborum. A needle was then introduced at the lateral end of the transducer and advanced until it was in the middle layer of the thoracolumbar fascia (i.e., the fascia that separates the quadratus lumborum from the latissimus dorsi and paraspinous muscle) close to the triangular structure of the LIFT. The local anesthetic was then injected intrafascially.

Cryoablation was performed under CT guidance (GE Revolution CT, Waukesha, WI, United States of America) and ultrasound (GE LOGIQ E10, Waukesha, WI, United States of America). Prior to ablation, single, portal venous phase, contrast-enhanced CT was conducted to better outline the lesion of interest. A variable number of cryoprobes were placed with 1–2 cm of spacing depending on the size and morphology of lesions, with single/multiple scans conducted for

Main points

- This is the first study to successfully employ the use of the quadratus lumborum block (QLB) during percutaneous renal cryoablation (PRC) for renal cell carcinoma (RCC).
- Hospital stays with QLB were reduced compared with prior studies that used the typical analgesic options of local anaesthetic and conscious sedation or general anaesthetic. This could transform PRC for RCC into an outpatient procedure instead of requiring overnight hospitalization.
- In this study, usage of QLB during PRC for RCC allowed for decreased demand for analgesics or sedation.

Table 2. Type and dosage of anesthetic used based on stage of procedure

	Anesthetic	Patient		
		1	2	3
Preprocedural sedation	Bupivacaine	20 mL of 0.25%	20 mL of 0.25%	20 mL of 0.25%
	Fentanyl (mcg)	50	50	100
	Midazolam (mg)	1	1	2
Intraprocedural sedation	None	N/A	N/A	N/A
Postprocedural analgesia	None	N/A	N/A	N/A

appropriate probe positioning. The double freeze–thaw protocol-consisting of cycles that involve a double freezing cycle of 10 min separated by a passive (9 min) and active (1 min) thaw session-was conducted. Unenhanced CT images at 5 and 10 min were captured to monitor ice ball growth and identify any vulnerable structures. More probes were used as needed if the ice ball did not fully cover the tumors. Following the ablation, further non-contrast CT was performed following probe removal to assess for any complications.

Results

Demographics and tumor characteristics

Three patients (n = 3) with an average age of 77.33 ± 2.19 y underwent cryoablation with QLB for small renal masses between October 2020 and October 2021 in this study. The mean volume of the tumors was 19.02 ± 7.85 cm³, and all three were located in the left kidney. Patients had prior biopsies completed with pathology consistent with RCC-two with clear cell subtypes and one with papillary subtypes.

Procedure and complications

Technical success with no major complications was achieved in all cases. One patient returned to the emergency department within 24 hours with postprocedural hematuria but required no additional intervention. No patient exhibited QLB-associated complications.

Anesthesia

On average, 66.67 ± 16.67 mcg of fentanyl and 1.33 ± 0.33 mg of midazolam were used to manage procedural anxiety prior to the QLB of 20 mL of bupivacaine solution. Two patients required only the standard 50 mcg of fentanyl and 1 mg of midazolam before the procedure, whereas one patient needed additional dosing for preprocedural anxiety and stress. No additional analgesics were needed during the cryoablation or during the 4 h postprocedural recovery period. Although no specific survey measurements were used to assess pain severity, patient

reporting was used as an indicator of pain management, and no patients reported discomfort or pain-related symptoms.

Discussion

With the increasing demand for regional anesthesia, QLB may reduce the need for CS and GA in the ablation of small renal masses. To the best of our knowledge, this is the first paper to evaluate the potential of QLB for pain management during percutaneous cryoablation in RCC. The QLB procedure used in our study consisted of a single administration of a small amount of bupivacaine solution prior to the placement of the cryoablation probe. Our patients received small loading doses of CS for preprocedural anxiety and did not receive additional intraprocedural or postprocedural doses. Other studies of PRC have used local anesthesia and CS induction; however, in our study, no patient reported postprocedural pain or discomfort, nor did any require intraprocedural CS titration, suggesting that the QLB procedure was effective. None of the participants in the current study who received QLB with CS required a conversion to GA for analgesia or experienced any QLB-related complications. In addition, the use of QLB allowed for reduced CS (fentanyl and midazolam) dosing.^{7,8} Complete procedural success with QLB was achieved in all three cases, and the average length of stay following cryoablation with QLB was approximately 4 h, which is shorter than previous studies using local anesthesia and CS.⁸ Thus, QLB may allow interventional radiologists to perform PRC on an outpatient basis instead of requiring an overnight hospital stay. The QLB procedure may be valuable in patients with contraindications to CS and/or comorbidities that make them poor surgical candidates. Thus, QLB is attractive as a safe, well-tolerated locoregional approach to anesthesia for PRC in RCC.

The limitations of this study include its small sample size and the lack of a comparison group to make a direct, objective comparison between different analgesics. Future studies are needed to confirm the benefit of QLB in PRC of T1 RCC.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Chan VW, Abul A, Osman FH, et al. Ablative therapies versus partial nephrectomy for small renal masses - A systematic review and meta-analysis. *Int J Surg*. 2022;97:106194. [\[CrossRef\]](#)
- Klatte T, Shariat SF, Remzi M. Systematic review and meta-analysis of perioperative and oncologic outcomes of laparoscopic cryoablation versus laparoscopic partial nephrectomy for the treatment of small renal tumours. *J Urol*. 2014;191(5):1209-1217. [\[CrossRef\]](#)
- Kunkle DA, Uzzo RG. Cryoablation or radiofrequency ablation of the small renal mass: a meta-analysis. *Cancer*. 2008;113(10):2671-2680. [\[CrossRef\]](#)
- Krohg A, Ullensvang K, Rosseland LA, Langesæter E, Sauter AR. The analgesic effect of ultrasound-guided quadratus lumborum block after cesarean delivery: a randomized clinical trial. *Anesth Analg*. 2018;126(2):559-565. [\[CrossRef\]](#)
- Ishio J, Komazawa N, Kido H, Minami T. Evaluation of ultrasound-guided posterior quadratus lumborum block for postoperative analgesia after laparoscopic gynecologic surgery. *J Clin Anesth*. 2017;41:1-4. [\[CrossRef\]](#)
- Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240(2):205-213. [\[CrossRef\]](#)
- Allaf ME, Varkarakis IM, Bhayani SB, Inagaki T, Kavoussi LR, Solomon SB. Pain control requirements for percutaneous ablation of renal tumors: cryoablation versus radiofrequency ablation--initial observations. *Radiology*. 2005;237(1):366-370. [\[CrossRef\]](#)
- Okhunov Z, Juncal S, Ordon M, et al. Comparison of outcomes in patients undergoing percutaneous renal cryoablation with sedation vs general anesthesia. *Urology*. 2015;85(1):130-134. [\[CrossRef\]](#)



Non-routine thrombectomy in pediatric arterial ischemic stroke

Sinan Balcı¹
 Nesibe Gevher Eroğlu-Ertuğrul^{2*}
 Ahmet Ziya Birbilen³
 Dilek Yalınzoğlu²
 Selman Kesici⁴
 Tefvik Karagöz⁵
 Anıl Arat¹

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

²Hacettepe University Faculty of Medicine,
Department of Child Health and Diseases, Division of
Pediatric Neurology, Ankara, Türkiye

³Hacettepe University Faculty of Medicine,
Department of Child Health and Diseases, Division of
Pediatric Emergency, Ankara, Türkiye

⁴Hacettepe University Faculty of Medicine,
Department of Child Health and Diseases, Division of
Pediatric Critical Care, Ankara, Türkiye

⁵Hacettepe University Faculty of Medicine,
Department of Child Health and Diseases, Division of
Pediatric Cardiology, Ankara, Türkiye

*Currently at Ankara City Hospital,
Ankara, Türkiye

Corresponding author: Anıl Arat

E-mail: anilarat@hotmail.com

Received 12 February 2024; revision requested 28 March
2024; accepted 05 May 2024.



Epub: 30.05.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242675

PURPOSE

Unlike in adults, the indications and techniques for mechanical thrombectomy for arterial ischemic stroke (AIS) in children are not clearly established. The medical and interventional management of children with acute large vessel occlusion may entail the modification of the standardized management of this condition in adults. We present six cases of children who underwent non-routine thrombectomy for AIS.

METHODS

We retrospectively reviewed the records of children diagnosed with AIS between 2015 and 2023 and evaluated patient characteristics, procedural technical data, and final clinical outcomes. Procedures deviating from the current definition and indications for AIS treatment in adults as well as previously reported pediatric thrombectomy cases were defined as non-routine thrombectomy.

RESULTS

Seven non-routine thrombectomy procedures in six children were included in the study. The National Institutes of Health Stroke Scale scores on admission ranged from 4 to 35; no procedure-related mortality or major neurologic morbidity occurred. One child died of causes related to the initial severe heart failure and stroke; otherwise, all the children had a modified Rankin scale score of 0 to 1 at follow-up. Unique clinical and procedural features in our case series included presentation with acute stent occlusion (two children), bilateral simultaneous internal carotid artery occlusions associated with a unilateral tandem middle cerebral artery (MCA) occlusion (one child), MCA occlusion caused by thromboembolism of the atrial myxoma (one child), and very distal (one child) or delayed thrombectomy (two children).

CONCLUSION

Modifications to the standard medical and interventional algorithms may be required for mechanical thrombectomy in children.

CLINICAL SIGNIFICANCE

Referral centers specialized in pediatric neurology, pediatric anesthesia, and pediatric intervention are optimal for treating children using mechanical thrombectomy and for modifying the treatment, if required.

KEYWORDS

Stroke, children, thrombectomy, thrombolysis, stent, aspiration

Arterial ischemic stroke (AIS) is a primary cause of long-term neurologic deficit and mortality in children despite its rarity (1.3–13:100.000).¹ Its etiology and pathophysiology in children differ from those in the adult population because adults and children differ in terms of arterial morphology and their particular response to ischemia.² The diagnosis of AIS is often challenging in children because of its low incidence, variability in clinical presentation, and wide spectrum of underlying risk factors.³

Evidence-based data and guidelines on revascularization treatments such as intravenous tissue plasminogen activator and mechanical thrombectomy have not yet been established

You may cite this article as: Balcı S, Eroğlu-Ertuğrul NG, Birbilen AZ, et al. Non-routine thrombectomy in pediatric arterial ischemic stroke.

Diagn Interv Radiol. 2025;31(3):237-248.

for children.⁴ However, following the worldwide acceptance of mechanical thrombectomy in adults with AIS, a growing number of cases have been reported in children. Recently, a multicenter study⁵ and a meta-analysis⁶ reported favorable neurologic outcomes, indicating that this procedure is safe and efficacious in children.

We reviewed the cases of six children treated with non-routine mechanical thrombectomy at a single pediatric tertiary referral center and evaluated these patients with regard to any unique, distinguishing characteristics that have not previously been reported in the relevant literature.

Methods

Our study was approved by the Hacettepe University Institutional Review Board (approval number: SBA 23/426 2023/09-06). The endovascular procedures were performed under emergency conditions with written informed consent signed by the parents, and the evaluations were conducted retrospectively; therefore, no additional consent to participate was required. The cases of patients with AIS admitted to our children's hospital for endovascular treatment between 2015 and 2023 were reviewed retrospectively to determine whether the procedures were performed according to the guidelines for adults. For cases with clinical suspicion of acute stroke, initial neuroimaging work-up included non-enhanced head computed tomography (CT) to exclude hemorrhagic stroke, followed by neck and head CT angiography (CTA), cranial time-of-flight magnetic resonance (MR) angiography, or diffusion-weighted imaging to evaluate cerebral vessels and the viability of the affected cerebral territory. Those diagnosed with acute intracranial large vessel occlusion were urgently evaluated by specialists in interventional neuroradiology, pediatric emergency

medicine, and pediatric neurology to determine if the patient would benefit from mechanical thrombectomy.

Between 2015 and 2023, 22 children presented with acute stroke caused by large vessel occlusion. Of these, seven with large core infarcts upon presentation identified through cross-sectional imaging and eight presenting more than 24 hours after the onset of stroke symptoms were excluded from the study. Among the remaining seven children, one underwent a straightforward endovascular mechanical thrombectomy, with the indications and techniques mirroring those for adult patients. After excluding the child receiving "routine" thrombectomy, six children aged between 6 and 17 years underwent "non-routine" mechanical thrombectomy (Figure 1).

The neurological status of the patients was evaluated using the National Institutes of Health Stroke Scale (NIHSS) on admission. Briefly, the NIHSS score (between 0 and 42) is determined as follows: 0: no stroke symptoms, 1–4: minor stroke, 5–15: moderate stroke, 16–20: moderate to severe stroke, 21–42: severe stroke. Revascularization at the end of the procedure was evaluated in accordance with the modified Thrombolysis in Cerebral Infarction (mTICI) score, and mTICI scores of 2B, 2C, or 3 were considered "successful revascularization." At follow-up, the modified Rankin scale (mRS), with scores ranging from 0 to 6, was used to determine the final neurologic status of the patients.^{7,8}

Informed consent was obtained from all families before the procedure. Endovascular procedures were performed under general anesthesia through transfemoral access.

Throughout the procedure, systemic anticoagulation was achieved by using intravenous heparin targeting an activated clotting time of between 200 and 250 s. Mechanical thrombectomy was performed through direct aspiration or with only stent retrievers or by a combination of the two. Systemic anticoagulation was continued for 24–36 hours after the procedure in the intensive care unit, except for case 2, in which anticoagulation was reversed with protamine at the end of the procedure because of the self-limited contrast extravasation on the post-procedural flat panel detector CT images. The five surviving children received long-term prophylactic anticoagulation or antiplatelet therapy after the procedures to address underlying risk factors (follow-up duration: 4 days to 31 months).

Results

Relevant clinical data from children treated with non-routine mechanical thrombectomy are presented in Table 1. All patients except the patient in case 5 presented with unilateral hemiparesis; one also had focal seizures and altered consciousness. The arterial occlusion was in the anterior circulation in five children (right sided in three, left sided in one, and bilateral in one) and in the posterior circulation in one. None received intravenous thrombolysis before mechanical thrombectomy, but the patient in case 4 had received enoxaparin sodium in hour 2 of the symptoms prior to the emergent transfer to our center. The mean NIHSS score was 13.3 (4–35), which is consistent with minor to moderate stroke, in five children. The patient in case 2 had an NIHSS score of 35 on neurologic examination upon admission, which

Main points

- Pediatric stroke differs from adult stroke in terms of etiology and underlying pathologies.
- The major technical aspects of mechanical thrombectomy for acute ischemic stroke with large vessel occlusion and the devices used for endovascular treatment are similar in both pediatric and adult patients.
- Endovascular treatment criteria and guidelines should be improved because neural tolerance to ischemia and collateral circulation dynamics are different in the pediatric population.

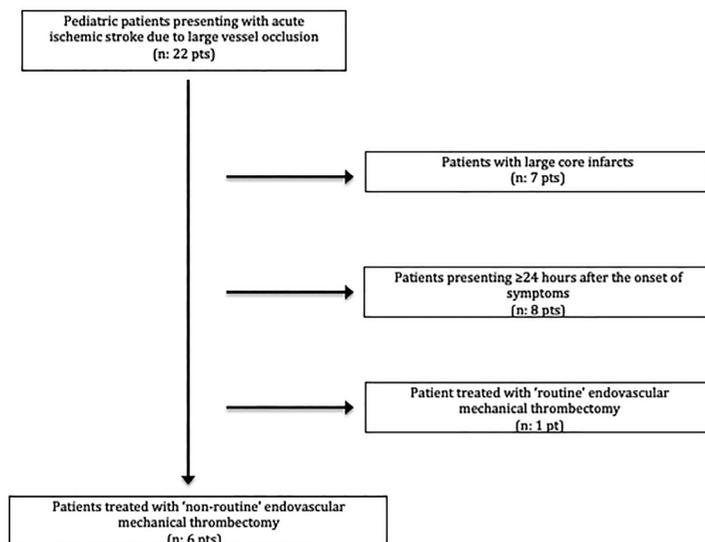


Figure 1. Patient selection flowchart.

was performed before emergent intubation. The mean symptom onset to recanalization time was 5.7 hours (3–9 hours). The technical data related to the procedures are summarized in Table 2, and the specific features of children with non-routine mechanical thrombectomy are explained below.

Case 1

This 6-year-old girl was born to consanguineous parents following an uneventful pregnancy and delivery. She was diagnosed with multiple intracranial aneurysms (Supplementary Figure 1a), one of which was a large (3.8 × 3.7 cm) dissecting cavernous internal carotid artery (ICA) aneurysm compressing the right optic nerve (Supplementary Figure 1b, c). She had proptosis, lateral gaze palsy, loss of vision in the right eye, and right optic nerve atrophy. After the diagnosis, the posterior circulation aneurysm was embolized. In a subsequent session, a flow diverter stent was inserted for the treatment of the carotid aneurysm. Antiplatelet therapy with a daily dose of 2.5 mg of oral prasugrel was started prior to the embolization of the

posterior circulation aneurysm after a VerifyNow assay identified the child as hyporesponsive to clopidogrel. Three days after the second treatment of the cavernous ICA aneurysm, the patient was admitted with confusion, left hemiparesis, mild facial paralysis, and dysarthria. At hour 2 of symptoms, her NIHSS score was 10, and CTA and digital subtraction angiography (DSA) identified total occlusion of the stents in the right ICA (Supplementary Figure 1d). Mechanical thrombectomy was performed immediately using a Solitaire device (Medtronic, Irvine, CA, USA) (Supplementary Figure 1e, f), and revascularization was achieved 3 hours after the onset of symptoms (Supplementary Figure 1g, h). At her subsequent follow-up, the right ICA eventually became occluded, and the right anterior circulation territory was reconstructed via the contralateral carotid and ipsilateral posterior circulations through the partially (and proximally) occluded flow diverter. The patient returned to her baseline neurologic state, remaining asymptomatic, and her mRS scores at discharge and at 31-month follow-up were both 0.

Case 2

This 10-year-old boy had no notable medical history other than mild learning difficulties. He presented with fatigue, vomiting, and diarrhea, which had continued for several days, and was diagnosed with gastroenteritis and administered antibiotics. After 10 days, he developed acute right hemiparesis, dysarthria, right focal seizures, and altered consciousness. Upon admission, he had hypotension and supraventricular arrhythmia. The Glasgow Coma scale score was 4. Emergent transthoracic echocardiography revealed dilatation of the right and left atria and systolic dysfunction. Cardiac thromboembolism was strongly suspected although no thrombus was detected on echocardiography. The patient underwent a coronary and head and neck CTA, which identified thrombus in the left atrial appendage (Supplementary Figure 2a), severe dilatation of all cardiac chambers, tandem occlusions of the right ICA at its origin and right MCA at the distal M1 segment (Supplementary Figure 2b), as well as occlusion of the left ICA terminus (Supplementary Figure 2c). The patient was

Table 1. Clinical and procedural summary of the patients

Procedure #	Age (y)/gender	Presenting time/symptom	NIHSS	Clot location	Risk factors/etiology	Follow-up after procedure (months)	Outcome (mRS)
1 (case 1)	6/F	2 hours/hemiparesis	10	Right ICA, inside the stent	Stent placement	31	0
2 (case 2)	10/M	3 hours/hemiparesis, seizure	>30	Right cervical ICA and MCA distal M1 segment	Dilated CMP	4 days	6
3 (case 2)	10/M	3 hours/hemiparesis, seizure	>30	Left ICA bifurcation	Dilated CMP	4 days	6
4 (case 3)	16/F	4 hours/hemiparesis	13	Right ICA bifurcation	Hypertrophic CMP → dilated CMP	24	0
5 (case 4)	16/M	7 hours/hemiparesis	4	Right MCA M1	Factor V Leiden, prothrombin, and heterozygous <i>MTHFR</i> mutation	12	1
6 (case 5)	11/M	More than 3 days/ataxia, vertigo, oculomotor nerve deficits	5	Basilar artery	Stent placement	5	0
7 (case 6)	17/F	3 hours/hemiparesis, aphasia	13	Left MCA M1	Atrial myxoma	3	1

NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin scale (0: no deficit, 6: death); F, female; M, male; ICA, internal carotid artery; MCA, middle cerebral artery; CMP, cardiomyopathy.

Table 2. Technical data of the procedures

Procedure	mTICI scores	Thrombectomy technique	Thrombectomy devices	Time to recanalization
1 (case 1)	3	Stent retriever	Solitaire	3 hours
2 (case 2)	2B	Aspiration & stent retriever	Catchview Maxi	4.5 hours*
3 (case 2)	0	Aspiration & stent retriever	Solitaire, Catchview Mini & Trevo Mini	-
4 (case 3)	2C	Aspiration & stent retriever	Catchview Maxi & Mini	7 hours
5 (case 4)	2C	Aspiration alone	Sofia 5F	9 hours
6 (case 5)	3	Stent retriever	Catchview	Not applicable
7 (case 6)	2C	Stent retriever	Solitaire	5 hours

*Revascularization of right internal carotid artery (ICA) and middle cerebral artery; left ICA could not be recanalized; mTICI, modified Thrombolysis in Cerebral Infarction.

immediately transferred to the angiography suite for mechanical thrombectomy. After documentation of the right cervical ICA and right MCA occlusions on the initial angiograms (Supplementary Figure 2d, e), the clot in the right cervical ICA was removed and the right MCA territory was recanalized with an mTICI score of 2B (Supplementary Figure 2f-h); however, the left ICA terminus could not be recanalized (Supplementary Figure 2i-n). A self-limiting extravasation occurred during the recanalization attempt of the left ICA, and a flat panel detector CT revealed focal subarachnoid hemorrhage and a hematoma with a maximal size of 8 mm at the left basal ganglia. At the follow-up imaging, the hema-

toma had not increased in size. The follow-up MR imaging (MRI) revealed signs of increased intracranial pressure. The patient died 4 days later as a result of cardiac dysfunction and cerebral hypoperfusion, which was most likely dilated cardiomyopathy.

Case 3

This 16-year-old girl had previously been diagnosed with hypertrophic cardiomyopathy and presented within the first 6 hours of symptom onset. A head and neck CTA and DSA revealed acute right ICA occlusion (Figure 2a, b). After angiography, the right ICA terminus was recanalized through throm-

boaspiration (Figure 2c, d). Subsequently, the M1 segment of the MCA and A2 segment of the anterior cerebral artery were recanalized after performing a stent retriever thrombectomy using a Catchview Maxi device (Balt, Montmorency, France) (Figure 2e, f). Finally, a “very distal” thrombectomy was performed using a Catchview Mini device (Balt) for the residual thrombus located at a parietal branch (M3/M4 segment) (Figure 2g, h), achieving a final mTICI score of 2C (Figure 2i, j). The follow-up MRI revealed focal acute infarction at the right basal ganglia (Figure 2k-n). At the 24-month follow-up, her follow-up mRS score was 0.

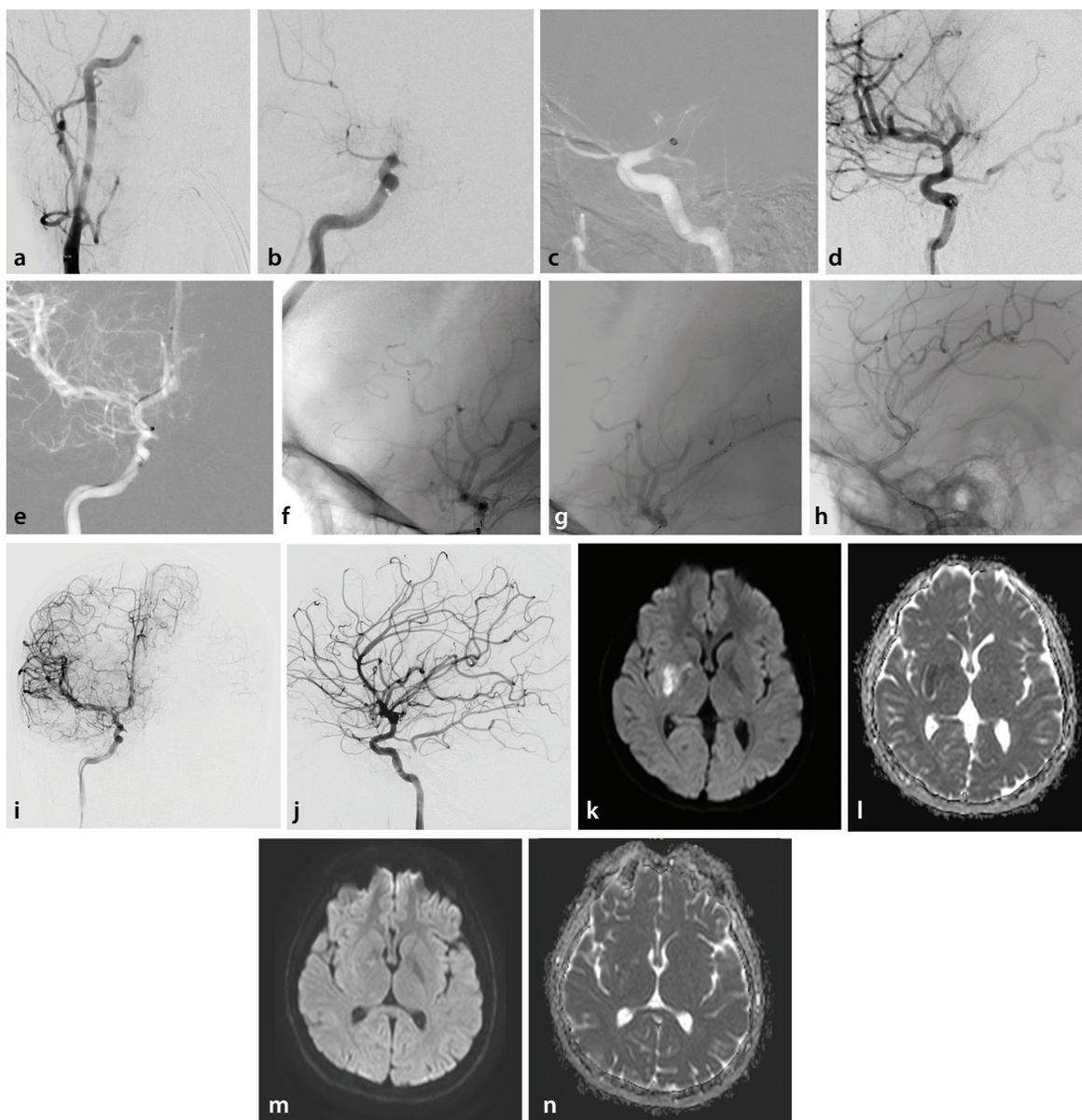


Figure 2. Preprocedural right carotid angiograms in anterior–posterior (AP) projection (a, b) reveal total occlusion of the right internal carotid artery (ICA) terminus. After recanalization of the supraclinoid ICA with aspiration thrombectomy (c), residual distal middle cerebral artery (MCA) and anterior cerebral artery (ACA) occlusions persist on the carotid angiogram (d). Both distal ACA (e, f) and distal MCA (g, h) occlusions were recanalized through stent retriever thrombectomy. Final AP (i) and lateral (j) carotid injections demonstrate the successful recanalization of the whole carotid artery territory. Diffusion-weighted image (k) and corresponding apparent diffusion coefficient (ADC) map (l) demonstrate acute infarction at the right basal ganglia. Follow-up magnetic resonance imaging (m, n) indicates regression of the diffusion restriction and consequent ADC pseudonormalization.

Case 4

This 16-year-old boy, who had a family history of thrombophilia (his mother and sister) as well as a father with myocardial infarction at the age of 40, was referred with left hemiparesis at hour 7 of the symptoms. He had received enoxaparin sodium 2 hours after symptom onset at another hospital before arriving at our center. His neurologic examination slightly improved during transfer, and his NIHSS score had decreased to 4 by the time of admission to the angiography suite. On the initial brain MRI, diffusion-weighted images and apparent diffusion coefficient maps revealed an acute infarction extending from the posterior right putamen to the corona radiata caused by a right proximal MCA occlusion. The groin was punctured at hour 8, and recanalization of an acute right M1 total occlusion was achieved 9 hours after the onset of symptoms (Supplementary Figure 3a-d). His mRS score was 1 (mild weakness in his left leg without disability) at the 12-month follow-up.

Case 5

This 11-year-old boy was treated at another hospital for a large, almost totally thrombosed, vertebrobasilar aneurysm (Supplementary Figure 4a) through the placement of telescopic flow diverters. The child was started on clopidogrel before the procedure, and this treatment was continued after the procedure. He was neurologically intact immediately after the endovascular procedure but developed oculomotor nerve deficits, ataxia, and vertigo during the postoperative period before discharge from the hospital. As a result of fluctuating neurologic symptoms, at the request of his family, the child was referred to our hospital after the onset of symptoms. He underwent mechanical thrombectomy within 24 hours of the last deterioration of his neurologic status. His NIHSS score was 5 on admission to the hospital and 4 at the angiography suite. A cerebral angiogram was obtained, which revealed near occlusion of the mid basilar artery secondary to a thrombus just distal to the flow diverter construct (Supplementary Figure 4b, c). Under a roadmap, a Rebar 18 microcatheter (Medtronic) was advanced through a 5F intermediate catheter, and the basilar artery was recanalized using a single pass of a Catchview device (Balt) (Supplementary Figure 4d). The child was switched from clopidogrel to prasugrel after the procedure. At the follow-up MRI, the mass effect of the partially thrombosed aneurysm on the neighboring brain

stem structures and focal ischemic foci at the pons and cerebellar hemispheres were evident (Supplementary Figure 4e-i). At his follow-up DSA at 5 months (Supplementary Figure 4j), both the flow diverter stent and basilar artery were patent and his mRS score was 0.

Case 6

This 17-year-old girl without a notable medical history presented with acute right hemiparesis, facial paralysis, and global aphasia. A mass-like lesion at the left atrium was visualized through echocardiography, indicating cardiac thromboembolism. A head and neck CTA demonstrated left MCA distal M1 segment occlusion. Partial occlusion at the distal M1 and proximal M2 segments of the MCA with significant slow antegrade flow was identified through angiography (Supplementary Figure 5a, b). After selective "distal" catheterization of the superior and inferior trunks and the temporal branch (Supplementary Figure 5c-e), successful recanalization was achieved through stent retriever thrombectomy using a Solitaire device (Medtronic) (Supplementary Figure 5f, g). Postoperative cardiac CT revealed a hypodense soft tissue mass at the left atrium suggestive of atrial myxoma (Supplementary Figure 5h). Her cardiac mass was operated on after mechanical thrombectomy on the same day, and the pathological evaluation results of both the cardiac mass and cerebral thrombus were consistent with myxoma. At the 3-month clinical follow-up, her mRS score was 1 and she had minimal aphasia, allowing almost complete communication and mild weakness in the right leg.

Discussion

Following extensive studies comparing endovascular treatment with medical management within 6 hours of stroke onset, mechanical thrombectomy has been accepted as the standard of care for adult patients with AIS secondary to intracranial large vessel occlusion.⁹⁻¹¹ Although the efficacy and safety of this treatment have not been clearly established in children, the number of pediatric case reports and series are increasing.¹²⁻¹⁴ Dicipinigaitis et al.¹⁵ extracted the data on patients with pediatric stroke from the National Inpatient Sample, identifying 190 children treated with mechanical thrombectomy with a favorable clinical outcome of 55.3%. Bhatia et al.¹⁶ compared 26 children with AIS who had undergone mechanical thrombectomy with 26 children with AIS who had received medical management, reporting improved

clinical outcomes among the mechanical thrombectomy group with an odds ratio of 3.76. A recent meta-analysis on pediatric mechanical thrombectomy revealed a successful recanalization rate of over 90% among 184 children, a positive clinical outcome (mRS ≤ 2) rate of 83.3% among 183 children, and a mortality rate of 3.2% among 184 children.¹⁷ These multicenter studies and case series with relatively high patient numbers suggest that pediatric thrombectomy may be a safe and effective treatment option for pediatric patients with AIS.

In general, thrombectomy for children has mirrored that for adults with respect to indications, technique, and to some extent, clinical evaluation scales. In the retrospective evaluation of the pediatric patients with stroke in this study, we noted that the patients had several unique and instructive characteristics, as all strayed from the usual indications, timing, and technique of routine mechanical thrombectomy. Cases 1 and 5 are examples of thrombectomy procedures performed for acute stent occlusions, which, to our knowledge, has not been reported before in the pediatric population. We are also unaware of an acute bilateral presentation or endovascular treatment in children, as in case 2. Moreover, in cases 3 and 6, distal and very distal (M3/4 segment of the MCA) thrombectomy was performed, which, in the adult population, is a controversial procedure¹⁸ and has not been reported in children. The literature contains few case reports of children treated with mechanical thrombectomy for intracranial large vessel occlusion caused by atrial myxoma.¹⁹ Finally, large vessel occlusion in case 4 was recanalized successfully beyond the first 8 hours of symptom onset. In the standard adult therapy, this would be possible only for patients meeting strict imaging-based indications.^{20,21}

Although the current endovascular armamentarium, which is adult oriented, was sufficient for the pediatric cases in this study,²² we faced various other technical challenges during treatment. One challenge, which affected two children with intracranial stent thrombosis, was the potential risk of flow diverter displacement, device intussusception, and arterial dissection or vasospasm during catheterization and thrombus retrieval. A similar salvage procedure was recently reported for acute in-stent occlusion in two adults.²³ However, mechanical thrombectomy for in-stent thrombosis has not been previously reported in children. The other challenge in our case series was related to bilateral acute ICA occlusion, associated

with a tandem occlusion on one side. These primary multivessel occlusions are rare, accounting for only 0.35% of adult thrombectomies, and they are frequently mortal.²⁴ We found just one report of thrombectomy for a primary multivessel occlusion in a child.²⁵ Nevertheless, multivessel occlusions associated with a tandem occlusion are extremely rare in adults,²⁴ and to our knowledge, they have not been reported in children. The compensation of bilateral ICA occlusion relies on collateral circulation through the vertebrobasilar system, an external carotid/ophthalmic anastomosis, or a combination of the two. In case 2 of our study, a 10-year-old patient with no notable history was diagnosed with dilated cardiomyopathy when he presented with acute stroke. Increased intracranial pressure indications were present, and the increased pressure possibly aggravated the ischemic insult.²⁶ Supraventricular tachycardia may also have contributed to the ischemia,²⁷ and tandem occlusions of the ICA and MCA on one side may have further hampered the collateral circulation. Routine isolated tandem occlusions, however, occur in up to 20% of patients with AIS in the adult population.²⁸ Acute tandem occlusion in pediatric patients is rare. In the literature, we were able to find a single report of a 13-year-old girl who underwent mechanical thrombectomy and had favorable clinical and radiological outcomes.²⁹ Recent studies in adults investigating prognoses of the subtypes of acute anterior circulation large vessel occlusions noted that, compared with unilateral MCA, patients with AIS with tandem ICA/MCA and bilateral occlusions or those with contralateral stenosis had less favorable outcomes.^{30,31} The most likely cause is the diminished collateral flow to the affected tissue. With a tandem occlusion as well as a contralateral acute ICA occlusion, our patient had an unfavorable prognosis on presentation.

Notably, some adult outcome measures (such as the 90-day outcome) may not be directly transferable to the pediatric population. Children with AIS have a much longer remaining life span than older adults with AIS. The extent of recovery resulting from brain plasticity and adaptation to stroke may be optimal in children, and a more aggressive approach may be warranted in pediatrics. It is postulated that neural injury at the early phases of brain development may be compensated more effectively because the pediatric brain is a dynamic environment with greater neuroplasticity.^{32,33} The pediatric brain responds to acute ischemia through

extensive neural network development and synaptogenesis, recruiting not only ipsilateral connections but also contralateral ones. Synaptogenesis is itself a dynamic process that includes synaptic pruning, eliminating weak synapses while enhancing stronger connections, a process that is most active before early adulthood.³⁴

In recent years, the critical time window for thrombectomy has been extended by up to 16–24 hours in a subpopulation of adult patients with rescuable penumbra.^{20,21} Moreover, pediatric cases with a time-window extension and favorable outcomes have been reported.³⁵ Accordingly, in the cohort in this study, case 4 was transferred to the angiography suite after the usual time window had passed, and in case 3, the groin puncture was at the upper limit of the time window when a very distal thrombectomy was performed.

As their underlying etiology, four patients had cardioembolic thrombus, one had thrombophilia, and two had intracranial aneurysms treated through flow diverter stent placement. With the exception of the latter etiology, these are known to be the most common risk factors for childhood AIS.^{36,37} With regard to the clinical results, we achieved functional independence with an mRS score ≤ 1 in five of our patients (83.3%) despite the drawbacks stated above. Given this more favorable outcome for stroke in children³⁸ and the similar patient outcomes in previous reports,⁶ we suggest that claims regarding the high risks of endovascular treatment for AIS in children (up to approximately 20 times of medical treatment as reported by Malik et al.³⁷) be reappraised, with treatment preferably provided by centers with sufficient pediatric experience.³⁸

Our study has limitations inherent to its retrospective nature as well as its small and heterogeneous patient cohort.

In conclusion, children that are likely to benefit from mechanical thrombectomy should be transferred urgently to centers with experience in pediatric neurology, anesthesia, and interventional neuroradiology for a case-by-case evaluation or modification of indications and methods for endovascular treatment.

Acknowledgments

We would like to thank the referring physicians and division of pediatric hematology as well as all medical staff involved in the care of our patients.

Footnotes

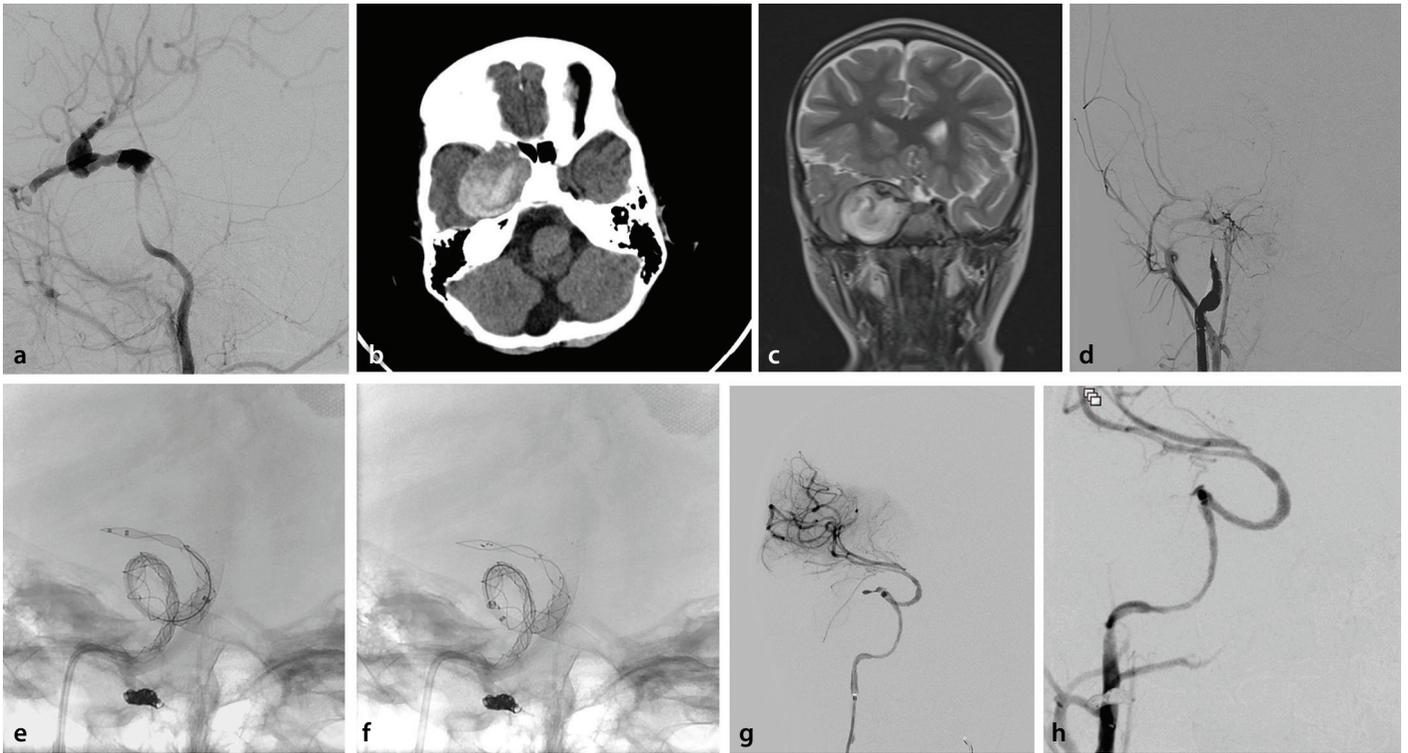
Conflict of interest disclosure

The authors declared no conflicts of interest.

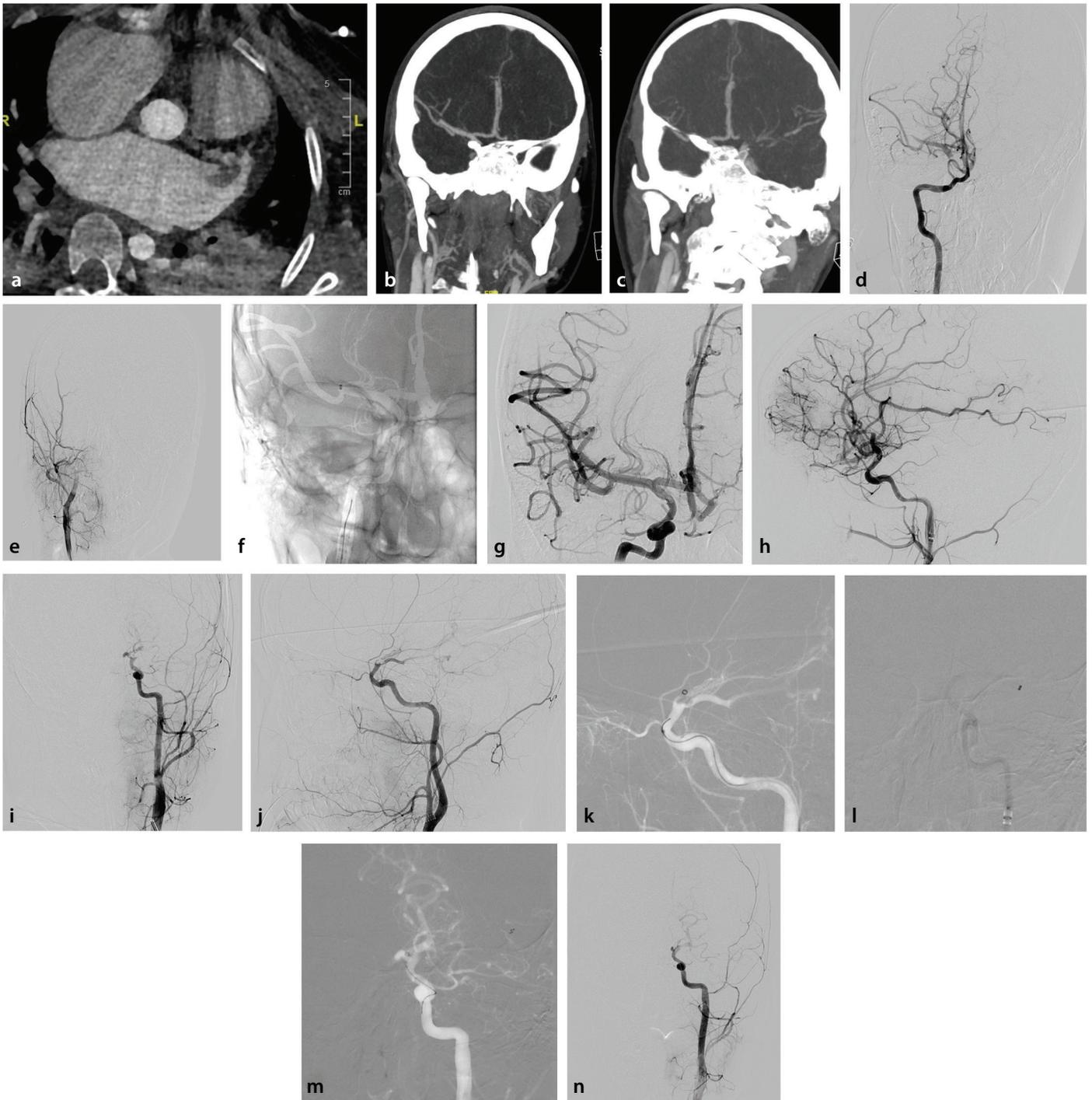
References

1. Mallick AA, O'Callaghan FJ. The epidemiology of childhood stroke. *Eur J Paediatr Neurol.* 2010;14(3):197-205. [\[CrossRef\]](#)
2. Wu C, Honarmand AR, Schnell S, et al. Age-related changes of normal cerebral and cardiac blood flow in children and adults aged 7 months to 61 years. *J Am Heart Assoc.* 2016;5(1):e002657. [\[CrossRef\]](#)
3. Mallick AA, Ganesan V, Kirkham FJ, et al. Diagnostic delays in paediatric stroke. *J Neurol Neurosurg Psychiatry.* 2015;86(8):917-921. [\[CrossRef\]](#)
4. Barry M, Hallam DK, Bernard TJ, Amli-Lefond C. What is the role of mechanical thrombectomy in childhood stroke? *Pediatr Neurol.* 2019;95:19-25. [\[CrossRef\]](#)
5. Sporns PB, Strater R, Minnerup J, et al. Feasibility, safety, and outcome of endovascular recanalization in childhood stroke: the save childs study. *JAMA Neurol.* 2020;77(1):25-34. [\[CrossRef\]](#)
6. Bhatia K, Kortman H, Blair C, et al. Mechanical thrombectomy in pediatric stroke: systematic review, individual patient data meta-analysis, and case series. *J Neurosurg Pediatr.* 2019:1-14. [\[CrossRef\]](#)
7. Meyer BC, Lyden PD. The modified National Institutes of Health Stroke Scale: its time has come. *Int J Stroke.* 2009;4(4):267-273. [\[CrossRef\]](#)
8. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J.* 1957;2(5):200-215. [\[CrossRef\]](#)
9. Berkhemer OA, Fransen PS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* 2015;372(1):11-20. Erratum in: *N Engl J Med.* 2015;372(4):394. [\[CrossRef\]](#)
10. Goyal M, Demchuk AM, Menon BK, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med.* 2015;372(11):1019-1030. [\[CrossRef\]](#)
11. Campbell BC, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med.* 2015;372(11):1009-1018. [\[CrossRef\]](#)
12. Madaelil TP, Kansagra AP, Cross DT, Moran CJ, Derdeyn CP. Mechanical thrombectomy in pediatric acute ischemic stroke: clinical outcomes and literature review. *Interv Neuroradiol.* 2016;22(4):426-431. [\[CrossRef\]](#)
13. Bhogal P, Hellstern V, AlMatter M, et al. Mechanical thrombectomy in children and adolescents: report of five cases and literature

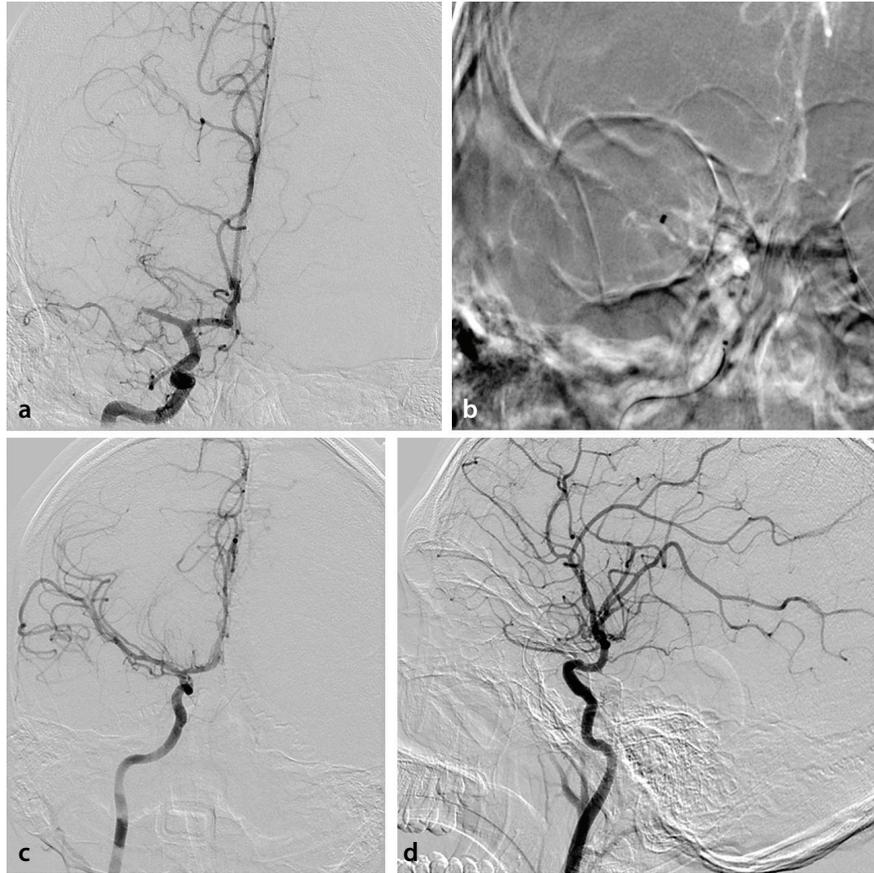
- review. *Stroke Vasc Neurol*. 2018;3(4):245-252. [\[CrossRef\]](#)
14. Sporns PB, Kemmling A, Hanning U, et al. Thrombectomy in childhood stroke. *J Am Heart Assoc*. 2019;8(5):e011335. [\[CrossRef\]](#)
 15. Dicipinigaitis AJ, Gandhi CD, Pisapia J, et al. Endovascular thrombectomy for pediatric acute ischemic stroke. *Stroke*. 2022;53(5):1530-1539. [\[CrossRef\]](#)
 16. Bhatia KD, Chowdhury S, Andrews I, et al. Association between thrombectomy and functional outcomes in pediatric patients with acute ischemic stroke from large vessel occlusion. *JAMA Neurol*. 2023;80(9):910-918. Erratum in: *JAMA Neurol*. 2023;80(9):1004. [\[CrossRef\]](#)
 17. Bilgin C, Ibrahim M, Azzam AY, et al. Mechanical thrombectomy for pediatric large vessel occlusions: a systematic review and meta-analysis. *Clin Neuroradiol*. 2023;33(3):635-644. [\[CrossRef\]](#)
 18. Peker A, Arsava EM, Topçuoğlu MA, Arat A. Catch plus thrombectomy device in acute stroke: initial evaluation. *J Neurointerv Surg*. 2017;9(12):1214-1218. [\[CrossRef\]](#)
 19. Tona C, Nosadini M, Pelizza MF, et al. Cardiac myxoma as a rare cause of pediatric arterial ischemic stroke: case report and literature review. *Neuropediatrics*. 2020;51(6):389-396. [\[CrossRef\]](#)
 20. Albers GW, Marks MP, Kemp S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med*. 2018;378(8):708-718. [\[CrossRef\]](#)
 21. Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2018;378(1):11-21. [\[CrossRef\]](#)
 22. Arat YO, Arat A, Aydin K. Angiographic morphometry of internal carotid artery circulation in turkish children. *Turk Neurosurg*. 2015;25(4):608-616. [\[CrossRef\]](#)
 23. Samaniego EA, Dandapat S, Roa JA, Zanaty M, Nakagawa D, Hasan DM. Mechanical thrombectomy of acutely occluded flow diverters. *Oper Neurosurg (Hagerstown)*. 2019;17(5):491-496. [\[CrossRef\]](#)
 24. Kaesmacher J, Meyer L, Styczen H, et al. Primary multivessel occlusions treated with mechanical thrombectomy: a multicenter analysis and systemic literature review. *Stroke*. 2020;51(9):232-237. [\[CrossRef\]](#)
 25. Gerstl L, Olivieri M, Heinen F, et al. Successful mechanical thrombectomy in a three-year-old boy with cardioembolic occlusion of both the basilar artery and the left middle cerebral artery. *Eur J Paediatr Neurol*. 2016;20(6):962-965. [\[CrossRef\]](#)
 26. Zhang J, Yang Y, Sun H, Xing Y. Hemorrhagic transformation after cerebral infarction: current concepts and challenges. *Ann Transl Med*. 2014;2(8):81. [\[CrossRef\]](#)
 27. Tu HT, Campbell BC, Christensen S, et al. Worse stroke outcome in atrial fibrillation is explained by more severe hypoperfusion, infarct growth, and hemorrhagic transformation. *Int J Stroke*. 2015;10(4):534-540. [\[CrossRef\]](#)
 28. Rubiera M, Ribo M, Delgado-Mederos R, et al. Tandem internal carotid artery/middle cerebral artery occlusion: an independent predictor of poor outcome after systemic thrombolysis. *Stroke*. 2006;37(9):2301-2305. [\[CrossRef\]](#)
 29. Parra-Farinas C, Dmytriw AA, Delgado-Alvarez I, et al. Primary endovascular treatment for acute ischemic stroke in teenage patients: a short case series. *Neuroradiology*. 2020;62(7):851-860. [\[CrossRef\]](#)
 30. Zhang K, Li T, Tian J, et al. Subtypes of anterior circulation large artery occlusions with acute brain ischemic stroke. *Sci Rep*. 2020;10(1):3442. [\[CrossRef\]](#)
 31. Maus V, Behme D, Borggrefe J, et al. Carotid artery stenosis contralateral to acute tandem occlusion: an independent predictor of poor clinical outcome after mechanical thrombectomy with concomitant carotid artery stenting. *Cerebrovasc Dis*. 2018;45(1-2):10-17. [\[CrossRef\]](#)
 32. Kolb B, Gibb R. Brain plasticity and behaviour in the developing brain. *J Can Acad Child Adolesc Psychiatry*. 2011;20(4):265-276. [\[CrossRef\]](#)
 33. Malone LA, Felling RJ. Pediatric stroke: unique implications of the immature brain on injury and recovery. *Pediatr Neurol*. 2020;102:3-9. [\[CrossRef\]](#)
 34. Petanjek Z, Judaš M, Šimic G, et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc Natl Acad Sci U S A*. 2011;108(32):13281-13286. [\[CrossRef\]](#)
 35. Kulhari A, Dorn E, Pace J, et al. Acute ischemic pediatric stroke management: an extended window for mechanical thrombectomy? *Front Neurol*. 2017;8:634. [\[CrossRef\]](#)
 36. Felling RJ, Sun LR, Maxwell EC, Goldenberg N, Bernard T. Pediatric arterial ischemic stroke: Epidemiology, risk factors, and management. *Blood Cells Mol Dis*. 2017;67:23-33. [\[CrossRef\]](#)
 37. Malik P, Patel UK, Kaul S, et al. Risk factors and outcomes of intravenous tissue plasminogen activator and endovascular thrombectomy utilization amongst pediatrics acute ischemic stroke. *Int J Stroke*. 2021;16(2):172-183. [\[CrossRef\]](#)
 38. Sun LR, Harrar D, Drocton G, et al. Mechanical thrombectomy for acute ischemic stroke: considerations in children. *Stroke*. 2020;51(10):3174-3181. [\[CrossRef\]](#)



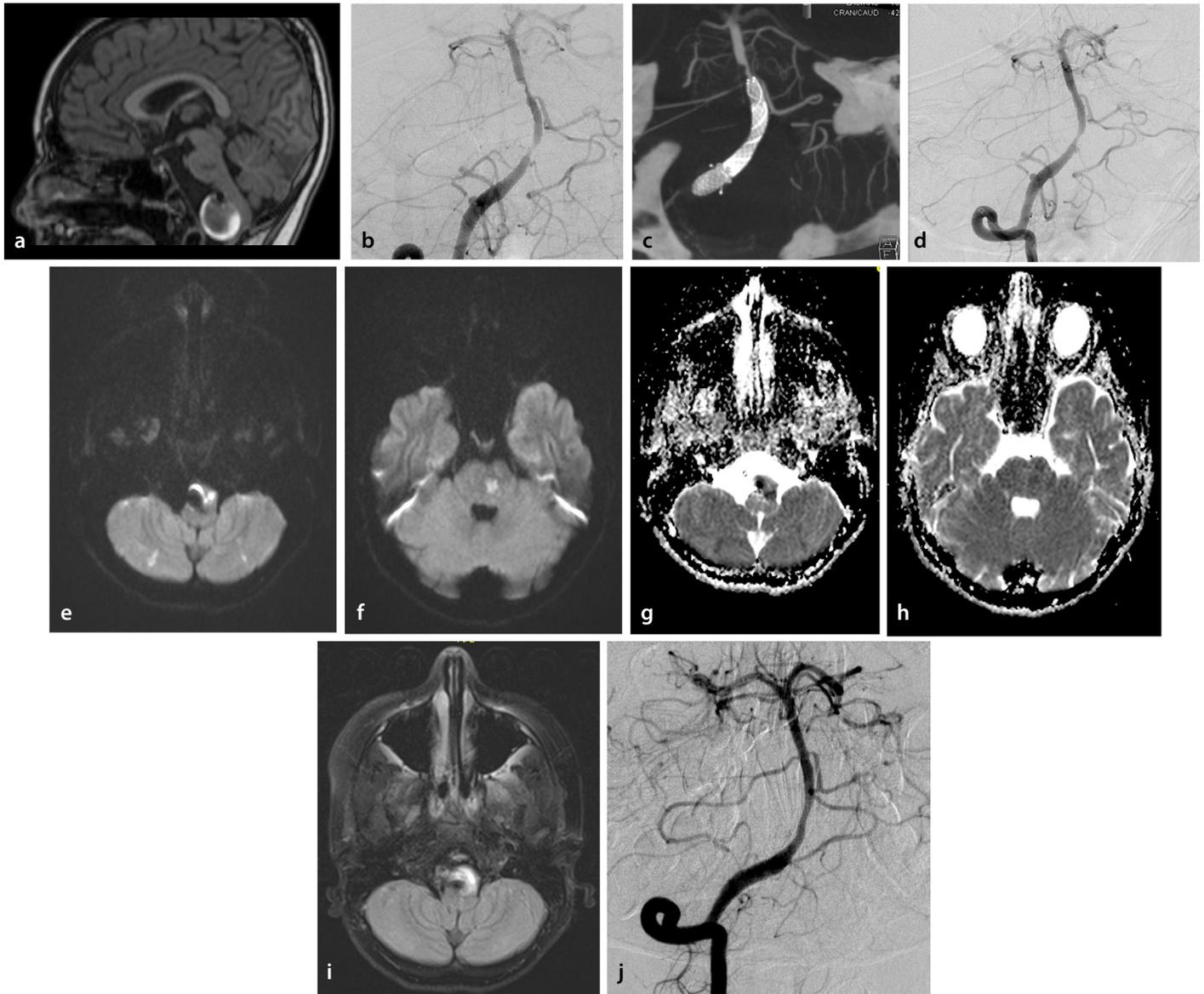
Supplementary Figure 1. Preoperative right carotid angiogram (a), axial CT (b) and coronal T2-weighted MR images (c) show a partially thrombosed giant cavernous ICA aneurysm distal to a preocclusive stenosis. Right carotid angiogram performed 72 hours after the endovascular treatment of the aneurysm with a flow diverter and stent (d) reveals acute stent occlusion with arterial occlusion at the distal cervical ICA. After passing the occlusion with a coaxial system of catheters (e), a stent retriever device was deployed (f) for thrombectomy. Control angiograms (g, h) demonstrate successful recanalization of the MCA, the ipsilateral ACA was supplied via the anterior communicating artery. CT, computed tomography; MR, magnetic resonance; ICA, internal carotid artery; MCA, middle cerebral artery.



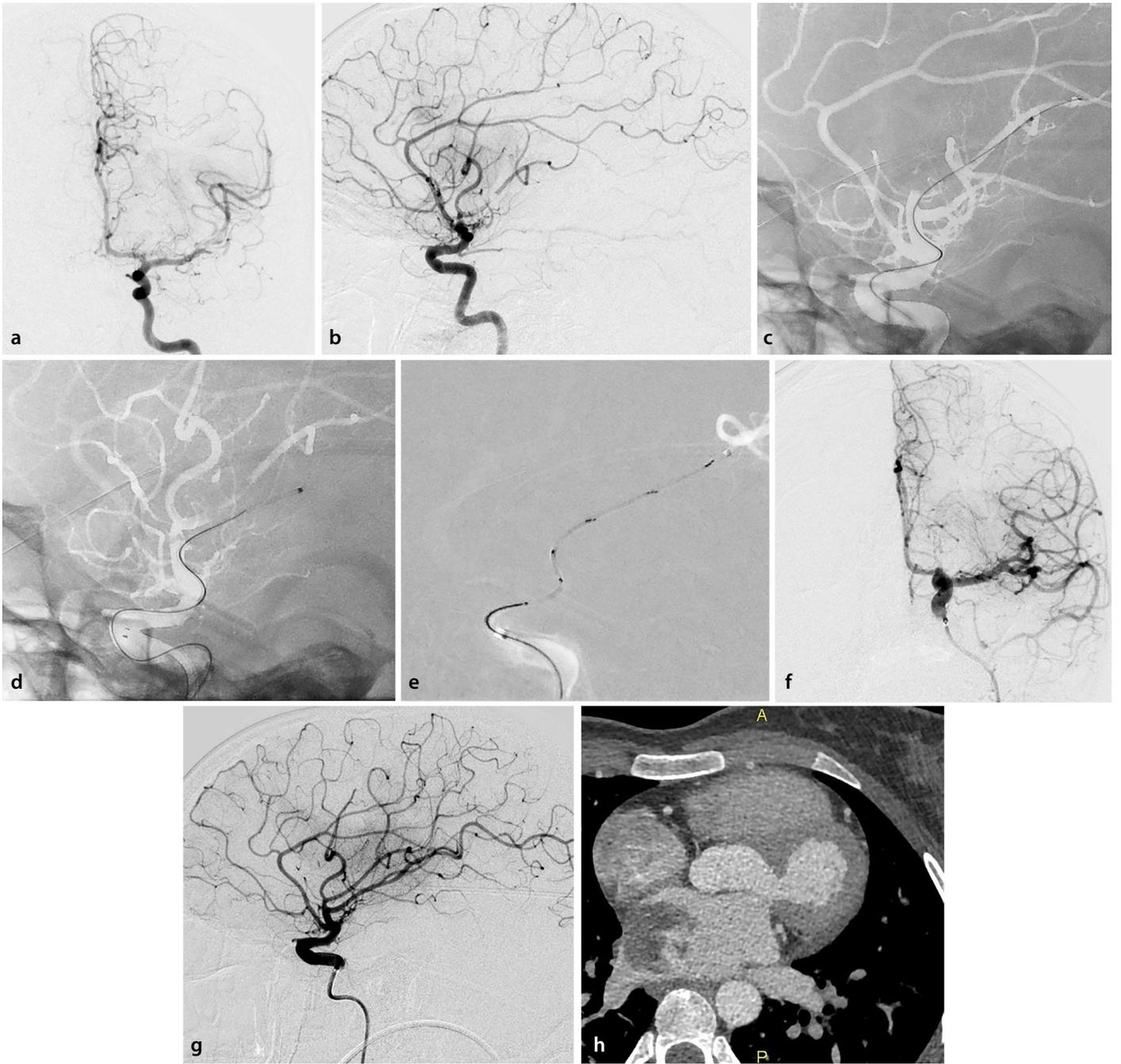
Supplementary Figure 2. A hypodense filling defect consistent with thrombus in the left atrial appendage (a), a right cervical ICA/MCA tandem occlusion (b) as well as a left ICA terminus occlusion (c) are seen on head and neck CT angiography. Vertebral angiogram (d) shows reconstruction of the right intracranial anterior circulation via the posterior communicating artery and the right MCA occlusion distal to the cervical occlusion is revealed. Right cervical ICA arteriogram (e) demonstrates the cervical occlusion. After cervical ICA recanalization, aspiration thrombectomy was undertaken (f) and resulted in recanalization of the right MCA territory as evident on AP and lateral projections of carotid angiograms (g, h). Then, following left ICA angiogram (i, j), aspiration (k, l) and stent retriever (m) thrombectomies were performed. Yet, as noted on the final left ICA angiograms (n) the attempts were futile. ICA, internal carotid artery; MCA, middle cerebral artery; CT, computed tomography.



Supplementary Figure 3. Right MCA occlusion is seen on the initial right carotid artery injection (a). After aspiration thrombectomy (b), final angiograms show recanalized MCA territory (c, d). MCA, middle cerebral artery.



Supplementary Figure 4. Sagittal pre-contrast T1-weighted MR image (a) demonstrates a large distal V4 segment vertebral artery aneurysm which was previously treated with flow diversion. Right vertebral injection (b) shows a pre-occlusive filling defect in the basilar artery distal to the flow diverter stent consistent with acute thrombosis. Maximum intensity projection reformatted image of 3D rotational angiogram (c) confirms the acute thrombus in the basilar artery. Postprocedural right vertebral angiogram (d) illustrates complete removal of the thrombus. Diffusion-weighted images (e, f) and corresponding ADC maps (g, h) reveal acute ischemic lesions at the brain stem and bilateral cerebellar hemispheres. Partially thrombosed aneurysm and compression of adjacent structures are appreciated on axial FLAIR image (i). Follow-up DSA at 5 months (j) shows an intact flow diverter and a patent basilar artery. MR, magnetic resonance; ADC, apparent diffusion coefficient; FLAIR, fluid-attenuated inversion recovery; DSA, digital subtraction angiography.



Supplementary Figure 5. Preprocedural AP (a) and lateral (b) projection carotid angiograms show partially occluded MCA. After superselective microcatheterization of both MCA trunks (c, d) and subsequent stent retriever thrombectomy (e), recanalization of the left MCA territory can be seen on AP (f) and lateral (g) projection angiograms. Axial contrast enhanced cardiac CT (h) of the patient reveals a hypodense filling defect in the left atrium. AP, anterior–posterior; MCA, middle cerebral artery.



Long-term outcomes of catheter-directed sclerotherapy for ovarian endometrioma

Jae Hwan Lee^{1-3*}

Jihyung Yoon^{4*}

Chong-Ho Lee^{1*}

Kun Yung Kim¹

Chang Jin Yoon¹⁻³

Minuk Kim⁵

Seul Ki Kim⁴

¹Seoul National University Bundang Hospital, Department of Radiology, Gyeonggi-do, South Korea

²Seoul National University College of Medicine, Department of Radiology, Seoul, South Korea

³Seoul National University Medical Research Center, Institute of Radiation Medicine, Seoul, Republic of Korea

⁴Seoul National University Bundang Hospital, Department of Obstetrics and Gynecology, Gyeonggi-do, South Korea

⁵Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Department of Radiology, Seoul, Republic of Korea

*Joint first authors

Corresponding author: Seul Ki Kim

E-mail: drksk80@gmail.com

Received 06 June 2024; revision requested 08 July 2024; accepted 17 July 2024.



Epub: 01.10.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242874

PURPOSE

Although favorable results have been reported on catheter-directed sclerotherapy (CDS) for ovarian endometrioma, a thorough evaluation of its long-term efficacy is lacking. This study evaluates the long-term efficacy and safety of CDS with 99% ethanol for treatment of ovarian endometrioma.

METHODS

Between January 2020 and February 2022, data from 33 consecutive patients with symptomatic ovarian endometriomas who underwent CDS were retrospectively evaluated. All patients underwent pre-procedural and 6- and 12-month post-procedural ultrasonography. To assess the effect on ovarian reserve, serum anti-Müllerian hormone (AMH) levels were measured before and after the procedure. Procedure-related complications were also assessed.

RESULTS

The mean volume of endometriomas decreased from 80.22 ± 66.43 to 0.73 ± 1.10 mL ($P < 0.001$), and the mean percentage of volume reduction was $98.99\% \pm 1.53\%$. No recurrences were observed during the follow-up period. In patients whose serum AMH levels were monitored for 1 year, no significant change in AMH level before and after CDS was observed (3.07 ± 1.81 vs. 2.72 ± 2.02 ng/mL, $P = 0.190$). One patient complained of moderate abdominal pain after CDS, which was conservatively managed.

CONCLUSION

CDS remained safe and effective in treating ovarian endometrioma at the 1-year follow-up. Ovarian function after CDS was well preserved.

CLINICAL SIGNIFICANCE

CDS is a safe and effective treatment option for patients with ovarian endometrioma without compromising ovarian function.

KEYWORDS

Endometrioma, interventional, ovary, percutaneous, sclerotherapy

Endometriosis is one of the most common health problems in women of reproductive age.^{1,2} Approximately 10% of reproductive aged women, 30% of women with infertility, and 82% of women with chronic pelvic pain have endometriosis.³ Endometrial tissue lining the wall of an ovarian cyst is defined as an ovarian endometrioma;⁴⁻⁹ it consists of fluid-filled sacs that develop from the accumulation of menstrual contents^{4,5,9,10} and is observed in approximately 45% of patients with endometriosis.^{1,5} Ovarian endometrioma is known to damage adjacent ovarian tissues because it contains reactive oxygen species, proteolytic enzymes, and inflammation-mediating factors.¹¹ This results in fibrosis, a decline in

stromal cells, and a decrease in vascularity, accompanied by follicular loss.¹⁰ A previous meta-analysis of 968 patients revealed that serum anti-Müllerian hormone (AMH), a surrogate marker of ovarian reserve, was lower in patients with ovarian endometrioma than in those without endometrial cysts.¹¹

Currently, the standard treatment for ovarian endometrioma has been laparoscopic cystectomy;^{1,4,5,9,12-14} however, parenchymal injury during surgical excision inevitably results in a decline in the ovarian reserve.^{1,9,12,13,15} Furthermore, relapse is relatively frequent; a pooled analysis of 23 studies involving 4,368 patients demonstrated that the recurrence rate of ovarian endometrioma following surgery was 21.5% and 50% at 2 and 5 years, respectively.¹⁶ Therefore, the use of non-surgical treatments such as aspiration or needle-directed sclerotherapy (NDS) has been advised to treat lesions while minimizing ovarian damage.^{4,6,7,10,17}

NDS uses needles to access the cyst, remove menstrual debris-like contents, and inject sclerosants into the cavity. However, it has inherent technical drawbacks, including needle instability during the process, poor aspiration/injection capability, and the possibility of needle dislodgement and spillage of endometrial tissue debris or sclerosant into the peritoneal cavity.⁶⁻⁸ Thus, the outcomes of NDS have been variable, and the recurrence rate following NDS is relatively high, up to 62.5%.^{10,17,18}

To overcome the limitations of NDS, catheter-directed sclerotherapy (CDS) was developed.⁶⁻⁸ In a small prospective study,⁶ the technical advantages translated into superior short-term clinical results with little harm to the ovarian reserve. However, only a few papers have reported on the procedure, and data regarding the long-term outcomes of CDS and recurrence after the procedure are lacking.⁶⁻⁸ Therefore, this study aims to evaluate the long-term efficacy and safety of CDS with 99% ethanol for ovarian endometriomas.

Main points

- Catheter-directed sclerotherapy (CDS) for ovarian endometrioma shows durable outcomes at the 1-year follow-up.
- CDS does not seem to result in a decline in ovarian function as measured by anti-Müllerian hormone.
- Procedure-related complications are minimal, suggesting a favorable safety profile for CDS.

Methods

Study design

The Seoul National University Bundang Hospital Institutional Review Board approved this retrospective study (decision no: B-2003-602-301; approved on: 05-10-2023) and waived the requirement for informed consent due to the study design. Data of patients who underwent CDS for ovarian endometrioma between January 2020 and February 2022 were obtained from the institution's electronic medical record system (date of access to patient data: 01/20/2023). The inclusion criteria for the CDS procedure were: (a) age >18 years; (b) symptoms suggesting endometriosis (dyspareunia, dysmenorrhea, and lower abdominal or pelvic pain) (c) largest diameter of cyst ≥ 3 cm; (d) ultrasound features suggestive of endometrioma; (e) no evidence of malignancy on contrast-enhanced magnetic resonance or computed tomography images; and (f) serum cancer antigen 125 (CA-125) levels <200 U/mL. Patients who lacked baseline serum AMH levels or were lost to follow-up less than 6 months after CDS were excluded (Figure 1). Following the CDS procedure, patients were required to continue taking dienogest or oral contraceptive pills for at least 2 years to prevent recurrence.¹⁸ During the follow-up period, serum AMH and CA-125 levels were monitored, and the largest diameter and mean volume of the lesion were measured using transvaginal ultrasonography.

Procedures

One board-certified radiologist (J.H.L) with 10 years of experience in interventional radiology and pelvic imaging performed the CDS. The procedure was performed as

previously described.^{6,7} Preprocedural ultrasonography was performed 1–2 weeks prior to CDS to evaluate the lesion characteristics, lesion size/volume, and access route (transvaginal or transabdominal). For transvaginal access, the target lesion was punctured using an 18-G 20 cm needle (Chiba biopsy needle, Bloomington, Cook, USA), followed by the placement of a 0.035-inch hydrophilic guidewire (Radifocus, Terumo, Japan) and 8.5-F drainage catheter (Dawson-Mueller Drainage catheter; Cook). After drainage of the contents and irrigation with normal saline, 2–3 cc of contrast was infused into the lesion to evaluate any signs of leakage or rupture.⁷ If there was no leakage or rupture, 99% ethanol was infused carefully at 25% of the drained volume, with a maximum dose of <100 mL.

The patient's position (clockwise rotation from supine to left decubitus, prone, and right decubitus) was then changed every 5 min. Finally, ethanol was aspirated and the catheter was removed. To rule out malignancy, the aspirated contents were sent to pathology for cytology evaluation.^{6,7} All patients underwent the procedure on the day of admission and were discharged the following day.

Follow-up

All patients underwent ultrasonography at 6 and 12 months after CDS to follow up on cyst size and any recurrence. To evaluate cyst size, the volume of the ovarian endometrioma was calculated using the formula for an ellipsoid, with the length and width of the lesion measured by ultrasound. Volume reduction was calculated by the percentage change in volume from pre-procedure to post-procedure measurements. Serum AMH levels were

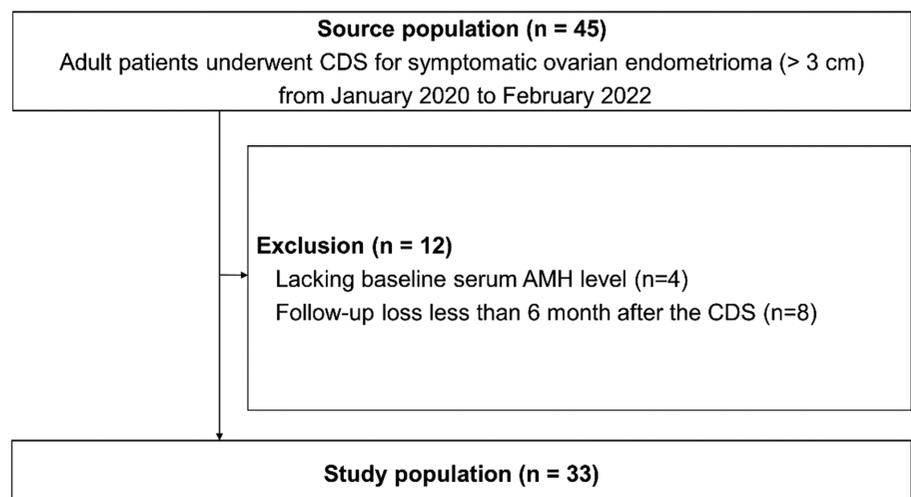


Figure 1. Patient flowchart. CDS, catheter-directed sclerotherapy; AMH, anti-Müllerian hormone.

tested 6 and 12 months after CDS to evaluate the effect of CDS on ovarian reserve. Serum CA-125 levels, which reflect endometrioma burden, were assessed at the same time points. Complications related to the procedure were recorded for each instance.

Statistical analysis

Data are presented as mean \pm standard deviation. Changes in endometrioma volume and serum AMH and CA-125 levels before and after CDS were analyzed using the paired t-test or Wilcoxon signed-rank test, depending on the normality of the variable. The PASW 18.0 software (IBM, Armonk, NY, USA) was used for all statistical analyses. A *P* value of <0.050 was considered statistically significant. Recurrence was defined as the return of symptoms and/or newly developed detectable endometrioma on follow-up ultrasound.

Results

In total, 33 of the 45 patients originally chosen were enrolled in this study; those without baseline serum AMH levels (*n* = 9) and those who were lost to follow-up less than 6 months after CDS (*n* = 3) were excluded. Table 1 summarizes the baseline characteristics of patients and lesions. The mean AMH level was 2.99 ± 2.16 ng/mL. The technical success rate of CDS was 100%. The hospitalization period for all patients was 2 days. One patient experienced moderate abdominal pain (visual analog scale score: 5), but it was resolved with conservative management. All cytological analyses of the aspirates were negative for malignant cells.

The largest diameter and mean volume of the endometriomas continuously decreased for 1 year after CDS (Table 2). The mean volume reduction percentage at 12 months after CDS was $98.99 \pm 1.54\%$. The mean follow-up period was 12 months (range: 8.36–17.5 months), with data focusing primarily on the 12-month outcomes. No recurrences were observed during the follow-up period.

The serum CA-125 level significantly decreased at 6 months and this was maintained at 1 year (*P* = 0.010); however, there was no significant difference in the serum AMH levels before and 1 year after CDS (*P* = 0.302) (Figure 2).

Discussion

This study aimed to evaluate the long-term efficacy and safety of CDS with 99% ethanol for the treatment of ovarian endo-

metriomas. We found that the endometrioma volume rapidly decreased in the first 6 months, with a 99% reduction maintained over a year. The trends in serum CA-125 levels were consistent with the changes in endometrioma volume, reflecting a decrease in the burden of disease. These findings align with previous reports.^{6,7} As the catheter is securely located, the risk of potential spillage of the contents is low, and patients can change their position to enhance the cyst wall exposure to ethanol.^{6,7,10}

Interestingly, the treated endometriomas had not recurred by the end of the follow-up period. No recurrence after CDS has been reported in the literature, including this study.^{6,7} In a review article comparing

11 studies on recurrence rates after ablation or cystectomy, recurrence rates over 1 year ranged from 4.4% to 37.0%.⁴ In previous articles on CDS, as well as in this study, patients were not prevented from continuing medication after the CDS treatment. Given the nature of ovarian endometriomas with frequent recurrence, CDS combined with hormonal treatment seems to suppress the recurrence of endometrioma effectively.

In this study, serum AMH, which reflects the ovarian reserve, was well preserved after CDS. Although laparoscopic cystectomy is the current standard treatment, it has several disadvantages, including the risk of general anesthesia, perilesional adhesions, and decline in ovarian function.^{1,7,15} In addition to

Table 1. Patient and lesion characteristics

Characteristic	Value
Age (years)	31.22 \pm 4.10
CA-125 (U/mL)	61.11 \pm 50.92
AMH (ng/mL)	2.99 \pm 2.16
Largest diameter (cm)	6.21 \pm 1.26
Mean volume (cm ³)	80.22 \pm 66.43
Unilateral endometrioma [n (%)]	32 (96.9%)

Data are presented as mean \pm standard deviation. CA-125, cancer antigen 125; AMH, anti-Müllerian hormone.

Table 2. Long-term outcomes of catheter-directed sclerotherapy

	Before CDS	6 months after CDS	12 months after CDS	<i>P</i> value
Largest diameter (cm)	6.21 \pm 1.26	1.65 \pm 1.03	0.86 \pm 0.94	$<0.010^*$ $<0.010^\dagger$
Mean volume (mL)	80.22 \pm 66.43	2.27 \pm 2.51	0.73 \pm 1.11	$<0.010^*$ $<0.010^\dagger$

Data are presented as mean \pm standard deviation. *Difference between the value before and 6 months after CDS; †Difference between the value before and 12 months after CDS. CDS, catheter-directed sclerotherapy.

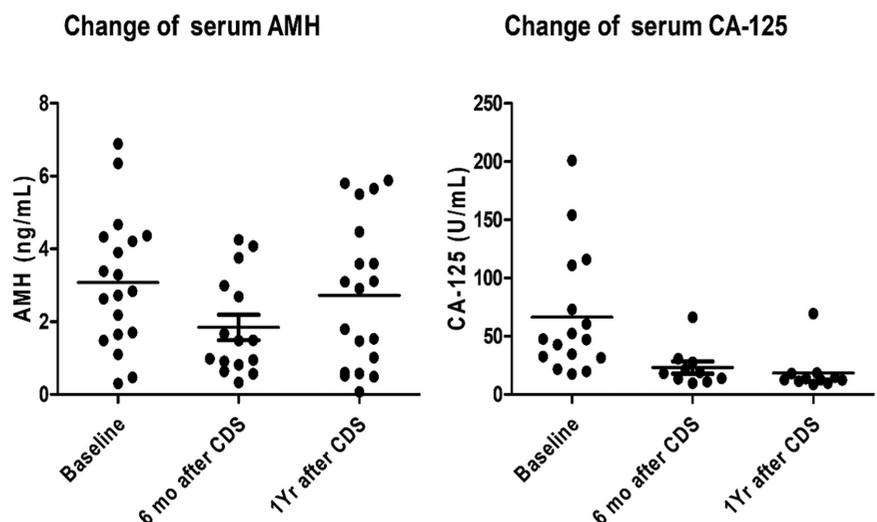


Figure 2. The serial change of serum AMH and CA-125 concentration after CDS. AMH, anti-Müllerian hormone; CA-125, cancer antigen 125; CDS, catheter-directed sclerotherapy; mo, month; Yr, year.

the inevitable removal of ovarian follicles adjacent to the endometrioma, hemostatic cauterization may add collateral damage to the ovarian circulation and cause further follicular loss.¹⁵ In CDS, catheters can be accurately placed in the lesion, and unintended injury to the adjacent healthy ovarian parenchyma can be minimized.

Previous studies have demonstrated AMH changes up to 6 months after the procedure, whereas this study included results up to 12 months, providing the most extended follow-up data available. Furthermore, this study has the advantage of having one skilled expert performing all procedures, eliminating the possibility of variation among different operators.

This study has several limitations. Due to its retrospective design, consistent and comparable data on intraprocedural and postprocedural pain were difficult to obtain. Nevertheless, medical records indicate that only one patient required additional analgesics for abdominal pain beyond conventional post-procedural management, suggesting the procedure was generally well-tolerated. Additionally, having a single operator perform all procedures and the absence of a control group limit the generalizability of the findings. Future prospective studies with standardized pain assessments and comparative analyses are needed.

In conclusion, CDS for ovarian endometrioma showed favorable safety and long-term outcomes without recurrence. The ovarian reserve was well preserved during the follow-up period of 1 year.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was granted by the Seoul National University Bundang Hospital research fund (06-2020-0200).

References

1. Hwang H, Chung YJ, Lee SR, et al. Clinical evaluation and management of endometriosis: guideline for Korean patients from Korean Society of Endometriosis. *Obstet Gynecol Sci*. 2018;61(5):553-564. [\[CrossRef\]](#)
2. Eskenazi B, Warner ML. Epidemiology of endometriosis. *Obstet Gynecol Clin North Am*. 1997;24(2):235-258. [\[CrossRef\]](#)
3. Fauconnier A, Fritel X, Chapron C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. *Gynecol Obstet Fertil*. 2009;37(1):57-69. [\[CrossRef\]](#)
4. Jee BC. Efficacy of ablation and sclerotherapy for the management of ovarian endometrioma: a narrative review. *Clin Exp Reprod Med*. 2022;49(2):76-86. [\[CrossRef\]](#)
5. Chapron C, Vercellini P, Barakat H, Vieira M, Dubuisson JB. Management of ovarian endometriomas. *Hum Reprod Update*. 2002;8(6):591-597. [\[CrossRef\]](#)
6. Han K, Seo SK, Kim MD, et al. Catheter-directed sclerotherapy for ovarian endometrioma: short-term outcomes. *Radiology*. 2018;289(3):854-859. [\[CrossRef\]](#)
7. Koo JH, Lee I, Han K, et al. Comparison of the therapeutic efficacy and ovarian reserve between catheter-directed sclerotherapy and surgical excision for ovarian endometrioma. *Eur Radiol*. 2021;31(1):543-548. [\[CrossRef\]](#)
8. Lee JK, Ahn SH, Kim HI, et al. Therapeutic efficacy of catheter-directed ethanol sclerotherapy and its impact on ovarian reserve in patients with ovarian endometrioma at risk of decreased ovarian reserve: a preliminary study. *J Minimally Invasive Gynecol*. 2022;29(1):317-323. [\[CrossRef\]](#)
9. Choi SH, Kim S, Lee SW, et al. Recurrence, reoperation, pregnancy rates, and risk factors for recurrence after ovarian endometrioma surgery: long-term follow-up of 756 women. *Yonsei Med J*. 2023;64(3):204-212. [\[CrossRef\]](#)
10. Cohen A, Almog B, Tulandi T. Sclerotherapy in the management of ovarian endometrioma: systematic review and meta-analysis. *Fertil Steril*. 2017;108(1):117-124. [\[CrossRef\]](#)
11. Sanchez AM, Viganò P, Somigliana E, Panina-Bordignon P, Vercellini P, Candiani M. The distinguishing cellular and molecular features of the endometriotic ovarian cyst: from pathophysiology to the potential endometrioma-mediated damage to the ovary. *Hum Reprod Update*. 2014;20:217-230. [\[CrossRef\]](#)
12. Muzii L, Di Tucci C, Di Feliciano M, et al. Antimüllerian hormone is reduced in the presence of ovarian endometriomas: a systematic review and meta-analysis. *Fertil Steril*. 2018;110(5):932-940. [\[CrossRef\]](#)
13. Hue H, Kim S, Choi J, et al. Previous ovarian surgery increases the risk of tubal factor infertility. *Clin Exp Obstet Gynecol*. 2020;47(1):84-88. [\[CrossRef\]](#)
14. Lee D, Kim SK, Lee JR, Jee BC. Management of endometriosis-related infertility: considerations and options. *Clin Exp Reprod Med*. 2020;47(1):1-11. [\[CrossRef\]](#)
15. Ata B, Turkgeldi E, Seyhan A, Urman B. Effect of hemostatic method on ovarian reserve following laparoscopic endometrioma excision; comparison of suture, hemostatic sealant, and bipolar desiccation. A systematic review and meta-analysis. *J Minim Invasive Gynecol*. 2015;22(3):363-372. [\[CrossRef\]](#)
16. Guo SW. Recurrence of endometriosis and its control. *Hum Reprod Update*. 2009;15(4):441-461. [\[CrossRef\]](#)
17. García-Tejedor A, Castellarnau M, Ponce J, Fernández ME, Burdio F. Ethanol sclerotherapy of ovarian endometrioma: A safe and effective minimal invasive procedure. Preliminary results. *Eur J Obstet Gynecol Reprod Biol*. 2015;187:25-29. [\[CrossRef\]](#)
18. Hsieh CL, Shiau CS, Lo LM, Hsieh TT, Chang MY. Effectiveness of ultrasound-guided aspiration and sclerotherapy with 95% ethanol for treatment of recurrent ovarian endometriomas. *Fertil Steril*. 2009;91(6):2709-2713. [\[CrossRef\]](#)



Combining transarterial chemoembolization, radiofrequency ablation, and iodine-125 seed implantation for recurrent hepatocellular carcinoma post-hepatectomy

Yong Zhong¹

Li Wang¹

Weibin Dan¹

Dan Liang²

¹Tongcheng County People's Hospital, Department of Oncology, Tongcheng, China

²Tongcheng County People's Hospital, Department of Ultrasound Medicine, Tongcheng, China

PURPOSE

This study aimed to evaluate the efficacy and safety of transarterial chemoembolization (TACE) combined with radiofrequency ablation (RFA) and iodine-125 (125I) seed implantation (TACE-RFA-125I) for recurrent hepatocellular carcinoma (HCC) after hepatectomy.

METHODS

The study retrospectively analyzed patients with recurrent HCC who received TACE-RFA-125I or TACE-RFA treatment in our institution between January 2013 and January 2023. Overall survival (OS), progression-free survival (PFS), and recurrence were compared between the two groups.

RESULTS

A total of 187 patients were enrolled in this study, with 105 in the TACE-RFA-125I group and 82 in the TACE-RFA group. There were 67 men and 15 women in the TACE-RFA group, with an average age of 55.4 ± 10.9 years, and 93 men and 12 women in the TACE-RFA-125I group, with an average age of 55.5 ± 10.7 years. The TACE-RFA-125I group exhibited a significantly improved survival benefit compared with the TACE-RFA group (median OS: 49 months vs. 32 months, $P < 0.001$; median PFS: 24 months vs. 16 months, $P < 0.001$). The univariate and multivariate analyses revealed that TACE-RFA-125I was a protective factor for OS and PFS. A total of 32 patients in the TACE-RFA group experienced recurrence during follow-up, with local recurrence in 12 cases, intrahepatic recurrence in 10 cases, and extrahepatic metastases in 10 cases. A total of 28 patients in the TACE-RFA-125I group experienced recurrence, 6 with local recurrence, 12 with intrahepatic recurrence, and 10 with extrahepatic metastases. No procedure-related deaths occurred in this study.

CONCLUSION

In patients with recurrent HCC, TACE-RFA-125I demonstrates promising tumor control and acceptable safety.

CLINICAL SIGNIFICANCE

This study provides promising clinical guidance for patients with recurrent HCC after hepatectomy and is expected to provide beneficial strategies for the treatment of this disease.

KEYWORDS

Hepatocellular carcinoma, hepatectomy, radiofrequency ablation, transarterial chemoembolization, iodine-125 seeds

Corresponding author: Weibin Dan, Dan Liang

E-mail: 13980444162@qq.com, 120750571@qq.com

Received 23 April 2024; revision requested 09 June 2024; accepted 22 July 2024.



Epub: 26.11.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242814

You may cite this article as: Zhong Y, Wang L, Dan W, Liang D. Combining transarterial chemoembolization, radiofrequency ablation, and iodine-125 seed implantation for recurrent hepatocellular carcinoma post-hepatectomy. *Diagn Interv Radiol.* 2025;31(3):253-258.

Hepatocellular carcinoma (HCC) is the third most common cause of cancer-related death worldwide, and treatment options remain limited.¹ The Barcelona Clinical Liver Cancer staging system is now a widely accepted tool for the treatment allocation of patients with HCC globally.² For patients with early-stage HCC, hepatectomy is one of the preferred curative treatments. Nonetheless, up to 70% of patients with HCC have reported tumor recurrence within 5 years of hepatectomy, of which 61.4%–83.3% of tumors recurred within 2 years.³ Until now, no consensus has been reached on the treatment of patients with HCC with recurrence after hepatectomy.

Previous studies have demonstrated that transarterial chemoembolization (TACE) combined with radiofrequency ablation (RFA) can cause extensive tumor necrosis and achieve significant survival benefits.^{4,5} The embolization of tumor vessels reduces tumor blood supply, which can reduce the influence of the heat sink effect on RFA and increase the ablation zone. However, for the ablation of larger tumors or tumors in high-risk locations, using RFA to completely kill tumors is challenging and often results in residual tumor and damage to the surrounding normal tissues, seriously affecting the prognosis of patients.^{6,7} Brachytherapy with iodine-125 (125I) seed implantation for the high-dose irradiation of focal lesions has been widely used in the treatment of HCC.⁸ Studies have demonstrated that 125I seeds can increase the efficacy of RFA in the treatment of HCC.⁹

To date, there has been limited data available reporting the treatment of recurrent HCC with TACE-RFA combined with 125I seeds (TACE-RFA-125I). Hence, the purpose of this study was to evaluate the efficacy and safety of TACE-RFA-125I in the treatment of recurrent HCC and to provide a more effective treatment strategy for recurrent HCC.

Main points

- Transarterial chemoembolization combined with radiofrequency ablation (RFA) and iodine-125 (125I) seed implantation have a positive curative effect on hepatocellular carcinoma (HCC) after hepatectomy.
- RFA and 125I seed implantation under the dual guidance of ultrasound and computed tomography are safe and effective.
- For the long-term survival of patients with HCC, all target lesions should be controlled as much as possible.

Methods

Patients

This study retrospectively analyzed the clinical data of 253 patients with HCC recurrence after hepatectomy who were treated with TACE-RFA or TACE-RFA-125I at our center between January 2013 and January 2023. This retrospective study was approved by the Ethics Committee of Tongcheng County People's Hospital (decision no: TC-IEC-013-01-04, date: 01/01/2024), which waived the need for written informed consent.

Based on the inclusion and exclusion criteria, 187 patients were eventually included in the study. The inclusion criteria were as follows: (a) recurrent HCC diagnosed using computed tomography (CT)/magnetic resonance (MR) after hepatectomy; (b) a solitary HCC 3.0 cm in diameter, or smaller or multiple (up to three) HCCs 3.0 cm in diameter or smaller; (c) Child–Pugh A or B. The exclusion criteria were as follows: (a) prior chemoradiotherapy or other similar treatment; (b) other accompanying malignancies; (c) incomplete perioperative clinical and imaging data, or the patient was lost to follow-up.

Transarterial chemoembolization

The TACE procedure was conducted according to the institutional standard protocol. First, a 5-F catheter (Cook, Bloomington, IN, USA) or 2.7-F microcatheter (Terumo, Japan) was inserted into the tumor-supplying arteries, and then 5–20 mL of suspension composed of lipiodol (Lipiodol Ultrafluido, Guerbet, France) and doxorubicin hydrochloride (Hisun Pharmaceutical, Zhejiang, China) was injected. Subsequently, an appropriate amount of gelatin sponge (300–700 μ m, Cook) was injected to completely embolize the tumor vessels. After embolization, reexamination angiography of the feeding artery was performed to confirm the devascularization. This procedure was performed by experienced physicians.

Radiofrequency ablation

In this study, RFA was usually administered 3–5 days after TACE. The electrode needle was inserted into the target lesion under ultrasound and CT guidance. The RITA 1500 generator (RITA Medical Systems, Mountain View, CA, USA) was then activated and ablation started. The choice of a single expandable electrode (≤ 2 cm) or a multi-hook probe (> 2 cm) depends on tumor size. To achieve a safe ablation range of 0.5–1.0 cm, multiple overlapping ablation zones are required. The

analgesia was conducted by local injection of 5 mL of 2% lidocaine and intravenous administration of 50–100 mg of a flurbiprofen axetil injection (Tide Pharmaceutical, Beijing, China).

Iodine-125 seed implantation

After RFA, CT scans were performed to assess possible residual tumor areas. The puncture needle was then inserted into the target area. To assess the number and total activity of the 125I seeds, a treatment planning system (TPS; HGGR300, Hokai Medical Instruments, Zhuhai, China) was used, with X-rays and γ -rays able to reach the intended target volume, including the tumors and 0.5–1 cm of the adjacent normal tissue. After implantation of the 125I seeds, CT scans were repeated to identify the location of the 125I seeds and possible complications, and dose validation was performed using TPS.

Follow-up

Abdominal contrast-enhanced CT/MR was performed 4–6 weeks after the initial TACE. The changes in liver and kidney function, routine blood, and tumor markers were also evaluated. If the enhanced CT/MR indicated residual tumor, the TACE, RFA, or 125I seed procedures were repeated. With no residual tumor, the patient had an outpatient review every 3 months. The study was followed up until January 2024.

Tumor response was evaluated using the Modified Response Evaluation Criteria in Solid Tumors. Complete response refers to the absence of enhancement in all target lesions; partial response is classified as at least a 30% decrease in the sum of the diameters of viable tumors; progressive disease is an increase of at least 20% in the sum of the diameters of target lesions; stable disease refers to any cases that do not qualify for either partial response or progressive disease. Local and intrahepatic tumor recurrence was assessed. Local recurrence was defined as residual tumor at the lesion edge on CT/MR images during follow-up, and intrahepatic recurrence was defined as a single new lesion in the liver at a distance of more than 2.0 cm from the target lesion. Overall survival (OS) was the time from initial TACE to death or the end of follow-up. Progression-free survival (PFS) was the time from initial TACE to tumor progression, patient death, or end of follow-up. Complications were recorded and assessed using the Common Terminology Criteria for Adverse Events version 5.0. In addition, postembolization syndrome, such as fever,

pain, nausea, and vomiting, was not considered a complication in itself but rather an expected outcome of embolization therapy.

Statistical analysis

For the statistical analyses, SPSS software (version 24.0; IBM, Armonk, NY, USA) was used. Discrete variables were represented by numbers with percentages, and a chi-square test was used to compare the differences. Continuous variables were presented as mean ± standard deviation, and the Student's t-test was used to compare the difference. OS and PFS were evaluated using the Kaplan–Meier method. The 95% confidence interval (CI) was calculated for median OS, median PFS, and hazard ratio. Multivariate Cox proportional hazards regression analysis was used to analyze the potential prognostic factors affecting OS and PFS. The statistical significance was two-tailed, and a *P* value <0.05 was considered statistically significant.

Results

Study population

Between January 2013 and January 2023, a total of 187 patients with recurrent HCC were enrolled in this study (Figure 1). There were 67 men and 15 women in the TACE-RFA group, with an average age of 55.4 ± 10.9 years, and 93 men and 12 women in the TACE-RFA-125I group, with an average age of 55.5 ± 10.7 years, with no significant difference between the two groups. In the TACE-RFA-125I group, a total of 2,188 seeds were implanted, an average of 20.8 ± 9.6 per patient. Detailed baseline data of the two groups of patients are presented in Table 1. The median follow-up time was 20 months (range, 15–59 months) in the TACE-RFA group and 30 months in the TACE-RFA-125I group (range, 17–85 months). By January 2024, 30 and 40 patients in the two groups had died, respectively.

Recurrence

A total of 32 patients in the TACE-RFA group experienced recurrence during follow-up, with local recurrence in 12 cases, intrahepatic recurrence in 10 cases, and extrahepatic metastases in 10 cases. A total of 28 patients in the TACE-RFA-125I group experienced recurrence, with 6 cases of local recurrence, 12 of intrahepatic recurrence, and 10 of extrahepatic metastases.

Overall survival

The median OS was 32 months (95% CI, 29.3–34.7) in the TACE-RFA group and 49

months (95% CI, 43.6–54.4) in the TACE-RFA-125I group (*P* < 0.001) (Figure 2). Univariate analyses indicated that tumor number and treatment method correlated with OS (*P* < 0.2) (Table 2). These factors were included in the multivariate analysis, which revealed that treatment method was an independent prognostic factor affecting patients' OS (Table 3).

Progression-free survival

The median PFS in the TACE-RFA group was 16 months (95% CI, 13.9–18.1), which was significantly lower than that in the TACE-RFA-125I group (24 months, 95% CI, 20.9–27.1, *P* < 0.001) (Figure 3). Univariate analyses revealed that Child–Pugh score, albumin, platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio, α-Fetoprotein level, and treatment method were related to patients' PFS (*P* < 0.2) (Table 2), and further multivariate analyses demonstrated

that albumin and treatment method were independent prognostic factors affecting patients' PFS (Table 4).

Complications

No procedure-related deaths occurred in this study. Post-embolization syndrome, such as pain, fever, nausea, and vomiting, was common in both groups, and symptoms improved or disappeared substantially after symptomatic treatment. Pneumothorax occurred in three patients in the TACE-RFA group and four patients in the TACE-RFA-125I group; all cases were mild with no discomfort. No migration of seeds from the liver to other organs was observed during follow-up.

Discussion

The synergy between radiation therapy and thermal ablation is reported to have "reciprocal zones of efficacy".¹⁰ Hyperther-

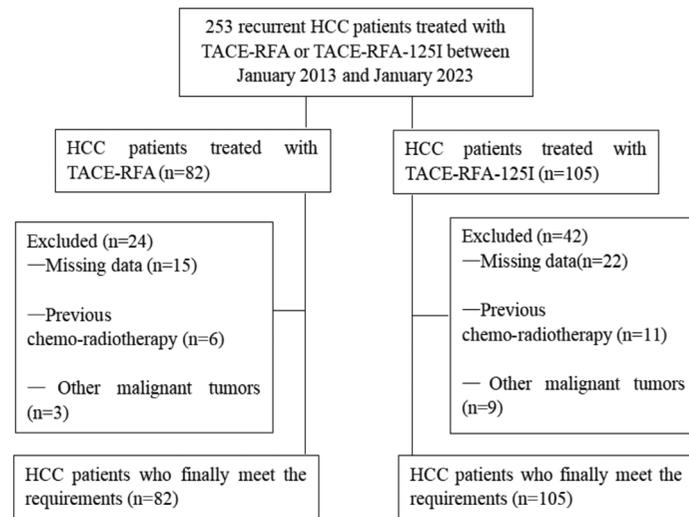


Figure 1. Flowchart of the screening procedure for patients with recurrent hepatocellular cancer. HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

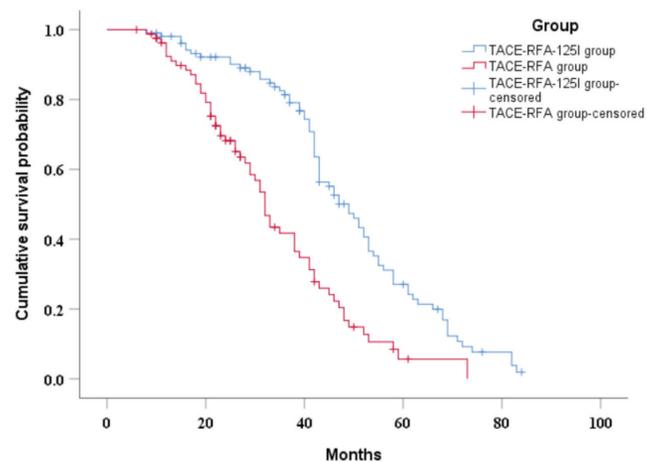


Figure 2. Kaplan–Meier curve of overall survival in patients with hepatocellular cancer. TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

mia leads to increased vasodilation and vascular permeability around the tumor, increasing oxygenation in the area and further improving the efficacy of radiotherapy. Furthermore, ablation and radiotherapy lead to

tumor necrosis, which stimulates the body's immunity and enhances the anti-tumor effect.¹¹ Additionally, the application of TACE in this study embolized tumor blood vessels and reduced the influence of the heat sink

effect.⁴ The deposition of lipiodol in tumors can also be used as a marker, and it is more conducive to RFA and seed treatment, especially for HCC in high-risk locations, reducing the risk of surgery and improving the curative effect. Hence, TACE-RFA-125I seeds could theoretically significantly enhance the therapeutic efficacy of HCC.

In our study, the efficacy of TACE-RFA-125I in the treatment of recurrent HCC was enhanced, primarily because the OS and PFS of patients were significantly improved compared with those of the TACE-RFA group. Liu et al.¹² reported that for recurrent HCC, the median OS and PFS of patients treated with TACE alone were 24 months and 10 months, respectively, which is significantly lower than the median OS and PFS of patients in our study. Zheng et al.¹³ compared the efficacy of TACE-RFA with that of repeat hepatectomy for recurrent HCC, demonstrating that the two groups of patients had similar OS and PFS. However, the efficacy of TACE-RFA-125I in this study was significantly superior to the results achieved by Zheng et al.¹³ This indicates that, as described by the theoretical advantages, the combination of the three treatment modalities produces a certain synergistic therapeutic effect, which can significantly improve the therapeutic effect on patients with recurrent HCC.

Brachytherapy with 125I has been reported to be effective in local tumor control.¹⁴ Chen et al.¹¹ revealed that RFA combined with 125I seed implantation significantly improved cumulative recurrence in patients with small HCC compared with RFA alone. Similarly, the results of this study demonstrated that the number of patients with local recurrence in the TACE-RFA-125I group was lower than that in the TACE-RFA group. The effective control of intrahepatic lesions is beneficial to the overall prognosis of patients. In the present study, the patients in the TACE-RFA-125I group had improved OS compared with the TACE-RFA group. Therefore, controlling all target lesions as much as possible is essential for the prognosis of patients with recurrent HCC.

Efficacy and safety are equally crucial for the treatment of recurrent HCC. In this study, there were no operation-related deaths or massive bleeding after puncture in the two groups of patients, and there were no significant fluctuations in the postoperative electrocardiogram monitoring of blood pressure and heart rate. To minimize treatment complications, the choice of puncture path and image-guidance tool is vital. Lin et al.¹⁵ per-

Table 1. Baseline characteristics

Characteristics	TACE-RFA-125I group (n = 105) (No, %; mean ± SD)	TACE-RFA group (n = 82) (No, %; mean ± SD)	P value
Age (years)	55.5 ± 10.7	55.4 ± 10.9	0.971
Gender			
Male	93 (88.6%)	67 (81.7%)	0.185
Female	12 (11.4%)	15 (18.3%)	
Child-Pugh score			
A	82 (78.1%)	67 (81.7%)	0.542
B	23 (21.9%)	15 (18.3%)	
Hepatitis			
Hepatitis B	86 (81.9%)	71 (86.6%)	0.387
Other	19 (18.1%)	11 (13.4%)	
Albumin (g/L)	37.9 ± 5.8	36.6 ± 5.1	0.117
PT(s)	13.8 ± 2.1	13.6 ± 1.2	0.894
TB (μmol/L)	22.4 ± 12.5	20.6 ± 11.3	0.169
AST (μmol/L)	40.3 ± 29.2	41.5 ± 30.2	0.471
ALT (μmol/L)	41.3 ± 30.2	41.8 ± 30.1	0.214
PLR	122.8 ± 97.7	137.1 ± 75.1	0.276
NLR	3.3 ± 3.5	3.2 ± 2.4	0.866
Tumor size (cm)	2.5 ± 0.8	2.3 ± 0.6	0.092
Tumor number			
1	85 (81.0%)	65 (79.3%)	0.774
2–3	20 (19.0%)	17 (20.7%)	
α-Fetoprotein level			
>400 ng/mL	45 (42.9%)	32 (39.0%)	0.597
≤400 ng/mL	60 (57.1%)	50 (61.0%)	

TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125; SD, standard deviation; PT, prothrombin time; TB, total bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; PLR, platelet-to-lymphocyte ratio; NLR, neutrophil-to-lymphocyte ratio.

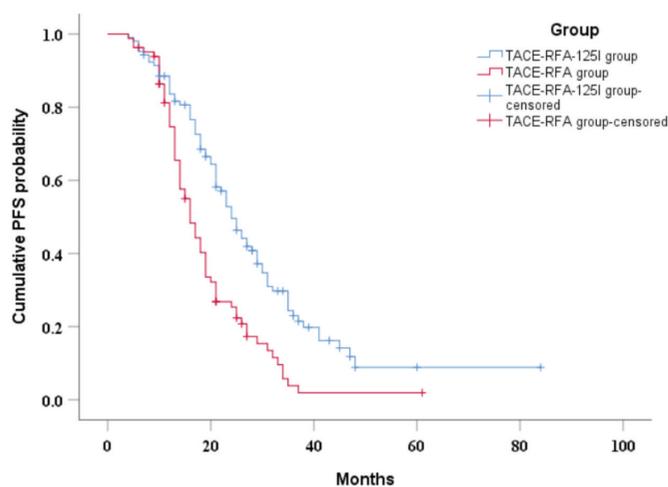


Figure 3. Kaplan–Meier curve of progression-free survival in patients with hepatocellular cancer. TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

Table 2. Univariate analysis of prognostic factors for overall survival and progression-free survival

Variables	OS		PFS	
	HR (95% CI)	P value	HR (95% CI)	P value
Age (years)	1.005 (0.988, 1.023)	0.553	1.005 (0.989, 1.021)	0.530
Gender				
Male	1		1	
Female	1.081 (0.663, 1.763)	0.754	1.066 (0.670, 1.695)	0.788
Child-Pugh score				
A	1			
B	0.808 (0.532, 1.225)	0.315	0.744 (0.497, 1.114)	0.151
Hepatitis				
Hepatitis B	1		1	
Other	0.897 (0.581, 1.387)	0.626	1.102 (0.721, 1.685)	0.653
Albumin (g/L)	0.991 (0.962, 1.021)	0.547	0.978 (0.951, 1.007)	0.131
PT (s)	0.950 (0.852, 1.058)	0.350	0.949 (0.866, 1.041)	0.270
TB (μmol/L)	1.004 (0.991, 1.016)	0.562	1.003 (0.990, 1.016)	0.663
AST (μmol/L)	1.124 (0.837, 1.273)	0.212	1.137 (0.875, 1.179)	0.392
ALT (μmol/L)	1.221 (0.925, 1.287)	0.372	1.026 (0.931, 1.126)	0.835
PLR	1.001 (0.999, 1.003)	0.205	1.001 (0.999, 1.003)	0.186
NLR	1.031 (0.977, 1.088)	0.266	1.046 (0.991, 1.103)	0.101
Tumor size	0.933 (0.727, 1.199)	0.589	0.967 (0.748, 1.250)	0.799
Tumor number				
1	1		1	
2–3	1.373 (0.926, 2.036)	0.115	0.983 (0.683, 1.413)	0.925
α-Fetoprotein level				
<400 ng/mL	1		1	
≥400 ng/mL	1.018 (0.931, 1.181)	0.256	1.365 (0.980, 1.901)	0.065
Treatment method				
TACE-RFA	1		1	
TACE-RFA-125I	0.411 (0.288, 0.586)	<0.001	0.493 (0.354, 0.687)	<0.001

OS, overall survival; PFS, progression-free survival; HR, Hazard ratio; CI, confidence interval; PT, prothrombin time; TB, total bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; PLR, platelet-to-lymphocyte ratio; NLR, neutrophil-to-lymphocyte ratio; TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

Table 3. Multivariate analysis of prognostic factors for overall survival

Variables	HR (95% CI)	P value
Tumor number		
1		
2–3	1.443 (0.973, 2.138)	0.068
Treatment method		
TACE-RFA	1	
TACE-RFA-125I	0.404 (0.284, 0.576)	<0.001

HR, Hazard ratio; CI, confidence interval; TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

formed RFA-125I seed therapy for HCC under MR guidance, but this procedure lasted a relatively long time and required magnetic compatible puncture equipment, which greatly limited its clinical application. Chen et al.¹⁶ applied CT-guided seed implantation and ablation therapy, but CT alone required the repeated adjustment of the needle tip position, and the researchers were unable

to observe the puncture path in real time. In our study, to improve the curative effect and reduce operation risk, all patients were first treated with TACE, and lipiodol played a role in further clarifying tumor location. In addition, we applied RFA-125I seed implantation under the guidance of ultrasound and CT, achieving a safe and positive therapeutic effect.

The main limitations of this study are its single-center and retrospective design. Therefore, prospective multicenter studies are necessary to verify our results. Recall bias is a potential limitation, possibly affecting the adverse effects assessment. In addition, the present study did not analyze the impact of follow-up treatment on outcomes or the incidence and severity of adverse events.

In conclusion, for recurrent HCC, TACE-RFA-125I seed therapy may be a promising treatment option. Moreover, ultrasound and CT-guided puncture are safe and reliable. The efficacy of this combination therapy still requires further validation in multicenter prospective studies.

Table 4. Multivariate analysis of prognostic factors for progression-free survival

Variables	HR (95% CI)	P value
Child–Pugh score		
A	1	0.088
B	1.102 (0.921, 1.271)	
Albumin	0.583 (0.409, 0.832)	0.003
PLR	1.000 (0.998, 1.002)	0.782
NLR	1.020 (0.958, 1.087)	0.529
α-Fetoprotein level		
<400 ng/mL	1	0.377
≥400 ng/mL	1.026 (0.924, 1.172)	
Treatment method		
TACE-RFA	1	0.003
TACE-RFA-125I	0.583 (0.409, 0.832)	

HR, Hazard ratio; CI, confidence interval; PLR, platelet-to-lymphocyte ratio; NLR, neutrophil-to-lymphocyte ratio; TACE, transarterial chemoembolization; RFA, radiofrequency ablation; 125I, iodine-125.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249. [\[CrossRef\]](#)
- Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. *J Hepatol.* 2021;76(3):681-693. [\[CrossRef\]](#)
- Kim J, Kang W, Sinn DH, et al. Substantial risk of recurrence even after 5 recurrence-free years in early-stage hepatocellular carcinoma patients. *Clin Mol Hepatol.* 2020;26(4):516-528. [\[CrossRef\]](#)
- Kim JW, Kim JH, Won HJ, et al. Hepatocellular carcinomas 2-3 cm in diameter: transarterial chemoembolization plus radiofrequency ablation vs. radiofrequency ablation alone. *Eur J Radiol.* 2012;81(3):189-193. [\[CrossRef\]](#)
- Ren YQ, Cao YY, Ma H, et al. Improved clinical outcome using transarterial chemoembolization combined with radiofrequency ablation for patients in Barcelona clinic liver cancer stage A or B hepatocellular carcinoma regardless of tumor size: results of a single-center retrospective case control study. *BMC Cancer.* 2019;19(1):983. [\[CrossRef\]](#)
- McWilliams JP, Yamamoto S, Raman SS, et al. Percutaneous ablation of hepatocellular carcinoma: current status. *J Vasc Interv Radiol.* 2010;21(8 Suppl):204-213. [\[CrossRef\]](#)
- Künzli BM, Abitabile P, Maurer CA. Radiofrequency ablation of liver tumors: actual limitations and potential solutions in the future. *World J Hepatol.* 2011;3(1):8-14. [\[CrossRef\]](#)
- Zhang Y, Wu YF, Yue ZD, et al. Iodine-125 implantation with transjugular intrahepatic portosystemic shunt for main portal vein tumor thrombus. *World J Gastrointest Oncol.* 2019;11(4):310-321. [\[CrossRef\]](#)
- Chen L, Ying XH, Zhang DK, et al. Iodine-125 brachytherapy can prolong progression-free survival of patients with locoregional recurrence and/or residual hepatocellular carcinoma after radiofrequency ablation. *Cancer Biother Radiopharm.* 2020;36(10):820-826. [\[CrossRef\]](#)
- Grieco CA, Simon CJ, Mayo-Smith WW, DiPetrillo TA, Ready NE, Dupuy DE. Percutaneous image-guided thermal ablation and radiation therapy: outcomes of combined treatment for 41 patients with inoperable stage I/II non-small-cell lung cancer. *J Vasc Interv Radiol.* 2006;17(7):1117-1124. [\[CrossRef\]](#)
- Chen K, Chen G, Wang H, et al. Increased survival in hepatocellular carcinoma with iodine-125 implantation plus radiofrequency ablation: a prospective randomized controlled trial. *J Hepatol.* 2014;61(6):1304-1311. [\[CrossRef\]](#)
- Liu Y, Ren Y, Ge S, et al. Transarterial chemoembolization in treatment-naïve and recurrent hepatocellular carcinoma: a propensity-matched outcome and risk signature analysis. *Front Oncol.* 2021;11:662408. [\[CrossRef\]](#)
- Zheng X, Ren Y, Hu H, Qian K. Transarterial chemoembolization combined with radiofrequency ablation versus repeat hepatectomy for recurrent hepatocellular carcinoma after curative resection: a 10-year single-center comparative study. *Front Oncol.* 2021;11:713432. [\[CrossRef\]](#)
- Tu JF, Ding YH, Chen L, et al. Iodine-125 brachytherapy prophylaxis after radiofrequency ablation cannot benefit patients in high risk of locoregional hepatocellular carcinoma recurrence. *Sci Rep.* 2017;7(1):3689. [\[CrossRef\]](#)
- Lin ZY, Chen J, Deng XF. Treatment of hepatocellular carcinoma adjacent to large blood vessels using 1.5T MRI-guided percutaneous radiofrequency ablation combined with iodine-125 radioactive seed implantation. *Eur J Radiol.* 2012;81(11):3079-3083. [\[CrossRef\]](#)
- Chen Z, Fu X, Qiu Z, et al. CT-guided ¹²⁵I brachytherapy for hepatocellular carcinoma in high-risk locations after transarterial chemoembolization combined with microwave ablation: a propensity score-matched study. *Radiol Oncol.* 2023;57(1):127-139. [\[CrossRef\]](#)



Comparing ultrasound-guided intra-articular injection and medial branch block for lumbar facet joint pain: a clinical study

Hui Zhao^{1-4*}
 Yun-Long Hou^{5*}
 Le-Hang Guo¹⁻⁴
 Qiao Wang¹⁻⁴
 Tian He¹⁻⁴
 Guang-Fei Gu⁶
 Li-Ping Sun¹⁻⁴
 Feng-Shan Jin¹⁻⁴

¹Tongji University Faculty of Medicine, Shanghai Tenth People's Hospital Center of Minimally Invasive Treatment for Tumor, Department of Medical Ultrasound, Shanghai, China

²Tongji University Faculty of Medicine, Ultrasound Research and Education Institute, Clinical Research Center for Interventional Medicine, Shanghai, China

³Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, Shanghai, China

⁴National Clinical Research Center for Interventional Medicine, Shanghai, China

⁵The First Hospital of Jiaying, Department of Medical Ultrasound, Zhejiang, China

⁶Shanghai Tenth People's Hospital Faculty of Medicine, Department of Orthopedics, Tongji University, Shanghai, China

*Joint first authors

Corresponding authors: Li-Ping Sun, Feng-Shan Jin

E-mails: sunliping_s@126.com, jfs198497@163.com

Received 22 March 2024; revision requested 15 April 2024; last revision received 08 July 2024; accepted 29 August 2024.



Epub: 27.11.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242765

PURPOSE

This study aims to compare the effectiveness of ultrasound-guided intra-articular (IA) injections with medial branch nerve blocks in treating lumbar facet joint pain.

METHODS

This retrospective study enrolled 94 patients clinically diagnosed with lumbar facet joint pain. Diagnostic blocks confirmed the diagnosis in 82 patients, evidenced by a pain visual analog score (VAS) reduction of at least 50% immediately following the injection. Of these, 42 were treated with ultrasound-guided IA injections (group 1), and 40 received ultrasound-guided medial branch blocks (group 2). Effective pain relief was defined as a VAS reduction of at least 50%.

RESULTS

Group 1 showed significantly higher pain relief rates compared with group 2 at both 1-month (54.76% versus 2.5%, $P < 0.001$) and 3-month (26.19% versus 5%, $P = 0.014$) follow-ups. Within group 1, patients aged 21–50 years experienced higher relief rates (81.25% at 1-month and 56.25% at 3-month follow-ups) compared with those over 50 (38.46% at 1-month and 7.69% at 3-month follow-ups), which was statistically significant ($P = 0.007$ at 1-month and $P = 0.001$ at 3-month follow-ups). Furthermore, in group 1, patients with sedentary jobs reported significantly greater pain relief (90.91% at 1-month and 81.82% at 3-month follow-ups) compared with those with non-sedentary jobs (41.94% at 1-month and 6.45% at 3-month follow-ups) ($P = 0.005$ at 1-month and $P < 0.001$ at 3-month follow-ups).

CONCLUSION

Ultrasound-guided IA injection provides better pain relief compared with medial branch nerve blocks. This method serves as a viable alternative for patients, especially younger and middle-aged patients with lumbar facet pain due to sedentary lifestyles.

CLINICAL SIGNIFICANCE

This study compared and analyzed the therapeutic effects of two different ultrasound intervention blockade methods on patients with lumbar facet joint pain, demonstrating that IA injection has a better pain relief effect and can be used as a pain relief method for patients who refuse radiofrequency therapy.

KEYWORDS

Lumbar facet joint pain, intra-articular injection, medial branch block, ultrasound-guided, pain visual analog score

The prevalence of chronic lower back pain (LBP) with various structural etiologies in the general population is about 80%.¹ Lumbar facet joint pain is caused by acute or chronic inflammation of the lumbar zygapophyseal joint and affects 15%–45% of patients with LBP.^{2,3} Regarding general treatment, conservative therapy includes exercise, physical methods, chiropractic care, and analgesics, which can help reduce surrounding muscle spasms and alleviate pain.⁴ Furthermore, facet joint pain relief can be achieved using therapeutic

interventions, including intra-articular (IA) injection, medial branch block (MBB), and radiofrequency ablation.^{5,6} IA injection involves a direct injection into the facet joint capsule, which can utilize the anti-inflammatory effect of hormones to alter inflammation in the joint. MBB alleviates pain by anesthetizing the medial branches of the posterior primary rami, the main nerves responsible for dual innervation to the joint.⁷ Both interventions are effective in facet joint pathology for managing LBP, but it is unclear which of them is superior.⁷ Various studies present differing views on the effectiveness of these treatments. One randomized study suggests that both approaches are equally ineffective in the long-term management of facet-joint-related back pain.⁸ Conversely, another study notes that both treatment options are beneficial,⁹ and a further prospective study indicates that IA injections are more effective than MBBs in patients with positive single-photon emission computed tomography (SPECT) scans.¹⁰

Ultrasound effectively visualizes soft tissue, neural structures, and vascular supply, offering an alternative guidance method for MBBs and IA injections. It is portable, facilitates intra-operative visualization, and does not require protective garments. Ultrasound-guided injection has widespread acceptance in regional anesthesia.¹¹ However, only a few studies have compared the long-term effects of ultrasound-guided IA injections and MBBs. This study aims to compare the effectiveness of ultrasound-guided IA and MBBs in treating lumbar facet joint pain.

Methods

Participants

The Shanghai Tenth People's Hospital Ethics Committee approved this study (protocol number: SHYS-IEC-5.0/22K203/P01, date: 26.09.2022). Before surgery, every patient was required to sign an informed consent form. These forms, along with the clinical and

imaging information pertinent to the study, were systematically archived. The hospital's medical records department will hold onto this data for 10 years. A diagram of the overall study design is shown in Figure 1.

A total of 94 patients diagnosed with lumbar facet joint pain were enrolled in this retrospective study between June 2021 and March 2022. A clinical doctor collected the patients' job types (sedentary or non-sedentary) and baseline pain visual analog score (VAS) before surgery through a questionnaire survey. Out of the 94 patients, 82 were confirmed to have lumbar facet joint pain through diagnostic blocks, which required their VAS score to decrease by at least 50% immediately after injection. Twelve patients who did not experience significant pain relief after treatment were excluded from the study. This study consisted of 40 men and 42 women, with a median age of 55.1 years and an average symptom duration of 8.5 weeks.

Among 82 patients, 42 received ultrasound-guided lumbar IA joint injections (group 1), whereas 40 received ultrasound-guided posterior MBB of lumbar spinal nerve roots (group 2). The primary outcome variable of this study was a 50% reduction in pain intensity from baseline measured at 1 month and 3 months post-operative. The pain extent was measured by a pain VAS. The evaluation at 1-month and 3-month follow-ups were recorded as short-

and medium-term efficacy of both treatment methods.

The inclusion criteria were as follows: (1) axial low back pain and tenderness during spinal extension and movement for 3 months or more, with an average back pain VAS score of more than 3 over the past week; (2) age ≥ 18 years; (3) failure to respond to more conservative therapy (e.g., physical therapy, integrative therapy, and pharmacotherapy). The exclusion criteria were as follows: (1) patients with a known specific etiology for low back pain, such as significant spinal stenosis or spondylolisthesis; (2) focal neurologic signs or symptoms; (3) patients with tuberculosis, lumbar spine tumor, osteomyelitis; (4) untreated coagulopathy; (5) poor compliance and inability to cooperate to complete treatment.

Equipment

Ultrasound guidance was conducted using a LOGIQ E9 with a 3-5MHz convex array transducer (GE Healthcare, Milwaukee, WI, USA). A 21-gauge puncture needle was used (KDL, Shanghai, China) for injection in this study.

Treatment procedure

The patient was placed in a prone position with a thin pillow on the abdomen to reduce lumbar anterior convexity. All procedures were performed by the same physician, who had more than 10 years of experience.

Main points

- For cases that did not respond to conservative treatment, interventional therapy is used for the treatment of lumbar facet joint pain.
- Ultrasound-guided intra-articular (IA) block is effective in relieving pain in the short term.
- When radiofrequency ablation is not feasible, ultrasound-guided IA block provides a viable means of pain management.

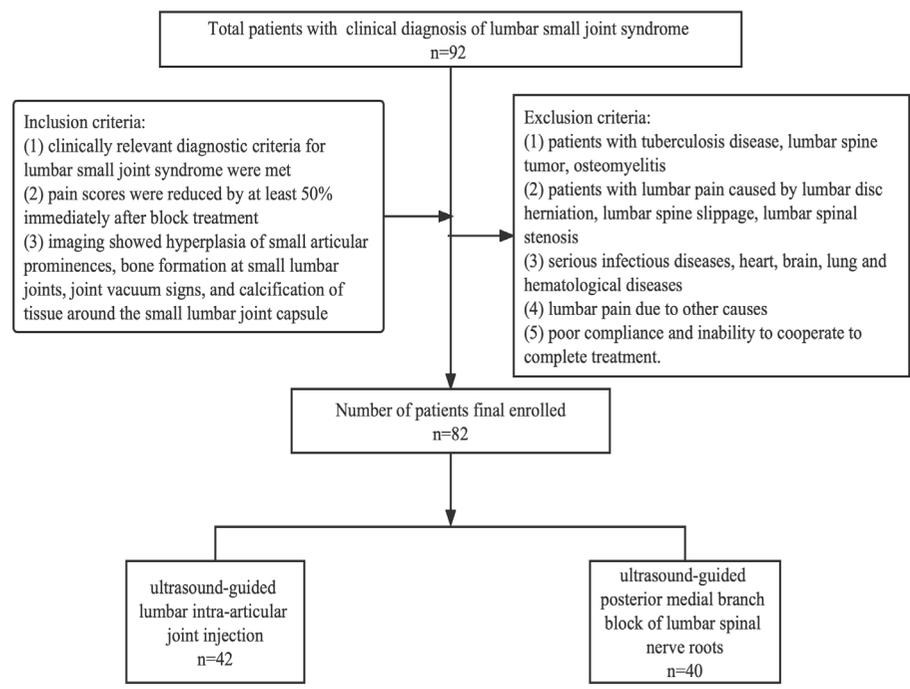


Figure 1. Study flow diagram.

rience in musculoskeletal system diseases ultrasound interventional therapy. First, locating the lumbar segment was essential. The transducer was placed in the midsagittal plane to visualize the spinous process of the lumbar vertebra, then moved 2 cm laterally to display the vertebral arch of the lumbar lamina and sacrum. The sacrum appeared as a continuous high-echo on ultrasound. Scanning from the sacrum, the short strip of high-echo above represented the L5 vertebral plate. Moving the probe upwards, the L4, L3, L2, and L1 vertebrae were displayed. On the parasagittal planes of the back, ultrasound clearly showed the high-echo superior and inferior articular processes with the low-echo articular cartilage between them. Then, with a rotated transducer, the transverse plane of the facet joints became visible, which was the target site for lumbar small joint injection (Figure 2).

The transducer was then moved downward to reveal the junction of the upper transverse process and the superior articular process, where the posteromedial branch of the nerve was located, which was the target for the posterior MBB (Figure 3).

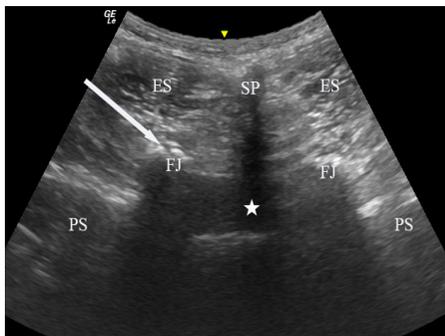


Figure 2. This figure shows the transverse plane of the facet joints. SP, spinous process; FJ, facet joint; arrows, needle; ES, erector spinae; PS, psoas; pentacle, spinal canal.

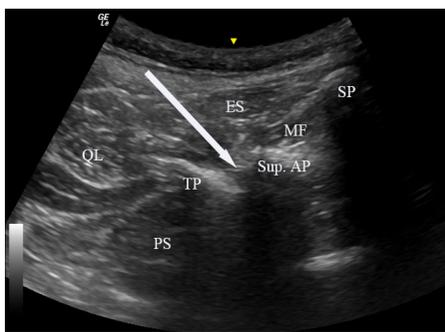


Figure 3. This figure shows the position of the posterior medial branch of the spinal nerve. SP, spinous process; TP, transverse process; Sup. AP, superior articular process; arrows, needle; ES, erector spinae; QL, quadratus lumborum; PS, psoas; MF, multifidus.

After sterile skin preparation and the placement of a fenestrated sterile drape, the needle was inserted in an “in-plane” manner under ultrasound guidance. Once the needle position was confirmed, 2 mL of solution was injected, containing 0.9 mL of 40 mg/mL triamcinolone acetonide, 0.2 mL of 2% lidocaine, and 0.9 mL of 0.9% normal saline. Following the procedure, the patient was observed in the recovery room for 30 minutes without experiencing any adverse effects before leaving. A blinded observer assessed patient pain scores at 30 minutes, 1 month, and 3 months post-treatment. A positive outcome was predefined as a 50% or greater reduction in VAS compared with baseline over the past week, indicating clinically meaningful improvement.⁹

Statistical analysis

All statistical analyses were performed using IBM, Armonk, NY, USA SPSS statistics version 25.0. Regarding demographic characteristics, continuous variables were presented as the mean and standard deviation and were compared between groups. Categorical variables were presented as frequencies and percentages

and analyzed using a chi-squared test. A value of $P < 0.05$ was considered statistically significant.

Results

Table 1 shows baseline demographic and clinical data by group assignment. There were no statistically significant differences in any variables at baseline between the groups. Treatment results following facet injection are shown in Table 2. The proportion of pain relief in group 1 (54.76%, 23/42) was significantly higher compared with that of group 2 (2.5%, 1/40) ($P < 0.001$), with a difference of 0.52 [95% (confidence interval) CI: 0.36–0.68] (Table 2). The proportion of pain relief at 3 months was 26.19% for group 1 and 5% for group 2, with a difference of 0.21 (95% CI: 0.06–0.36). There was a significant difference between the groups ($P = 0.014$). The findings reveal that IA injections, compared with posterior medial branch nerve blocks, provided superior short-term (1-month) pain relief for facet joint disorder. However, both interventions showed limited long-term (3-month) pain relief efficacy, suggesting that additional therapeutic strategies or follow-up interventions may be necessary to

Table 1. Demographic characteristics of patients in both groups

Group	Intra-articular injection group (n = 42)	Posterior medial branch block group (n = 40)	P
Age (years)	57.4 ± 17.4	52.8 ± 15.8	0.218
Gender	Men (19) Women (23)	Men (21) Women (19)	0.511
Duration of illness (months)	9.0 ± 6.5	7.9 ± 5.3	0.419
Visual analog score	6.48 ± 1.0	6.98 ± 0.8	0.052

Table 2. Treatment results and disposition after facet block

Group	Intra-articular injection group n = 42	Posterior medial branch block group n = 40	P	
1-month	Effective treatment (number, %)	23 (54.76%)	1 (2.5%)	<0.001
	Age <50	13/16 (81.25%)*	1/18 (5.56%)	
	Age ≥50	10/26 (38.46%)	0/22 (0%)	
	Sedentary job type	10/11 (90.91%)*	1/17 (5.88%)	
	Non-sedentary job type	13/31 (41.94%)	0/23 (0%)	
	Visual analog score (mean ± SD)	3.36 ± 1.17	6.5 ± 1.43	<0.001
3-month	Effective treatment (number, %)	11 (26.19%)	2 (5%)	0.014
	Age <50	9/16 (56.25%)*	2/18 (11.11%)	
	Age ≥50	2/26 (7.69%)	0/22 (0%)	
	Sedentary job type	9/11 (81.82%)*	1/17 (5.88%)	
	Non-sedentary job type	2/31 (6.45%)	1/23 (4.35%)	
	Visual analog score (mean ± SD)	4.45 ± 1.66	6.7 ± 1.31	<0.001

* $P < 0.05$ within group analysis between different age stratification. # $P < 0.05$ within group analysis between different job type stratification. SD, standard deviation.

maintain or improve pain management over longer periods.

Upon stratifying the analysis by age, it was observed that for young individuals aged 21–50 years in group 1, the effectiveness rates at 1 month and 3 months were 81.25% and 56.25%, respectively. In contrast, among the elderly population aged over 50 years, the effectiveness rates were significantly lower, at 38.46% at 1 month and 7.69% at 3 months ($P_{1\text{-month}} = 0.007$; $P_{3\text{-month}} = 0.001$).

In the analysis by job type, sedentary workers showed significantly higher pain relief rates (90.91% at 1 month, 81.82% at 3 months) after IA block therapy compared with non-sedentary workers (41.94% at 1 month, 6.45% at 3 months). Notably, in group 2, only 2 young men, aged 31 and 29, showed effective outcomes—one with a non-sedentary job and the other with a sedentary job. Age and job type may influence short-term efficacy, with middle-aged and young patients in sedentary jobs more likely to benefit from this treatment.

Discussion

The main finding in this retrospective study is that ultrasound-guided IA injection shows significant posttreatment improvement in pain relief compared with MBBs at 1 month. However, neither technique produced useful sustained pain remissions at 3 months. This is similar to the findings of Robert Cornelius Marks.¹²

Based on fundamental scientific studies, it has been observed that inflammatory mediators are present in and around degenerative facet joints.^{13,14} This presence indicates the potential role of inflammation in causing pain and discomfort. Consequently, it suggests that IA steroid injections may provide short-term pain relief by alleviating synovitis or inflammation related to osteoarthritis through their anti-inflammatory effects.¹⁵

In our stratified analysis by age and job type, we discovered that middle-aged and young individuals, as well as those with sedentary work, are more likely to achieve better pain relief from IA injections. The underlying cause of low back pain in these individuals is primarily attributed to their sedentary lifestyle, which exerts prolonged pressure on the lumbar facet joint, triggering an inflammatory response and stimulating the lumbar spinal nerves around the facet joint. Consequently, administering hormones can reduce the inflammatory response, leading to therapeutic results.⁴ A study suggests that

IA injection is more effective among patients exhibiting an inflammatory process in the facet joint on SPECT imaging.¹⁰ Although our current study did not account for the impact of positive ultrasound signs on efficacy, this aspect will be thoroughly analyzed in our forthcoming study. The research results indicate that injection therapy has limited effectiveness for the elderly and non-sedentary individuals. These individuals usually have significant osteophytes or other degenerative facet joint diseases. IA steroid injections cannot change the structural pressure problem or any other problems of facet joints.

Ultrasound-guided MBB had little apparent therapeutic effect in the short or medium term. We speculate that this may be because the nerve trunk is temporarily blocked while inflammation affecting the peripheral nerves remains. When conduction is restored, the pain reappears.

This study is subject to several limitations. First, its retrospective design introduces the possibility of bias, and the non-randomized allocation of patient treatments could result in selection bias. To mitigate this, patients in the study were enrolled continuously according to strict inclusion and exclusion criteria. Furthermore, no significant statistical differences were observed in the general clinical information across the two data groups. In future studies, we aim to conduct randomized, controlled, and prospective research to further assess the effectiveness of ultrasound-guided block therapy in treating small joint disorders. Second, the relatively small sample size and brief follow-up period limit our ability to evaluate long-term treatment outcomes. Future research will aim to enlarge the sample size and extend the follow-up period.

In conclusion, ultrasound-guided IA injections are more effective than medial branch nerve blocks in patients with lumbar facet joint pain, particularly for non-degenerative lumbar facet joint pain caused by sedentary behavior in young and middle-aged patients. MBBs appear to have no significant therapeutic effect in the short or medium term. For patients who cannot undergo surgery or radiofrequency therapy, ultrasound-guided IA injection of small joints can be chosen to alleviate pain.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Fast A. Low back disorders: conservative management. *Arch Phys Med Rehabil*. 1988;69(10):880-891. [\[CrossRef\]](#)
2. Van de Kelft E, Verleye G, Van de Kelft AS, Melis K, Van Goethem J. Validation of topographic hybrid single-photon emission computerized tomography with computerized tomography scan in patients with and without nonspecific chronic low back pain. A prospective comparative study. *Spine J*. 2017;17(10):1457-1463. [\[CrossRef\]](#)
3. Rahimzadeh P, Faiz HR, Baghaee AR, Nader ND. Sonoanatomic indices of lumbar facet joints in patients with facetogenic back pain in comparison to healthy subjects. *J Clin Anesth*. 2017;36:67-71. [\[CrossRef\]](#)
4. Seo JH, Baik SW, Ko MH, et al. Comparing the efficacy of combined treatment with medial branch block and facet joint injection in axial low back pain. *Pain Res Manag*. 2021;2021:1343891. [\[CrossRef\]](#)
5. Ogsbury JS 3rd, Simon RH, Lehman RAW. Facet “denervation” in the treatment of low back syndrome. *Pain*. 1977;3(3):257-263. [\[CrossRef\]](#)
6. Koh WU, Kim SH, Hwang BY, et al. Value of bone scintigraphy and single photon emission computed tomography (SPECT) in lumbar facet disease and prediction of short-term outcome of ultrasound guided medial branch block with bone SPECT. *Korean J Pain*. 2011;24(2):81-86. [\[CrossRef\]](#)
7. Bogduk N. International Spinal Injection Society guidelines for the performance of spinal injection procedures. Part 1: Zygapophysial joint blocks. *Clin J Pain*. 1997;13(4):285-302. [\[CrossRef\]](#)
8. Cohen SP, Doshi TL, Constantinescu OC, et al. Effectiveness of lumbar facet joint blocks and predictive value before radiofrequency denervation: the facet treatment study (FACTS), a Randomized, Controlled Clinical Trial. *Anesthesiology*. 2018;129(3):517-535. Erratum in: *Anesthesiology*. 2018;129(3):618. [\[CrossRef\]](#)
9. Cohen SP, Bhaskar A, Bhatia A, et al. Consensus practice guidelines on interventions for lumbar facet joint pain from a multispecialty, international working group. *Reg Anesth Pain Med*. 2020;45(6):424-467. [\[CrossRef\]](#)
10. Ackerman WE 3rd, Ahmad M. Pain relief with intraarticular or medial branch nerve blocks in patients with positive lumbar facet joint SPECT imaging: a 12-week outcome study. *South Med J*. 2008;101(9):931-934. [\[CrossRef\]](#)
11. Wang D. Image guidance technologies for interventional pain procedures: ultrasound, fluoroscopy, and CT. *Curr Pain Headache Rep*. 2018;22(1):6. [\[CrossRef\]](#)
12. Marks RC, Houston T, Thulbourne T. Facet joint injection and facet nerve block: a randomised comparison in 86 patients with chronic

- low back pain. *Pain*. 1992;49(3):325-328. [\[CrossRef\]](#)
13. Willburger RE, Wittenberg RH. Prostaglandin release from lumbar disc and facet joint tissue. *Spine (Phila Pa 1976)*. 1994;19(18):2068-2070. [\[CrossRef\]](#)
14. Igarashi A, Kikuchi S, Konno S, Olmarker K. Inflammatory cytokines released from the facet joint tissue in degenerative lumbar spinal disorders. *Spine (Phila Pa 1976)*. 2004;29(19):2091-2095. [\[CrossRef\]](#)
15. Lakemeier S, Lind M, Schultz W, et al. A comparison of intraarticular lumbar facet joint steroid injections and lumbar facet joint radiofrequency denervation in the treatment of low back pain: a randomized, controlled, double-blind trial. *Anesth Analg*. 2013;117(1):228-235. [\[CrossRef\]](#)



The value of dual-energy computed tomography angiography-based virtual monoenergetic imaging for evaluations after cerebral aneurysm clipping

Zhihua Lu¹
 Suying Wu¹
 Feijian Wu¹
 Qingdong Jin²
 Qingjing Huang¹
 Baoteng Zhang¹

¹The First Hospital of Putian City, Department of Radiology, Putian, China

²The First Hospital of Putian City, Department of Neurosurgery, Putian, China

PURPOSE

This study aimed to research the optimal energy range of dual-energy computed tomography angiography (DECTA)-based virtual monoenergetic imaging (VMI) for evaluations after cerebral aneurysm clipping.

METHODS

Sixty patients who underwent DECTA after cerebral aneurysm clipping were analyzed retrospectively. Conventional computed tomography angiography (CTA) was compared with VMIs at 60, 70, 80, 90, and 100 keV. The mean attenuation and standard deviation values within the regions of interest placed in the brain parenchyma and arteries with the worst artifact were measured, respectively. The Δ CT and artifact index (AI) values were calculated to assess the artifact severity. The contrast-to-noise ratio (CNR) was calculated to assess vascular contrast. Two radiologists assessed brain parenchyma and cerebrovascular scores qualitatively using a five-point Likert scale.

RESULTS

Quantitative analysis showed that the artifacts of VMIs were significantly reduced compared with conventional CTA ($P \leq 0.014$), except for the Δ CT and AI of 60 keV and the Δ CT of 70 keV. However, there was no significant difference in the vascular contrast on VMIs compared with conventional CTA, except for the CNR of 60 keV ($P = 0.008$). In qualitative analysis, the proportions of brain parenchyma scores and cerebrovascular scores ≥ 4 on the VMIs of 70 and 80 keV were higher than those of conventional CTA and other VMIs.

CONCLUSION

For the patients who underwent DECTA after cerebral aneurysm clipping, the 70–80 keV VMIs are expected to be the optimal energy range for balancing clip artifacts and visibility of adjacent vessels.

CLINICAL SIGNIFICANCE

Studying the optimal energy range of DECTA-based VMI for post-operative assessment of aneurysm clipping can reduce metal artifacts in images and increase vascular contrast. This facilitates the follow-up of patients after aneurysm clipping, offers timely and accurate detection of postoperative complications, provides assistance to clinicians in diagnosis and treatment, and improves patient prognosis.

KEYWORDS

Dual-energy, computed tomography angiography, virtual monoenergetic imaging, aneurysm clipping, metal artifact

Corresponding author: Baoteng Zhang

E-mail: 519145686@qq.com

Received 14 August 2024; revision requested 09 September 2024; accepted 29 October 2024.



Epub: 16.12.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242975

You may cite this article as: Lu Z, Wu S, Wu F, Jin Q, Huang Q, Zhang B. The value of dual-energy computed tomography angiography-based virtual monoenergetic imaging for evaluations after cerebral aneurysm clipping. *Diagn Interv Radiol.* 2025;31(3):264-273.

Cerebral aneurysms are localized weakened pouches formed by pathological rupture of the internal elastic lamina and media of the arterial wall.¹ The incidence of unruptured cerebral aneurysms in the general population is 3.6%–6%,² and they are mostly asymptomatic or have non-specific symptoms. Ruptured cerebral aneurysms are a neurosurgical emergency with potentially devastating consequences, which may be accompanied by complications, such as subarachnoid hemorrhage, hydrocephalus, cerebral parenchymal hemorrhage, and epidural hematoma.² Surgical clipping can safely treat ruptured and unruptured cerebral aneurysms.³ However, there are still risks, such as incomplete clipping, residue or recurrence of the aneurysm, distal occlusion of the parent artery, cerebral vasospasm, and cerebral hemorrhage.^{4–6} As a non-invasive and simple technique, computed tomography angiography (CTA) is widely used to evaluate treated aneurysms. Early postoperative evaluation helps identify complications in time and serves as a basis for late follow-up.⁷ However, surgically implanted clips produce metal artifacts, affecting the observation of fine structures, such as peripheral brain parenchyma and vessels. Therefore, it is essential to minimize metal artifacts.

In recent years, several studies have applied different techniques to reduce metal artifacts in patients with treated cerebral aneurysms, among which metal artifact reduction (MAR) algorithms and dual-energy computed tomography angiography (DECTA)-based virtual monoenergetic imaging (VMI) can effectively improve image quality and reduce metal artifacts.^{8–12} Low-keV monoenergetic images can improve density resolution and enhance the contrast between vessels and brain parenchyma; however, the noise and metal artifacts are worsened. In contrast, high-keV monoener-

getic images can reduce metal artifacts, but the contrast is also reduced. Therefore, finding the optimal energy range to balance clip artifacts and visibility of adjacent vessels is crucial. To our knowledge, there are very few studies to address this issue.^{10–12} For instance, a skull phantom study¹⁰ showed that 120keV VMI significantly reduced metal artifacts and improved visibility of adjacent vessels compared with non-corrected DECTA, but they did not study the value of other energy levels. Furthermore, Dunet et al.¹¹ demonstrated the best compromise between MAR and relative contrast-to-noise ratio (CNR) was obtained at 70–75 keV for gemstone spectral imaging (GSI) DECTA after cerebral aneurysm clipping. However, their results were not compared with conventional CTA.

In the current study, we comprehensively evaluate the effects of artifacts in the head, tail, and middle part of the implanted clip on peripheral vessels and brain parenchyma by comparing VMIs with conventional CTA to find the optimal energy range of DECTA-based VMI for evaluation after cerebral aneurysm clipping.

Methods

Population and design

One hundred twelve patients who underwent aneurysm surgical clipping in the First Hospital of Putian City between March 2019 and July 2023 were collected. The inclusion criteria were as follows: (1) early follow-up DECTA performed after surgical clipping (the interval between examination and operation was <4 weeks); (2) no pregnancy, no history of iodine contrast allergy, and no severe cardiopulmonary and renal dysfunction. The exclusion criteria were as follows: (1) different scanners and different concentrations of contrast media; (2) poor image quality or incomplete clinical data; (3) multiple aneurysms or use of multiple metal clips. Figure 1 shows the flowchart of inclusion and exclusion criteria for the study. This retrospective study was approved by the First Hospital of Putian City Hospital Ethics Committee on December 28, 2022 (protocol no: 2022-100), and patient written consent was waived.

Image acquisition and reconstruction

All patients underwent DECTA on a third-generation dual-source computed tomography (CT) scanner (SOMATOM Force, Siemens Medical Solutions, Forchheim, Germany). The patients were in the supine position. The scanning range was from the 1 cm

level below the skull base to the top of the skull, from foot to head. Scanning parameters were as follows: tube voltages 80 kVp and 150 kVp, pitch 0.7 mm, field of view 200 × 200 mm. Automated tube current modulation (CareDose4D, Siemens Medical Solutions) was adopted. Reconstruction parameters were as follows: Matrix 512 × 512, layer thickness 1.0 mm, spacing 0.7 mm, convolution kernel Qr40.

Ultravist (370 mg/mL; Bayer-Schering Healthcare, Berlin, Germany) was injected into the anterior cubital vein with a double-barrel high-pressure syringe at a dose of 1.0 mL/kg and rate of 4.5 mL/s, followed by a 40 mL saline flush at the same rate. The acquisition was started automatically 4 seconds after the peak time when the attenuation value of the ascending aorta reached 100 HU.

After the scanning, the automatically reconstructed 80 kV, Sn150 kV, and blended 115 kV images were transmitted to a Siemens workstation (Syngo.via VB10B) for post-processing. The Monoenergetic module was used to reconstruct the VMIs every 10 keV from 60 to 100 keV. The blended 115 kV image using a ratio of 50/50 from the 80 kV and Sn150 kV data is approximately equivalent to a conventional 120 kV acquisition and was considered as the reference. Finally, six groups of images were obtained for subsequent analysis.

Objective evaluation of image quality

Selection of regions of interest

Objective image analysis was performed by a board-certified neuroradiologist with 11 years of experience in brain imaging. A 10 mm² circular region of interest (ROI) was placed within the center of one occipital lobe parenchyma (OLP) as the reference OLP, avoiding vessels, calcification, malacic foci, and hemorrhage.¹³ Another ROI (as large as possible) was placed in the basilar artery as the reference BA, avoiding vascular wall calcification and skull base artifacts. Regions of interest A, B, and C between 6 and 8 mm² (mean 7.2 mm²) were placed in the brain parenchyma, with the worst artifacts around the head, tail, and middle part of the implanted clip, respectively. Regions of interest D and E were placed in the vessels with the worst artifacts around the head/tail and middle part of the implanted clip, respectively. These ROIs were copied in the same locations on both conventional CTA and VMIs (Figure 2). These ROIs' attenuation

Main points

- Metal artifacts on the tail of the clip were more serious than those on the head on both conventional computed tomography angiography (CTA) and virtual monoenergetic imaging (VMI).
- VMI can significantly reduce metal artifacts compared with conventional CTA, except for 60 keV.
- Metal artifacts of clips near the skull base were higher than those away from the skull base on both conventional CTA and VMIs.
- The 70–80 keV VMIs are expected to be the optimal energy range for balancing clip artifacts and visibility of adjacent arteries.

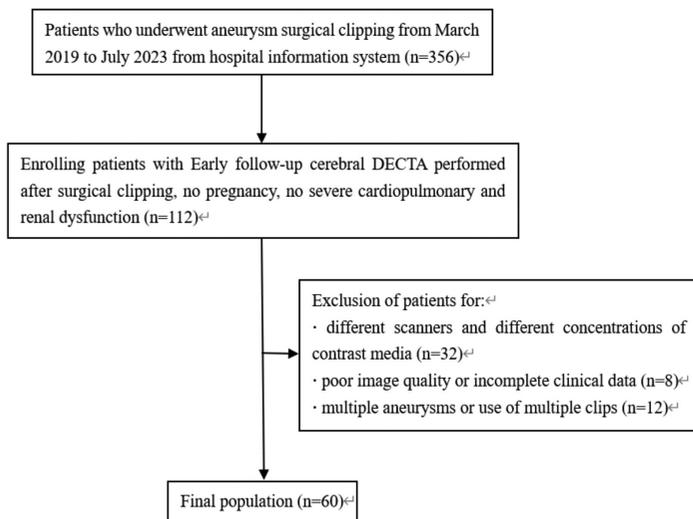


Figure 1. The flowchart of inclusion and exclusion criteria for the study. DECTA, dual-energy computed tomography angiography.

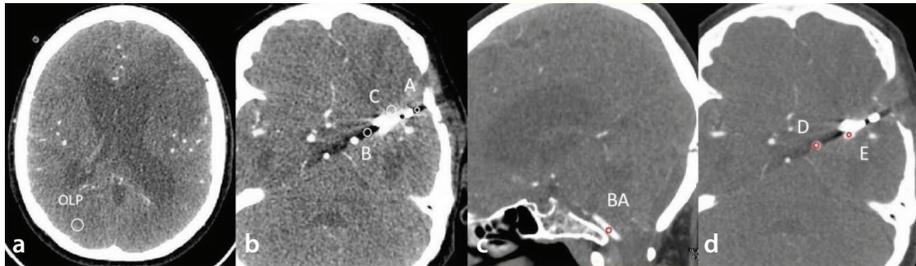


Figure 2. An example of manually designating regions of interest in conventional computed tomography angiography images at different slices or different window widths and levels. (a) The occipital lobe parenchyma is drawn in the right occipital lobe, avoiding vessels, malacic foci, and hemorrhage. (b) Draw A, B, and C in the areas with the worst artifacts in the metal clip's head, tail, and middle part, respectively. (c) BA is located in the basilar artery. (d) D and E are located in the vessels with the worst artifacts in the tail and middle part of the metal clip, respectively.

and standard deviation (SD) values were recorded. Each ROI was measured three times to calculate the average value for further analysis. In addition, images of 30 patients were randomly selected, and the ROIs were measured by another well-trained neurosurgeon in a blinded method to assess the inter-reader agreement.

Quantitative analysis

The ΔCT values between A, B, C, and OLP and between D, E, and BA were calculated using the following formula:

$$\Delta CT = |CT_{\text{artifact}} - CT_{\text{reference}}| \quad (1)$$

The artifact index (AI) values of A, B, C, D, and E and the CNR values of D and E were calculated using the following formula⁹:

$$AI = \sqrt{SD_{\text{artifact}}^2 - SD_{\text{reference}}^2} \quad (2)$$

$$CNR = \frac{|CT_{\text{artifact}} - CT_{\text{reference}}|}{\sqrt{(SD_{\text{artifact}}^2 - SD_{\text{reference}}^2) / z}} \quad (3)$$

Artifact severity was assessed quantitatively by ΔCT (1) and AI (2). Vascular contrast was assessed quantitatively by CNR (3).

Subjective evaluation of image quality

A five-point Likert scale⁸ was used to evaluate the artifact severity of brain parenchyma and the contrast of adjacent arteries subjectively (Supplementary Table S1). Subjective image analysis was carried out independently by two more well-trained radiologists 1 and 2.

Statistical analysis

The statistical analysis was performed using SPSS statistics version 25.0 (IBM, Chicago, Illinois, USA). For the quantitative analysis, inter-reader and intra-reader agreements

regarding attenuation and SD values of ROIs were calculated using intraclass correlation coefficients (ICCs). All ICCs between 0.75 and 1.00 were interpreted as excellent agreement. A comparison of data from the six groups of images was performed by one-way analysis of variance (ANOVA). According to the homogeneity of variance, the least significant difference or Tamhane's T2 test was used to compare ΔCT , AI, and CNR between conventional CTA and VMIs. A graphical comparison was used to illustrate the results. In addition, to evaluate whether the clips' location may affect the metal artifact severity of conventional CTA and VMIs, as previously suggested,¹¹ ΔCT and AI were also compared between patients with clips near the skull base (internal carotid artery or communicating arteries) and patients with clips away from the skull base (anterior cerebral artery, middle cerebral artery, or posterior cerebral artery) in the six groups of images using the Mann-Whitney U test. To avoid the influence of different orientations of clips, the worst ΔCT and AI values from ROI A, B, and C were selected for artifact severity comparison.

For the qualitative analysis, the inter-reader agreement of subjective scores (brain parenchyma scores and cerebrovascular scores) were calculated using Cohen's kappa coefficients, with values of 0.41–0.60 representing a moderate agreement, 0.61–0.80 a strong agreement, and 0.81–1.00 an almost perfect agreement. Every disagreement was resolved by consensus between two radiologists before further analysis. The chi-squared test was used to compare subjective scores ≥ 4 proportions between conventional CTA and VMIs. A *P* value of <0.05 was considered statistically significant.

Results

Finally, 60 patients (of whom the mean age was 55.15 ± 9.24 years and 23 were men) underwent surgical clipping, and 60 aneurysms were included in this study (one metal clip per aneurysm). In summary, seven different clips (Yasargil, Braun Medical, Tuttlingen, Germany) made of titanium were included, and the blade length was 7.0–13.7 mm (Supplementary Table S2). The dose-length product (DLP) was 158.00 ± 31.45 mGy·cm, the computed tomography dose index (CTDIvol) was 8.75 ± 1.74 mGy, and the effective dose (ED) was 0.33 ± 0.07 mSv according to ICRP 103 conversion coefficients.¹³ The radiation dose in the current study is lower than that in previous cerebral DECTA-related studies (DLP: 280–685 mGy·cm, CTDIvol: 13–42 mGy, ED: 0.59–1.44 mSv).^{11,12,14,15} The interval be-

tween postoperative DECTA and surgery was 16.77 ± 5.64 days. Supplementary Table S3 summarizes patient characteristics as well as aneurysm location and postoperative complications.

Quantitative analysis

The inter-reader and intra-reader ICCs of attenuation and SD values in conventional CTA and VMIs are summarized in Supplementary Table S4. All ICCs showed excellent agreement (inter-reader ICCs: 0.892–0.985, intra-reader ICCs: 0.911–0.995). The $\Delta CT_{A'}$, $\Delta CT_{B'}$, $\Delta CT_{C'}$, $AI_{A'}$, $AI_{B'}$, and $AI_{C'}$ of the brain parenchyma and the $\Delta CT_{D'}$, $\Delta CT_{E'}$, $AI_{D'}$, $AI_{E'}$, $CNR_{D'}$, and $CNR_{E'}$ of the cerebral vessels were all in concordance with the normal or slightly skew distribution. The *P* values of one-way ANOVA were all <0.05 , except for $CNR_{D'}$. There were significant differences for $\Delta CT_{A'}$, $\Delta CT_{B'}$, and $\Delta CT_{C'}$ of all VMIs compared with those of conventional CTA ($P \leq 0.002$). They were all lower as the keV level increased, and the lowest values were 44.39 ± 24.40 , 60.97 ± 34.42 ,

and 30.28 ± 10.91 , respectively, at 100 keV (Table 1, Figure 3a). Except for 60 keV, there were significant differences for $AI_{A'}$, $AI_{B'}$, and $AI_{C'}$ of all VMIs compared with those of conventional CTA ($P \leq 0.014$). They were all lower as the keV level increased, and the lowest values were 17.83 ± 10.50 , 23.93 ± 16.90 , and 15.92 ± 5.81 , respectively (Table 2, Figure 3b). The ΔCT and *AI* of clips near the skull base were higher than those of clips away from the skull base on both conventional CTA and VMIs, with significant differences ($P \leq 0.031$), except for the *AI* of 90 and 100 keV (Supplementary Table S5).

Except for $\Delta CT_{D'}$ and $\Delta CT_{E'}$ of 60 keV and $\Delta CT_{E'}$ of 70 keV, the $\Delta CT_{D'}$ and $\Delta CT_{E'}$ of all VMIs were significantly different from those of conventional CTA ($P \leq 0.003$). They basically decreased with the increase in keV level, and the lowest values were 28.47 ± 17.72 and 51.63 ± 28.55 , respectively (Table 3, Figure 4a). The $AI_{D'}$ and $AI_{E'}$ of all VMIs differed significantly from those of conventional CTA ($P < 0.001$). They all decreased with the increase

in keV level, and the lowest values were 21.20 ± 6.05 and 29.88 ± 12.14 , respectively (Table 4, Figure 4b). However, there was no significant difference in $CNR_{D'}$ and $CNR_{E'}$ between all VMIs and conventional CTA, except for the $CNR_{E'}$ of 60 keV ($P = 0.008$) (Table 4, Figure 4c).

Qualitative analysis

The Cohen's kappa coefficients of brain parenchyma and cerebrovascular scores of radiologists 1 and 2 in conventional CTA and VMIs are summarized in Supplementary Table S6. All kappa coefficients showed strong or almost perfect agreement (0.633–0.832). In conventional CTA and 60 to 100 keV VMIs, the proportions of brain parenchyma scores ≥ 4 were 6.67%, 6.67%, 61.67%, 80.00%, 20.00%, and 5.00%, respectively. Except for 60 and 100 keV, the brain parenchyma scores of other keV levels were significantly higher than those of conventional CTA ($P \leq 0.032$). The proportions of cerebrovascular scores ≥ 4 were 1.67%, 50.00%, 95.00%, 73.30%, 13.33%, and 0.00%, respectively. Except for

Table 1. $\Delta CT_{A'}$, $\Delta CT_{B'}$, and $\Delta CT_{C'}$ in the five groups of VMIs compared with conventional CTA

Group	$\Delta CT_{A'}$ (HU)	<i>P</i> value	$\Delta CT_{B'}$ (HU)	<i>P</i> value	$\Delta CT_{C'}$ (HU)	<i>P</i> value
Conventional CTA	164.95 ± 46.62	-	251.47 ± 106.01	-	101.30 ± 24.54	-
60 keV	199.18 ± 53.78	<0.001	302.41 ± 122.58	0.001	114.17 ± 28.95	0.001
70 keV	136.90 ± 38.33	<0.001	205.66 ± 83.51	0.002	79.32 ± 20.60	<0.001
80 keV	98.68 ± 32.21	<0.001	141.50 ± 59.28	<0.001	57.92 ± 17.34	<0.001
90 keV	67.11 ± 27.95	<0.001	94.86 ± 42.74	<0.001	41.17 ± 12.95	<0.001
100 keV	44.39 ± 24.40	<0.001	60.97 ± 34.42	<0.001	30.28 ± 10.91	<0.001

VMI, virtual monoenergetic imaging; CTA, computed tomography angiography; HU, hounsfield unit.

Table 2. $AI_{A'}$, $AI_{B'}$, and $AI_{C'}$ in the five groups of VMIs compared with conventional CTA

Group	$AI_{A'}$ (HU)	<i>P</i> value	$AI_{B'}$ (HU)	<i>P</i> value	$AI_{C'}$ (HU)	<i>P</i> value
Conventional CTA	40.89 ± 19.92	-	78.23 ± 49.72	-	34.43 ± 11.13	-
60 keV	43.85 ± 21.85	0.311	87.26 ± 57.49	0.197	37.62 ± 13.21	0.066
70 keV	31.87 ± 15.97	0.002	61.01 ± 39.38	0.014	27.65 ± 9.60	<0.001
80 keV	24.89 ± 13.20	<0.001	44.30 ± 27.42	<0.001	21.90 ± 8.24	<0.001
90 keV	20.64 ± 11.20	<0.001	31.77 ± 20.28	<0.001	18.12 ± 6.78	<0.001
100 keV	17.83 ± 10.50	<0.001	23.93 ± 16.90	<0.001	15.92 ± 5.81	<0.002

VMI, virtual monoenergetic imaging; CTA, computed tomography angiography; HU, hounsfield unit.

Table 3. $\Delta CT_{D'}$ and $\Delta CT_{E'}$ in the five groups of VMIs compared with conventional CTA

Group	$\Delta CT_{D'}$ (HU)	<i>P</i> value	$\Delta CT_{E'}$ (HU)	<i>P</i> value
Conventional CTA	94.97 ± 56.76	-	94.60 ± 53.70	-
60 keV	99.53 ± 58.61	0.681	101.42 ± 59.21	0.441
70 keV	67.44 ± 41.56	0.003	78.02 ± 44.47	0.061
80 keV	51.29 ± 33.55	<0.001	64.22 ± 35.49	0.001
90 keV	44.95 ± 72.41	<0.001	66.13 ± 60.11	0.001
100 keV	28.47 ± 17.72	<0.001	51.63 ± 28.55	<0.001

VMI, virtual monoenergetic imaging; CTA, computed tomography angiography; HU, hounsfield unit.

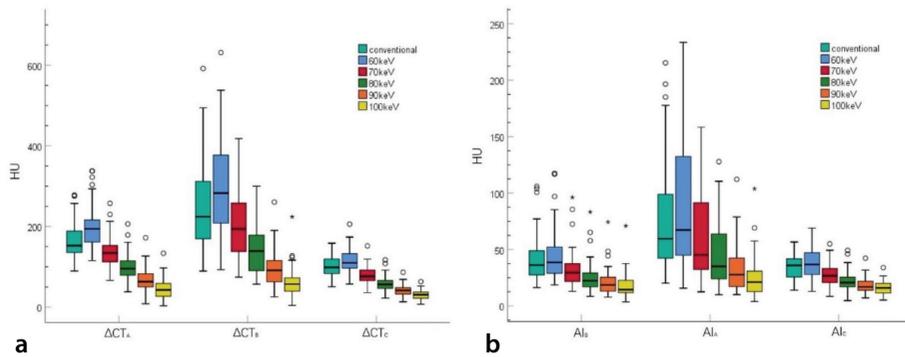


Figure 3. The box plots of the distribution of brain parenchyma parameters in conventional computed tomography angiography and virtual monoenergetic imaging with different keV levels. (a) The distribution of ΔCT_A , ΔCT_B , and ΔCT_C in six groups of images. (b) The distribution of AI_A , AI_B , and AI_C in six groups of images.

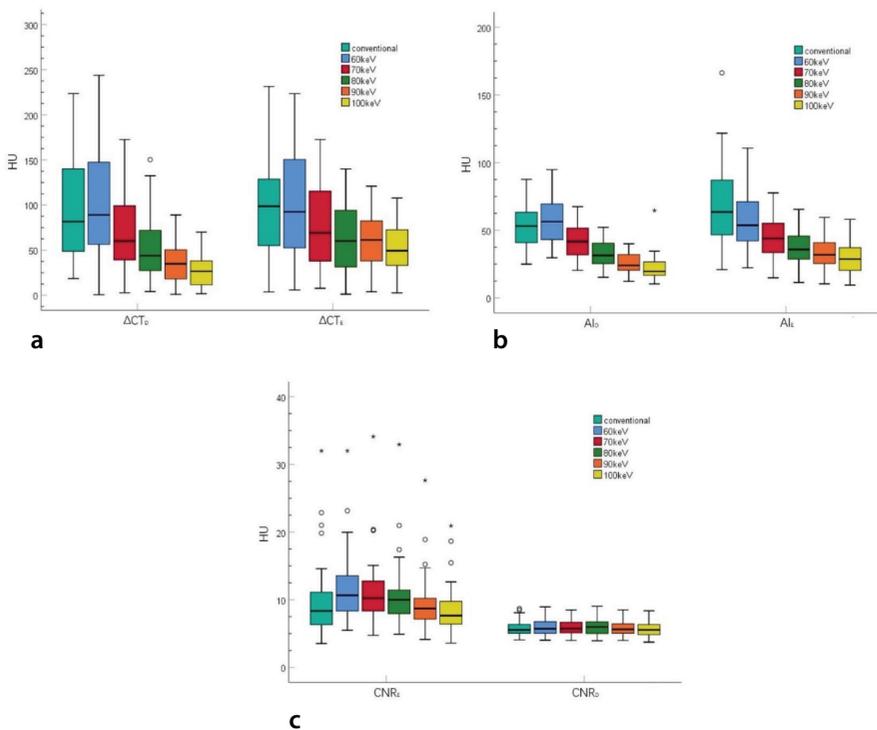


Figure 4. The box plots of the distribution of cerebrovascular parameters in conventional computed tomography angiography and virtual monoenergetic imaging with different keV levels. (a) The distribution of ΔCT_D and ΔCT_E in six groups of images. (b) The distribution of AI_D and AI_E in six groups of images. (c) The distribution of CNR_D and CNR_E in six groups of images.

100 keV, the cerebrovascular scores of other keV levels were significantly higher than those of conventional CTA ($P \leq 0.015$). Results are shown in Table 5. Figure 5 shows the conventional CTA and VMIs of a patient after aneurysm clipping.

Discussion

The dominant findings of our study can be summarized as follows: (1) metal artifacts were more serious on the tail of the clip than those on the head on both conventional CTA and VMIs; (2) VMIs can significantly reduce metal artifacts compared with conventional CTA except for 60 keV, whether at the head, tail or middle part of the clip; (3) metal artifacts of clips near the skull base were higher than those away from the skull base on both conventional CTA and VMIs; (4) the 70–80 keV VMIs showed the best compromise between clip artifact reduction and contrast vessel visibility.

There are many causes of the artifacts of implanted metal clips, including beam hardening, photon starvation, scattering, noise, and non-linear distribution effects, mainly characterized by the appearance of light and dark stripes around the clip.^{16–18} In this study, we found that there were different numbers and lengths of dark stripes in the head and tail of the clip, and the dark stripes in the tail were more serious than those in the head. The ΔCT_B value of the tail was larger than the ΔCT_A value of the head in both conventional CTA and VMIs. This may be related to the shape and position of the metal clip after implantation. The aneurysm clip is scissor-like; the shape of the head is quasi-round or similar to the number “8,” and the tail is thin and long after clipping, resulting in the tail absorbing more low-energy X-ray photons and, hence, a more obvious beam hardening effect. It is characterized by a wider range of low-density dark stripes. In addition, we found that there are multiple scattered patchy or coronal high-density artifacts around the middle part of the metal clip, which may be related to noise, photon

Table 4. AI_D , AI_E , CNR_D , and CNR_E in the five groups of VMIs compared with conventional CTA

Group	AI_D (HU)	<i>P</i> value	AI_E (HU)	<i>P</i> value	CNR_D	<i>P</i> value	CNR_E	<i>P</i> value
Conventional CTA	51.94 ± 13.81	-	67.94 ± 29.52	-	5.70 ± 1.06	-	9.62 ± 4.89	-
60 keV	57.41 ± 15.59	0.007	56.50 ± 19.02	<0.001	5.92 ± 1.19	0.313	11.64 ± 4.39	0.008
70 keV	42.94 ± 11.40	<0.001	44.51 ± 14.58	<0.001	5.94 ± 1.17	0.271	11.03 ± 4.34	0.061
80 keV	32.47 ± 9.09	<0.001	37.06 ± 12.56	<0.001	5.98 ± 1.22	0.195	10.34 ± 4.14	0.336
90 keV	25.56 ± 7.09	<0.001	33.09 ± 12.21	<0.001	5.84 ± 1.20	0.527	9.27 ± 3.60	0.646
100 keV	21.20 ± 6.05	<0.001	29.88 ± 12.14	<0.001	6.72 ± 1.20	0.957	8.47 ± 3.19	0.128

VMI, virtual monoenergetic imaging; CTA, computed tomography angiography; HU, hounsfield unit.

starvation, and similar effects. It affects the observation of the brain parenchyma around the clip and may cause false positives of cerebral hemorrhage, overestimates of the extent of cerebral hemorrhage, or coverage of cerebral infarction or subarachnoid hemorrhage, resulting in misdiagnosis or missed diagnosis. Meanwhile, it also affects the observation of the vessel around the clip and interferes with the diagnosis of residual or recurrent aneurysms, vascular stenosis, or vasospasm.

X-ray tubes generate photons with different energy levels. When they encounter met-

al clips with high attenuation coefficients, more low-energy photons are absorbed, whereas high-energy photons penetrate. This unequal proportion of absorption characteristics causes beam-hardening artifacts and reduces image quality. The ideal way to eliminate this artifact is for the tube to output photons of the same energy; however, such a device cannot be manufactured with current technology. Dual-energy CT virtually calculates the attenuation value of each voxel at different keV levels through high-energy and low-energy scanning modes, thus generating VMIs. In this retrospective study,

we researched the ability of VMI to reduce metal artifacts and enhance vascular contrast. We found that VMI can significantly reduce the metal artifacts compared with conventional CTA. The quantitative analysis showed that the Δ CT and AI at the head, tail, or middle part of the clip decreased significantly on VMIs except for 60 keV, meaning significant MAR. These results were similar to the qualitative analysis, showing a significant improvement in the proportions of brain parenchyma score ≥ 4 in 70, 80, and 90 keV. Several studies over the past decade⁸⁻¹² investigated the values of DECT-based VMI or MAR algorithms to reduce artifacts in clipped or coiled aneurysms. As previously mentioned, a skull phantom study¹⁰ showed that VMI significantly reduced metal artifacts compared with non-corrected images, which was similar to our results, and performed better when combined with the iterative MAR algorithm. MAR algorithms were designed to reduce artifacts caused by metal implants at low energy while preserving good image quality. Bier et al.⁸ found that iterative MAR algorithms improved non-enhanced CT image quality after clipping or coiling but did not improve single-energy CTA image quality and reduced adjacent vessel contrast in 30% of cases. They supposed that this was caused by novel artifacts generated by the over-correction of the algorithm. Furthermore, the competing interests of high-energy and low-energy images are a limitation of DECTA-based VMI. Thus, we thought that they were complementary, and a few studies have proved this.^{10,11} Furthermore, MAR algorithms are available for both single-energy and dual-energy CT acquisition, whereas VMI requires a dual-energy CT scanner, which can be expensive.

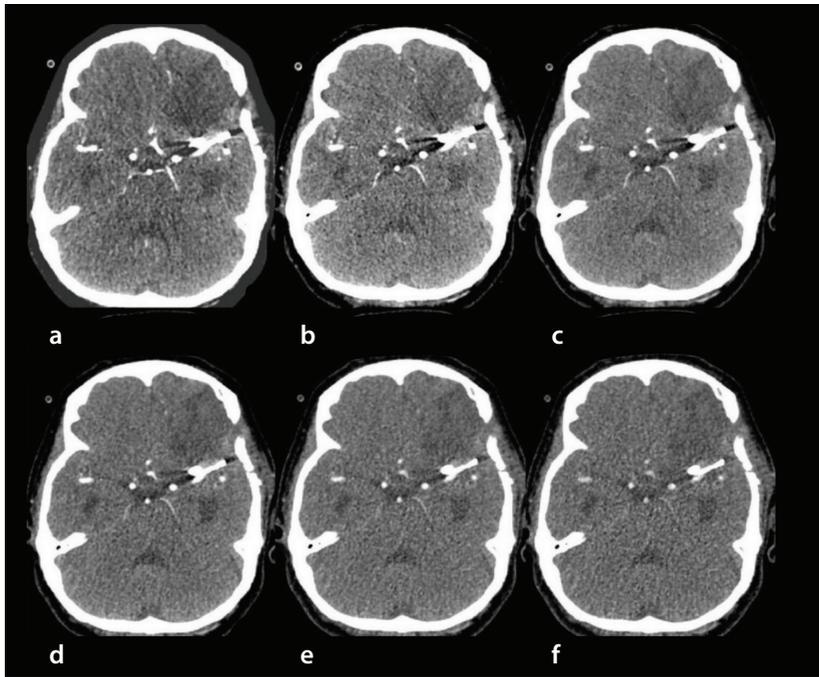


Figure 5. A 61-year-old woman underwent surgical clipping for a left middle cerebral artery aneurysm. Axial reformatted of a blended 115 kV image (a) and virtual monoenergetic imaging at 60 (b), 70 (c), 80 (d), 90 (e), and 100 keV (f) are shown. Virtual monoenergetic imaging at 70 (c) and 80 keV (d) shows the best compromise between the reduction of artifacts and contrast vessel visibility.

Table 5. Results from the qualitative analysis in the five groups of VMIs compared with conventional CTA									
Group (n, %)	Points	1	2	3	4	5	≥ 4	χ^2	P value
Brain parenchyma score	Conventional CTA	0 (0)	3 (5)	53 (88.33)	4 (6.67)	0 (0)	4 (6.67)	-	-
	60 keV	0 (0)	6 (10)	50 (83.3)	4 (6.67)	0 (0)	4 (6.67)	0.00	1.000
	70 keV	0 (0)	0 (0)	23 (38.33)	36 (60)	1 (1.67)	37 (61.67)	40.35	<0.001
	80 keV	0 (0)	1 (1.67)	11 (18.33)	48 (80)	0 (0)	48 (80)	65.70	<0.001
	90 keV	0 (0)	0 (0)	48 (80)	12 (20)	0 (0)	12 (20)	4.62	0.032
	100 keV	0 (0)	7 (11.67)	50 (83.3)	3 (5)	0 (0)	3 (5)	0.15	0.697
Cerebrovascular score	Conventional CTA	0 (0)	38 (63.33)	21 (35.0)	1 (1.67)	0 (0)	1 (1.67)	-	-
	60 keV	0 (0)	0 (0)	30 (50)	30 (50)	0 (0)	30 (50)	36.58	<0.001
	70 keV	0 (0)	0 (0)	3 (5)	56 (93.33)	1 (1.67)	57 (95)	104.65	<0.001
	80 keV	0 (0)	0 (0)	16 (26.67)	43 (71.67)	1 (1.67)	44 (73.3)	44.71	<0.001
	90 keV	0 (0)	2 (3.33)	50 (83.33)	8 (13.33)	0 (0)	8 (13.33)	5.89	0.015
	100 keV	0 (0)	17 (28.33)	43 (71.67)	0 (0)	0 (0)	0 (0)	1.01	0.315

VMI, virtual monoenergetic imaging; CTA, computed tomography angiography.

Due to the different methods and positions of metal clip implantation, the cerebral arteries of some patients are not necessarily affected by the artifacts of the head and tail of the clip simultaneously; therefore, we selected the vessel with the worst artifacts as ROI D. With the increase in keV level, ΔCT_D , ΔCT_E , AI_D , and AI_E gradually decreased, indicating that the vascular artifacts gradually decreased. Nevertheless, except for the CNR_E of 60 keV, there was no significant difference in CNR_D and CNR_E between VMIs and conventional CTA, indicating no significant increase in vascular contrast. In qualitative analysis, the proportions of cerebrovascular score ≥ 4 in VMIs were significantly higher than those of conventional CTA, except for 100 keV. This difference may be because CNR reflects the actual discernibility of vascular contrast in the presence of noise, whereas subjective scores evaluate the vascular contrast by comparing it with the surrounding brain parenchyma. This needs further investigation.

High-keV VMIs can reduce beam hardening artifacts,¹⁹ a phenomenon that has also been verified in our study. However, the contrast of high-keV images decreased, and the density resolution of vessels and brain parenchyma diminished. Our results showed that the ΔCT and AI values of 90 and 100 keV were lower than those of 70 and 80 keV, but the proportions of brain parenchyma and cerebrovascular scores ≥ 4 were also significantly lower than those of 70 and 80 keV. The proportions of brain parenchyma and cerebrovascular scores ≥ 4 of 70 keV were 61.67% and 95%, respectively, and the proportions of 80 keV were 80% and 73.3%, respectively, which were higher than those of conventional CTA, 60 keV, 90 keV, and 100 keV. Therefore, we believe that 70–80 keV is the optimal energy range for evaluation after cerebral aneurysm clipping. Dunet et al.¹¹ reported that the best compromise between MAR and relative CNR was obtained at 70–75 keV for GSI DECTA, but they did not compare their results with conventional CTA as we did, and the relative CNRs of the contralateral middle cerebral artery and internal carotid artery far from the clip were not representative of the visibility of the vessels adjacent to the clip. In addition, we found that metal artifacts of clips near the skull base were higher than those away from the skull base, whereas their results showed that the clips' location and number did not influence the ability of GSI with or without MAR to reduce metal artifacts. This problem needs further study.

There were several limitations in our study. First, due to the relatively small patient

cohort, there is no classification of different materials and sizes of clips; for example, cobalt alloy clips are known to produce more artifacts than titanium clips.²⁰ In addition, we only consider the patients treated by one clip, whereas it seems obvious that metal artifacts increase with the number of clips. Therefore, it is necessary to increase the sample size to further study the application value of the DECTA-based VMI technique for clips of different materials, sizes, and numbers. Second, this study is retrospective, so blended 115 kV images were used instead of conventional CTA. There may be some differences between them, although previous studies have reported that blended images are similar to conventional CT.²¹ In addition, it is unethical to perform two CTA scans (single-energy and dual-energy) in the same patient. Finally, we did not compare the VMI with other MAR techniques.

In conclusion, our research demonstrated that for patients who underwent DECTA after cerebral aneurysm clipping, the 70–80 keV VMIs are expected to be the optimal energy range for balancing metal artifacts and cerebrovascular visibility. The clip artifacts of 70 keV images are lower than those of conventional CTA, and the cerebrovascular contrast can meet the clinical evaluation. Images at 80 keV can further reduce clip artifacts and better show the changes in the surrounding brain parenchyma. We recommend using them together.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was supported by the Fujian Provincial Science and Technology Department, Fujian, China (grant number 2023J011727), and the Putian University, Fujian, China (grant number 2022107).

References

1. Texakalidis P, Sweid A, Mouchtouris N, et al. Aneurysm formation, growth, and rupture: the biology and physics of cerebral aneurysms. *World Neurosurg.* 2019;130:277-284. [CrossRef]
2. Cianfoni A, Pravata E, De Blasi R, Tschuor CS, Bonaldi G. Clinical presentation of cerebral aneurysms. *Eur J Radiol.* 2013;82(10):1618-1622. [CrossRef]
3. Abila AA, Lawton MT. Predictors of complications with unruptured middle

cerebral artery aneurysm clipping in a surgically treated series of 416 patients: a clip first approach is still best. *World Neurosurg.* 2015;84(4):884-885. [CrossRef]

4. Brown MA, Parish J, Guandique CF, et al. A long-term study of durability and risk factors for aneurysm recurrence after microsurgical clip ligation. *J Neurosurg.* 2017;126(3):819-824. [CrossRef]
5. Piao J, Luan T, Qu L, Yu J. Intracranial post-clipping residual or recurrent aneurysms: current status and treatment options (Review). *Med Int (Lond).* 2021;1(1):1. [CrossRef]
6. Yu LB, Fang ZJ, Yang XJ, Zhang D. Management of residual and recurrent aneurysms after clipping or coiling: clinical characteristics, treatments, and follow-up outcomes. *World Neurosurg.* 2019;122:838-846. [CrossRef]
7. Zachenhofer I, Cejna M, Schuster A, Donat M, Roessler K. Image quality and artifact generation post-cerebral aneurysm clipping using a 64-row multislice computer tomography angiography (MSCTA) technology: a retrospective study and review of the literature. *Clin Neurol Neurosurg.* 2010;112(5):386-391. [CrossRef]
8. Bier G, Bongers MN, Hempel JM, et al. Follow-up CT and CT angiography after intracranial aneurysm clipping and coiling-improved image quality by iterative metal artifact reduction. *Neuroradiology.* 2017;59(7):649-654. [CrossRef]
9. Pan YN, Chen G, Li AJ, et al. Reduction of metallic artifacts of the post-treatment intracranial aneurysms: effects of single energy metal artifact reduction algorithm. *Clin Neuroradiol.* 2019;29(2):277-284. [CrossRef]
10. Winkhofer S, Hinzpeter R, Stocker D, et al. Combining monoenergetic extrapolations from dual-energy CT with iterative reconstructions: reduction of coil and clip artifacts from intracranial aneurysm therapy. *Neuroradiology.* 2018;60(3):281-291. [CrossRef]
11. Dunet V, Bernasconi M, Hajdu SD, Meuli RA, Daniel RT, Zerlauth JB. Impact of metal artifact reduction software on image quality of gemstone spectral imaging dual-energy cerebral CT angiography after intracranial aneurysm clipping. *Neuroradiology.* 2017;59(9):845-852. [CrossRef]
12. Mocanu I, Van Wettene M, Absil J, Bruneau M, Lubicz B, Sadeghi N. Value of dual-energy CT angiography in patients with treated intracranial aneurysms. *Neuroradiology.* 2018;60(12):1287-1295. [CrossRef]
13. Waaijer A, Prokop M, Velthuis BK, Bakker CJ, de Kort GA, van Leeuwen MS. Circle of Willis at CT angiography: dose reduction and image quality—reducing tube voltage and increasing tube current settings. *Radiology.* 2007;242(3):832-839. [CrossRef]
14. Christner JA, Kofler JM, McCollough CH. Estimating effective dose for CT using

- dose-length product compared with using organ doses: consequences of adopting International Commission on Radiological Protection publication 103 or dual-energy scanning. *AJR Am J Roentgenol.* 2010;194(4):881-889. [\[CrossRef\]](#)
15. Yun SY, Heo YJ, Jeong HW, et al. Dual-energy CT angiography-derived virtual non-contrast images for follow-up of patients with surgically clipped aneurysms: a retrospective study. *Neuroradiology.* 2019;61(7):747-755. [\[CrossRef\]](#)
 16. Bucher AM, Wichmann JL, Schoepf UJ, et al. Quantitative evaluation of beam-hardening artefact correction in dual-energy CT myocardial perfusion imaging. *Eur Radiol.* 2016;26(9):3215-3222. [\[CrossRef\]](#)
 17. Wellenberg RHH, Hakvoort ET, Slump CH, Boomsma MF, Maas M, Streekstra GJ. Metal artifact reduction techniques in musculoskeletal CT-imaging. *Eur J Radiol.* 2018;107:60-69. [\[CrossRef\]](#)
 18. Mori I, Machida Y, Osanai M, Iinuma K. Photon starvation artifacts of X-ray CT: their true cause and a solution. *Radiol Phys Technol.* 2013;6(1):130-141. [\[CrossRef\]](#)
 19. Katsura M, Sato J, Akahane M, Kunimatsu A, Abe O. Current and novel techniques for metal artifact reduction at CT: practical guide for radiologists. *Radiographics.* 2018;38(2):450-461. [\[CrossRef\]](#)
 20. van der Schaaf I, van Leeuwen M, Vlassenbroek A, Velthuis B. Minimizing clip artifacts in multi CT angiography of clipped patients. *AJNR Am J Neuroradiol.* 2006;27(1):60-66. [\[CrossRef\]](#)
 21. Marin D, Boll DT, Mileto A, Nelson RC. State of the art: dual-energy CT of the abdomen. *Radiology.* 2014;271(2):327-342. [\[CrossRef\]](#)

Supplementary Table S1. Five-point Likert scale for evaluation of the artifacts severity of brain parenchyma and the cerebrovascular visibility

Points	Brain parenchyma score	Cerebrovascular score
1	Severe artifacts, mostly not diagnostic	Severe artifacts, mostly not diagnostic
2	Poor image quality, partly non-diagnostic	Reduced vessel delineation, partly non-diagnostic
3	Moderate image quality, limitedly diagnostic	Partly limited vessel contrast, limitedly diagnostic
4	Good image quality, enough for diagnostic	Good vessel delineation, enough for diagnostic
5	Perfect image quality, no artifacts	Excellent vessel delineation, no artifacts

Supplementary Table S2. Details of clips used in the study

No.	Model	Material	Blade length (mm)	Shape	Maximum opening (mm)	Closing force (N)	Closing force (g)	Specification
1	FT740T	Titanium	7.0	Straight	6.2	1.47	150	Yasargil, Braun Medical, Tuttlingen, Germany
2	FT742T	Titanium	6.5	Curved	6.0	1.47	150	Yasargil, Braun Medical, Tuttlingen, Germany
3	FT750T	Titanium	9.0	Straight	7.0	1.77	180	Yasargil, Braun Medical, Tuttlingen, Germany
4	FT752T	Titanium	8.3	Curved	6.8	1.77	180	Yasargil, Braun Medical, Tuttlingen, Germany
5	FT760T	Titanium	11.0	Straight	7.8	1.77	180	Yasargil, Braun Medical, Tuttlingen, Germany
6	FT762T	Titanium	10.2	Curved	7.5	1.77	180	Yasargil, Braun Medical, Tuttlingen, Germany
7	FT782T	Titanium	13.7	Curved	8.7	1.96	200	Yasargil, Braun Medical, Tuttlingen, Germany

Supplementary Table S3. Study patients' characteristics (n = 60)

Characteristics	Mean ± SD or n
Age (years)	55.15 ± 9.24
Male/female	23/37
Aneurysm localization (left/right)	
Anterior communicating artery	17
Anterior cerebral artery	2/1
Middle cerebral artery	17/6
Posterior cerebral artery	1/0
Internal carotid artery	8/6
Posterior communicating artery	0/2
Postoperative complication	
None	9
Cerebral hemorrhage	20
Cerebral infarction	7
Encephaledema	33
Hydrocephalus	12
Epidural or subdural hematoma	20
Subarachnoid hemorrhage	12
Cerebral vasospasm	8
Residual aneurysm	2

SD, standard deviation.

Supplementary Table S4. Inter-reader and Intra-reader ICCs regarding attenuation and SD values of ROIs in the six groups of images

Parameters		Conventional CTA	60 keV	70 keV	80 keV	90 keV	100 keV
Attenuation	ROI A	0.901/0.912	0.953/0.924	0.940/0.936	0.921/ 0.995	0.981/0.981	0.978/0.946
	ROI B	0.893/0.951	0.898/0.979	0.972/0.970	0.982/0.931	0.940/0.944	0.937/0.945
	ROI C	0.906/0.956	0.921/0.945	0.907/0.952	0.971/0.967	0.977/0.916	0.934/0.930
	ROI D	0.926/0.945	0.936/0.992	0.923/0.957	0.960/0.943	0.923/0.920	0.942/0.925
	ROI E	0.967/0.934	0.935/0.929	0.898/0.925	0.965/0.918	0.956/0.995	0.936/0.911
	OLP	0.925/0.937	0.977/0.935	0.943/ 0.911	0.950/0.955	0.956/0.970	0.902/0.969
	BA	0.907/0.990	0.907/0.981	0.923/0.937	0.934/0.942	0.914/0.990	0.892/0.913
SD	ROI A	0.907/0.953	0.949/0.937	0.923/0.974	0.938/0.957	0.934/0.985	0.911/0.929
	ROI B	0.932/0.947	0.906/0.967	0.981/0.953	0.937/0.918	0.975/0.918	0.892/0.965
	ROI C	0.899/0.990	0.954/0.927	0.959/0.988	0.963/0.923	0.931/0.945	0.901/0.940
	ROI D	0.948/0.918	0.909/0.976	0.907/0.961	0.940/0.939	0.904/0.985	0.942/0.978
	ROI E	0.972/0.924	0.934/0.972	0.953/0.945	0.939/0.955	0.923/0.930	0.953/0.957
	OLP	0.912/0.916	0.972/0.972	0.928/0.989	0.955/0.940	0.985/0.929	0.926/0.960
	BA	0.974/0.966	0.928/0.977	0.928/0.929	0.943/0.982	0.971/0.986	0.980/0.980

Please note that these values were inter-reader ICC/intra-reader ICC. ROI, region of interest; CTA, computed tomography angiography; SD, standard deviation; OLP, occipital lobe parenchyma; BA, basilar artery; ICC, intraclass correlation coefficient.

Supplementary Table S5. Comparison of the severity of metal artifacts between clips near or away from the skull base in the six groups of images

	ΔCT_{near} (HU)	ΔCT_{away} (HU)	Z	P	AI _{near} (HU)	AI _{away} (HU)	Z	P
Conventional CTA	288.51 ± 103.21	208.27 ± 80.27	-3.252	0.001	93.80 ± 49.66	62.63 ± 42.24	-2.756	0.006
60 keV	347.88 ± 115.83	249.64 ± 101.54	-3.252	0.001	104.76 ± 57.39	71.25 ± 48.83	-2.483	0.013
70 keV	239.27 ± 79.55	168.64 ± 66.31	-3.397	0.001	73.76 ± 39.62	51.07 ± 33.50	-2.451	0.014
80 keV	164.50 ± 59.43	119.28 ± 45.45	-2.804	0.005	53.54 ± 28.80	38.70 ± 23.04	-2.163	0.031
90 keV	112.42 ± 45.83	81.81 ± 29.60	-2.980	0.003	39.23 ± 22.11	30.16 ± 16.55	-1.730	0.084
100 keV	74.92 ± 37.98	56.27 ± 21.32	-2.451	0.014	30.82 ± 18.67	25.83 ± 13.47	-0.961	0.336

ΔCT_{near} , ΔCT of clips near the skull base; ΔCT_{away} , ΔCT of clips away from the skull base; AI_{near}, artifact index of clips near the skull base; AI_{away}, artifact index of clips away from the skull base; Z/P, statistics/significance of Mann-Whitney U test; HU, hounsfield unit; CTA, computed tomography angiography.

Supplementary Table S6. Cohen's kappa coefficients of the scores between radiologists 1 and 2 in the six groups of images

Score	Conventional CTA	60 keV	70 keV	80 keV	90 keV	100 keV
Brain parenchyma score	0.633	0.771	0.666	0.725	0.832	0.691
Cerebrovascular score	0.797	0.694	0.706	0.691	0.694	0.697

CTA, computed tomography angiography.



Maximum standardized uptake value-to-tumor size ratio in fluorodeoxyglucose F18 positron emission tomography/computed tomography: a simple prognostic parameter for non-small cell lung cancer

Soo Jeong Kim¹
 Koeun Lee¹
 Hyun Joo Lee²
 Du-Young Kang³
 Young Hwan Kim¹

¹Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Department of Nuclear Medicine, Seoul, Republic of Korea

²Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Department of Pathology, Seoul, Republic of Korea

³Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Department of Thoracic and Cardiovascular Surgery, Seoul, Republic of Korea

PURPOSE

By correcting the effect of tumor size on metabolic activity, the maximum standardized uptake value-to-tumor size (SUV_{max} :tumor size) ratio on fluorodeoxyglucose F18 positron emission tomography (^{18}F -FDG PET)/computed tomography (CT) scans can be a prognostic parameter of non-small cell lung cancer (NSCLC). The current study evaluates the prognostic value of SUV_{max} :tumor size ratio on pretreatment ^{18}F -FDG PET/CT scans in patients with NSCLC. Furthermore, the SUV_{max} :tumor size ratio is compared with other established PET parameters.

METHODS

This study included 108 patients with NSCLC who underwent pretreatment ^{18}F -FDG PET/CT scans and curative lung surgery. The associations between the SUV_{max} :tumor size ratio and other conventional PET parameters were investigated. The recurrence-free survival according to the SUV_{max} :tumor size ratio was also analyzed. In addition, the SUV_{max} :tumor size ratio was compared according to postoperative pathologic findings.

RESULTS

In total, 72 (66.7%) of the 108 participants presented with adenocarcinoma (ADC). Nineteen (17.6%) patients experienced recurrence during a median follow-up period of 32.3 months. The median SUV_{max} :tumor size ratio was 2.37 (1.23 for ADCs and 3.90 for other histologic types). The SUV_{max} :tumor size ratio was associated with SUV_{max} and mean SUV, as well as metabolic tumor volume and total lesion glycolysis. Patients with an SUV_{max} :tumor size ratio higher than the median had a worse recurrence outcome than those with an SUV_{max} :tumor size ratio lower than the median. Participants with ADC who presented with lymphovascular invasion had a higher SUV_{max} :tumor size ratio than those without. The presence of lymph node metastasis and advanced histologic grade were associated with a high SUV_{max} :tumor size ratio in patients with ADC.

CONCLUSION

The SUV_{max} :tumor size ratio on pretreatment ^{18}F -FDG PET/CT scans was associated with aggressive tumor behavior and poor outcome in NSCLCs, particularly ADC.

CLINICAL SIGNIFICANCE

The SUV_{max} :tumor size ratio on pretreatment ^{18}F -FDG PET/CT scans has a prognostic value in patients with NSCLCs, especially ADC.

KEYWORDS

Cancer, fluorodeoxyglucose, lung, marker, positron emission tomography, prognosis

Corresponding author: Young Hwan Kim

E-mail: yohan2727@naver.com

Received 08 May 2024; revision requested 07 July 2024; accepted 24 August 2024.



Epub: 01.10.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242837

You may cite this article as: Kim SJ, Lee K, Lee HJ, Kang DY, Kim YH. Maximum standardized uptake value-to-tumor size ratio in fluorodeoxyglucose F18 positron emission tomography/computed tomography: a simple prognostic parameter for non-small cell lung cancer. *Diagn Interv Radiol.* 2025;31(3):274-279.

Despite advancement in prevention, screening, and management in recent decades, non-small cell lung cancer (NSCLC) is among the leading causes of cancer-related mortality worldwide.¹ Fluorodeoxyglucose F18 (¹⁸F-FDG) positron emission tomography/computed tomography (PET/CT) is a widely used imaging tool for NSCLC management. The role of ¹⁸F-FDG PET/CT scanning in determining the status of solitary pulmonary nodules (malignant or not), staging of lung cancer, planning of radiation therapy, and evaluating treatment response is well established.²

The ¹⁸F-FDG PET/CT technique is used for prognostic prediction in NSCLC. Studies have shown that ¹⁸F-FDG uptake is related to prognosis. Among these studies, one revealed the presence of histological invasion in early-stage adenocarcinoma (ADC).³ In addition, in a meta-analysis of surgically resected NSCLC, a high standardized uptake value (SUV) and other metabolic parameters [e.g., metabolic tumor volume (MTV) and total lesion glycolysis (TLG)] were poor prognosis factors of disease-free survival and overall survival.⁴ However, the prognostic significance of the maximum SUV (SUV_{max}), a representative parameter of ¹⁸F-FDG uptake in tumors, has not as yet been completely elucidated.⁵ The inconsistent results can be attributed to various confounding factors affecting the SUV_{max}.

The SUV_{max}-to-tumor size ratio (SUV_{max}:tumor size) on ¹⁸F-FDG PET/CT scans was introduced to assess the metabolic activity of NSCLC by correcting the effect of tumor size, an established prognostic factor, on the SUV. Studies have shown that it has a prognostic value in patients with NSCLC.^{6,7} Moreover, this indicator does not correspond to the fact that thresholds are not easy to apply when measuring metabolic parameters in the early stages of NSCLC because of the generally low SUV values. However, studies supporting

the use of this indicator in clinical settings are still lacking.

The current study investigates the prognostic value of the SUV_{max}:tumor size ratio on pretreatment ¹⁸F-FDG PET/CT scans in patients with NSCLC. Furthermore, the SUV_{max}:tumor size ratio is compared with other established PET parameters.

Methods

Participants

This study recruited 131 consecutive patients who underwent an ¹⁸F-FDG PET/CT scan between March 2020 and December 2021 and lung surgery within 30 days after examination at our institution. Among the patients, 13 with benign postoperative pathologic results, four with small cell lung cancer, three who received neoadjuvant treatment, two with distant metastasis at the time of surgery, and one with myxofibrosarcoma were excluded from the analysis. Finally, 108 patients were included in the study (Figure 1).

The Kangbuk Samsung Hospital Institutional Review Board (decision no: KBSMC 2024-02-040, date: 03/05/2024) approved this study. The need for written informed consent from the participants was waived.

Assessment of the medical records of the patients

The medical records of the patients were assessed for clinical and tumor characteristics, including pathologic findings after surgery, type of treatment, clinical follow-up results, and recurrence diagnosis. Cancer stage was based on the American Joint Committee on Cancer Staging Manual, 8th edition. Lung cancer recurrence was diagnosed via

pathologic examination or imaging results assessed by the attending physician. Recurrence-free survival (RFS) was defined as the date of recurrent lesion detection or last follow-up date from the date of surgery

Fluorodeoxyglucose F18 positron emission tomography/computed tomography scan acquisition

The patients fasted for at least 6 h, and each patient's blood glucose level was <200 mg/dL at the time of ¹⁸F-FDG injection. PET/CT scan images were obtained approximately 60 min after the intravenous administration of ¹⁸F-FDG at a dose of 2.96 MBq/kg. The Discovery MI system (GE Healthcare, Milwaukee, WI, USA) was used for ¹⁸F-FDG PET/CT scan examination. Following the CT scan (120 kVp; 10–80 mA; section thickness: 3.75 mm), emission PET was performed from the thigh to the skull base (1.5 min per bed). The PET images were reconstructed using the Bayesian penalized likelihood reconstruction algorithm with attenuation-corrected images.

Image analyses

The PET/CT scan images were reviewed on a dedicated workstation (AW; GE Healthcare, Chicago, IL, USA) by a nuclear medicine physician who can recognize lesions that would be surgically resected. A volume of interest was drawn over lung cancer lesions with an ¹⁸F-FDG uptake greater than that of surrounding background activity. The semi-automatic method was used to delineate the boundaries of the tumor with the SUV-based contouring software (volume viewer software on GE AW 4.7). The SUV threshold was set to 2.5. Data on the SUV_{max}, mean SUV (SUV_{mean}), and MTV were recorded. TLG was defined as the product of MTV and

Main points

- Fluorodeoxyglucose F18 positron emission tomography (¹⁸F-FDG PET)/computed tomography (CT) is a widely used imaging tool for non-small cell lung cancer (NSCLC) management.
- The maximum standardized uptake value (SUV_{max})-to-tumor size ratio on ¹⁸F-FDG PET/CT scans was introduced to assess the metabolic activity of NSCLC by correcting the effect of tumor size.
- The SUV_{max}:tumor size ratio on ¹⁸F-FDG PET/CT scans was associated with aggressive tumor behavior and poor outcome in NSCLCs, particularly adenocarcinoma.

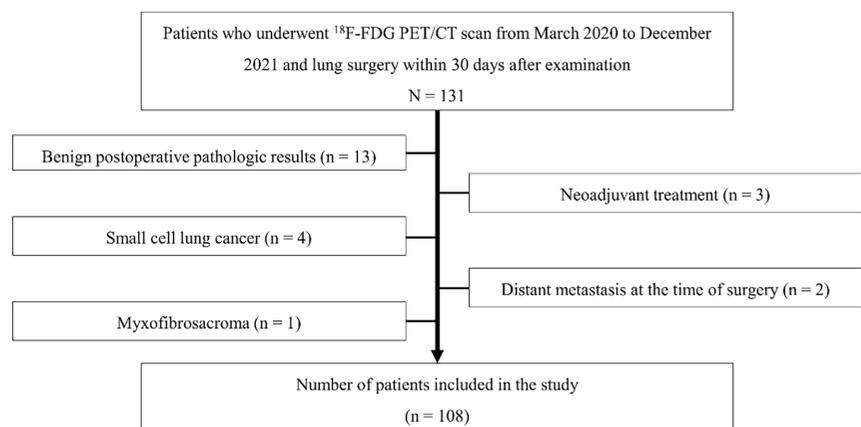


Figure 1. Flowchart of the study participants. F18-FDG, fluorodeoxyglucose F18; PET/CT, positron emission tomography/computed tomography.

SUV_{mean}. If the SUV_{max} of a lung cancer lesion was <2.5, only SUV_{max} was measured, and other volumetric PET parameters were omitted. The longest diameter of the lung cancer lesion on ¹⁸F-FDG PET/CT scan images was also measured. The SUV_{max}:tumor size ratio was calculated by dividing the SUV_{max} of the lung cancer lesion by the longest diameter of the lesion.

Statistical analysis

The clinical characteristics and PET/CT scan parameters of the participants were compared according to histologic findings using the Mann–Whitney U test, chi-square test, or Fisher’s exact test. The association between the SUV_{max}:tumor size ratio and other PET parameters was evaluated using Kendall’s tau-b correlation coefficient. Survival analysis was performed using the Kaplan–Meier method with subgroup survival estimates compared using the log-rank test. The Cox proportional hazards regression model was used to analyze the prognostic impact of variables. Finally, the SUV_{max}:tumor size ratio was compared according to postoperative pathologic findings. In the analysis using SUV_{max}:tumor size ratios in which patients had lung cancer lesions without a discernable ¹⁸F-FDG uptake were treated as zero.

Jamovi version 2.3.28 was used to perform statistical analysis. A *P* value of <0.05 was considered statistically significant.

Results

Characteristics of the participants

Table 1 shows the clinical characteristics of all patients. Their median age was 66 years (range: 22–83) and 62 (57.4%) were men. In total, 72 (66.7%) of the 108 participants were diagnosed with ADC, whereas 36 patients presented with histologic findings other than ADC. The findings included squamous cell carcinoma [*n* = 21 (58.3%)], pleomorphic carcinoma [*n* = 9 (25.0%)], large-cell carcinoma [*n* = 4 (11.1%)], adenosquamous carcinoma [*n* = 1 (2.8%)], and atypical carcinoid [*n* = 1 (2.8%)]. Patients diagnosed with ADC had a lower median age than those with other histologic types. Furthermore, the proportion of male patients diagnosed with ADC was lower than that of male patients with other histologic types. Overall, 58 (53.7%) of the 108 patients were diagnosed with stage IA or lower lung cancer, and most of them were diagnosed with ADC. In total, 94 (87.0%) of the 108 patients underwent lobectomy, whereas 24 (22.2%) received adjuvant treatment. In

total, 19 (17.6%) patients developed recurrence during a median follow-up period of 32.3 months.

Fluorodeoxyglucose F18 positron emission tomography/computed tomography scan parameters

The ¹⁸F-FDG uptake in 101 (93.5%) of the 108 patients with lung cancer was discernable. The median SUV_{max} was 5.71 and the median tumor size was 2.35 cm. The median SUV_{max}:tumor size ratio was 2.37 (range: 0–25.4). More specifically, the median SUV_{max}:tumor size ratios were 1.23 for ADCs

and 3.90 for other histologic types. Binomial logistic regression analysis of the preoperative SUV_{max}:tumor size ratio for histologic type other than ADC returned an odds ratio of 2.13 [95% confidence interval (CI): 1.56–2.91]. The area under curve was 0.872, the sensitivity was 50.0%, and the specificity was 86.1 when the cut-off value was 0.5. There were 31 (28.7%) patients with an SUV_{max} of <2.5, and all these patients were diagnosed with ADC. The volumetric PET parameters of 76 patients were measured. However, in one patient, the volumetric PET parameters could not be measured due to image data issues. Table 2 shows the volumetric PET pa-

Table 1. Characteristics of the participants

		All patients (<i>n</i> = 108)	Patients with adenocarcinoma (<i>n</i> = 72)	Patients with other histologic types (<i>n</i> = 36)	<i>P</i> value
Age	Years, median (range)	66 (22–83)	65 (22–80)	68.5 (47–83)	0.042 ^a
	Sex	Male	62 (57.4%)	30 (41.7%)	32 (88.9%)
Female		46 (42.6%)	42 (58.3%)	4 (11.1%)	
Type of surgery	Lobectomy	94 (87.0%)	61 (84.7%)	33 (91.7%)	0.378 ^c
	Segmentectomy	7 (6.5%)	6 (8.3%)	1 (2.8%)	0.420 ^c
	Wedge resection	7 (6.5%)	5 (6.9%)	2 (5.6%)	1.000 ^c
		0	1 (0.9%)	1 (1.4%)	0
Stage	IA1	15 (13.9%)	15 (20.8%)	0	0.002 ^c
	IA2	28 (25.9%)	23 (31.9%)	5 (13.9%)	0.044 ^b
	IA3	14 (13.0%)	9 (12.5%)	5 (13.9%)	1.000 ^c
	IB	19 (17.6%)	11 (15.3%)	8 (22.2%)	0.372 ^b
	IIA	4 (3.7%)	0	4 (11.1%)	0.011 ^c
	IIB	16 (14.8%)	9 (12.5%)	7 (19.4%)	0.338 ^a
	IIIA	10 (9.3%)	4 (5.6%)	6 (16.7%)	0.081 ^c
	IIIB	1 (0.9%)	0	1 (2.8%)	0.333 ^c
Tumor grade	1	13 (12.0%)	12 (16.7%)	1 (2.8%)	0.056 ^c
	2	37 (34.3%)	27 (37.5%)	10 (27.8%)	0.316 ^b
	3	41 (38.0%)	27 (37.5%)	14 (38.9%)	0.888 ^b
	4	4 (3.7%)	0	4 (11.1%)	0.011 ^c
	Unspecified	13 (12.0%)	6 (8.3%)	7 (19.4%)	0.120 ^c
Lymphovascular invasion	Present	29 (26.9%)	12 (16.7%)	17 (47.2%)	<0.001 ^b
Adjuvant treatment		24 (22.2%)	12 (16.7%)	12 (33.3%)	0.050 ^b
	Chemotherapy	18 (16.7%)	11 (15.3%)	7 (19.4%)	0.584 ^b
	Radiotherapy only	1 (0.9%)	0	1 (2.8%)	0.333 ^c
	Concurrent chemoradiotherapy	5 (4.6%)	1 (0.4%)	4 (11.1%)	0.041 ^c
Recurrence	Yes	19 (17.6%)	7 (9.7%)	12 (33.3%)	0.002 ^b
	Censored	89 (82.4%)	65 (90.3%)	24 (66.7%)	
Recurrence-free survival	Months, median (range)	30.3 (0.6–46.9)	32.1 (1.5–46.9)	26.9 (0.6–42.7)	0.096 ^a

^aMann–Whitney U test; ^bchi-square test; ^cFisher’s exact test.

rameters. Each PET parameter distribution was biased, with a significant frequency in the near-zero interval (Figure 2).

The SUV_{max} :tumor size ratio was found to be correlated with SUV_{max} and SUV_{mean} (Kendall's tau-b = 0.678 and 0.495, respectively; $P < 0.001$) and MTV and TLG (Kendall's tau-b = 0.191 and 0.243; $P = 0.015$ and 0.002, respectively). However, the correlation strength of MTV and TLG was not as strong as that of SUV_{max} or SUV_{mean} . Similar results were obtained for the ADC subgroup. Nevertheless, the SUV_{max} :tumor size ratio was not significantly associated with MTV or TLG in other patients with NSCLC.

Maximum standardized uptake value-to-tumor size ratio and recurrence-free survival analysis

Patients with an SUV_{max} :tumor size ratio higher than the median (2.37) had a worse recurrence outcome than those with an SUV_{max} :tumor size ratio lower than the median (Figure 3). In patients with ADC, an SUV_{max} :tumor size ratio of >1.23 was consistently associated with worse outcomes (Figure 4). Based on univariate analysis, the hazard ratio of a high SUV_{max} :tumor size ratio for recurrence was 6.01 (1.75–20.65, $P = 0.004$) in all participants and 6.22 (0.75–51.68, $P = 0.091$) in patients with ADC ($n = 72$). A high SUV_{max} :tumor size ratio was associated with poor prognosis based on the multivariate analysis using clinical variables, albeit that it did not reach statistical significance (Table 3).

Association between the maximum standardized uptake value-to-tumor size ratio and aggressive tumor behavior based on pathologic findings following surgery

The preoperative ^{18}F -FDG PET/CT scan parameters according to postoperative pathologic results were compared in patients with ADC and those with other histologic types. In the ADC group, patients with lymphovascular invasion had a higher SUV_{max} :tumor size ratio than those without (3.53 vs. 0.98, $P < 0.001$). Patients with lymph node metastasis had a higher SUV_{max} :tumor size ratio than those without (3.70 vs. 1.04, $P = 0.006$). An advanced histological grade was associated with a high SUV_{max} :tumor size ratio in ADC (Figure 5). In patients with ADC, the odds ratios of the SUV_{max} :tumor size ratio were 1.57 (1.14–2.30, $P = 0.010$) for lymphovascular invasion, 1.42 (0.99–2.09, $P = 0.050$) for lymph node metastasis, and 1.92 (1.33–2.98, $P = 0.001$) for histologic grade 3 lesions.

Unlike the ADC group, the other histologic type group showed no significant differences in SUV_{max} :tumor size ratio according to the lymphovascular invasion (3.80 vs. 4.50, $P = 0.346$), the presence of lymph node metastasis (3.77 vs. 5.18, $P = 0.220$) and histologic grade 3/4 lesions (4.15 vs. 4.82, $P = 0.387$).

Discussion

The clinical implications for SUV_{max} :tumor size ratio on pretreatment ^{18}F -FDG PET/CT in surgically resected NSCLC were evaluated. The SUV_{max} :tumor size ratio was found to be

associated with conventional metabolic PET parameters in patients with ADC. An SUV_{max} :tumor size ratio higher than the median was associated with worse recurrence outcomes. Patients with ADC who presented with lymphovascular invasion, lymph node metastasis, or histologic grade 3 lesions based on the pathologic results had a higher SUV_{max} :tumor size ratio than those without. However, the results did not significantly differ in the non-ADC group.

In addition to SUV_{max} , the volumetric PET parameters, including MTV and TLG, for the

Table 2. Fluorodeoxyglucose F18 positron emission tomography/computed tomography scan parameters in patients with lung cancer lesions

	All patients (n = 108)	Patients with adenocarcinoma (n = 72)	Patients with other histologic types (n = 36)	P value
Median (range)				
Tumor size	2.35 (0.90–8.90)	2.15 (0.90–7.20)	2.90 (1.10–8.90)	0.006 ^a
No ^{18}F -FDG uptake	7 (6.5%)	7 (9.7%)	0	0.093 ^b
SUV_{max}	5.71 (0.69–63.48)	3.10 (0.69–5.22)	12.1 (5.22–63.48)	$<0.001^a$
SUV_{max} /tumor size	2.37 (0–25.4)	1.23 (0–9.25)	3.90 (2.34–25.39)	$<0.001^a$
$SUV_{max} < 2.5$	31 (28.7%)	31 (43.1%)	0	$<0.001^b$
SUV_{mean}	4.37 (2.56–18.48)	3.46 (2.56–8.28)	5.20 (3.49–18.48)	$<0.001^a$
MTV (cm ³)	4.21 (0.02–229)	2.08 (0.02–82.4)	10.7 (0.61–229)	$<0.001^a$
TLG	19.2 (0.05–2249)	7.60 (0.05–617)	49.8 (2.11–2249)	$<0.001^a$
Correlation coefficient of SUV_{max} -to-tumor size ^c				
SUV_{max}	0.678 (<0.001)	0.709 (<0.001)	0.713 (<0.001)	
SUV_{mean}	0.495 (<0.001)	0.539 (<0.001)	0.654 (<0.001)	
MTV	0.191 (0.015)	0.247 (0.025)	-0.150 (0.381)	
TLG	0.243 (0.002)	0.292 (0.008)	-0.062 (0.719)	

^aMann-Whitney U test; ^bFisher's exact test; ^cKendall's tau-b (P value). SUV_{max} , maximum standardized uptake; SUV_{mean} , mean standardized uptake; MTV, metabolic tumor volume; TLG, total lesion glycolysis; ^{18}F -FDG, fluorodeoxyglucose F18.

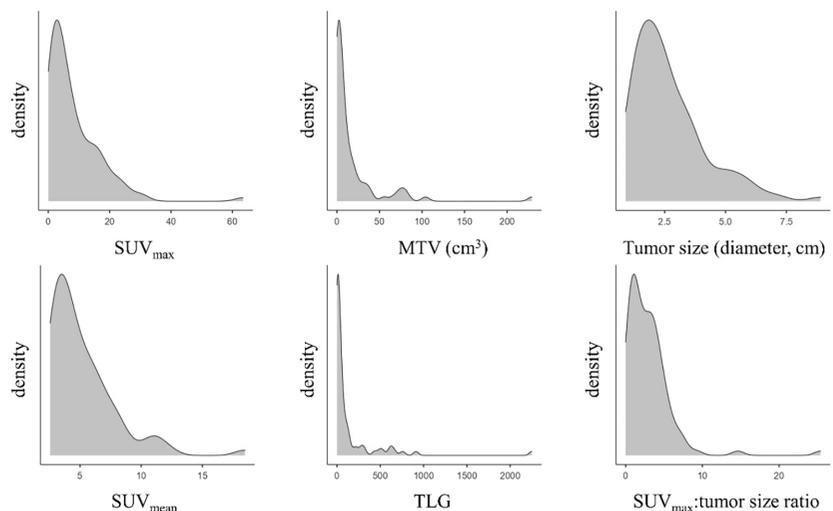


Figure 2. The fluorodeoxyglucose F18 positron emission tomography/computed tomography scan-derived parameters of patients with non-small cell lung cancer ($n = 108$) were distributed considerably close to zero. The volumetric parameters, metabolic tumor volume and total lesion glycolysis, were evident. SUV_{max} , maximum standardized uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

prognostic stratification of NSCLC have been evaluated.⁸ However, there is no clear standard method for segmentation to measure the parameters, and the various methods affect the measurements.⁹ An SUV threshold of 2.5 is widely used to delineate the tumor, and this value was also adopted in this study. However, other methods can be utilized. A specific percentage of SUV_{max} as a threshold or a gradient method without threshold are also frequently used to define tumors. However, all these methods have a common weakness in tumors with a generally low ¹⁸F-FDG uptake, such as ADC.¹⁰ In the present study, the SUV_{max} of 31 (43.1%) of 72 patients with ADCs was <2.5. Hence, volumetric PET parameters were not measured in these patients. In contrast, the SUV_{max}:tumor size ratio could be measured in most cases and was correlated with the volumetric parameters in ADCs.

In this study, an SUV_{max}:tumor size ratio higher than the median was associated with a poor RFS. This result is similar to that of previous studies on SUV_{max}:tumor size ratio.^{6,7} Furthermore, in patients with ADC, a higher SUV_{max}:tumor size ratio was consistently related to worse outcomes. This study did not reach the level of prognostic prediction modeling via multivariate analysis. However, this outcome can be achieved in further studies with a larger number of participants and an extended follow-up period.

A higher SUV_{max}:tumor size ratio was associated with unfavorable postoperative pathologic findings in ADCs. This is consistent with the explanation that an increase in ¹⁸F-FDG uptake is related to the aggressive characteristics of tumor cells.¹¹ Tumor cell density can be another factor affecting the SUV_{max}:tumor size ratio and is associated with tumor grade in lung ADC.¹² However, in the present study, no significant association was found in the remaining NSCLC group except ADC. The lack of association could not be explained based on the small number of participants alone. It seems reasonable to distinguish ADC from the other histologic types when evaluating the prognostic value of ¹⁸F-FDG uptake in NSCLC.

Interestingly, in a meta-analysis on the prognostic value of ¹⁸F-FDG PET/CT scanning for surgically resected NSCLC, the results presented by classifying the values into lower and higher than an SUV_{max} of 6.0 had somewhat different patterns.⁴ The adjusted hazard ratio of the SUV_{max} for disease-free survival was 4.63 (2.53–8.48) in the sub-threshold group and was higher than 1.68 (95% CI: 1.07–2.63) in the above threshold group.

Therefore, the association between the tumor ¹⁸F-FDG uptake and prognosis may weaken following disease progress. Furthermore, acidosis can reduce ¹⁸F-FDG uptake by inhibiting aerobic glycolysis with cancer progression.^{13,14}

The current study has several limitations. First, this retrospective study focused on patients with NSCLC who underwent surgical resection, meaning selection bias could have affected the results. The potential value of predicting histologic tumor type by preop-

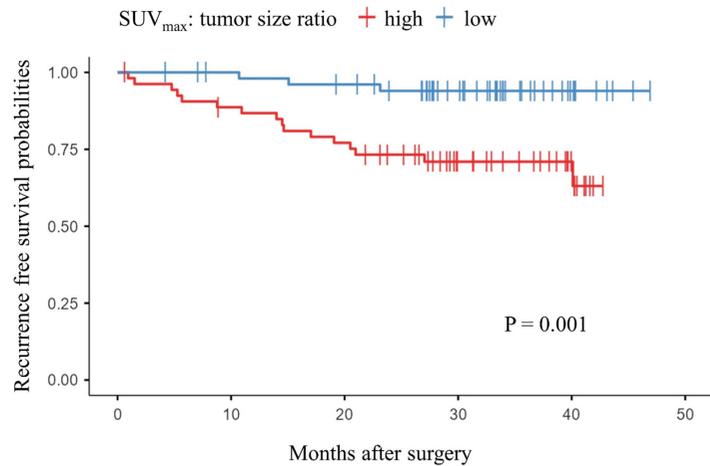


Figure 3. Patients with a maximum standardized uptake value (SUV_{max})-to-tumor size ratio higher than the median (2.37) had a significantly worse recurrence outcome than those with an SUV_{max}-to-tumor size ratio lower than the median ($P = 0.001$).

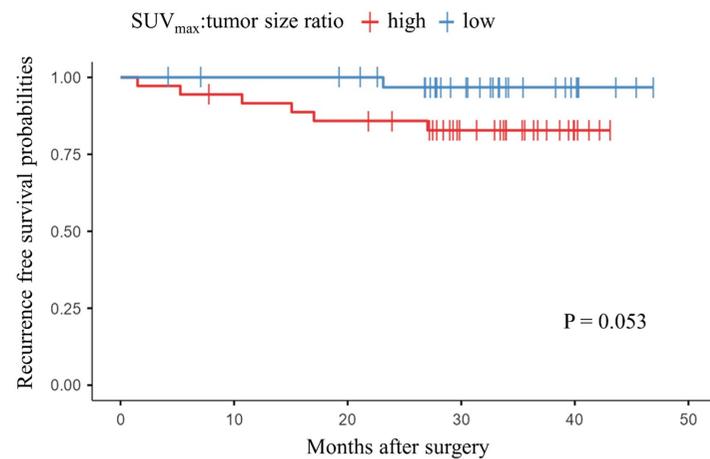


Figure 4. Patients with adenocarcinoma who presented with a maximum standardized uptake value (SUV_{max})-to-tumor size ratio higher than the median (1.23) tended to have a worse recurrence outcome than those with an SUV_{max}-to-tumor size ratio lower than the median ($P = 0.053$).

Table 3. Recurrence-free survival in patients with non-small cell lung cancer who underwent surgical resection based on a multivariate analysis

Variables	HR	95% CI	<i>P</i> value
Male sex	0.59	0.18–1.98	0.397
Age ≥70 years	0.51	0.18–1.49	0.221
pT3 or pT4	3.60	1.28–10.13	0.015
pN1 or pN2	1.14	0.36–3.57	0.824
Histology, not adenocarcinoma	2.09	0.57–7.67	0.264
SUV _{max} -to-tumor size ratio greater than the median (2.37)	3.62	0.89–14.67	0.072

HR, hazard ratio; CI, confidence interval; SUV_{max}, maximum standardized uptake.

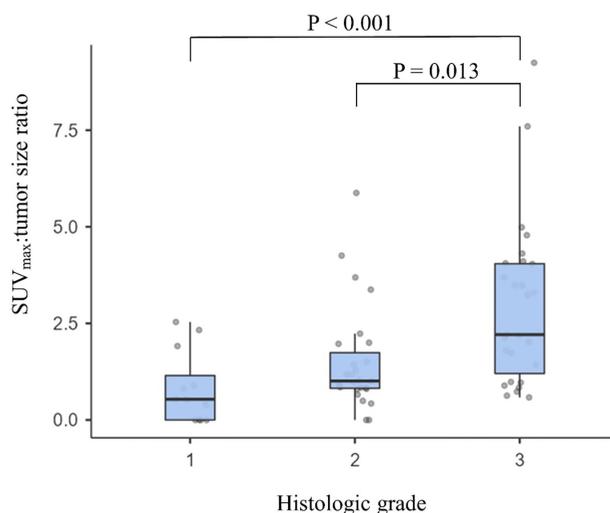


Figure 5. The boxplot of maximum standardized uptake value (SUV_{max})-to-tumor size ratio according to histologic grade in 72 patients with ADC showed a higher level of SUV_{max}-to-tumor size ratio for more advanced tumor grade. ADC, adenocarcinoma.

erative SUV_{max}:tumor size ratio could be additionally evaluated using a larger number of participants with various histologic tumor types in the future. In addition, various factors might have influenced the clinical judgement of physicians regarding treatments, such as the type of treatment modality and adjuvant treatment following surgery. Among them, ¹⁸F-FDG PET/CT scan findings might have had an influence to an uncontrolled extent. Second, the follow-up duration of this study was limited. Thus, overall survival could not be evaluated. Controlled studies with a larger number of patients and extended follow-up periods should be conducted to validate the prognostic value of the SUV_{max}:tumor size ratio in NSCLC. Compared with other complicated parameters, the SUV_{max}:tumor size ratio can be advantageous for multicenter research as it involves less variation based on the institution.

In conclusion, the SUV_{max}-to-tumor size ratio on ¹⁸F-FDG PET/CT scanning was associated with aggressive tumor behavior and poor outcomes in NSCLCs, particularly ADC.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med.* 2020;41(1):1-24. [\[Crossref\]](#)
- Groheux D, Quere G, Blanc E, et al. FDG PET-CT for solitary pulmonary nodule and lung cancer: literature review. *Diagn Interv Imaging.* 2016;97(10):1003-1017. [\[Crossref\]](#)
- Shao X, Shao X, Niu R, Xing W, Wang Y. A simple prediction model using fluorodeoxyglucose-PET and high-resolution computed tomography for discrimination of invasive adenocarcinomas among solitary pulmonary ground-glass opacity nodules. *Nucl Med Commun.* 2019;40(12):1256-1262. [\[Crossref\]](#)
- Liu J, Dong M, Sun X, Li W, Xing L, Yu J. Prognostic value of 18F-FDG PET/CT in surgical non-small cell lung cancer: a meta-analysis. *PLoS One.* 2016;11(1):e0146195. [\[Crossref\]](#)
- Kaseda K. Recent and current advances in FDG-PET imaging within the field of clinical oncology in NSCLC: a review of the literature. *Diagnostics (Basel).* 2020;10(8):561. [\[Crossref\]](#)
- Stiles BM, Nasar A, Mirza F, et al. Ratio of positron emission tomography uptake to tumor size in surgically resected non-small cell lung cancer. *Ann Thorac Surg.* 2013;95(2):397-404. [\[Crossref\]](#)
- Chen F, Yao Y, Ma C, et al. Ratio of maximum standardized uptake value to primary tumor size is a prognostic factor in patients with advanced non-small cell lung cancer. *Transl Lung Cancer Res.* 2015;4(1):18-26. [\[Crossref\]](#)
- Im HJ, Pak K, Cheon GJ, et al. Prognostic value of volumetric parameters of (18)F-FDG PET in non-small-cell lung cancer: a meta-analysis. *Eur J Nucl Med Mol Imaging.* 2015;42(2):241-251. [\[Crossref\]](#)
- Zhuang M, Garcia DV, Kramer GM, et al. Variability and repeatability of quantitative uptake metrics in (18)F-FDG PET/CT of non-small cell lung cancer: impact of segmentation method, uptake interval, and reconstruction protocol. *J Nucl Med.* 2019;60(5):600-607. [\[Crossref\]](#)
- Hicks RJ. The value of the standardized uptake value (SUV) and metabolic tumor volume (MTV) in lung cancer. *Semin Nucl Med.* 2022;52(6):734-744. [\[Crossref\]](#)
- Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science.* 2009;324(5930):1029-1033. [\[Crossref\]](#)
- Lin S, Samsundar JP, Bandari E, et al. Digital quantification of tumor cellularity as a novel prognostic feature of non-small cell lung carcinoma. *Mod Pathol.* 2023;36(3):100055. [\[Crossref\]](#)
- Vaupel P, Multhoff G. Revisiting the Warburg effect: historical dogma versus current understanding. *J Physiol.* 2021;599:1745-1757. [\[Crossref\]](#)
- Peppicelli S, Andreucci E, Ruzzolini J, Bianchini F, Calorini L. FDG uptake in cancer: a continuing debate. *Theranostics.* 2020;10(7):2944-2948. [\[Crossref\]](#)



Consolidation in the right middle lobe in pediatric bronchial–pulmonary artery shunt: radiology’s Aunt Minnie?

Chi Wang^{1*}
 Rongchang Wu^{2*}
 Zihan Wang³
 Shuai Ma¹
 Xinyu Yuan¹
 Yuchun Yan¹
 Yun Peng²

¹Capital Institute of Pediatrics, Department of Radiology, Beijing, China

²Beijing Children’s Hospital, National Center for Children’s Health, Capital Medical University, Department of Radiology, Beijing, China

³Beijing Friendship Hospital, Capital Medical University, Department of Radiology, Beijing, China

PURPOSE

By retrospectively studying the chest computed tomography (CT) data of children with bronchial artery (BA)–pulmonary artery fistula, this study summarizes the characteristic imaging features of the disease and provides imaging support for the diagnosis and clinical treatment of these children.

METHODS

Digital subtraction angiography and CT angiography data were collected from 74 children with pulmonary hemorrhage following BA embolization. Bronchial–pulmonary shunt was present in 30 cases.

RESULTS

Of the 74 children with pulmonary hemorrhage in this study, seven exhibited signs of consolidation in the middle lobe of the right lung, and bronchial–pulmonary shunt existed in all of them. A total of 30 children with BA–pulmonary artery shunt (PAS) had BA tortuosity and thickening. Regarding primary BA-PAS, the middle lobe and lower lobe of the right lung were involved in 94.1% (16) of the children. Those with a fistula located in the middle lobe of the right lung accounted for 58.8% (10 cases), of which 40.0% (four cases) presented consolidation. In this study, 41.2% (seven) of the children with primary BA-PAS exhibited no abnormal changes on chest CT, and 58.8% (10 cases) exhibited abnormal changes in the unilateral lung.

CONCLUSION

For children with pulmonary hemorrhage who have consolidation in the right middle lobe, the formation of BA-PAS should be anticipated. The possibility of primary BA-PAS should not be disregarded in children with pulmonary hemorrhage with tortuosity and dilation of BAs, despite no apparent abnormalities on lung CT, or ground-glass density or consolidation on only one side.

CLINICAL SIGNIFICANCE

The chest CT of patients with pulmonary hemorrhage showed consolidation of the right middle lobe of the lung, which was highly likely to indicate BA-PAS.

KEYWORDS

Computed tomography, X-ray, child, bronchial artery–pulmonary artery shunt

*Joint first authors

Corresponding author: Yuchun Yan, Yun Peng

E-mail: yanyuchun@aliyun.com, ppengyun@hotmail.com

Received 28 June 2024; revision requested 03 August 2024; accepted 09 September 2024.



Epub: 21.10.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242908

Pulmonary hemorrhage occurs when blood enters the alveoli or airways, which is rare and may be insidious among children.¹ Clinical symptoms in children are variable and can range from asymptomatic to bloody sputum or hemoptysis.² However, it is challenging to detect pulmonary hemorrhage or hemoptysis in younger children because they usually swallow bloody sputum or because the bleeding is minimal. Common causes of pulmonary hemorrhage in children include idiopathic hemosiderosis, pulmonary cystic fibrosis, pulmonary vascular malformation, primary heart disease, hematological disease, and tumors.³

The bronchial artery (BA) arises from the T5–6 level of the thoracic aorta. The BAs outside the descending thoracic aorta at the T5–6 level are defined as ectopic BAs, also known as vagal

You may cite this article as: Wang C, Wu R, Wang Z, et al. Consolidation in the right middle lobe in pediatric bronchial–pulmonary artery shunt: radiology’s Aunt Minnie? *Diagn Interv Radiol.* 2025;31(3):280-284.

BA⁴. BAs normally receive only approximately 1% of the heart's blood output to maintain airway and lung function. They supply blood to the bronchi, esophagus, visceral pleura, thoracic aorta, pulmonary arteries, nerves, pulmonary veins, and lymph nodes in the thoracic cavity.⁵ They do not normally participate in gas exchange. The BA is connected to the pulmonary artery by anastomosis of several microvessels at the level of the alveolar and respiratory bronchioles. In children with pulmonary hemorrhage, the BA and its anastomotic site may dilate to form a BA–pulmonary artery shunt (PAS),⁶ also termed bronchial–pulmonary arterial fistula,⁷ which further exacerbates bleeding symptoms. The normal pressure in the pulmonary artery is approximately 13 mmHg, whereas the pressure in the BA can be as high as 95 mmHg. The marked difference in pressure leads to the rupture of the BA-PAS, causing pulmonary hemorrhage. A BA-PAS can also be primary, and some researchers believe that it is related to fetal hypoxia in utero.⁸ Primary BA-PAS is an exclusive diagnosis in the clinic. Children with primary BA-PAS are often untreated and require BA embolization to improve clinical symptoms. Bronchial arteriography is an invasive test that is often performed in patients with massive hemoptysis.⁹ The interventional treatment for primary and secondary BA-PAS is the same. The formation of BA-PAS is a major indication of BA embolization.¹⁰

Early detection of BA-PAS on computed tomography (CT) and timely BA embolization can effectively alleviate patients' symptoms and improve their prognosis. We hypothesize that there are characteristic findings on CT imaging that increase the probability of BA-PAS diagnosis in children that may otherwise be missed. We thus reviewed 74 BA-PAS cases from the National Medical Center of Pediatrics in China. If our hypothesis is prov-

en to be correct, children with BA-PAS could benefit from this study.

Methods

Data acquisition

This study was a single-center retrospective observational study conducted between March 1, 2016 and March 31, 2020. The study was approved by the Ethics Committee of Beijing Children's Hospital (decision no: IEC-C-008-A08-V.05.1, date: 25.12.2018). Being a retrospective study, it was approved by the ethics committee without informed consent.

The CT images and clinical data of 334 children with pulmonary hemorrhage were reviewed. A total of 74 children who underwent bronchial arteriography for pulmonary hemorrhage were enrolled in the study. The inclusion criteria were as follows: (1) clinical diagnosis of pulmonary hemorrhage; (2) patients who underwent chest CT; (3) patients who underwent BA digital subtraction angiography (DSA); and (4) patients who underwent chest CT angiography (CTA). The exclusion criteria included: (1) patients with pulmonary hemorrhage of unclear etiology; (2) poor-quality CT scan images for evaluation; (3) patients with poor adherence or unable to cooperate adequately with the examination because of their illness or other reasons; and (4) patients with lung contusion and laceration.

Scan protocol

All DSA scans were performed using a Siemens AXIOM Artis. Patients were under general anesthesia during the operation, where a percutaneous femoral artery puncture using a modified Seldinger technique was performed. Super-selective contrast was administered using the hand method (10–15 mL, 2–3 mL/s). The contrast agent was non-ionic (Ultravist® 370). Once the location of the BA was established, super-selective intubation was performed using a 3F coaxial microcatheter on a guidewire first principle. Angiography was performed repeatedly, and important branches, such as the spinal and esophageal arteries, were avoided during the procedure. The main size of the embolization material was approximately 300–500 μm , and the size of the selected particles was adjusted appropriately according to whether there was an arteriovenous fistula. If the BA shares a trunk with the internal thoracic artery, protective embolization of the distal artery can be performed using a spring coil. Following the manual identification of

the BA, angiography was performed using a high-pressure syringe. The presence of a pulmonary artery image in the arterial phase indicated BA-PAS; therefore, embolization treatment was performed.

All CTA scans were performed using the Discovery CT750 HD system (GE Healthcare). During scanning, the left or right median cubital vein was selected for puncture and indwelling to confirm smooth venous access and leakage. Following the preparation, the patient was taken to the CT scan room and was placed in the supine position with the feet advanced. The tube voltage used was 80–100 kVp, which is low, and the tube current was configured using automatic tube current regulation technology. The tube current range was 20–500 mA, and the noise index (NI) was set according to the weight of the child: NI: 11 for 0–16 kg, NI: 13 for 16–35 kg, and NI: 15 for >35 kg. The speed of the tube was 0.5 seconds with a pitch of 0.992:1. The contrast agent used for enhanced scanning was iophorol (320 mgI/mL) (Hengrui Medicine, China), with 1.2–1.6 mL/kg of the contrast agent administered according to the child's weight, and the injection was completed within 15 seconds using a single-cylinder high-pressure syringe. The arterial phase scanning was started at 17 seconds, and the intravenous scanning was performed at 50 seconds. The original image obtained was reduced by 70% weighted adaptive statistical iterative reconstruction-V (GE Healthcare), a practical standard algorithm, without any additional core algorithm. The image slices were reconstructed at 5 and 0.625 mm.

Image evaluation

All images were read by two senior pediatric radiologists. The sites of the lesion and imaging features were recorded, including consolidation, ground-glass opacity (GGO), and reticular opacity. A consensus was reached following discussion.

Statistical analysis

All statistical analyses were performed using Jeffrey's Amazing Statistics Program (version 0.18.1). All continuous variables that conformed to the normal distribution were expressed by mean \pm standard deviation ($\bar{x} \pm s$). The chi-squared test and Fisher's exact test were used to compare CT features between the BA-PAS and non-BA-PAS groups. The level of significance was set at α : 0.05.

Main points

- Chest computed tomography (CT) images of pulmonary hemorrhage in children are varied and may even be negative.
- Primary bronchial artery–pulmonary artery shunt (BA-PAS) typically manifests as ground-glass opacity or consolidation in one side of the lung.
- Not all patients with BA-PAS develop right middle lobe consolidation.
- The formation of BA-PAS should be considered when the right middle lobe of the lung is seen on chest CT in children with pulmonary hemorrhage.

Results

Demographics and clinical measures of participants

A total of 74 children with pulmonary hemorrhage were included in this study, including 32 boys and 42 girls, ranging in age from 1 to 12 years, with a mean age of 7.43 ± 3.59 years. Of the 74 children, 36 had diffuse alveolar hemorrhage, including 15 boys and 21 girls. In addition, there were 30 children with BA-PAS, 17 with primary BA-PAS and 13 with secondary BA-PAS, including 12 boys and 18 girls.

Of the 74 children with pulmonary hemorrhage, seven exhibited signs of consolidation in the middle lobe of the right lung, and all of them had BA-PAS, including four cases of primary BA-PAS and three cases of secondary BA-PAS. Consolidation was not observed on chest CT in the other children; BA-PAS in the right middle lobe occurred in 17 cases (10 primary BA-PAS and 7 secondary BA-PAS). BA tortuosity and thickening were observed in all 30 children with BA-PAS.

Correlation analysis

The CT findings were divided into three categories: normal, GGO, and consolidation (Figure 1). All imaging features are presented in Tables 1 and 2. The odds ratios of normal, GGO, and consolidation were 0.533 ($P = 0.238$), 1.050 ($P = 0.918$), and 2.250 ($P = 0.141$), respectively (Table 3).

Discussion

The blood in the pulmonary artery is mainly used for air exchange. The BAs supply blood to the lung parenchyma. When abnormalities occur in the lung parenchyma, the blood supply to the BAs increases, leading to lumen thickening and even BA-PAS, which can be primary or secondary. Many diseases can cause BA-PAS, including tumors, diffuse alveolar damage, tumors, and bronchiectasis. Primary BA-PAS intestinal manifestations of recurrent pulmonary hemorrhage are often difficult to distinguish from idiopathic pulmonary hemosiderosis on imaging, and hormone therapy is not effective. Because BA-PAS can be a complication of many diseases, primary BA-PAS is often treated as an exclusive diagnosis. Interventional treatment of BA-PAS is both effective and necessary. As an invasive procedure, angiography needs to be applied with caution. Identification of BA-PAS on CT images is important for both radiologists and pulmonologists.

In this study, of the 74 children with pulmonary hemorrhage, seven exhibited signs of consolidation in the middle lobe of the right lung, and BA-PAS was present in all of them, including four cases of primary BA-PAS and three cases of secondary BA-PAS. Moreover, BA-PAS was observed in 30 children (17 primary BA-PAS cases and 13 secondary BA-PAS cases), and BA tortuosity and thickening were observed in all of them (Figure 2). Regarding primary BA-PAS, the middle lobe and lower lobe of the right lung were

involved in 94.1% (16) of the children. Those with a fistula located in the middle lobe of the right lung accounted for 58.8% (10 cases), of which 40.0% (four cases) had consolidation in the middle lobe of the right lung. In this study, 41.2% (seven) of the children with primary BA-PAS exhibited no abnormal changes on chest CT, and 58.8% (10 cases) exhibited abnormal changes in the unilateral lung.

The threshold for BA thickening in adults is 2 mm,¹¹ and there is no corresponding standard in children, as their arteries are generally compared with the contralateral BAs. In this study, all children with BA-PAS exhibited signs of tortuous and enlarged bronchial arteries on the affected side. Primary BA-PAS lesions are predominantly found in the right lower lobe of the lung, and among children with lesions in this region, approximately 50% exhibit consolidation of the right middle lobe. In this study, while not all children diagnosed with BA-PAS presented with consolidation of the right middle lobe, none of the children classified as Non-BA-PAS showed any signs of such consolidation. Therefore, when there is consolidation in the right middle lobe of the lung in children with pulmonary hemorrhage, bronchial arteriography should be performed promptly to confirm the presence of BA-PAS. If the condition is present, embolization should be performed promptly. In 41.2% of the children with primary BA-PAS, there was no apparent hemorrhage on the chest CT scan. In 58.8% of the children with primary BA-PAS, there was only GGO and consolidation shadows in one lung or lobe. In children with pulmonary hemorrhage who have no abnormalities on lung CT or only unilateral lesions, primary BA-PAS should be strongly suspected. Right middle lobe consolidation in secondary BA-PAS may be related to pulmonary deposition during pulmonary hemorrhage. Non-right middle lobe consolidation, GGO, and normal CT findings were not statistically different between the BA-PAS and non-BA-PAS groups of children with pulmonary hemorrhage.

The accumulation of hemosiderin damages lung tissue and leads to fibrosis.¹² This chronic pathological process leads to dilatation of the BAs and the formation of secondary BA-PAS. The right middle lobe bronchus is susceptible to near or total obstruction because its intraluminal diameter is smaller than that of other lobar bronchi.¹³ Furthermore, the fissures of the middle lobe and lingula isolate these segments from collateral ventilation, reducing blood clearance.¹⁴

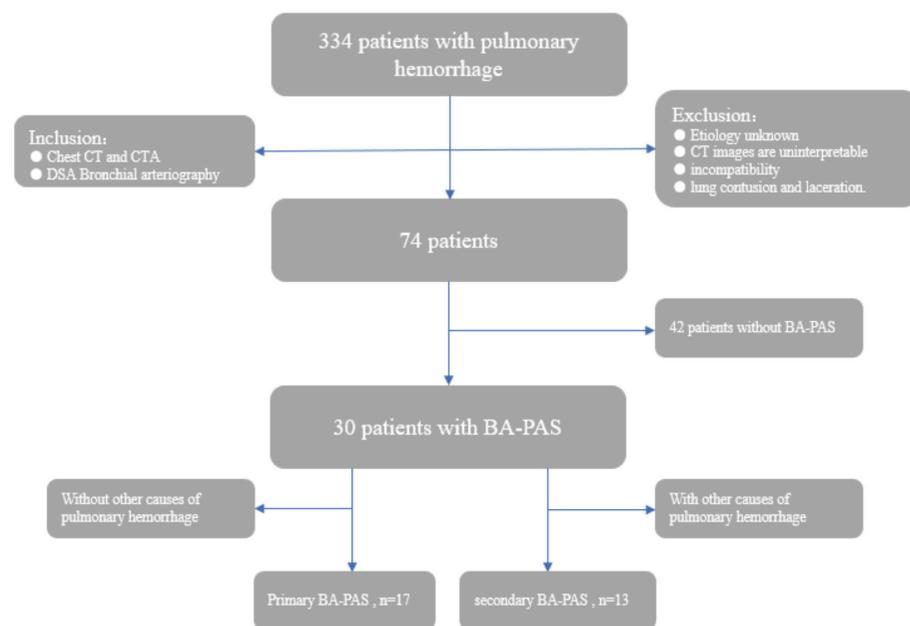


Figure 1. Summary of patient recruitment and exclusions. CT, computed tomography; CTA, computed tomography angiography; DSA, digital subtraction angiography; BA-PAS, bronchial artery–pulmonary artery shunt.

Table 1. Data of children with primary BA-PAS

Number	Age (years)	Sex	Location of BA-PAS	CT imaging
1	14	F	RML	RML consolidation
2	13	F	RML	RML, RLL consolidation
3	13	F	RML	Normal
4	12	F	RLL	RLL GGO
5	12	M	RML	Normal
6	11	F	RML	Normal
7	10	M	RLL, RUL	Normal
8	10	F	RLL	Normal
9	8	F	RLL	Right lung multiple GGOs
10	8	F	RUL	Normal
11	7	M	RML	Normal
12	6	F	RML	RML consolidation
13	5	F	RML	RML consolidation; bilateral multiple GGOs
14	3	M	RLL, RUL	Right lung multiple GGOs
15	3	F	RLL	RLL consolidation
16	5	M	RLL	RLL GGO
17	4	M	RUL	Right lung multiple GGOs

BA-PAS, bronchial artery–pulmonary artery shunt; CT, computed tomography; RML, right middle lobe; RLL, right lower lobe; GGO, ground-glass opacity; RUL, right upper lobe; F, female; M, male.

Table 2. Data of children with secondary BA-PAS

Number	Age (years)	Sex	Location of BA-PAS	Clinical diagnosis	CT imaging
1	14	F	RML	DAH	LML GGO
2	12	M	RLL	CF	Bilateral multiple patchy consolidation and GGO
3	9	F	RML, RUL	DAH	Bilateral multiple GGOs
4	9	M	RUL, RLL	IPH	Bilateral multiple GGOs
5	8	M	RLL	Pneumonia	RML consolidation
6	7	F	RML	HT	RML consolidation
7	7	M	RML	Mucoepidermoid carcinoma of main airway	RML consolidation
8	4	F	RML	DAH	Bilateral multiple GGOs
9	3	M	LLL	Pneumonia	LLL consolidation
10	2	F	RUL, RML, RLL	Hemangioma of the right main bronchus	Right lung multiple GGOs
11	2	F	RLL	IPH	Bilateral multiple GGOs
12	2	M	RLL, LLL	IPH	Bilateral multiple GGOs
13	1	M	RLL	Pneumonia	RLL multiple GGOs

BA-PAS, bronchial artery–pulmonary artery shunt; CT, computed tomography; RML, right middle lobe; DAH, diffuse alveolar hemorrhage; LML, left middle lobe; GGO, ground-glass opacity; RUL, right upper lobe; RLL, right lower lobe; IPH, idiopathic pulmonary hemosiderosis; HT, hereditary telangiectasia; LLL, left lower lobe; CF, cystic fibrosis; F, female; M, male.

Table 3. Comparison of the BA-PAS and non-BA-PAS groups of children with hemorrhage

CT findings	BA-PAS (n = 30)			Non-BA-PAS (n = 44)	Odds ratio	P value
	Primary (n = 17) (%)	Secondary (n = 13) (%)	Total (%)			
Normal	7 (41.2)	0	7 (23.3)	16 (36.4)	0.533	0.238
GGO	5 (29.4)	9 (69.2)	14 (46.7)	20 (45.4)	1.050	0.918
Consolidation	5 (29.4)	5 (38.5)	10 (33.3)	8 (18.2)	2.250	0.141
RML consolidation	4 (23.6)	3 (23.1)	7 (23.3)	0	-	0.070*

*Fisher's exact test; BA-PAS, bronchial artery–pulmonary artery shunt; CT, computed tomography; GGO, ground-glass opacity; RML, right middle lobe.

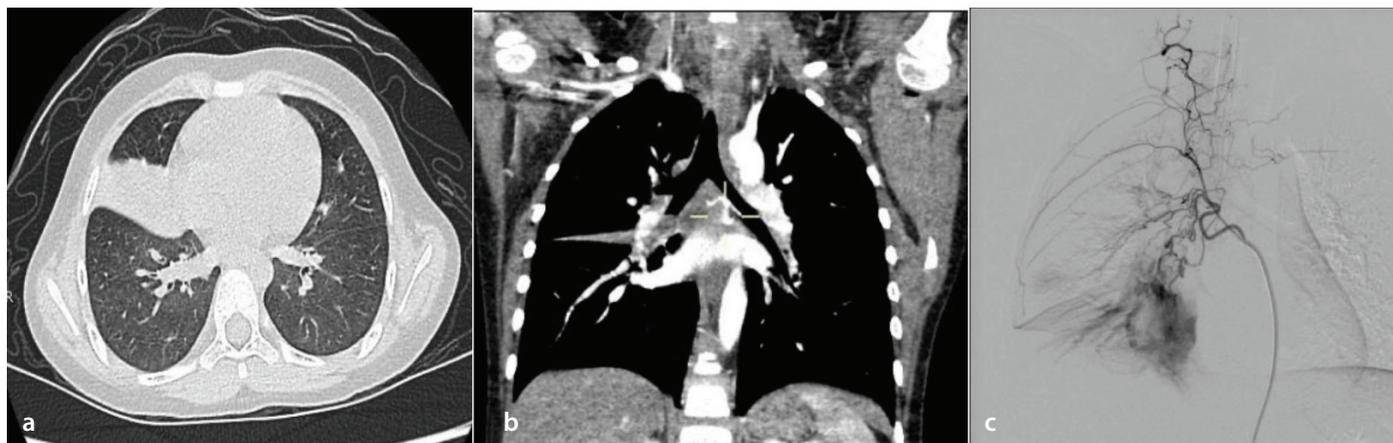


Figure 2. Bronchial artery (BA)–pulmonary artery shunt in a 5-year-old girl: (a) computed tomography (CT) scan indicated a patchy dense shadow in the middle lobe of the right lung; (b) enhanced CT multiplanar reformation indicated tortuous and thickened BAs; (c) digital subtraction angiography–bronchial arteriography indicated pulmonary artery expansion.

The right middle lobe bronchus diverges from the main bronchus at an acute angle, which also leads to obstruction of blood clearance.¹⁵ All of these factors cause blood accumulation in the middle lobe of the right lung, leading to BA-PAS. The cause of consolidation in the middle lobe of the right lung is the same as that of the syndrome in the middle lobe of the right lung. Hemorrhage caused by BA-PAS and blood retention in the middle lobe of the right lung together constitute consolidation in the middle lobe of the right lung on chest CT.

In the authors' experience, some adults with pulmonary hemorrhage and BA-PAS were also observed to have consolidation in the right middle lobe of the lung. Whether there are hemodynamic factors in the middle lobe of the right lung that lead to the formation of BA-PAS needs to be further investigated. The diagnosis of secondary BA-PAS without consolidation in the middle lobe of the right lung is often complex. Close follow-up is required, and the presence of BA-PAS should be suspected if symptoms of pulmonary hemorrhage are not effectively controlled following a period of symptomatic treatment.

In conclusion, BA-PAS can be classified into primary and secondary types. Chest CT images of pulmonary hemorrhage in children are varied and may even be negative. Primary BA-PAS typically manifests as GGO or consolidation in one side of the lung. Regardless of type, there is consolidation in the right middle lobe of the lung on chest CT in children with pulmonary hemorrhage. At this time, bronchial arteriography should be actively performed to confirm the diagnosis.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This research was funded by Beijing Hospitals Authority's Ascent Plan (DFL20221002).

References

- Martínez-Martínez MU, Oostdam DAH, Abud-Mendoza C. Diffuse alveolar hemorrhage in autoimmune diseases. *Curr Rheumatol Rep.* 2017;19(5):27. [\[CrossRef\]](#)
- Alexander GR. A retrospective review comparing the treatment outcomes of emergency lung resection for massive haemoptysis with and without preoperative bronchial artery embolization. *Eur J Cardiothorac Surg.* 2014;45(2):251-255. [\[CrossRef\]](#)
- Godfrey S. Pulmonary hemorrhage/hemoptysis in children. *Pediatr Pulmonol.* 2004;37(6):476-484. [\[CrossRef\]](#)
- Cauldwell EW, Siekert RG. The bronchial arteries; an anatomic study of 150 human cadavers. *Surg Gynecol Obstet.* 1948;86(4):395-412. [\[CrossRef\]](#)
- Walker CM, Rosado-de-Christenson ML, Martínez-Jiménez S, Kunin JR, Wible BC. Bronchial arteries: anatomy, function, hypertrophy, and anomalies. *Radiographics.* 2015;35(1):32-49. [\[CrossRef\]](#)
- Pang Y, Hu D, Dang Y, Huang S, Qin L, Li M. Bronchial artery-pulmonary artery shunt by *Apiotrichum mycotoxinivorans* infection in a recurrent hemoptysis case. *Infect Drug Resist.* 2022;15:4611-4615. [\[CrossRef\]](#)
- Ley S, Kreitner KF, Morgenstern I, Thelen M, Kauczor HU. Bronchopulmonary shunts

in patients with chronic thromboembolic pulmonary hypertension: evaluation with helical CT and MR imaging. *AJR Am J Roentgenol.* 2002;179(5):1209-1215. [\[CrossRef\]](#)

- Jacheć W, Tomasik A, Kurzyńska M, et al. The multiple systemic artery to pulmonary artery fistulas resulting in severe irreversible pulmonary arterial hypertension in patient with previous history of pneumothorax. *BMC Pulm Med.* 2019;19(1):80. [\[CrossRef\]](#)
- Panda A, Bhalla AS, Goyal A. Bronchial artery embolization in hemoptysis: a systematic review. *Diagn Interv Radiol.* 2017;23(4):307-317. [\[CrossRef\]](#)
- Zhao Q, Zhao L, He L, et al. Bronchial artery embolization in pediatric pulmonary hemorrhage: A single-center experience. *J Vasc Interv Radiol.* 2020;31(7):1103-1109. [\[CrossRef\]](#)
- Esparza-Hernández CN, Ramírez-González JM, Cuéllar-Lozano RA, et al. Morphological analysis of bronchial arteries and variants with computed tomography angiography. *Biomed Res Int.* 2017;2017:9785896. [\[CrossRef\]](#)
- Lara AR, Schwarz MI. Diffuse alveolar hemorrhage. *Chest.* 2010;137(5):1164-1171. [\[CrossRef\]](#)
- Mi W, Zhang C, Wang H, et al. Measurement and analysis of the tracheobronchial tree in Chinese population using computed tomography. *PLoS One.* 2015;10(4):e123177. [\[CrossRef\]](#)
- Lindskog GE, Spear HC. Middle-lobe syndrome. *N Engl J Med.* 1955;253(12):489-495. [\[CrossRef\]](#)
- Livingston GL, Holinger LD, Luck SR. Right middle lobe syndrome in children. *Int J Pediatr Otorhinolaryngol.* 1987;13(1):11-23. [\[CrossRef\]](#)