EISSN 1305-3612

DIR

Diagnostic and Interventional Radiology

TSR
1924
TURKISH SOCIETY
OF RADIOLOGY

dirjournal.org

VOLUME **31**
ISSUE **6**
November 2025

DIR
Diagnostic and Interventional
Radiology

Official Journal of the
Turkish Society of Radiology

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

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

Serap Gültekin, MD
Department of Radiology, Gazi University Faculty of Medicine, Ankara, Türkiye
ORCID ID: 0000-0001-6349-3998

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

Evrim 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, FCIRSE
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

Evrim 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 Aydıngöz, 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)

DIR
Diagnostic and Interventional
Radiology

Official Journal of the
Turkish Society of Radiology

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

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 (MEDLINE abbreviation: Tani Girisim Radyol), the current title's exact Turkish translation.

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

The editorial and publication processes of the journal are shaped in accordance with the guidelines of the ICMJE, WAME, CSE, COPE, EASE, and NISO.
Diagnostic and Interventional Radiology is indexed in **SCI-Expanded, Pubmed/Medline, Pubmed Central, TUBITAK ULAKBIM TR Index, Scopus, DOAJ, EBSCO, HINARI, EMBASE, CINAHL, ProQuest, Gale and CNKI..**

The journal is published online.

Owner: 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

galenos
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: November 2025

EISSN 1305-3612

International scientific journal published bimonthly.

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

TSR 1924
TURKISH SOCIETY OF RADIOLOGY

Diagnostic and Interventional Radiology

Volume 31 • Issue 6

November 2025

www.dirjournal.org

Contents

ABDOMINAL IMAGING

ARTIFICIAL INTELLIGENCE AND INFORMATICS

BREAST IMAGING

CARDIOVASCULAR IMAGING

INTERVENTIONAL RADIOLOGY

MUSKOLOSKELETAL IMAGING

NEURORADIOLOGY

PEDIATRIC RADIOLOGY

RADIOLOGY PHYSICS

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

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253218

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Detection of common bile duct dilatation on magnetic resonance cholangiopancreatography by deep learning

Hilal Er Ulubaba[1]
Rukiye Çiftçi[2]
İpek Atik[3]
Osman Furkan Karakuş[4]

[1]İnönü University Faculty of Medicine, Department of Radiology, Malatya, Türkiye

[2]Gaziantep İslam Science and Technology University Faculty of Medicine, Department of Anatomy, Gaziantep, Türkiye

[3]Gaziantep İslam Science and Technology University Faculty of Engineering and Natural Sciences, Department of Electrical Electronics Engineering, Gaziantep, Türkiye

[4]Yıldız Technical University, Department of Computer Engineering, İstanbul, Türkiye

PURPOSE
This study aims to detect common bile duct (CBD) dilatation using deep learning methods from artificial intelligence algorithms.

METHODS
To create a convolutional neural network (CNN) model, 77 magnetic resonance cholangiopancreatography (MRCP) images without CBD dilatation and 70 MRCP images with CBD dilatation were used. The system was developed using coronal maximum intensity projection reformatted 3D-MRCP images. The ResNet50, DenseNet121, and visual geometry group models were selected for training, and detailed training was performed on each model.

RESULTS
In the study, the DenseNet121 model showed the best performance, with a 97% accuracy rate. The ResNet50 model ranked second, with a 96% accuracy rate.

CONCLUSION
CBD dilatation was detected with high performance using the DenseNet CNN model. Once validated in multicenter studies with larger datasets, this method may help in diagnosis and treatment decision-making.

CLINICAL SIGNIFICANCE
Deep learning algorithms can aid clinicians and radiologists in the diagnostic process once technical, ethical, and financial limitations are addressed. Fast and accurate diagnosis is crucial for accelerating treatment, reducing complications, and shortening hospital stays.

KEYWORDS
Artificial intelligence, bile duct dilatation, choledocholithiasis, convolutional neural network, magnetic resonance cholangiopancreatography

The common bile duct (CBD) is approximately 7–10 cm long and 3–6 mm in diameter, and its diameter increases slightly with age. The CBD functions as a duct that stores bile produced in the liver and empties it into the small intestine. Bile, a digestive fluid, plays an important role in breaking down fats during the digestive process. CBD dilation refers to a condition in which the bile ducts dilate beyond the normal range. This condition can be caused by a variety of factors, including bile duct obstruction, the presence of gallstones, infections affecting the bile ducts, or other diseases that block the flow of bile, causing the bile ducts to dilate.[1] CBD dilatation is a common clinical symptom that can result from various conditions, including pancreatobiliary tumors, choledocholithiasis, and periampullary diverticula.[2,3] Non-pathological causes of bile duct dilatation, such as advanced age, previous surgical interventions, and chronic narcotic use, are also widely recognized.[4] Although patients with CBD dilatation may present with colicky pain, fever, jaundice, and other clinical

Corresponding author: Hilal Er Ulubaba

E-mail: erhilal44@yahoo.com

Received 08 January 2025; revision requested 24 January 2025; last revision received 25 February 2025; accepted 05 April 2025.

Epub: 05.05.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253218

You may cite this article as: Er Ulubaba H, Çiftçi R, Atik İ, Karakuş OF. Detection of common bile duct dilatation on magnetic resonance cholangiopancreatography by deep learning. *Diagn Interv Radiol.* 2025;31(6):532-538.

symptoms, a considerable proportion of patients remain asymptomatic. Nevertheless, the degree of dilation in the bile duct often serves as an essential indicator for assessing disease severity and guiding treatment selection.[5] Bile duct dilatation is a condition that can negatively affect the digestive system and can cause greater health problems if left untreated. Early diagnosis and appropriate treatment are important for preventing this problem and maintaining health.[6]

Diagnostic modalities for CBD dilatation include transabdominal ultrasonography (US), computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), and endoscopic retrograde cholangiopancreatography (ERCP). Among these, MRCP plays a well-established role in investigating biliary disorders and serves as a non-invasive alternative to ERCP.[7]

Treatment often becomes necessary in cases of CBD dilatation caused by stones. Early detection of CBD dilatation is crucial for enabling the timely initiation of treatment modalities. Artificial intelligence applications (AIAs) are increasingly utilized across various domains and are gradually being integrated into healthcare.[7] They facilitate disease diagnosis for physicians and shorten the time to treatment initiation. This can provide opportunities for early treatment and improve treatment success rates.[8]

In the healthcare industry, machine learning techniques are becoming increasingly common. As the term suggests, machine learning allows algorithms to learn and extract meaningful representations from data in a semi-automatic manner. Early diagnosis and treatment of biliary dilatation enable the detection of abnormalities in the biliary tract in a shorter time and shorten the treatment period.

Deep learning is one of the artificial learning approaches based on convolutional neural network (CNN) models, also known as multi-layer neural networks. These models are advanced feedforward neural networks widely used in image analysis, natural language processing, and other complex image classification problems. They are uniquely capable of identifying and interpreting patterns from images and text.[9,10]

Main points

- Common bile duct (CBD) dilation was detected with high performance from magnetic resonance cholangiopancreatography images.
- CBD dilation was also detected with high accuracy using the DenseNet deep learning algorithm.
- Artificial intelligence can assist clinicians and radiologists in the early and accurate diagnosis of bile duct dilation.

Deep learning is widely used in automatic image segmentation and also in medical image processing. Due to its superior performance, deep learning has emerged as the most popular technology.[11,12] Considering the importance of deep learning, this paper aims to analyze bile duct dilatation using ResNet50, DenseNet121, and visual geometry group (VGG) models.

This study aims to determine the diagnostic performance of deep learning algorithms in detecting CBD dilatation caused by bile duct stones using MRCP images.

Methods

This was a retrospective study, and ethical approval was obtained from Gaziantep Islam Science and Technology University in 2024 with decision number 2024/465. In this study, MRCP images of individuals aged 18–65 years were retrospectively analyzed. The images were obtained using a 1.5 Tesla magnetic resonance imaging system (Siemens Vision-Symphony Upgrade, Erlangen, Germany) and reviewed by a radiologist with approximately 13 years of clinical experience. The presence of stone-associated CBD dilatation was evaluated using coronal maximum intensity projection (MIP) reformatted 3D-MRCP images. A total of 147 MRCP images were included in this study, comprising 77 images from the normal group and 70 from patients with bile duct dilatation caused by choledocholithiasis. The dataset consisted of 147 MRCP images collected from [source]. The images were categorized based on the presence of stone-associated bile duct dilatation. Preprocessing steps included contrast enhancement, normalization, and resizing to 224 × 224 pixels. The exclusion criteria included individuals with cholecystectomy, images with artifacts in which the CBD was not visible, patients with tumors in the bile ducts, and those without a confirmed diagnosis of choledocholithiasis through clinical evaluation or ERCP.

MRCP artifacts, technical and reconstruction-related artifacts, gas-related artifacts, and other fixed fluids (such as those in the duodenum or ascites) may result from the overlap of the CBD. Poor spatial resolution may also limit interpretation due to artifacts. Examination of thin sections and multiple planes reduces these issues. However, coronal MIP reformatted 3D-MRCP images were used. Therefore, if the CBD was not visible in coronal MIP reformatted 3D-MRCP images, those images were excluded.

Patients with detected stones and dilatation on MRCP along with a confirmed diagnosis of stones via ERCP were assigned to the CBD dilatation group. Patients with normal bile ducts on MRCP who clinically improved without requiring ERCP or were found to have no stones on ERCP were assigned to the normal group.

To enhance generalization, five-fold cross-validation was employed. In this approach, the dataset was divided into five subsets, with four subsets used for training and one for validation in each iteration. This ensured that the model learned from different portions of the data, enhancing robustness. To mitigate overfitting, dropout layers (rate: 0.3) were applied, and L2 weight regularization was used. Additionally, early stopping was employed to prevent excessive training on the same patterns.

Study design

The images, stored in Digital Imaging and Communications in Medicine format, were transferred to a personal workstation using the Horos Medical Image Viewer (Version 3.0; Horos Project, Annapolis, MD, USA). The images were analyzed using a deep learning approach, a subset of AI. Initially, all images underwent preprocessing steps such as cropping, padding, smoothing, and resizing to fit the input layer of the deep learning architecture. The processed images were then fed into the input layer of the architecture. In the subsequent step, features were extracted from the fully connected layers of the architecture. These features were saved in a file and used as training and test data for classifier methods.[13]

Pre-trained CNN models were employed as feature extractors. The features were obtained from the fully connected layers of each model, with a total of 1,000 features per model. Training and testing processes were conducted using all features, as well as subsets selected through feature selection, to analyze the impact of feature selection on system performance. Various numbers of features were tested to assess their effect on system success.

CNNs operate on specific-sized patches of images, processing them piece by piece through comparison. Filters are spatially small along the width and height but extend through the entire depth of the input image. They are designed to detect specific feature types in the input image. In the convolutional layer, the filter/kernel is moved to every possible position over the input matrix. Element-wise multiplication is performed between the filter and the patch of the input image, followed by summation. This process is repeated for every possible position of the filter on the input image matrix, enabling feature detection at any location in the image.[14]

Deep learning algorithms

This study employed a deep learning-based process to classify images into normal and choledocholithiasis-associated CBD dilatation groups. In the initial stage, raw datasets were prepared, consisting of normal and CBD dilatation images. During the data preprocessing phase, the images were resized, and data augmentation techniques were applied. The ResNet50, DenseNet121, and VGG models were selected for the training process, with detailed training conducted on each model. The models' performance was evaluated using metrics such as accuracy, F1 score, and the confusion matrix. Finally, the results were compared, and the model with the best performance was identified. The process flowchart is presented in Figure 1.

Dataset

This study utilized a dataset comprising a total of 147 images categorized into two classes: normal and stone-associated CBD dilatation. Of these, 77 images belonged to the normal group, and 70 images were classified as CBD dilatation. The images were subjected to various preprocessing techniques to prepare them for deep learning models. During preprocessing, all images were resized to 224 × 224 pixels, and pixel values were normalized between 0 and 1. The dataset was split into two subsets: 80% for training and 20% for testing. Additionally, to enhance model generalization and mitigate overfitting, five-fold cross-validation was applied, ensuring that different subsets of the data were used for training and validation in each iteration.

Data augmentation

To increase the diversity of the dataset and prevent model overfitting during training, data augmentation techniques were applied. Data augmentation techniques included rotation (–20° to +20°), horizontal and vertical flipping, zoom (0–10%), brightness adjustment, and contrast normalization. These transformations improved model robustness. The following data augmentation methods were used:

1. **Horizontal flipping:** Random flipping of images along the horizontal axis.

2. **Rotation:** Random rotation of images between 0° and 30°.

3. **Brightness adjustment:** Random modification of image brightness levels.

4. **Random cropping:** Cropping specific portions of the images.

These data augmentation techniques were dynamically applied during each epoch to enhance variability during training. By applying augmentation, the total number of images increased to 735.

Image examples with data augmentation

Figure 2 presents examples of images processed with data augmentation techniques.

1. **Original image:** This represents the raw image with no data augmentation applied.

2. **Horizontal flipping:** The image is flipped along the horizontal axis, swapping the left and right sides.

3. **Rotation:** The image is rotated randomly between 0° and 30°.

4. **Brightness adjustment:** The brightness level of the image is either increased or decreased.

5. **Random cropping:** A random portion of the image is cropped and then resized to the original dimensions.

Models

In this study, we selected three widely used CNN architectures—ResNet50, DenseNet121, and VGG16—to evaluate their effectiveness in detecting stone-associated CBD dilatation. Each model was chosen based on its specific strengths in feature extraction, classification accuracy, and computational efficiency.

- **ResNet50:** This model was selected for its residual learning framework, which effectively addresses the vanishing gradient problem in deep networks. The use of residual connections allows for deeper architectures while maintaining efficient feature propagation, making it well-suited for complex medical image classification.

- **DenseNet121:** This model was incorporated due to its dense connectivity mechanism, which enables improved gradient flow and feature reuse, leading to enhanced model efficiency. Compared with ResNet50, DenseNet121 requires fewer parameters while maintaining high classification accuracy, making it advantageous for medical imaging tasks where computational efficiency is crucial.

- **VGG16:** VGG16 was included as a baseline model, as it has been extensively used in medical image analysis studies. Its simple yet effective architecture allows for a controlled comparison against more advanced CNNs while providing insights into the relative advantages of deep feature extraction techniques.

By selecting these three architectures, this study provided a comprehensive evaluation of different deep learning models, ensuring a balanced comparison of accuracy, computational cost, and real-world applicability for CBD dilatation detection.

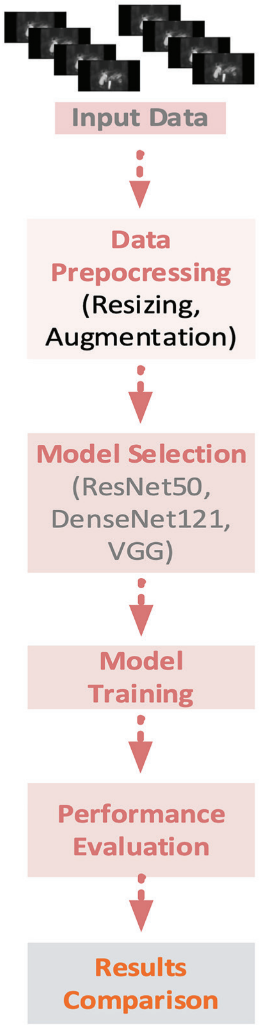


Figure 1. Flux diagram. VGG, visual geometry group.

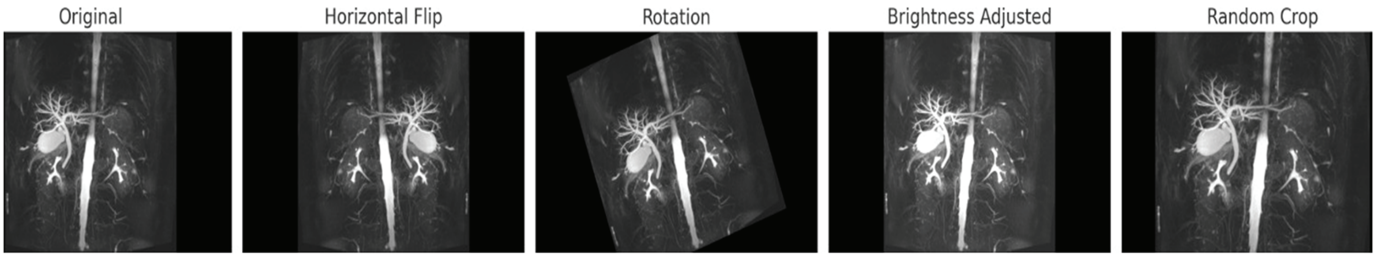


Figure 2. Augmented MRCP images.

Training process

The models were trained for a total of 100 epochs, with training and validation losses monitored during each epoch. The Adam optimization algorithm was used to update model weights. The cross-entropy loss function, suitable for multi-class classification problems, was employed for loss calculation. The learning rate was set at 0.001, and the mini-batch size was chosen as 32. To ensure optimal model performance, key hyperparameters were carefully selected and fine-tuned during the training phase. The models were trained using the Adam optimizer with a learning rate of 0.0001, where beta1: 0.9 and beta2: 0.999. The batch size was set to 16, balancing computational efficiency with stable convergence. The models were trained for 50 epochs, applying categorical cross-entropy as the loss function due to its effectiveness in multi-class classification problems.

Additionally, learning rate decay was incorporated using a step decay strategy, gradually reducing the learning rate to prevent the model from getting stuck in local minima and to enhance convergence stability. These hyperparameter choices were made based on preliminary experiments to optimize model accuracy while preventing overfitting. The detailed tuning of these parameters contributed to the improved generalizability of the deep learning models for detecting CBD dilatation.

Performance evaluation

The models' performance was evaluated using metrics commonly employed in classification problems, such as precision, recall, and F1 score. These metrics analyze the prediction success of the model from different perspectives:

- **Precision:** This is the ratio of true positive (TP) predictions to the total positive predictions. It measures the impact of false positive (FP) predictions.

- **Recall:** This is the ratio of correctly classified positive samples to the total actual positive samples. It measures the impact of false negative (FN) predictions.

- **F1 Score:** This metric aims to balance precision and recall, representing their harmonic mean. It is particularly useful when there is an imbalance in data distribution between classes.

The formulas for these metrics are provided in Equations 1–5, and the evaluation results are presented in Table 1.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

In Formulas 1–5, the following abbreviations are used:

- **TP**: The number of correctly predicted positive cases.

- **FP**: The number of cases incorrectly predicted as positive.

- **FN**: The number of cases incorrectly predicted as negative.

- **TN (true negatives)**: The number of correctly predicted negative cases.

These formulas are derived from standard classification metric definitions and are widely used in deep learning-based medical image classification.[14]

Results

To address learning challenges in deep networks, this model was evaluated using the 50-layer ResNet50 architecture, which employs residual connections. The ResNet model achieved high performance with a minimum loss value of 0.0926, precision of 0.9642, recall of 0.9705, and F1 score of 0.9663.

The VGG16 architecture, characterized by a fixed structure of 3 × 3 filters, was also evaluated. Its minimum loss value was measured at 0.6752. However, it demonstrated lower performance than the other models, with a precision of 0.2166, recall of 0.5102, and F1 score of 0.3023.

The DenseNet121 model, which connects each layer to all preceding layers to enhance learning efficiency, achieved a minimum loss value of 0.1384. Its performance metrics were a precision of 0.9722, recall of 0.9615, and F1 score of 0.9657.

The loss values throughout the training process are presented in Figure 3, and the changes in precision, recall, and F1 score are illustrated in Figure 4.

The analysis revealed that the deep learning-based ResNet, VGG, and DenseNet models demonstrated high performance in classifying normal and dilated images, with the ResNet and DenseNet models showing low loss values and high precision, recall, and F1 scores. Specifically, DenseNet emerged as an effective model in terms of learning efficiency. In contrast, the VGG model performed relatively poorly compared with the others. The data augmentation techniques and training processes used in this study supported the models' general feature learning and enhanced classification success.

Discussion

In this study, which aimed to determine the diagnostic performance of CBD dilation in patients with choledocholithiasis from MRCP images using the CNN method, the highest accuracy was achieved with the DenseNet algorithm at 97%. MRCP can effectively show changes in the bile ducts; however, MRCP images can have many artifacts and low resolution, which may complicate diagnosis. In contrast, deep learning methods

Table 1. Performance evaluation metrics

	Model	Accuracy	Precision	Recall	Specificity	F1-score
ResNet	ResNet50	0.9682	0.9642	0.9705	0.9661	0.9663
VGG	VGG16	0.3923	0.2166	0.5102	0.3513	0.3023
DenseNet	DenseNet121	0.9701	0.9722	0.9615	0.9772	0.9657

VGG, visual geometry group.

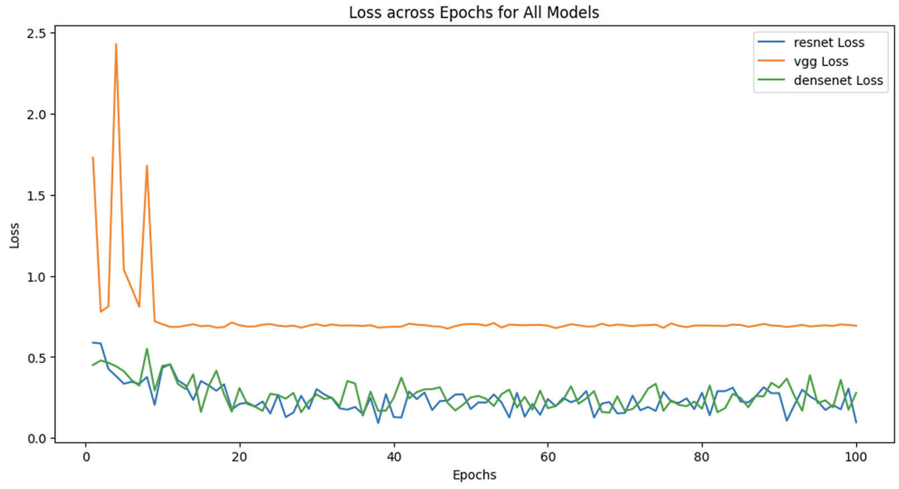


Figure 3. Loss changes during the training process of models.

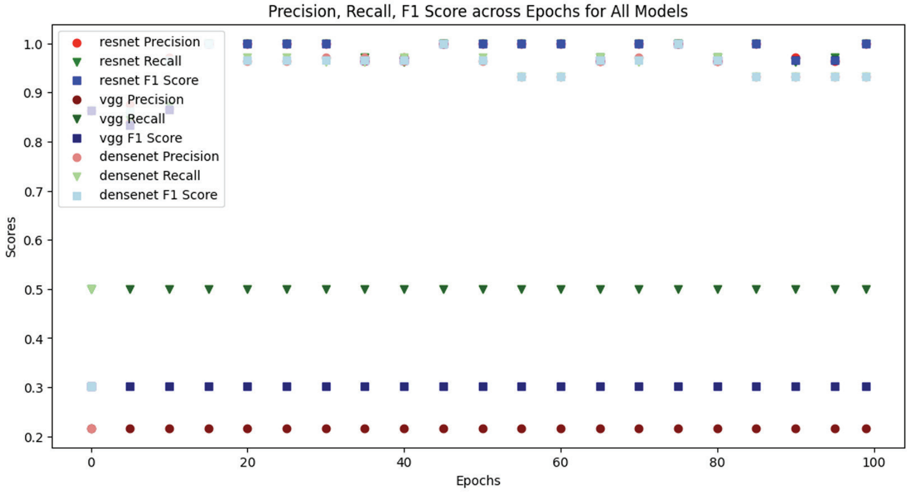


Figure 4. Performance evaluation metrics of the models.

can assist clinicians and radiologists quickly and independently in the diagnostic process. Our study also included images with artifacts and low resolution in which the CBD could be seen. These images can help the clinician determine the indication for ERCP treatment. Early detection of stone-related CBD dilation is crucial in accelerating treatment, preventing complications related to stones, and reducing hospitalization and intensive care unit stay times. This also reduces complications that may arise from prolonged hospital stays.

To better contextualize the effectiveness of deep learning models in detecting stone-associated CBD dilatation, we conducted a comparative analysis with existing non-deep learning diagnostic techniques. Traditional methods, such as radiologist-based manual assessment, have long been the gold standard in clinical practice due to their high accuracy and interpretability. However, these assessments are highly dependent on the radiologist's expertise and experience, making them prone to inter-observer variability and subjective interpretation.

Another conventional approach involves feature engineering-based machine learning models, such as support vector machines and random forest classifiers, which extract handcrafted features from MRCP images. Although these methods can improve classification accuracy, their reliance on manually selected features limits their ability to generalize across different datasets and imaging conditions. Additionally, conventional image processing techniques, such as thresholding and edge detection, have been explored for bile duct segmentation and anomaly detection but often struggle with complex variations in anatomy and image artifacts.

In contrast, our proposed deep learning approach leverages CNNs to automatically extract high-level features, eliminating the need for manual feature selection and minimizing human bias. Unlike traditional methods, CNN-based models can learn hierarchical representations of CBD dilatation patterns directly from raw images, leading to superior classification accuracy and robustness. Furthermore, deep learning models have the potential for real-time application, enabling automated and consistent diagnoses without requiring extensive human intervention. Although CNN models may lack the inherent interpretability of traditional techniques, methods such as gradient-weighted class activation mapping (Grad-CAM) visualization can enhance explainability by highlighting the regions of interest in MRCP images.

Overall, our comparative analysis highlights the advantages of deep learning models in terms of automation, scalability, and reproducibility while recognizing the strengths of traditional diagnostic approaches in interpretability and clinical trustworthiness. Future research could explore hybrid approaches that combine deep learning with traditional radiology methods to further enhance diagnostic performance. As AI rapidly advances and integrates into the healthcare field, this high-performing method can help reduce the workload of radiologists and improve diagnostic accuracy in collaboration with clinicians. Despite the high diagnostic value of the current model, a small margin of error remains. This is because the deep learning model evaluates based on a single, predetermined image, unlike radiologists.[9,15]

There are several studies in the literature that use AIAs to analyze bile duct pathologies from MRCP images. Ringe et al.[16] detected primary sclerosing cholangitis with 95% sensitivity using machine learning algorithms on coronal MIP 3D-MRCP images. Hou et al.[17] detected choledocholithiasis with 95% accuracy from thick-section 2D MRCP images. Sun et al.[18] also detected choledocholithiasis with 93.48% accuracy using 3D-MRCP images. In our study, a CNN model was developed for the diagnosis of stone-related CBD dilation using 3D-MRCP images, achieving high precision, recall, and F1 score. Among the three models used, DenseNet demonstrated the highest performance, with precision, recall, and F1 score values of 0.9722, 0.9615, and 0.9657, respectively.

There are some limitations in our study. The first is that it only includes patients with stone-related CBD dilation and a normal group. Including patients with CBD dilation due to other pathologies, such as tumors and strictures, would have made the detection of stone-related dilation more valuable. Another limitation is that it is a single-center study and includes a small number of patients. Additionally, we did not differentiate CBD dilatation according to severity, so we did not specifically determine the accuracy rate for patients with a borderline CBD diameter. Therefore, we do not know if our method will perform well on MRCP images obtained from different centers with various devices and parameters. These limitations should be addressed, and larger studies should be conducted with this method.

Although we think that deep learning will be useful in diagnosis, there are many technical, ethical, and financial challenges to overcome. For example, incorrect predictions can be made due to image artifacts and noise. Deep learning practitioners need to manually correct image noise to create an effective dataset. This process is both time-consuming and costly. Another challenge is the requirement for large datasets to build an effective deep learning system. This is both costly and may raise ethical concerns. Additionally, it may not be possible to create large datasets for infrequent diseases.[19]

In conclusion, a high-performance method for diagnosing CBD dilatation in patients with choledocholithiasis using the DenseNet CNN model from MRCP images has been described. However, the number of patients in our study was small, and only patients with choledocholithiasis were included. We hope to validate this method with a larger multicenter dataset in future studies. Additionally, future studies could include images of other pathologies, such as tumors, stenosis, and inflammation, that cause biliary tract dilatation to examine whether differential diagnosis can be made in addition to dilatation detection using this method. After addressing these deficiencies, the DenseNet CNN model may become a valuable tool for detecting CBD dilatation and determining its pathogenesis in the future.

To evaluate the impact of artifacts on model performance, we analyzed images with varying levels of noise and low resolution. The results indicate that although the model performed well on high-quality images, accuracy was slightly reduced in images with major artifacts. This finding highlights the need for artifact-specific preprocessing methods.

To further evaluate the reliability of our model, we calculated confidence intervals for the reported performance metrics and conducted statistical analyses using the bootstrap method. Additionally, five-fold cross-validation was applied to assess the model's generalizability across different data partitions. These analyses confirmed the robustness and consistency of our model, strengthening its potential for clinical applications.

Performance comparisons among ResNet50, DenseNet121, and VGG16 revealed that DenseNet121 exhibited superior sensitivity and classification accuracy. DenseNet121 consistently outperformed the other models, particularly in sensitivity, indicating a higher ability to correctly identify stone-associated CBD dilatation.

This study highlights the potential of deep learning in detecting stone-associated CBD dilatation but has limitations. The single-center dataset limits generalizability, requiring multicenter validation. Future research should include other etiologies, such as tumors and strictures, for broader clinical use. Image artifacts and MRCP variations remain challenges, necessitating improved preprocessing. Additionally, explainability methods such as Grad-CAM and multi-modal imaging integration (CT, US) could enhance diagnostic accuracy and clinical applicability.

Ethical considerations

The integration of deep learning models into clinical decision-making raises important ethical considerations, particularly regarding bias, human–AI collaboration, and regulatory challenges. One of the primary concerns in AI-driven diagnostics is algorithmic bias, which may arise if the training dataset is not sufficiently diverse or representative of different patient populations. A biased dataset can lead to disparities in diagnostic accuracy, particularly among underrepresented demographic groups. Therefore, future studies should prioritize multicenter and demographically diverse datasets to ensure fair and unbiased AI decision-making.

As for human–AI collaboration, although deep learning models demonstrate high accuracy in detecting CBD dilatation, they should be regarded as assistive tools rather than replacements for radiologists. The final clinical decision should always involve expert verification, ensuring that AI-generated diagnoses are contextualized within the broader clinical picture. Methods such as Grad-CAM visualization and other explainability techniques can enhance trust in AI models by providing interpretable insights into the decision-making process.

Additionally, the deployment of AI in medical imaging is subject to regulatory and transparency challenges. Ensuring compliance with healthcare standards and AI governance frameworks is essential for widespread clinical adoption. Future research should focus on developing transparent, explainable AI models that align with ethical and legal standards, thereby increasing both clinician and patient trust in automated diagnostic systems. By addressing these ethical concerns, we aim to contribute to the responsible development and implementation of AI-driven radiology applications, ensuring that deep learning models are used in a manner that prioritizes fairness, safety, and clinical efficacy.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Toy Ş, Şenol D, Çiftçi R, Sevgi S, Seçgin Y, Yıldırım İO. Morphometric examination of the hepatobiliary duct system in healthy individuals and patients with cholelithiasis: a radio-anatomic magnetic resonance cholangiopancreatography study. *Cukurova Med J*. 2023;48(2):706-714. [Crossref]
2. Ham JH, Yu JS, Choi JM, Cho ES, Kim JH, Chung JJ. Peri-ampullary duodenal diverticulum: effect on extrahepatic bile duct dilatation after cholecystectomy. *Clin Radiol*. 2019;74(9):735. [Crossref]
3. Holm AN, Gerke H. What should be done with a dilated bile duct? *Curr Gastroenterol Rep*. 2010;12(2):150-156. [Crossref]
4. Benjaminov F, Leichtman G, Naftali T, Half EE, Konikoff FM. Effects of age and cholecystectomy on common bile duct diameter as measured by endoscopic ultrasonography. *Surg Endosc*. 2013;27(1):303-307. [Crossref]
5. Radmard AR, Khorasanizadeh F, Poustchi H, et al. Prevalence and clinical outcomes of

common bile duct dilation in patients who use opium. *Am J Med Sci*. 2018;356(1):39-46. [Crossref]

6. Abi Nader C, Vetil R, Wood LK, et al. Automatic detection of pancreatic lesions and main pancreatic duct dilatation on portal venous CT scans using deep learning. *Invest Radiol*. 2023;58(11):791-798. [Crossref]
7. Sugimoto Y, Kurita Y, Kuwahara T, et al. Diagnosing malignant distal bile duct obstruction using artificial intelligence based on clinical biomarkers. *Sci Rep*. 2023;13(1):3262. [Crossref]
8. Diaz-Martinez J, Pérez-Correa N. Postcholecystectomy duodenal ınjuries, their management, and review of the literature. *Euroasian J Hepatogastroenterol*. 2024;14(1):44-50. [Crossref]
9. Çiftçi R, Dönmez E, Kurtoğlu A, Eken Ö, Samee NA, Alkanhel RI. Human gender estimation from CT images of skull using deep feature selection and feature fusion. *Sci Rep*. 2024;14(1):16879. [Crossref]
10. Luna LH. Test of the minimum supero-inferior femoral neck diameter as a sex predictor in a contemporary documented osteological collection from Portugal. *Forensic Sci Res*. 2024:owae045. [Crossref]
11. Shoaib MA, Lai KW, Chuah JH, et al. Comparative studies of deep learning segmentation models for left ventricle segmentation. *Front Public Health*. 2022;10:981019. [Crossref]
12. Wu X, Zhang YT, Lai KW, Yang MZ, Yang GL, Wang HH. A novel centralized federated deep fuzzy neural network with multi-objectives neural architecture search for epistatic detection. *IEEE Transactions on Fuzzy Systems*. 2025;33(1):94-107. [Crossref]
13. Atik I. Classification of electronic components based on convolutional neural network Architecture. *Energies*. 2022;15(7):2347. [Crossref]
14. Atik I. COVID-19 case forecast with deep learning BiLSTM approach: the Turkey case. Int. Int J Mech Eng. 2022;7(1):6307-6314. [Crossref]
15. Kurtoğlu A, Eken Ö, Çiftçi R, et al. The role of morphometric characteristics in predicting 20-meter sprint performance through machine learning. *Sci Rep*. 2024;14(1):16593. [Crossref]
16. Ringe KI, Vo Chieu VD, Wacker F, et al. Fully automated detection of primary sclerosing cholangitis (PSC)-compatible bile duct changes based on 3D magnetic resonance cholangiopancreatography using machine learning. *Eur Radiol*. 2021;31(4):2482-2489. [Crossref]
17. Hou JU, Park SW, Park SM, Park DH, Park CH, Min S. Efficacy of an artificial neural network algorithm based on thick-slab magnetic resonance cholangiopancreatography images for the automated diagnosis of common bile duct stones. *J Gastroenterol Hepatol*. 2021;36(12):3532-3540. [Crossref]
18. Sun K, Li M, Shi Y, et al. Convolutional neural network for identifying common bile duct stones based on magnetic resonance cholangiopancreatography. *Clin Radiol*. 2024;79(7):553-558. [Crossref]
19. Dhar T, Dey N, Borra S, Sherratt RS. Challenges of deep learning in medical image analysis-improving explainability and trust. *IEEE Transactions on Technology and Society*. 2023;4(1):68-75. [Crossref]

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253283

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Readout-segmented echo-planar imaging and conventional single-shot echo-planar imaging for determining cervical cancer image quality, lymphovascular space invasion, and lymph node metastasis status: a comparative study

Huizhen Song[1,2]
Jiao Bai[2]
Yu Wang[2]
Juan Xie[2]
Yunzhu Wu[3]
Jian Shu[2,4]

[1]Chengdu Seventh People's Hospital (Affiliated Cancer Hospital of Chengdu Medical College), Department of Radiology, Chengdu, China

[2]The Affiliated Hospital of Southwest Medical University, Department of Radiology, Luzhou, China

[3]Nanjing University of Information Science and Technology, School of Artificial Intelligence, AI in Medicine, Nanjing, China

[4]Medical Imaging Key Laboratory of Sichuan Province, North Sichuan Medical College, Nanchong, China

Corresponding author: Jian Shu

E-mail: shujiannc@163.com

Received 21 February 2025; revision requested 05 April 2025; accepted 05 May 2025.

Epub: 23.06.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253283

PURPOSE

Diffusion-weighted imaging (DWI) using single-shot echo-planar imaging (ss-EPI) is prone to artifacts, geometric distortion, and T2* blurring. Readout-segmented echo-planar imaging (rs-EPI) may improve image quality in the DWI of cervical cancer (CC). This study aimed to compare the image quality between rs-EPI and ss-EPI DWI in CC and to evaluate whether the apparent diffusion coefficient (ADC) values of ss-EPI (ssADC) and rs-EPI (rsADC) can differentiate the status of lymphovascular space invasion (LVSI) and lymph node metastasis (LNM).

METHODS

This prospective study included 69 patients with CC who underwent ss-EPI and rs-EPI DWI before surgery. Qualitative reader scores, signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and ADC values derived from ss-EPI and rs-EPI were compared. The differences in ADC values were analyzed in patients who were (a) LNM-positive (LNM+, n = 17) and LNM-negative (LNM–, n = 52); (b) LVSI-positive (LVSI+, n = 33) and LVSI-negative (LVSI–, n = 36).

RESULTS

The rs-EPIs of CC had higher subjective image quality scores and a lower SNR than ss-EPI (all $P < 0.001$); no significant differences existed between rs-EPI and ss-EPI for either CNR or ADC (CNR, $P = 0.313$; ADC, $P = 0.949$; P > 0.05 for all). The rsADC and ssADC of the LNM+ group were substantially lower than those of the LNM– group (rsADC, $P = 0.000$; ssADC, $P = 0.000$; $P < 0.001$ for all); the areas under the receiver operating characteristic curve were 0.855 and 0.851, respectively. However, there were no differences in ADC values between the LVSI+ and LVSI– groups (rsADC, $P = 0.271$; ssADC, $P = 0.200$; $P > 0.05$ for all).

CONCLUSION

Over a similar scan time, rs-EPI improves the qualitative image quality of DWI significantly more than ss-EPI and has good diagnostic accuracy for LNM status in CC. However, neither could predict the LVSI status.

CLINICAL SIGNIFICANCE

Readout-segmented EPI improves the qualitative image quality of DWI and has good diagnostic accuracy for LNM status in CC, compared with conventional ss-EPI. It is more inclined to qualitative analysis of CC foci and provides a better scheme when choosing the DWI sequence scanning strategy for CC.

KEYWORDS

Echo-planar imaging, cervical cancer, signal-to-noise ratio, diffusion-weighted imaging, lymphovascular space invasion, lymph node metastasis

You may cite this article as: Song H, Bai J, Wang Y, Xie J, Wu Y, Shu J. Readout-segmented echo-planar imaging and conventional single-shot echo-planar imaging for determining cervical cancer image quality, lymphovascular space invasion, and lymph node metastasis status: a comparative study. *Diagn Interv Radiol.* 2025;31(6):539-546.

Despite its extensive screening programs, cervical cancer (CC) remains the fourth most common cancer among women worldwide.[1] Lymphovascular space invasion (LVSI) refers to the presence of tumor cells in the lymphatic vessels and/or within the blood vessels. Although not involved in CC staging, it can predict lymph node metastasis (LNM) and is used as an independent prognostic factor for recurrence and survival.[2,3] The latest International Federation of Gynecology and Obstetrics 2018 staging system noted that the treatment efficacy of stage IB2 and stage IIA1 early CC surgery and chemoradiation is consistent, and the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology recommends that patients with LNM classified as stage IIIC cancer, who are candidates for external beam radiation therapy, undergo brachytherapy and chemotherapy instead of surgical procedures.[3,4] If the risk factors for LNM and LVSI are identified after surgery, adjuvant chemoradiotherapy needs to be further strengthened. However, combined surgery and chemoradiotherapy is associated with increased adverse effects and greater complications.[5] Accurate evaluation of the LVSI and LNM before surgery can reduce the complications of both surgery and chemoradiation, avoid excessive surgery, and prevent depriving young women of childbirth.

Around the edges of the tumor focus are functional lymphatic channels, along which the primary tumor spreads to the regional lymph nodes.[6] The peripheral tumor tissue between the tumor and normal tissues is rich in tumor stem cells and likely to determine tumor invasiveness.[7] However, the anterior bladder and posterior gas-containing rectum are challenging anatomical regions, making cervical diffusion-weighted imaging (DWI) at 3.0T magnetic resonance (MR) susceptible to artifacts and geometric distortion, especially in these tumor marginal areas. The conventional sequence used for DWI–single-shot echo-planar imaging (ss-EPI)–has various limitations that may induce geometric distortion, T2 blurring effect, and susceptibility to artifacts.[8] Through shortening the echo time (TE) and echo spacing along the k-space readout trajectory, readout-segmented echo-planar imaging (rs-EPI) DWI sequences may overcome these shortcomings.[9,10]

Main points

- The anterior bladder and posterior gas-containing rectum are challenging anatomical regions, making cervical diffusion-weighted imaging (DWI) at 3.0T magnetic resonance susceptible to artifacts, geometric distortion, and T2* blurring, especially in cervical cancer (CC) marginal areas.
- Readout-segmented echo-planar imaging (rs-EPI) DWI can produce clear anatomic details and reduce ghost artifacts and distortion in DWI of CC, achieving higher subjective and qualitative diagnostic value.
- No difference was found between rs-EPI and single-shot echo-planar imaging (ss-EPI) with respect to apparent diffusion coefficient values pertaining to the edges of CC lesions.
- rs-EPI and ss-EPI had similar good diagnostic accuracy in the prediction of lymph node metastasis status in CC.

This study compares the subjective and objective image quality of rs-EPI and ss-EPI in CC and determines the feasibility and effectiveness of the marginal area apparent diffusion coefficient (ADC) values from ss-EPI and/or rs-EPI for evaluating the invasive status of the tumor into the lymphovascular space and lymph nodes in patients with CC.

Methods

Patients

This study was approved by the Institutional Ethics Committee of The Affiliated Hospital of Southwest Medical University, and informed consent was obtained from all the patients (May 2021; decision number of the ethics committee approval: KY2021-04101). A total of 130 consecutive female patients with pathologically diagnosed CC underwent an MR examination and subsequent surgery as primary surgery at the hospital between June 2021 and March 2023. The MR imaging (MRI) examination included ss-EPI DWI and rs-EPI DWI. Of the patients initially included in the study, 46 were excluded due to the following criteria: 1) history of preoperative therapy (neoadjuvant chemotherapy, radiotherapy, or conization) (n = 36); 2) incomplete histopathological data (n = 7); 3) the diameter of each lesion was <1 cm on the MRI (n = 15); 4) rare histological types of CC (big cell neuroendocrine carcinoma, clear cell carcinoma) (n = 3). The remaining 69 cases were finally divided into an LVSI+ versus LVSI- patients' group and an LNM+ versus LNM- patients' group, respectively. The inclusion and exclusion criteria are summarized in Figure 1.

Magnetic resonance imaging protocols

The MR examinations were performed on a 3T MR scanner (MAGNETOM Prisma, Siemens Healthineers, Forchheim, Germany). Conventional MRI was acquired, including the following sequences: axial and sagittal T2W imaging [repetition time (TR)/TE, 4,000/89 ms]; axial contrast-enhanced T1W imaging with fat saturation (TR/TE, 600/10 ms). Before injecting the contrast agent, ss-EPI and rs-EPI DWI were performed with comparable imaging parameters (Table 1). Patients who were menstruating or had an intrauterine device in the uterine cavity at the time of MRI acquisition were excluded.

Qualitative evaluation of image quality

All database analyses were performed on a workstation (Syngo.via; Siemens Healthineers). Qualitative image evaluation was performed by two independent radiologists (reader 1: BZ, with 8 years of experience;

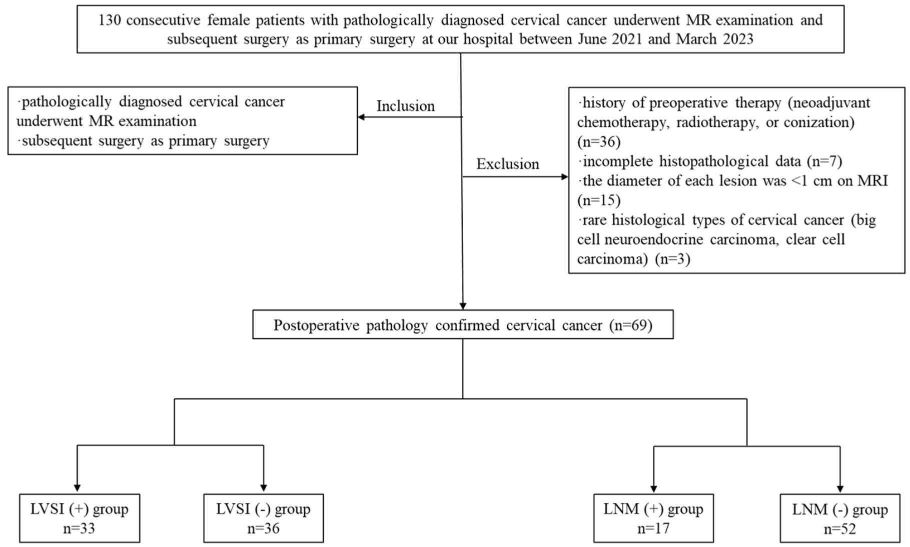


Figure 1. Flowchart showing the inclusion and exclusion criteria for patient selection. LVSI, lymphovascular space invasion; LNM, lymph node metastasis; MRI, magnetic resonance imaging.

reader 2: JC, with 15 years of experience) who were blinded to the corresponding DWI scanning sequence parameters. The two observers anonymized and randomly distributed all DWI parameters. Taking axial and sagittal T2W images and axial contrast-enhanced T1W images as a reference, the image anatomical details, geometric distortion and ghosting artifacts, lesion conspicuity, and overall image quality were assessed using a 5-point scale (1 = non-diagnostic, 2 = poor, 3 = general, 4 = good, 5 = excellent) according to the research by Zhang et al.[11]

Quantitative evaluation of images

The quantitative analysis included the measurement of the signal-to-noise ratio (SNR), CNR, and ADC values. All regions of interest (ROIs) were manually delineated by a radiologist (FW) who had 20 years of experience involving pelvic MRI. Lesion ROIs were placed on axial oblique DWIs with a b-value of 800 s/mm^2 guided by T2W images and contrast-enhanced T1W images. An irregular lesion ROI included the tumor edge on the biggest image slice, which encompassed the peripheral area within a 5-10-mm radius of the edge of high-intensity tumors and avoided obvious necrotic, higher signal artifacts and deformation areas. Background and tissue ROIs were selected as a circular drawing with an area of approximately 5 mm^2. All ROIs were measured twice, and the average was calculated. In the results, signal $(S)_{lesion}$ is the mean signal intensity inside the lesion ROI, standard deviation $(SD)_{background}$ is the standard deviation of the background noise ROI, S_{tissue} depicts the mean signal intensity of the musculus glutaeus medius tissue ROI, and SD_{lesion} and SD_{tissue} represent the standard deviation of the lesion ROI and musculus glutaeus medius tissue ROI, respectively.

Next, a well-matched copy of the cervical lesion ROI was generated automatically at the corresponding location of each ADC image. An example of a cervical lesion ROI acquired on rs-EPI is shown in Figure 2 (b = 800 s/mm^2).

Statistical analysis

Statistical analysis was performed using SPSS, version 25.0 and MedCalc, version 15.2.2. A two-tailed *P* value of <0.05 was considered statistically significant for all analyses. Regarding subjective image quality, intra-observer agreement was assessed using kappa statistics. The kappa values were in the range of 0–1.00 and interpreted as follows: 0.40 = poor; 0.41–0.60 = moderate; 0.61–0.80 = good; >0.81 = excellent.[12] The qualitative parameters (reader score) between the two DWI protocols were compared using the Wilcoxon signed-rank test. The Shapiro–Wilk test was conducted to assess the normal distribution of all continuous variables. The quantitative parameters (SNR, CNR, ADC) were compared between ss-EPI and rs-EPI using the paired Student's t-test (when conforming to a normal distribution) or the Wilcoxon signed rank test. The consistency of ADC values between ss-EPI and rs-EPI was estimated using the intraclass correlation (ICC) coefficient with a two-way analysis of variance with a random-effects model. An ICC of 0.75–1.00 indicated excellent agreement, 0.60–0.74 good agreement, 0.40–0.59 fair agreement, and <0.4 poor agreement.[13]

For further analysis, differences in ADC values were analyzed between (a) the LNM+ and LNM– groups and (b) LVSI+ and LVSI– using Student's t-test (when data were normally distributed) or the Mann–Whitney test. The diagnostic performance of the two ADC values was described by receiver operating characteristic (ROC) analysis. The area under the ROC curve (AUC) values were also calculated, and the DeLong test was used to compare the differences among the AUCs.

Table 1. Imaging parameters for ss-EPI and rs-EPI sequences

Parameters	ss-EPI	rs-EPI
Diffusion mode	Three-scan trace	Three-scan trace
Parallel imaging	GRAPPA	GRAPPA
Fat-suppressed	Fat sat; strong	Fat sat; strong
TR (ms)	6,400	6,400
TE (ms)	65	53
Partial fourier	0.625	0.625
Field of view (mm × mm)	320 × 320	320 × 320
Matrix	160 × 160	160 × 160
Number of slices	25	25
Slice thickness (mm)	3	3
Intersection gap (%)	20	20
Phase-encoding direction	Anteroposterior	Anteroposterior
Echo spacing (ms)	0.64	0.32
Bandwidth (Hz/Px)	2,840	925
EPI factor	136	113
Number of readout segments	1	5
b-value (s/mm^2)	50, 800	50, 800
Average	1, 5	1, 1
Scan time (min:sec)	2:20	2:18

ss-EPI, single-shot echo-planar imaging; rs-EPI, readout-segmented echo-planar imaging; GRAPPA, generalized autocalibrating partially parallel acquisition; TR, repetition time; TE, echo time.

Results

Patient characteristics

A total of 69 patients (mean age: 47.5 years; age range: 31–66 years) with CC were enrolled in this retrospective study. Of these, 17/52 patients were staged as LNM+/LNM–; in the LNM+ group, the lymph node size was <10 mm for all patients. Among all patients, 33/36 patients were staged as LVSI+/LVSI– according to the pathological findings. Table 2 shows the demographic characteristics of patients recruited in this study.

Comparison of subjective visual scores

Comparisons of the rs-EPI and ss-EPI quality based on a 5-point scoring system are shown in Table 3. Intra-observer agreement of the anatomic structure, artifacts and distortion, lesion conspicuity, and overall imaging quality were good or excellent in two DWI sequences (kappa statistics: ss-EPI 0.777, 0.726, 0.883, 0.787; rs-EPI 0.731, 0.879, 0.692, 0.705). For both readers, the rs-EPI group achieved significantly better scores in each aspect than the ss-EPI group (all *P* < 0.001, Table 3). Examples of rs-EPI advantages are presented in Figure 3.

Comparison of quantitative image quality

The mean and standard deviation of CNR and SNR are listed in Table 4. The SNR value of the rs-EPI was lower than that of the ss-EPI DWI (445.28 ± 107.33 vs. 138.60 ± 47.80, *P* <

Figure 2. Plots **(a-d)** show the schematic of drawing a CC lesion region of interest (ROI). **(a)** In the sagittal T2W image, the red dashed line region is the biggest image slice depicting cancer foci; **(b)** single-shot echo-planar imaging diffusion-weighted image with b-values of 800 s/mm^2 with an irregular ROI; **(c)** the corresponding apparent diffusion coefficient image with the ROI from **(b)**; **(d)** the corresponding axial contrast-enhanced T1W image. CC, cervical cancer.

Table 2. Patient characteristics

Characteristics	Patients	
Total number	69	
Age (years)	47.5 (31–66)	
Histological subtype	n	%
Squamous cell carcinoma	62	89.86
Adenocarcinoma	7	10.14
FIGO stage (2018)		
IB	32	46.38
IIA	20	28.99
°C	17	24.64
LNM status		
LNM+	17	24.64
LNM-	52	75.36
LVSI status		
LVSI+	33	47.83
LVSI-	36	52.17

FIGO, International Federation of Gynecology and Obstetrics 2018; LNM, lymph node metastasis; LVSI, lymphovascular space invasion.

0.001). In terms of CNR, there was no significant difference between the ss-EPI and rs-EPI (5.11 ± 1.55 vs. 4.89 ± 1.60, $P = 0.313$).

Diffusion-weighted imaging quantitative parameters

There was no significant difference between the ADC values for ss-EPI and rs-EPI DWI ($P = 0.949$, Table 4). The consistencies of ADC values between ss-EPI and rs-EPI were in complete agreement (ICC: 0.886, $P < 0.001$). Table 5 shows the ADC values corresponding to LVSI and LNM. In both ss-EPI and rs-EPI DWI, the ADC value of ss-EPI (ssADC) and that of rs-EPI (rsADC) were significantly lower in the LNM+ group than in the LNM- group ($P = 0.000, 0.000$, respectively), and there were no significant differences in the ssADC and rsADC values between the LVSI+ and LVSI- groups ($P = 0.271, 0.200$, respectively).

The results of the ROC analyses for ADC values used to distinguish LNM+ from LNM– are shown in Figure 4 and Table 5. The mean AUCs were 0.851 for ssADC and 0.855 for rsADC, and there was no significant difference between the two AUCs (DeLong test: $Z = 0.871$, $P = 0.163$).

Discussion

Comparison of image quality between readout-segmented echo-planar imaging and single-shot echo-planar imaging

The echo spacing of rs-EPI was reduced to 0.32 ms compared with 0.64 ms for ss-EPI in this study. Intra-reader agreement of subjective visual estimation evaluated using kappa was good or excellent, which confirmed the reliability of the investigation, and we demonstrated that the anatomic detail, ghosting artifacts, geometric distortion, and lesion conspicuity of CC DWI based on rs-EPI were significantly superior to ss-EPI. This is consistent with previous findings from abdomen pelvic research, which confirmed improved subjective visual assessments of image quality in rs-EPI for CC, endometrial carcinoma, rectal cancer, renal cancer, pelvic cancer, and sacroiliitis.[10,11,14-17] However, the quantitative image quality (SNR, CNR) of rs-EPI was not satisfactory. We found that the SNR on rs-EPI was significantly lower than on ss-EPI, which is consistent with previous studies.[18-22] No significant difference in the CNR of rs-EPI and ss-EPI was observed, which agrees with previous studies.[18,20-22]

The evaluation of the objective image quality between rs-EPI and ss-EPI continues to be a controversial topic. As the SNR depends to a large extent on the specific protocol followed, we adjusted the two DWI technique parameters of TR values, fields of view, matrices, slice thicknesses, and gaps to coincide for the facilitation of the quantitative comparison. Theoretically, a shorter TE results in an increased SNR. In principle, the shortened TE of rs-EPI was not enough to offset the increase in the SNR caused by

Table 3. Comparison of qualitative parameters between ss-EPI and rs-EPI

Parameters	Reader 1			Reader 2		
	ss-EPI mean ± SD	rs-EPI mean ± SD	*P* value	ss-EPI mean ± SD	rs-EPI mean ± SD	*P* value
Anatomic structure	3.12 ± 0.54	4.43 ± 0.58	<0.001	3.33 ± 0.63	4.47 ± 0.61	<0.001
Artifacts and distortion	3.06 ± 0.38	4.65 ± 0.48	<0.001	3.08 ± 0.34	4.61 ± 0.57	<0.001
Lesion conspicuity	3.71 ± 0.50	4.22 ± 0.77	<0.001	3.73 ± 0.53	4.27 ± 0.76	<0.001
Overall imaging quality	3.45 ± 0.50	4.41 ± 0.64	<0.001	3.35 ± 0.48	4.45 ± 0.50	<0.001

ss-EPI, single-shot echo-planar imaging; rs-EPI, readout-segmented echo-planar imaging; SD, standard deviation.

the increase in the ss-EPI average number to 5. Through setting the scanning parameters, we contribute to the similar scanning time of the two techniques (2 min 20 s vs. 2 min 18 s). We found that such imaging parameter settings and the results reached were similar to previous studies.[18-22] Recently, most images from rs-EPI tend to have higher SNRs and CNRs than the images from ss-EPI but with the expense of a comparatively longer scan time. However, in clinical practice, saving scanning time is an important task considering the numerous MR examinations performed daily. In the present study, the similar short scanning time of the two techniques (2 min 20 s vs. 2 min 18 s), under the premise of reasonable lower SNR of rs-EPI, was not representative of decreased definition or image quality; the SNR and CNR only refer to a part of the quantitative index of the image quality. rs-EPI technology mainly helps achieve higher resolution and reduces susceptibility artifacts and T2* blurring compared with ss-EPI technology. Under the premise of reasonable SNR and CNR, we need rs-EPI DWIs with these visual advantages to improve the accuracy of the early assessment of CC lesions.

In addition, the simultaneous multi-slice (SMS) acquisition scheme may be a potential solution to ensure a higher SNR and CNR while taking a shorter scan time. Instead of a consecutive excitation, the SMS technique concurrently excites multiple slices to decrease scan time.[23] This method has been successfully applied to MRI scanning of the kidney, liver, pancreas, breast, and rectum.[24-28] In the future, we will investigate the combination of SMS and readout-segmented techniques for cervix uteri DWI.

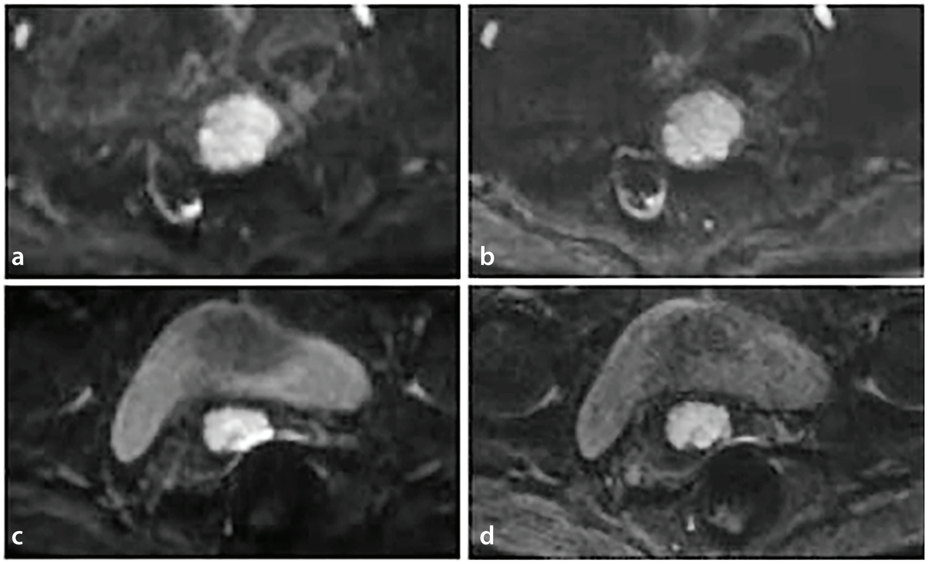


Figure 3. Plots **(a, b)** show a 51-year-old woman with cervical cancer (CC); **(a)** the single-shot echo-planar imaging (ss-EPI) shows blurred anatomical details and a diffusion-restricted lesion with blurred edges, making it unclear whether the lesion broke the outer edge of the cervical wall; **(b)** distinct boundary between the rectum, uterus, and diffusion-restricted lesion in readout-segmented echo-planar imaging (rs-EPIs). Plots **(c, d)** show a 54-year-old woman with CC; **(c)** ss-EPI demonstrating obvious artifacts shown as significantly enhanced local signal and evident distortion between the rectum and lesion; **(d)** minor artifacts and distortion in rs-EPIs.

Table 4. Comparison of quantitative parameters between ss-EPI and rs-EPI

Parameter	ss-EPI mean ± SD	rs-EPI mean ± SD	t value	*P* value
SNR	445.28 ± 107.33	138.60 ± 47.80	22.749	<0.001
CNR	5.11 ± 1.55	4.89 ± 1.60	1.016	0.313
ADC (×10^{-3} mm^2/s)	1.01 ± 0.17	1.02 ± 0.17	-0.064	0.949

SNR, signal to noise ratio; CNR, contrast to noise ratio; ADC, apparent diffusion coefficient; ss-EPI, single-shot echo-planar imaging; rs-EPI, readout-segmented echo-planar imaging; SD, standard deviation.

Application of apparent diffusion coefficient in the study of cervical cancer

In this study, rs-EPI showed clearer anatomical details, fewer artifacts and deformation, and higher lesion conspicuity; furthermore, the focal edge ROI outlined on the corresponding DWI truly reflected the tumor edge. We found that the ADC values obtained using ss-EPI and rs-EPI were not significantly different. As previously reported, the new DWI technique did not affect ADC quantification; the ADC values of rs-EPI DWI were as reliable as conventional ss-EPI but with better image quality.[14,17,29,30] Only a few studies found the opposite result; Wisner's study on breast tumors revealed that within benign lesions, malignant lesions, and normal tissue, the mean ADC measurements were lower with rs-EPI than with ss-EPI.[14] According to the authors, this was due to T2* blurring on rs-EPI. However, Xu et al.[21] and Zhao et al.[22] considered the distortion and artifacts of ss-EPIs to be a more reasonable explanation, as no adjacent normal tissue with high ADC value could be found in the sinonasal or orbital region. In the present study, drawing an irregular ROI avoided obvious necrotic, higher signal artifacts and deformation areas on ADC maps of CC, which may explain the lack of differences in ADC values between rs-EPI and ss-EPI DWI sequences.

This study also found that 17 cases eventually developed lymph node metastases and were finally classified as stage °C, and their lymph node size was <10 mm. In general, node size is the MRI criterion in distinguishing lymphatic metastasis from non-lymphatic metastasis; the threshold diameter is 10 mm in the short axis.[31] However, the size of metastatic, hyperplastic, and normal lymph nodes can overlap, and the small lymph nodes could undergo micro-metastases.[32] Williams et al.[33] verified via histological analysis that 54.5% of metastatic nodes were <10 mm in patients with gynecologic malignancy. Clearly, the assessment of lymph node size alone increased the false-negative rate of LNM. In the present study, the inclusion criteria included surgical radical treatment that was abandoned due to suspected LNM based on imaging findings. Specifically, cases with irregular lymph node morphology or a short axis diameter >10 mm on MRI were excluded. In essence, the 17 cases of lymph node metastases that were missed by MRI and not confirmed as positive by LNM exhibited more subtle pathological significance that is not readily apparent to the naked eye. The measured ADC values in these cases can more accurately reflect the internal tumor cell proliferation and tumor aggressiveness within the lesion. This suggests potential value in differentiating ADC values between the LNM+ and LNM- cases. Although the number of lymph node-positive cases was limited to 17, this sample size remains statistically acceptable for analysis.

We validated that both marginal region rsADC and ssADC values of CC could predict lymph node metastatic status (*P* = 0.000, 0.000, respectively), in agreement with a previous study;[34] both rsADC and ssADC values had a good diagnostic efficacy (AUC: 0.855, 0.851, respectively). In addition, neither of these two ADC values could distinguish the LVSI status of CC, in agreement with previous studies.[35,36] LVSI-positive lesion tumors are characterized by their high invasiveness and a complex microenvironment, which includes an increased density of cancer cells, a high nuclear-cytoplasmic ratio, and reduced extracellular space. Additionally, the vascular blood flow within these tumors is marked by high perfusion. These features collectively illustrate the biological characteristics of tumor heterogeneity. The ADC value serves as a quantifier for the degree of diffusion movement of water molecules, which correlates with cell density in biological tissues. However, it is important to note that, in contrast to cell density, the vascular properties of tumors exert an opposing influence on ADC values, highlighting the distinct physical and biological differences between diffusion and perfusion in tumor tissues;[37] in other words, the heterogeneous ADC values are insufficient for evaluating the LVSI+ status in CC. Even with the higher image quality of rs-EPI, identifying the LVSI status using the ADC from the tumor edge region remains challenging. Nevertheless, a study by Yang et al.[38] concluded that minimum-ADC values could predict LVSI in CC; the ROI of the tumor location was plotted layer by layer on DWIs, with a total of 20–30 ROIs placed (40–50 mm^2), ultimately selecting the ROI with the smallest ADC value. Yang et al.[38] explained that minimum-ADC values reflect a higher density of tumor cells, which may be more sensitive to tumor proliferation. Therefore, minimum ADC values could serve as a feasible and objective parameter to reduce ADC variability. In contrast, Cheng et al.[39] investigated the value of ADC histogram analysis based on whole tumor volume for the preoperative prediction of LVSI. The conclusion drawn is that ADCmax, ADCrange, ADC90, ADC95, and ADC99 were significantly lower in the LVSI+ group than in the LVSI– group (all *P* values <0.05). This finding may be related to the inclusion of hemorrhagic, necrotic, and cystic regions in the volume of interest (VOI). It is evident that different ADC-derived parameters exhibit variations in their ability to identify LVSI, which depend on the heterogeneity represented in the measured ADCs and are closely related to the ROI or VOI. The original intention of our study was to use the tumor margin area, which is prone to artifacts, as the monitoring reference point. The obtained value represented an average ADC value of the margin area, differing from the ADC measurement values of various regions within the tumor or the ADC histogram of the entire tumor volume.

DWI improves uterine tumor detection and characterization and the visualization of small implants in peritoneal carcinomatosis.[40] To improve the accuracy of a qualitative assessment of CC, such as judging parametrial invasion, DWIs should be included.[41] On this basis, the present study found that advantage should be taken of the higher image quality of rs-EPI, as it is, to some extent, more inclined to the subjective and qualitative judgment of CC, providing a better choice for making DWI sequences in the future.

This study has several limitations. First, some cervical tumors with small lesion diameters (<10 mm) were excluded because of insufficient image pixels for analysis. Analyses of advanced CC (stages IIB–IVA) were also limited because most patients with advanced CC were excluded from concurrent chemo-

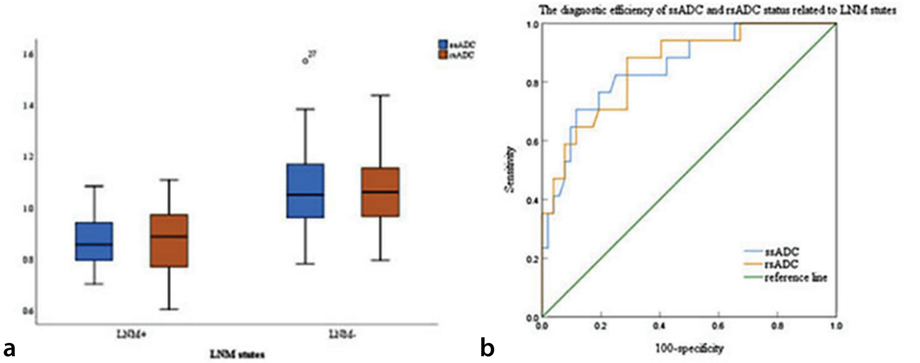


Figure 4. Plots **(a, b)** show boxplots of apparent diffusion coefficient values and receiver operating characteristic curves, respectively. ADC, apparent diffusion coefficient; LNM, lymph node metastasis.

Table 5. Diagnostic efficiency of apparent diffusion coefficient values in ss-EPI and rs-EPI with respect to LNM and LVSI status

ADC value ($\times 10^{-3}$ mm^2/s)	LNM status		*P* value	AUC	95% CI	LVSI status		*P* value
	LNM+ mean ± SD	LNM- mean ± SD				LVSI+ mean ± SD	LVSI- mean ± SD	
ssADC	0.87 ± 0.11	1.07 ± 0.16	0.000	0.851	0.744–0.925	0.99 ± 0.20	1.04 ± 0.13	0.271
rsADC	0.86 ± 0.14	1.07 ± 0.15	0.000	0.855	0.754–0.956	0.99 ± 0.21	1.05 ± 0.13	0.200

ADC, apparent diffusion coefficient; ssADC, apparent diffusion coefficient value of ss-EPI; rsADC, apparent diffusion coefficient value of rs-EPI; AUC, area under the receiver operating characteristic curve; LNM, lymph node metastasis; LVSI, lymphovascular space invasion; ss-EPI, single-shot echo-planar imaging; rs-EPI, readout-segmented echo-planar imaging; SD, standard deviation; CI, confidence interval.

radiotherapy without surgery or receiving neoadjuvant chemoradiotherapy prior to the surgery. Therefore, there were potential selection biases intrinsic to this retrospective single-center study. Second, the ROI in this study was delineated manually along the tumor margins on diffusion-weighted images. This approach may inadvertently introduce errors and deviations in determining tumor boundaries, and it is also notably time-consuming. Therefore, automated or semi-automated segmentation methods may represent a more efficient alternative. Future research should focus on enhancing comparative studies of these methods to develop a more effective research strategy. Third, although the evaluation of LVSI status in our study cohort was negative, we need to measure different ADCs (including tumor ADC, mini-ADC, and mini-ADC ratio) (32) and further estimate the relationship between ADC and LVSI status in large-cohort multi-center studies.

In conclusion, Over a similar scan time, rs-EPI significantly improves the qualitative image quality of DWI and has good diagnostic accuracy for LNM status in CC compared with conventional ss-EPI DWI, especially for the margin of CC lesions. However, neither could predict the LVSI status. Including rs-EPI DWI in routine clinical protocols for MRI may be a better choice.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This study was supported by the Opening Project of Medical Imaging Key Laboratory of Sichuan Province (grant MIKLSP202106) and the popularization and application project of Chengdu Seventh People's Hospital-Chengdu Medical College joint scientific research (grant 2022LHIYPJ-03).

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* . 2018;68(6):394-424. Erratum in: *CA Cancer J Clin*. 2020;70(4):313. [Crossref]
2. Padera TP, Kadambi A, di Tomaso E, et al. Lymphatic metastasis in the absence of functional intratumor lymphatics. Science. 2002;296(5574):1883-1886. [Crossref]
3. Bhatla N, Aoki D, Sharma DN, Sankaranarayanan R. Cancer of the cervix uteri. *Int J Gynaecol Obstet*. 2018;143(Suppl 2):22-36. [Crossref]
4. Koh WJ, Abu-Rustum NR, Bean S, et al. Cervical Cancer, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *Natl Comp Canc Netw*. 2019;17(1):64-84. [Crossref]
5. Landoni F, Maneo A, Colombo A, et al. Randomised study of radical surgery versus radiotherapy for stage Ib-IIa cervical cancer. *Lancet*. 1997;350(9077):535-540. [Crossref]
6. Nathanson SD. Insights into the mechanisms of lymph node metastasis. *Cancer*. 2003;98(2):413-423. [Crossref]
7. Singh N, Arif S. Histopathologic parameters of prognosis in cervical cancer--a review. *Int J Gynecol Cancer*. 2004;14(5):741-750. [Crossref]
8. Wenkel E, Geppert C, Schulz-Wendtland R, et al. Diffusion weighted imaging in breast MRI: comparison of two different pulse sequences. *Acad Radiol*. 2007;14(9):1077-1083. [Crossref]
9. Porter DA, Heidemann RM. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magn Reson Med*. 2009;62(2):468-475. [Crossref]
10. Qian W, Chen Q, Zhang Z, Wang H, Zhang J, Xu J. Comparison between readout-segmented and single-shot echo-planar imaging in the evaluation of cervical cancer staging. *Br J Radiol*. 2019;92(1094):20180293. [Crossref]
11. Zhang H, Huang H, Zhang Y, et al. Diffusion-weighted MRI to assess sacroiliitis: improved image quality and diagnostic performance of readout- segmented echo- planar imaging (EPI) over conventional single-shot EPI. *AJR Am J Roentgenol* .2021;217(2):450-459. [Crossref]
12. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360-363. [Crossref]
13. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. [Crossref]
14. Xie M, Ren Z, Bian D, et al. High resolution diffusion-weighted imaging with readout segmentation of long variable echo-trains for determining myometrial invasion in endometrial carcinoma. *Cancer Imaging*. 2020;20(1):66. [Crossref]
15. Xia CC, Pu J, Zhang JG, et al. Readout-segmented echo-planar diffusion-weighted MR for the evaluation of aggressive characteristics of rectal cancer. *Sci Rep*. 2018;8(1):12554. [Crossref]
16. Thian YL, Xie W, Porter DA, Weileng Ang B. Readout-segmented echo-planar imaging for diffusion-weighted imaging in the pelvis at 3T-A feasibility study. *Acad Radiol*. 2014;21(4):531-537. [Crossref]
17. Friedli I, Crowe LA, Viallon M, et al. Improvement of renal diffusion-weighted magnetic resonance imaging with readout-segmented echo-planar imaging at 3T. *Magn Reson Imaging*. 2015;33(6):701-708. [Crossref]
18. Bogner W, Pinker-Domenig K, Bickel H, et al. Readout-segmented echo- planar imaging improves the diagnostic performance of diffusion-weighted MR breast examinations at 3.0 T. *Radiology*. 2012;263(1):64-76. [Crossref]
19. Yeom KW, Holdsworth SJ, Van AT, et al. Comparison of readout-segmented echo-planar imaging (EPI) and single-shot EPI in clinical application of diffusion-weighted imaging of the pediatric brain. *AJR Am J Roentgenol*. 2013;200(5):W437-W443. [Crossref]
20. Xu XQ, Liu J, Hu H, et al. Improve the image quality of orbital 3-T diffusion-weighted magnetic resonance imaging with readout-segmented echo-planar imaging. *Clin Imaging*. 2016;40(4):793-796. [Crossref]
21. Xu X, Wang Y, Hu H, et al. Readout-segmented echo-planar diffusion-weighted imaging in the assessment of orbital tumors: comparison with conventional single-shot echo-planar imaging in image quality and diagnostic performance. *Acta Radiol*. 2017;58(12):1457-1467. [Crossref]
22. Zhao M, Liu Z, Sha Y, et al. Readout-segmented echo- planar imaging in the evaluation of sinonasal lesions: a comprehensive comparison of image quality in single-shot echo-planar imaging. *Magn Reson Imaging*. 2016;34(2):166-172. [Crossref]
23. Moeller S, Yacoub E, Olman CA, et al. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med*. 2010;63(5):1144-1153. [Crossref]
24. Song SE, Woo OH, Cho KR, et al. Simultaneous multislice readout-segmented echo planar imaging for diffusion-weighted MRI in patients with invasive breast cancers. *J Magn Reson Imaging*. 2021;53(4):1108-1115. [Crossref]
25. Kenkel D, Barth BK, Piccirelli M, et al. Simultaneous multislice diffusion-weighted imaging of the kidney: a systematic analysis of image Quality. *Invest Radiol*. 2017;52(3):163-169. [Crossref]
26. Obele CC, Glielmi C, Ream J, et al. Simultaneous multislice accelerated free-breathing diffusion-weighted imaging of the liver at 3T. *Abdom Imaging*. 2015;40(7):2323-2330. [Crossref]
27. Taron J, Martirosian P, Kuestner T, et al. Scan time reduction in diffusion-weighted imaging of the pancreas using a simultaneous multislice technique with different acceleration factors: How fast can we go? *Eur Radiol*. 2018;28(4):1504-1511. [Crossref]

28. Huang H, Zhou M, Gong T, Wang Y. Feasibility of high-resolution readout-segmented echo-planar imaging with simultaneous multislice imaging in assessing rectal cancer. *Abdom Radiol (NY)*. 2023;48(7):2258-2269. [Crossref]

29. Byeon J, Kim JY, Cho AH. Readout-segmented echo-planar imaging in diffusion- weighted MR imaging of acute infarction of the brainstem and posterior fossa: comparison of single-shot echo-planar diffusion-weighted sequences. *Clin Imaging*. 2015;39(5):765-769. [Crossref]

30. Yao F, Huang M, Li J, Gao X. Yao F, Huang M, Li J, Gao X. Readout-segmented diffusion weighted imaging of the testis at 3.0 T: comparison with single-shot echo-planar imaging. *Abdom Radiol (NY)*. 2023;48(6):2131-2138. [Crossref]

31. Choi HJ, Kim SH, Seo SS, et al. MRI for pretreatment lymph node staging in uterine cervical cancer. *AJR Am J Roentgenol*. 2006;187(5):W538-W543. [Crossref]

32. Balleyguier C, Sala E, Da Cunha T, et al. Staging of uterine cervical cancer with MRI: guidelines of the European Society of Urogenital Radiology. *Eur Radiol*. 2011;21(5):1102-1110. [Crossref]

33. Williams AD, Cousins C, Soutter WP, et al. Detection of pelvic lymph node metastases in gynecologic malignancy: a comparison of CT, MR imaging, and positron emission tomography. *AJR Am J Roentgenol*. 2001;177(2):343-348. [Crossref]

34. Mi HL, Suo ST, Cheng JJ, et al. The invasion status of lymphovascular space and lymph nodes in cervical cancer assessed by mono-exponential and bi-exponential DWI-related parameters. *Clin Radiol*. 2020;75(10):763-771. [Crossref]

35. Li S, Liu J, Zhang F, et al. Novel T2 mapping for evaluating cervical cancer features by providing quantitative T2 maps and synthetic morphologic images: a preliminary study. *J Magn Reson Imaging*. 2020;52(6):1870-1871. [Crossref]

36. Li S, Zhang Z, Liu J, et al. The feasibility of a radial turbo-spin-echo T2 mapping for preoperative prediction of the histological grade and lymphovascular space invasion of cervical squamous cell carcinoma. *Eur J Radiol*. 2021;139:109684. [Crossref]

37. Zhang Q, Ouyang H, Ye F, et al. Multiple mathematical models of diffusion-weighted imaging for endometrial cancer characterization: correlation with prognosis- related risk factors. *Eur J Radiol*. 2020;130:109102. [Crossref]

38. Yang W, Qiang JW, Tian HP, Chen B, Wang AJ, Zhao JG. Minimum apparent diffusion coefficient for predicting lymphovascular invasion in invasive cervical cancer. *J Magn Reson Imaging*. 2017;45(6):1771-1779. [Crossref]

39. Cheng JM, Luo WX, Tan BG, Pan J, Zhou HY, Chen TW. Whole-tumor histogram analysis of apparent diffusion coefficients for predicting lymphovascular space invasion in stage IB-IIA cervical cancer.*Front Oncol*. 2023;13:1206659. [Crossref]

40. Otero-García MM, Mesa-Álvarez A, Nikolic O, et al. Role of MRI in staging and follow-up of endometrial and cervical cancer: pitfalls and mimickers. *Insights Imaging*. 2019;10(1):19. [Crossref]

41. Addley H, Moyle P, Freeman S. Diffusion-weighted imaging in gynaecological malignancy. *Clin Radiol*. 2017;72(11):981-990. [Crossref]

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.243102

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Pix2Pix generative-adversarial network in improving the quality of T2-weighted prostate magnetic resonance imaging: a multi-reader study

Yeliz Başar[1]
Mustafa Said Kartal[2]
Mustafa Ege Seker[3]
Deniz Alis[4]
Delal Seker[5]
Müjgan Orman[1]
Sabri Şirolu[6]
Serpil Kurtcan[1]
Aydan Arslan[7]
Nurper Denizoğlu[8]
İlkay Öksüz[9]
Ercan Karaarslan[4]

[1] Acıbadem Healthcare Group, Department of Radiology, İstanbul, Türkiye

[2] Cumhuriyet University Faculty of Medicine, Sivas, Türkiye

[3] University of Wisconsin-Madison, School of Medicine, Department of Radiology, Madison, USA

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

[5] Dicle University Faculty of Engineering, Department of Electrical-Electronics Engineering, Diyarbakır, Türkiye

[6] University of Health Sciences Türkiye, Şişli Hamidiye Etfal Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

[7] Ümraniye Training and Research Hospital, Clinic of Radiology, İstanbul, Türkiye

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

[9] İstanbul Technical University Faculty of Engineering, Department of Computer Engineering, İstanbul, Türkiye

Corresponding author: Deniz Alis

E-mail: drdenizalis@gmail.com

Received 28 October 2024; revision requested 23 November 2024; last revision received 22 January 2025; accepted 02 February 2025.

Epub: 05.05.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.243102

PURPOSE

To assess the performance and feasibility of generative deep learning in enhancing the image quality of T2-weighted (T2W) prostate magnetic resonance imaging (MRI).

METHODS

Axial T2W images from the prostate imaging: cancer artificial intelligence dataset (n = 1,476, biologically males; n = 1,500 scans) were used, partitioned into training (n = 1300), validation (n = 100), and testing (n = 100) sets. A Pix2Pix model was trained on original and synthetically degraded images, generated using operations such as motion, Gaussian noise, blur, ghosting, spikes, and bias field inhomogeneities to enhance image quality. The efficacy of the model was evaluated by seven radiologists using the prostate imaging quality criteria to assess original, degraded, and improved images. The evaluation also included tests to determine whether the images were original or synthetically improved. Additionally, the model's performance was tested on the in-house external testing dataset of 33 patients. The statistical significance was assessed using the Wilcoxon signed-rank test.

RESULTS

Results showed that synthetically improved images [median score (interquartile range) 4.71 (1)] were of higher quality than degraded images [3.36 (3), $P = 0.0001$], with no significant difference from original images [5 (1.14), $P > 0.05$]. Observers equally identified original and synthetically improved images as original (52% and 53%), proving the model's ability to retain realistic attributes. External testing on a dataset of 33 patients confirmed a significant improvement ($P = 0.001$) in image quality, from a median score of 4 (2.286)–4.71 (1.715).

CONCLUSION

The Pix2Pix model, trained on synthetically degraded data, effectively improved prostate MRI image quality while maintaining realism and demonstrating both applicability to real data and generalizability across various datasets.

CLINICAL SIGNIFICANCE

This study critically assesses the efficacy of the Pix2Pix generative-adversarial network in enhancing T2W prostate MRI quality, demonstrating its potential to produce high-quality, realistic images indistinguishable from originals, thereby potentially advancing radiology practice by improving diagnostic accuracy and image reliability.

KEYWORDS

Deep learning, generative artificial intelligence, prostate, magnetic resonance imaging, prostate imaging quality

You may cite this article as: Başar Y, Kartal MS, Seker ME, et al. Pix2Pix generative-adversarial network in improving the quality of T2-weighted prostate magnetic resonance imaging: a multi-reader study. *Diagn Interv Radiol.* 2025;31(6):547-565.

The prostate imaging reporting and data system (PI-RADS) and its updates prescribe best practices for the acquisition and interpretation of prostate magnetic resonance imaging (MRI) scans,[1] emphasizing minimum technical requirements to ensure scan quality, which is crucial for the accurate detection of clinically significant prostate cancer. However, adherence to PI-RADS guidelines does not invariably guarantee high-quality MRI scans, as evidenced by various studies.[2-4]

Deep learning (DL)-based reconstruction techniques can speed up image acquisition and improve quality beyond traditional MRI methods.[5-8] Nonetheless, variables such as patient characteristics, equipment quality, and the expertise of the radiology team can still result in suboptimal images.[3,9] Moreover, DL-based reconstruction typically requires newer scanner models and significant initial investments, limiting its accessibility. However, limited research has been conducted on applying DL techniques to enhance prostate MRI quality. Existing studies are often constrained by the use of single-center datasets and proprietary scoring systems, which can affect the reproducibility of their outcomes.[10]

In this study, we employed a generative-adversarial network (GAN) model, Pix2Pix, to enhance the quality of axial T2-weighted (T2W) prostate MRI. We used a large-scale, multi-center, and publicly available dataset, prostate imaging: cancer artificial intelligence (PI-CAI),[11] allowing us to overcome some of the limitations noted in previous studies. Image quality was evaluated by multiple readers from different centers using a scoring system adopted from the newly introduced Prostate imaging quality (PI-QUAL):[12] which provided a standardized assessment method. We also examined the realism of the generated images and tested the model's performance on the in-house external testing dataset to evaluate its generalizability.

Main points

- The Pix2Pix generative-adversarial network (GAN) significantly improved T2-weighted prostate magnetic resonance imaging (MRI) quality while maintaining realism.
- Synthetically improved images scored higher than degraded ones when compared with the original images.
- External testing confirmed significant image quality improvement.
- Radiologists could not distinguish between original and synthetically improved images.
- The study demonstrated GANs' potential for realistic, high-quality prostate MRI enhancement.

Methods

Study sample

The Acıbadem University and Healthcare Institution's Medical Research Ethics Committee approved this retrospective study and waived the requirement for informed consent for the retrospective collection, analysis, and presentation of anonymized medical data (date: 11.02.2021, decision no: 2021-03/12).

This study utilized the publicly available PI-CAI training dataset, which consisted of 1,500 bi-parametric prostate MRI scans obtained from 1,476 biologically male individuals at 4 tertiary academic centers in the Netherlands and Norway between March 2015 and January 2018. The data from these four centers were stratified across the training, validation, and internal testing sets to ensure representation from each center in all data partitions. The examinations were stratified into three distinct groups: a development set (1400 scans, 1300 training, and 100 validation) and a testing set (100 exams). This stratification was done with careful consideration to ensure that scans from the same patient were not included across the development and testing sets. The flowchart of the study is given in Figure 1.

We also included the in-house dataset of 33 bi-parametric MRI examinations from 33 biologically male individuals as an external testing dataset in this study. The overall workflow of the study is shown in Figure 2.

Bi-parametric magnetic resonance imaging examinations

All bi-parametric MRI scans of the PI-CAI dataset were conducted using 1.5T units (n = 82) from Siemens (Aera and Avanto models, Siemens Healthcare, Erlangen, Germany) and Philips (Achieva and Intera models, Philips Healthcare, Eindhoven, the Netherlands), as well as 3T units (n = 1418) from Siemens (Skyra, TrioTim, and Prisma models, Siemens Healthcare, Erlangen, Germany) and Philips (Ingenia model, Philips Healthcare, Eindhoven, the Netherlands). These scans utilized surface coils and adhered to the PI-RADS V2 guidelines. Additional specifications regarding the MRI protocols used for the study sample are detailed in.[11]

The examinations of the in-house testing dataset were performed using 1.5T units from Siemens (Avanto-fit, Siemens Healthcare, Erlangen, Germany). These scans were also performed with surface coils. For this study, only axial T2W images were used for further analysis. Table 1 shows the imaging protocol for the in-house testing dataset.

Synthetic data creation

In this study, a crucial step involved the creation of a robust training dataset by applying clinically relevant MRI artifacts to create realistic low-quality T2W images. For this purpose, TorchIO library[13] was used: a powerful tool specifically designed for data augmentation in medical imaging. A variety of techniques were employed to simulate commonly encountered artifacts, including motion, Gaussian noise, blur, ghosting, spikes, and bias field inhomogeneities (detailed in Supplementary S1).

All images were normalized and resized to uniform dimensions to facilitate consistent neural network training, ensuring each image had intensity values within a specific range for optimal input standardization.

The training set included images manipulated with each artifact individually, as well as in specific combinations, enabling the Pix2Pix model to learn from a wide variety of possible artifact scenarios and improve its ability to generalize across different image quality corruption. Internal testing was

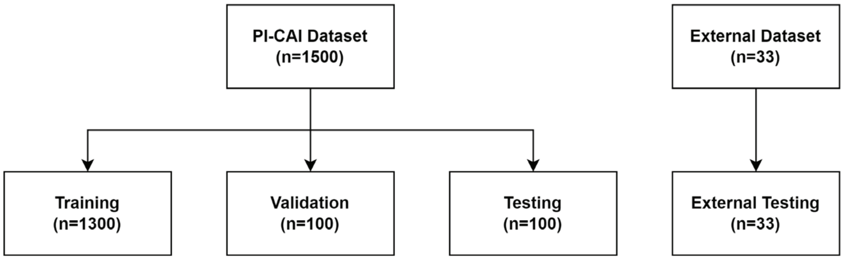


Figure 1. Flowchart of the study. PI-CAI, prostate imaging: cancer artificial intelligence.

created using either a single artifact or a predefined combination of multiple artifacts. This setup allowed for a controlled evaluation of the model's performance in enhancing images with known quality issues. Detailed descriptions of the data pre-processing are given in Supplementary Document S1.

Pix2Pix model

The Pix2pix model, a conditional GAN, was utilized for image-to-image translation using paired images to improve accuracy. It consisted of a generator, employing a U-Net architecture to maintain anatomical features in medical images, and a discriminator, which used a PatchGAN classifier to focus on high-frequency details and realism by evaluating small patches within the images.[14,15]

The training involved an adversarial process where the generator tried to create increasingly realistic images, whereas the discriminator improved at detecting synthetically improved images. The process was governed by a combined loss function: adversarial loss ensured the images were visually indistinguishable from the original ones, and L1 loss maintained structural integrity, reducing blurring and preserving crucial details. This setup enhanced the model's ability to produce clinically useful MRI images while retaining essential diagnostic features.

The Pix2pix model was trained using the Adam optimizer with a learning rate of 0.0002, focusing on 200 epochs where the L1 loss was emphasized initially to enhance accuracy. The training involved an adversarial setup where the generator aimed to produce images close to the original ones by minimizing both L1 and adversarial loss, whereas the discriminator sought to identify whether the image patches were original or synthetic, aiming to maximize the adversarial loss. A detailed description of the model is given in Supplementary Document S1.

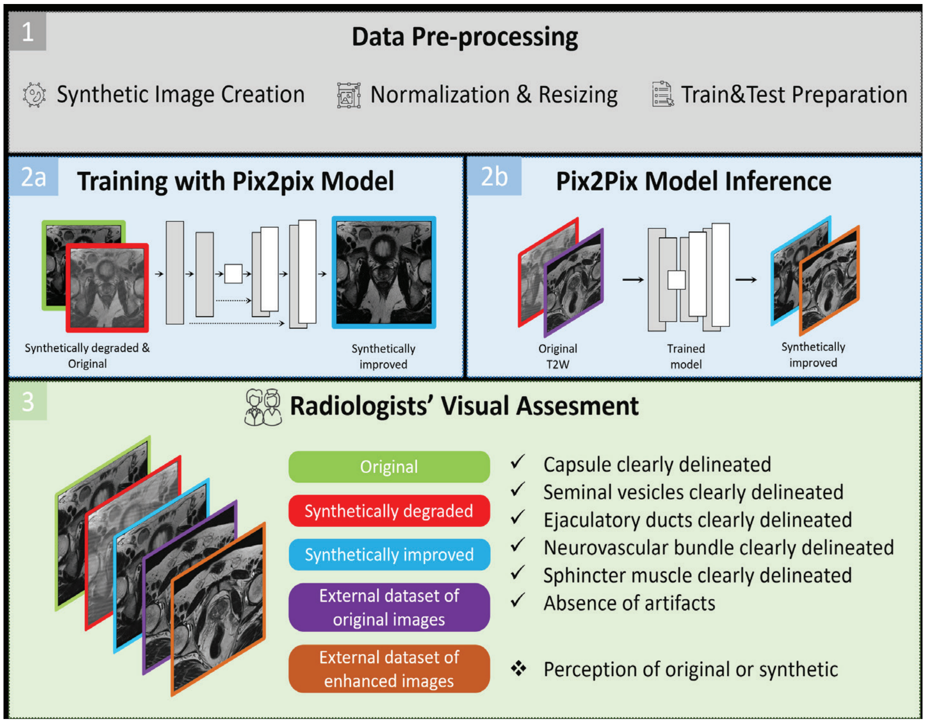


Figure 2. This figure illustrates the three main stages of the study. First, during data pre-processing, synthetically degraded T2-weighted images were created to mimic real-world artifacts, then normalized and resized along with the original images before being split into training and testing sets. Next, in the model training and inference stage, the Pix2Pix model was trained on paired original and synthetically degraded images **(2a)**, with the generator learning to produce improved images from degraded inputs while the discriminator differentiated between original and synthetically improved images. The trained model was then applied to an external in-house testing dataset, generating improved versions of the original images **(2b)**. Finally, radiologists visually assessed a set of images, including original, synthetically degraded, synthetically improved, and improved images from the external in-house testing dataset, evaluating them based on predefined criteria encompassing anatomical delineation, artifact presence, and perceived realism.

Table 1. Prostate T2-WI acquisition parameters of the in-house external testing dataset

Parameters	Values
TR (ms)	5454
TE (ms)	101
FOV (mm)	268 × 400
Matrix size	206 × 512
Slice thickness (mm)	4
Slice gap (mm)	0.6
Flip angle	150°
Temporal resolution (s)	-

T2WI, T2-weighted imaging; TR, repetition time; TE, echo time; FOV, field of view.

The model's performance during training was monitored using the mean absolute error (MAE) calculated between the reconstructed images and the original high-quality images. The model with the lowest MAE on the validation data was selected as the best-performing model for subsequent evaluation and application to the test set.

The best-performing model was then applied to synthetically degraded internal testing data (n = 100) and original images from the in-house external testing dataset (n = 33) to assess performance, as detailed in the subsequent sections.

Study readers

Seven readers participated in the analysis of the scans for this study. Reader 1, an expert prostate radiologist, interpreted over 300 cases annually for more than 10 years. Readers 2–7 were basic prostate readers, each handling 150–200 cases per year for 2–7 years. The classification of the readers adhered to the consensus statement of the European Society of Urogenital Radiology.[16] Readers one and three were from the same center, whereas the others were based in various other hospitals, ranging from academic to non-academic settings.

Assessment criteria

The evaluation by the readers was adopted from the visual assessment criteria proposed in the PI-QUAL for T2W imaging.[12] Specifically, the readers assessed the clarity with which they could delineate the capsule, seminal vesicles, ejaculatory ducts, neurovascular bundle, and sphincter muscle, awarding one point for each positively identified structure (i.e., if the structure could be seen clearly). Additionally, they awarded one point in the absence of artifacts and zero points if artifacts were present. Thus, the total score for each examination ranged from zero (the worst quality) to six points (the best quality).

Before the reading sessions, several online meetings were conducted to familiarize the readers with the PI-QUAL criteria through examples from published papers[17] and to acquaint them with the reading platform. The primary aim of these sessions was to enhance their understanding of the PI-QUAL.

We used only axial T2W images for the reading sessions, as the model employed in the current study was designed to work with axial T2W images. Although this may be considered a limitation, it was consistent with earlier work, which primarily focused on axial images as they were the primary sequence used in PI-RADS assessments.

Case reading sessions

The readers used a dedicated workstation equipped with a 6-megapixel diagnostic color monitor (Radiforce RX 660, EIZO) and a dedicated browser-based platform (https://matrix.md.ai). All reviewed images were in the Digital Imaging and Communications in Medicine format.

Initially, 7 readers evaluated 300 T2W series in the internal testing set of 100 patients, consisting of 100 original, 100 synthetically degraded, and 100 synthetically improved series. The readers independently assessed the cases in a random order to minimize bias, not knowing which images were original, degraded, or improved. They assigned points to each examination based on the previously described criteria and judged whether the images were original or synthesized.

Subsequently, to further evaluate the model's performance and its ability to enhance image quality on real data, the readers assessed the scans in the in-house external testing dataset of 33 patients, which included 33 original and 33 synthetically improved T2W series.

Statistical analysis

Statistical analyses were performed using the SciPy library in Python version 3. Continuous variables were presented using medians and interquartile ranges, whereas categorical and ordinal variables were presented with frequencies and percentages.

The structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) were used as quantitative metrics to assess image quality. The SSIM evaluated perceptual similarity by comparing luminance, contrast, and structure between images, with a range from −1 to 1, where 1 indicated perfect similarity. The PSNR measured the ratio between the maximum possible signal value and the distortion introduced, expressed in decibels, where higher values indicated better quality.

For comparing image quality assessments across original, synthetically degraded, and synthetically improved images, pairwise comparisons were conducted using the Friedman test and post-hoc Durbin-Conover test due to the matched nature of the data. For the pairwise comparison of the in-house external testing dataset, the Wilcoxon signed-rank test was used.

To evaluate the performance of radiologists in correctly identifying original versus synthetically improved images, accuracy was calculated. To analyze the differences in radiologists' ability to detect synthetically improved versus original images, McNemar's test was used. A *P* value less than 0.05 was considered statistically significant.

Results

Image quality assessment

We included 100 examinations in the testing set, each paired with their synthetically degraded and improved versions from the PI-CAI testing set. The PSNR and SSIM values of synthetically improved images [PSNR: 28.79 (32.54), SSIM: 0.92 (0.16)] were statistically significantly higher than those of the degraded image forms [PSNR: 24.87 (15.27), SSIM: 0.78 (0.13)] (PSNR: $P < 0.001$, SSIM: $P < 0.001$).

During the random blinded assessment, the observers gave median scores of 5 (1.14) to the original images, 3.36 (3) to the synthetically degraded images, and 4.71 (1) to the synthetically improved images ($P = 0.0001$). Pair-wise comparisons revealed that original images had a significantly higher median quality score than the synthetically degraded images ($P < 0.0001$). Likewise, synthetically improved images also had a higher image quality than synthetically degraded images ($P < 0.0001$). No statistically significant difference was found between the median image quality of the original and synthetically improved images ($P = 0.37$) (Figure 3a). A detailed breakdown of each reader's median scores for original, synthetically degraded, and synthetically improved images is given in Supplementary Document S2.

Figure 4 shows a representative example of original, synthetically degraded, and synthetically improved images. More representative examples can be found in Supplementary Document 2.

Original vs. synthetic assessment

We evaluated whether the observers could discriminate between original and synthetically improved T2W images using a majority voting scheme from the PI-CAI testing set. In this test, the observers identified 52% of the original and 53% of the synthetically improved images as original, with no statistical difference, indicating that the observers could not reliably discriminate between original and synthetically improved images ($P = 0.62$). A detailed breakdown of each reader's assessments on whether the images are original or synthetic is provided in Supplementary Document S2.

External testing

We evaluated whether the proposed model could also improve original images from the in-house external testing dataset. This set consisted of T2W images of 33 patients from the in-house center, where prostate images were obtained using a 1.5T scanner. The observers gave a median score of 4 (2.286) for the original images in the in-house external testing dataset. The median image quality score for this dataset was statistically lower than that for the original T2W images from the PI-CAI testing set ($P = 0.009$).

The proposed model improved the image quality of the original images from 4 (2.2) to 4.71 (1.7), demonstrating a statistically significant improvement ($P = 0.001$) (Figure 3b). Notably, after the improvement, we found no statistical difference in median image quality between the original images from the PI-CAI dataset [median: 5 (1.14)] and the synthetically improved images from the in-house dataset [median: 4.71 (1.7)] ($P = 0.16$). A detailed breakdown of each reader's median scores for original and synthetically improved images for the in-house external testing dataset is given in Supplementary Document S2.

Figure 5 shows representative examples of original and synthetically improved images of a patient along with observers' ratings from the in-house external testing dataset. More representative examples can be found in Supplementary Document S2.

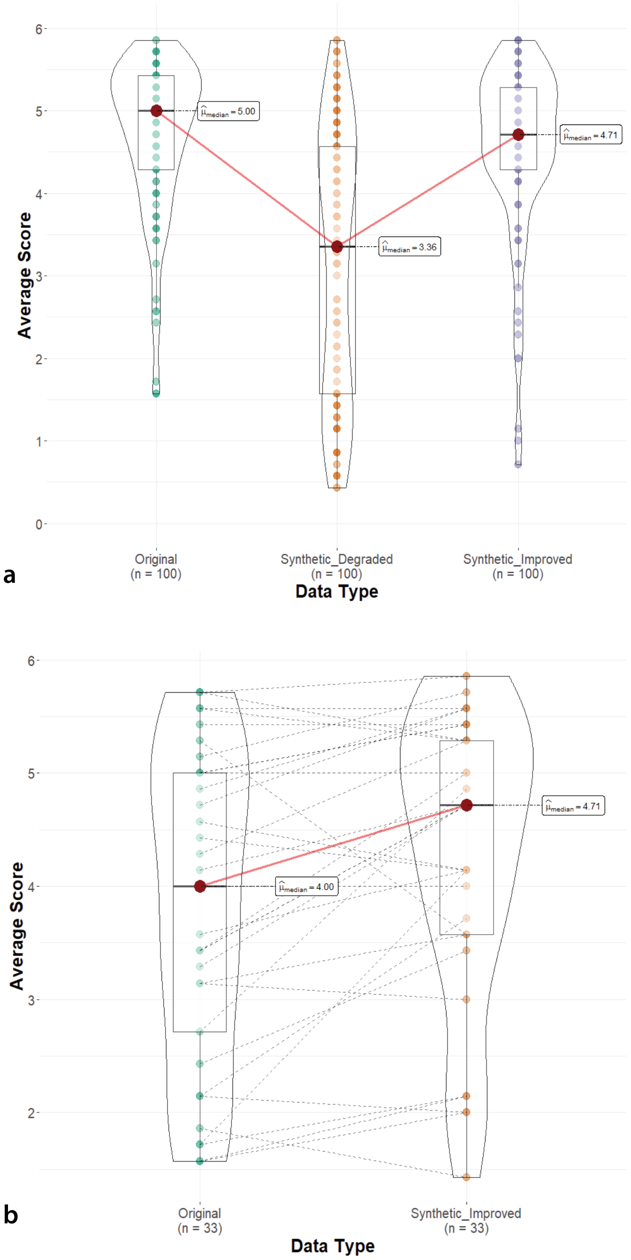


Figure 3. Comparison of the radiologists' scores for original, synthetically degraded, and synthetically improved T2-weighted images **(a)**. Comparison of the radiologists' average score for original and synthetically improved prostate magnetic resonance imaging scans from the in-house external testing dataset **(b)**.

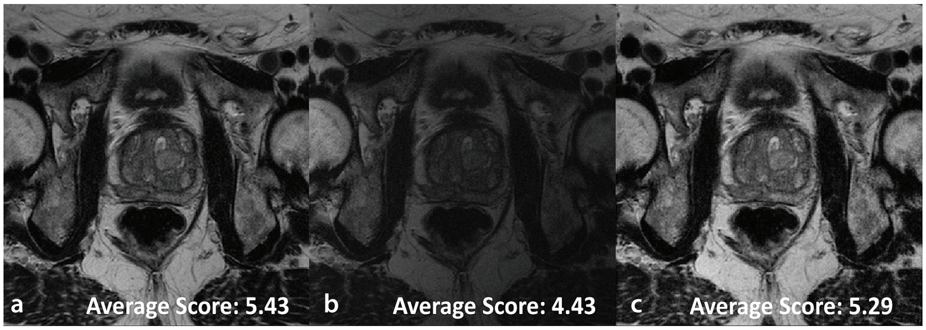


Figure 4. A representative prostate magnetic resonance imaging scan **(a)** original image with the average score of 5.43, **(b)** synthetically degraded image with the average score of 4.43, and **(c)** a synthetically improved image with the average score of 5.29.

Discussion

We found that the Pix2Pix model significantly improved the quality of synthetically degraded images evidenced by quantitative metrics and assessments of multiple readers with different experience levels from different institutions following the criteria adopted from the PI-QUAL. Notably, the synthetically improved images showed no statistical difference in image quality compared with the original images.

We further tested the performance of the proposed model on an external testing dataset, where it substantially increased the image quality. This demonstrates that the model not only works across different datasets but is also effective in improving image quality for original images that have not been synthetically manipulated. This finding is promising as it suggests that DL models can be trained on available datasets without the need for actual poor-quality prostate MRIs. It is important to note that the PI-CAI dataset is derived from centers in the Netherlands and Norway. This geographical restriction could limit the generalizability of our findings to other populations. Future studies should include data from more diverse geographical regions.

Another important finding was that the readers were not unambiguously able to discriminate original from synthetically improved images regardless of their experience levels, showing the proposed model did not only improve image quality but was also able to generate realistic looking images without introducing over-smoothness or plastic appearance.

Our findings diverge from those of the study by Belue et al.[10], where the authors observed no qualitative improvement and the readers mostly opted for original images over synthetically improved images evidenced by expert radiologists. Belue et al.[10] utilized a Cycle-GAN model and tested it using paired original images of both poor and good quality from the same patients. Moreover, they employed bespoke qualitative criteria, which they acknowledged as a significant limitation of their study.[10] We propose that by systematically incorporating a variety of artifacts, our model may better learn the representations of both poor- and good-quality images, thereby effectively transforming poor-quality images into good-quality ones in a realistic manner.

The tendency of DL methods in over-smoothing diagnostic images has also been documented in studies using

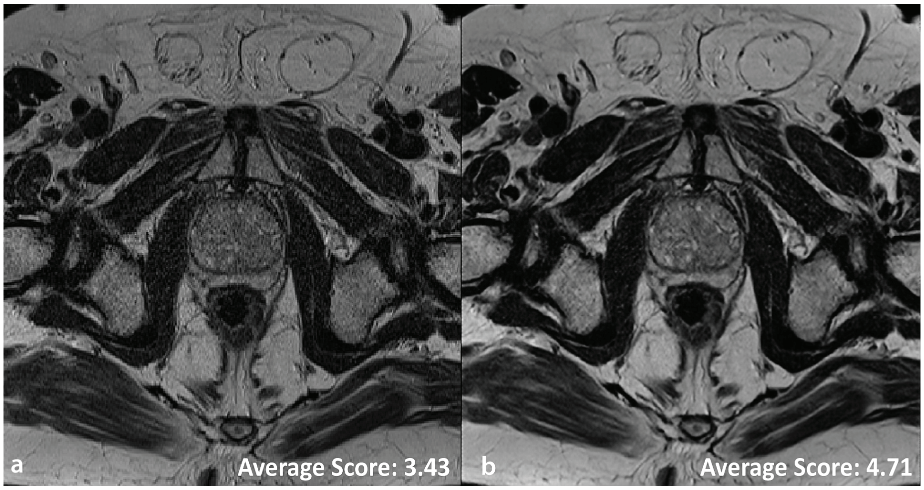


Figure 5. A representative prostate magnetic resonance imaging scan from the in-house external testing dataset, (**a**) an original image with an average radiologists' assessment score of 3.43, and (**b**) a synthetically improved image with an average radiologists' assessment score of 4.71.

DL-based reconstruction methods.[18] This smoothness can cause radiologists to feel uncertain about their interpretations, fearing potential loss of diagnostic information, such as the disruption of lesion appearance or visibility.[18,19] In contrast, our Pix2Pix model, trained on a meticulously prepared dataset, successfully generated realistic images, addressing these concerns by maintaining critical image details essential for accurate diagnosis. The training data included various levels of corruption for each augmentation as well as a combination of these augmentations with the corresponding good quality data. Including a combination of ghosting, spike artifacts, and bias field inhomogeneities with general Gaussian blur and noise in the training regime of the Pix2Pix model increased the robustness of our model against over-smoothing. However, our study did not explicitly evaluate the impact of image enhancement on lesion detection or characterization, which represented an essential area for future investigation.

In reflecting on the methods and results of our study, particularly in terms of experts identifying whether the images were original or synthetic, it is crucial to acknowledge the potential impact of bias. To minimize bias, we did not show the readers both the original and synthetic images simultaneously. Instead, the images were presented in a random order, and the readers were asked to determine their authenticity. A potential limitation is that readers one and three were from the same institution. Although this could introduce bias, the inclusion of readers from other centers helped mitigate this potential issue. Future work could incorporate strategies such as stratified sampling based on institutional affiliation to further address this. Intriguingly, the results suggested that the readers were essentially guessing, indicating no clear distinction between the original and synthetically improved images. However, this design may have inadvertently introduced another form of bias.

Knowing the study's objective—to assess the realism of synthetically generated images—likely predisposed the readers to scrutinize each image more critically. This awareness could have heightened sensitivity to any minor imperfections, predisposing the readers to identify these as indicators of synthetic origin. Admittedly, it is virtually impossible to completely isolate this information from the readers since the core of our evaluation involved discerning the nature of the images, thus directly revealing the study's design.

We openly acknowledge that the design of our study might have influenced the readers' judgments. Recognizing this does not diminish the validity of our findings but rather enhances the transparency and integrity of our analysis. This situation underscores the need for further research to quantify and adjust for such bias, ensuring that the conclusions drawn are robust and applicable in real-world diagnostic settings. This will help in developing methodologies that better emulate the blind assessments typically conducted in clinical practice.

Several other limitations to our study warrant acknowledgment. First, our model was limited to axial T2W images and excluded other crucial sequences. Future studies could explore enhancing image quality across all sequences and integrating them into a single DL pipeline for more effective improvements.[20] Although our study employed PI-QUAL V1, we acknowledge that V2.0 was released during our study period. Future studies should utilize the updated version for assessment.

Second, we used Pix2Pix due to its established use and relatively lower computational demands compared with the state-of-the-art diffusion denoising probabilistic models, which required significantly more resources. Future work will include applying advanced architectures, including transformers and diffusion models for image enhancement.

In conclusion, we demonstrated that a GAN model, Pix2Pix, trained on synthetically degraded axial T2W prostate MRI, can substantially improve image quality as evidenced by quantitative metrics and assessments from multiple readers with varying levels of experience following PI-QUAL criteria, showing no statistical difference in image quality compared with the original images. Additionally, the readers were unable to distinguish between original and synthetic images, indicating that the model did not introduce any unnatural appearance. Furthermore, the same model was able to improve image quality in an external testing dataset of original images, demonstrating its generalizability across datasets and its capability to improve both original and synthetically degraded images.

Acknowledgements

This paper was produced with support from the 1001 Science and Technology Grant Program National Program of TUBITAK (Project No: 122E022). However, the entire responsibility for the publication rests with the author. The financial support received from TUBITAK does not imply that the content of the publication is scientifically endorsed by TUBITAK.

Footnotes

Conflict of Interest

Deniz Alis is the CEO and co-founder of Hevi AI Health Tech. None of Hevi AI's products were used or mentioned in the current work. Furthermore, this paper did not use any commercially available DL software. Other authors have nothing to disclose.

References

1. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging

reporting and data system version 2. *Eur Urol*. 2019;76(3):340-351. [CrossRef]

2. Esses SJ, Taneja SS, Rosenkrantz AB. Imaging facilities' adherence to PI-RADS v2 minimum technical standards for the performance of prostate MRI. *Academic Radiology*. 2018;25(2):188-195. [CrossRef]
3. Burn PR, Freeman SJ, Andreou A, Burns-Cox N, Persad R, Barrett T. A multicentre assessment of prostate MRI quality and compliance with UK and international standards. *Clinical Radiology*. 2019;74(11):894. [CrossRef]
4. Sackett J, Shih JH, Reese SE, et al. Quality of Prostate MRI: Is the PI-RADS Standard Sufficient? *Acad Radiol*. 2021;28(2):199-207. [CrossRef]
5. Ueda T, Ohno Y, Yamamoto K, et al. Deep learning reconstruction of diffusion-weighted MRI improves image quality for prostatic imaging. *Radiology*. 2022;303(2):373-381. [CrossRef]
6. Gassenmaier S, Afat S, Nickel D, Mostapha M, Herrmann J, Othman AE. Deep learning-accelerated T2-weighted imaging of the prostate: reduction of acquisition time and improvement of image quality. *Eur J Radiol*. 2021;137:109600. [CrossRef]
7. Wang X, Ma J, Bhosale P, et al. Novel deep learning-based noise reduction technique for prostate magnetic resonance imaging. *Abdom Radiol*. 2021;46(7):3378-3386. [CrossRef]
8. Lebel RM. Performance characterization of a novel deep learning-based MR image reconstruction pipeline. 2020. [CrossRef]
9. Lin Y, Yilmaz EC, Belue MJ, Turkbey B. Prostate MRI and image quality: it is time to take stock. *Eur J Radiol*. 2023;161:110757. [CrossRef]
10. Belue MJ, Harmon SA, Masoudi S, et al. Quality of T2-weighted MRI re-acquisition versus deep learning GAN image reconstruction: a multi-reader study. *Eur J Radiol*. 2024;170:111259. [CrossRef]
11. Saha A, Twilt JJ, Bosma JS, et al. The PI-CAI challenge: public training and development dataset. Published online May 5, 2022. [CrossRef]
12. Giganti F, Allen C, Emberton M, Moore CM, Kasivisvanathan V, PRECISION study group. Prostate imaging quality (PI-QUAL): a new quality control scoring system for multiparametric magnetic resonance imaging of the prostate from the PRECISION trial. *Eur Urol Oncol*. 2020;3(5):615-619. [CrossRef]
13. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs Biomed*. 2021;208:106236. [CrossRef]
14. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer Assisted Intervention*; 2015;234-241. [CrossRef]
15. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. [CrossRef]
16. de Rooij M, Israël B, Tummers M, et al. ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *Eur Radiol*. 2020;30(10):5404-5416. [CrossRef]
17. Giganti F, Dinneen E, Kasivisvanathan V, et al. Inter-reader agreement of the PI-QUAL score for prostate MRI quality in the NeuroSAFE PROOF trial. *Eur Radiol*. 2022;32(2):879-889. [CrossRef]
18. Lee KL, Kessler DA, Dezonie S, et al. Assessment of deep learning-based reconstruction on T2-weighted and diffusion-weighted prostate MRI image quality. *Eur J Radiol*. 2023;166:111017. [CrossRef]
19. Barrett T, Lee KL, de Rooij M, Giganti F. Update on Optimization of Prostate MR Imaging Technique and Image Quality. *Radiologic Clinics of North America*. 2024;62(1):1-15. [CrossRef]
20. Karagoz A, Seker ME, Yergin M, et al. Prostate Lesion estimation using prostate masks from biparametric MRI. *arXiv*. Published online January 11, 2023. [CrossRef]

Supplementary Document S1

1. data pre-processing: creating a robust training set for prostate MRI enhancement

A crucial step in this study involved the creation of a robust training dataset comprised of T2W images realistically mimicking various MRI artifacts. This approach, utilizing the TorchIO library for medical image augmentation, aimed to enhance the robustness and generalizability of the trained model by exposing it to a wide range of image degradations commonly encountered in clinical practice.[1]

1.1. Motion artifacts: simulating the impact of patient movement

Motion artifacts leads to blurring or ghosting that can obscure anatomical details.[2] To make our deep learning model more robust to these artifacts, we simulated realistic motion during the MRI scan. This simulation involves randomly generating a series of small movements, representing the kind of positional shifts a patient might make.

Each simulated movement is mathematically represented as a 3D transformation. These transformations include both rotation (turning) and translation (shifting) components, and their parameters are randomly varied to create a diverse range of plausible motions. To ensure that these simulated movements don't unrealistically displace the prostate from its average position within the image, each movement is adjusted using a "de-meaning" process. This process ensures that the simulated motion primarily degrades image quality through blurring or ghosting, rather than causing a significant shift in the prostate's overall location.

The simulated movements are then applied to the original image data, which is resampled to maintain smooth image features despite the introduced motion. Finally, the motion-corrupted image is synthesized by processing the image data in k-space, the frequency domain representation of the MRI signal. This process incorporates the temporal dynamics of the simulated movements, resulting in a realistic depiction of a motion-affected T2-weighted prostate MRI.

1.2. Gaussian noise: simulating inherent acquisition noise

Inherent noise is an unavoidable part of MRI acquisition, creating random fluctuations in signal intensity that can obscure subtle details within prostate images. To make our synthetically degraded prostate images more realistic, we incorporated random Gaussian noise, simulating this inherent noise.

Instead of using fixed noise levels, we varied the amount of noise added to each image. This mimics the range of noise levels encountered in real-world prostate MRI scans. For each image, the parameters controlling the noise distribution were randomly selected, ensuring our model is exposed to a variety of noise profiles during training.

This random noise is added to each individual voxel within the image. The amount of noise added is determined by scaling a random value with a randomly chosen standard deviation and shifting it by a randomly selected mean. This process creates realistic noise patterns, reflecting the noise characteristics inherent to real-world prostate MRI.

1.3. Gaussian blur: simulating loss of sharpness in prostate MRI

Blurring is often caused by factors like imperfect scanner focus or slight patient movements. This loss of sharpness can make it difficult to see fine anatomical details, potentially affecting diagnosis. To prepare our model for this real-world challenge, we introduced Gaussian blur into our synthetic image degradation process.

We didn't apply the same amount of blurring to each image. Instead, we randomly varied the degree of blurring, mimicking the range of sharpness variations seen in real prostate MRIs. This exposes our model to a wider range of blurring artifacts during training, making it more robust to blurry images in real-world scenarios.

The blurring is implemented by convolving each image with a Gaussian filter. The size of this filter, which controls how much blurring is applied, is randomly chosen and scaled based on the resolution of each image. This ensures the blurring effect is appropriately applied relative to the size of the details in the prostate image.

1.4. Ghosting artifacts: simulating periodic motion effects

Ghosting artifacts are often caused by rhythmic motions like blood flow or bowel movement. These artifacts appear as faint copies or "ghosts" of anatomical structures, shifted along a specific direction in the image. To simulate ghosting artifacts, we manipulate the image data in its frequency domain representation, known as k-space.

In k-space, periodic motions like those causing ghosting affect specific frequency bands. We simulate this by selectively suppressing the strength of certain frequencies in k-space. The amount of suppression controls the intensity of the ghosting effect, while the spacing between the suppressed frequencies determines how often the ghosting pattern repeats. By adjusting these parameters, we can create a wide range of ghosting artifacts with varying appearances.

Once the k-space data has been modified to include the simulated ghosting, we transform the data back to its original spatial representation, resulting in a prostate MRI image containing realistic ghosting artifacts.

1.5. Spike artifacts: simulating radio-frequency interference

Spike artifacts, often called herringbone or corduroy artifacts, can create unwanted stripes in MRI images. These artifacts are caused by radio-frequency interference during the scan, which introduces spikes in the k-space representation of the image.

To simulate these artifacts, we directly add spikes to the k-space data of the prostate MRI. Each spike's location in k-space determines the direction and frequency of the stripe pattern that will appear in the final image.

We randomize both the number and location of these spikes to simulate the unpredictable nature of real-world spike artifacts. The number of spikes controls how severe the artifact is, while their random locations create stripes in various directions and positions within the image. The intensity of each spike is also randomized, resulting in stripes with varying prominence. After adding the spikes to the k-space data, we transform the data back to its normal spatial representation, creating a prostate MRI image with realistic spike artifacts.

1.6. Bias field inhomogeneities: simulating magnetic field imperfections

Bias field inhomogeneity is a common artifact in MRI, caused by imperfections in the scanner's magnetic field.[3] This artifact creates gradual changes in image brightness across the prostate, making some areas appear brighter or darker than others, even if the tissues are the same. This can make it harder to distinguish between different tissues and interpret the image accurately.

We simulated this artifact using a mathematical model based on polynomials. This model can create smooth, gradual changes in image brightness similar to those seen in real bias field artifacts. The model uses a set of coefficients to control the intensity variations, and by randomly generating these coefficients, we create a variety of bias field patterns.

The complexity of these patterns can be adjusted by changing the order of the polynomial used in the model. A higher-order polynomial allows for more intricate and spatially varying intensity changes. The generated bias field pattern is then applied to the original prostate MRI image, simulating the effect of magnetic field imperfections.

1.7. Combining noise types: enhancing training realism

To create a more challenging and realistic training scenario, each image in the training and validation sets was augmented with every type of synthetic degradation individually (motion, Gaussian noise, blur, ghosting, spike artifacts, bias field inhomogeneity). This ensured the model learned to handle each artifact in isolation.

In addition to individual augmentations, combined noise augmentations were also applied to enhance the model's robustness and generalizability. For 20% of the images in the training and validation sets, we randomly selected a combination of two or more of the aforementioned noise types and applied them together to the same image. This approach aimed to simulate the more complex and diverse noise profiles that are representative of real-world clinical scenarios where multiple artifacts can co-occur. The specific combinations of noise types were randomly selected for each image, ensuring a wide variety of combined degradations within the training data.

For the test set, to evaluate the model's ability to generalize to unseen combinations of artifacts, these noise types were combined randomly with varying ratios. This rigorous testing procedure helped assess the model's performance under conditions that more closely reflect real-world clinical prostate MRI.

2. The Pix2Pix model: image-to-image translation for prostate MRI enhancement

To address the challenge of improving the quality of degraded T2-weighted prostate MRI images, we employed the Pix2Pix model, a conditional Generative Adversarial Network (cGAN) renowned for its efficacy in image-to-image translation tasks.[4] Unlike Cycle-GAN, which relies on unpaired data from two different domains, Pix2Pix leverages paired images for training. This makes it particularly suitable for our study, where we have access to ground truth data in the form of original, high-quality images corresponding to the synthetically degraded images.

The Pix2Pix architecture comprises two key components: a generator and a discriminator, which are simultaneously trained in an adversarial manner. The generator, structured as a U-Net,[5] takes the degraded T2-weighted image as input and endeavors to generate a high-quality image that closely resembles the original, artifact-free image. The U-Net architecture, with its encoder-decoder structure and skip connections, is particularly advantageous in medical image processing. The encoder progressively downsamples the input image, extracting features at multiple scales, while the decoder upsamples the encoded representation, reconstructing the output image. The skip connections, linking corresponding encoder and decoder layers, facilitate the direct flow of low-level information across the network, preserving crucial anatomical details and preventing excessive blurring often associated with traditional encoder-decoder networks.

The discriminator, on the other hand, employs a PatchGAN classifier,[4] which evaluates individual N x N image patches rather than the entire image at once. This patch-based approach encourages the generator to focus on generating realistic high-frequency details crucial for maintaining the clarity and realism of medical images. During training, the generator and discriminator are engaged in a continuous adversarial loop. The generator strives to produce increasingly realistic images to deceive the discriminator, while the discriminator becomes more adept at discerning real images from the synthetic images generated by the generator. The proposed Pix2Pix architecture is given in Figure 1.

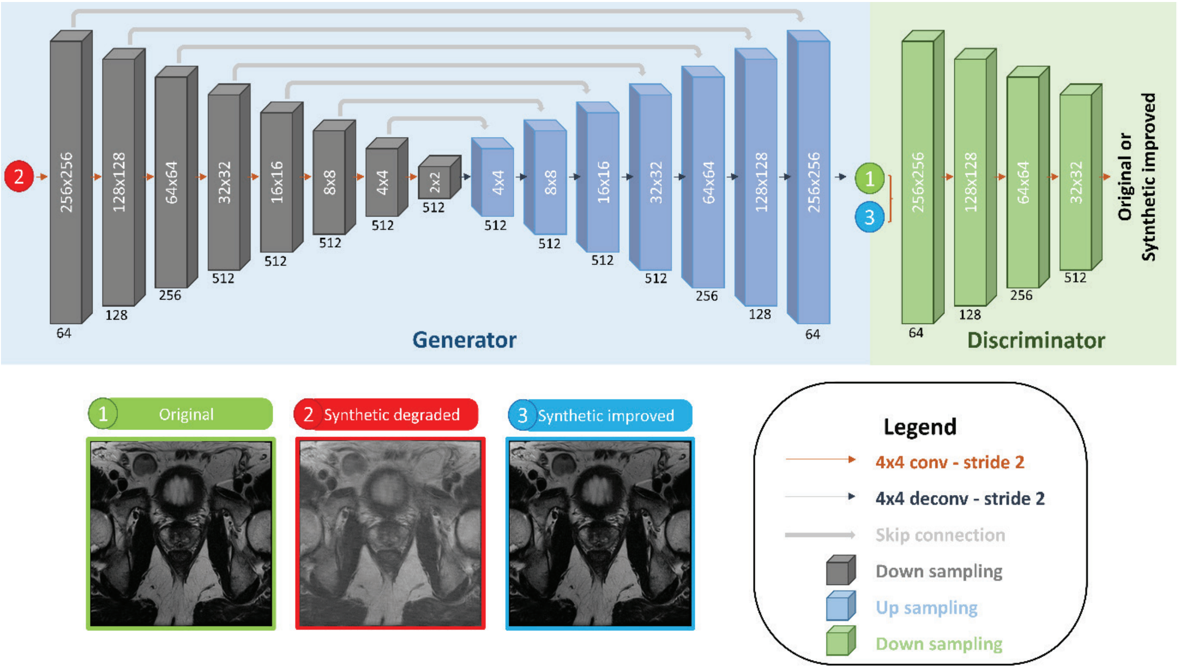


Figure 1. The proposed Pix2Pix architecture in current work.

A U-Net architecture represented by gray and blue blocks generate a synthetically improved image in generator. PatchGAN architecture shown in green blocks evaluate image patches to differentiate between original and synthetic images in discriminator and promotes high-frequency detail generation in the synthetic images. The image examples at the bottom of Figure 1. showcase (1) the original image, (2) the synthetically degraded image, and (3) the synthetically improved image respectively. Moreover, the numbers on the blocks denote the number of feature maps at each layer, and the image dimensions are displayed above each block. Each convolution operations, skip connections, down and up sampling operations are attended to corresponding arrows and blocks with specified colors under legend section in Figure 1.

This adversarial process is guided by a combined loss function encompassing both an adversarial loss and an L1 loss. The adversarial loss, determined by the discriminator's ability to classify image patches as real or fake, ensures that the generated images are visually indistinguishable from the real images in the training dataset. The L1 loss, calculated as the mean absolute error between the generated image and the ground truth target image, promotes structural similarity, preventing excessive blurring and preserving the anatomical integrity crucial for accurate diagnosis.

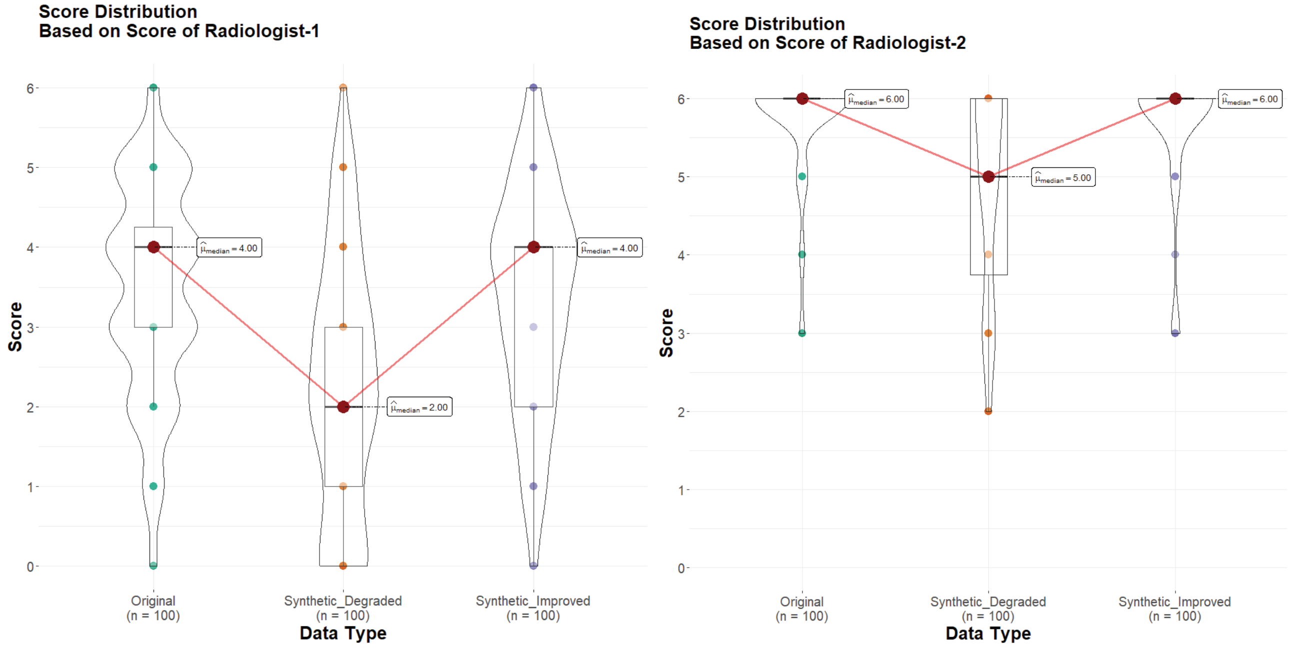
We optimized the Pix2Pix model using the Adam optimizer with a learning rate of 0.0002 for both the generator and the discriminator. The model was trained for 200 epochs, with a heavier emphasis placed on the L1 loss during the early stages of training. This prioritizes the generation of structurally accurate images over purely visually realistic ones, particularly crucial in the context of medical imaging, where diagnostic accuracy is paramount.

References

1. Pérez-García F, Sparks R, Ourselin S. TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed. 2021;208:106236. [CrossRef]
2. Shaw R, Sudre C, Ourselin S, Cardoso MJ. MRI k-Space motion artefact augmentation: model robustness and task-specific uncertainty, in: Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning. PMLR. 2019:427-436. Accessed July 1, 2024. [CrossRef]
3. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging. 1999;18(10):885-896. [CrossRef]
4. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2018. [CrossRef]
5. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015:234-241. [CrossRef]

Supplementary Document S2

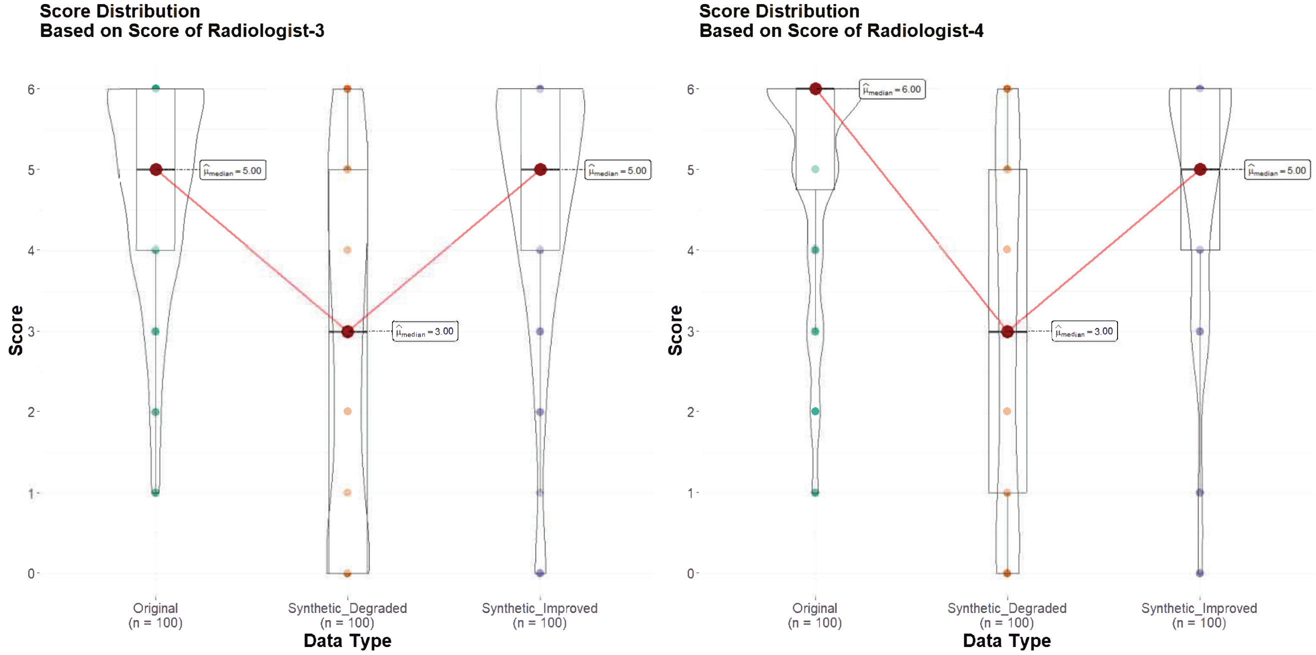
Comparison of radiologists' score for original, synthetic degraded, and synthetic improved prostate MRI images.



Score Distribution Based on Score of Radiologist-1

Score Distribution Based on Score of Radiologist-2

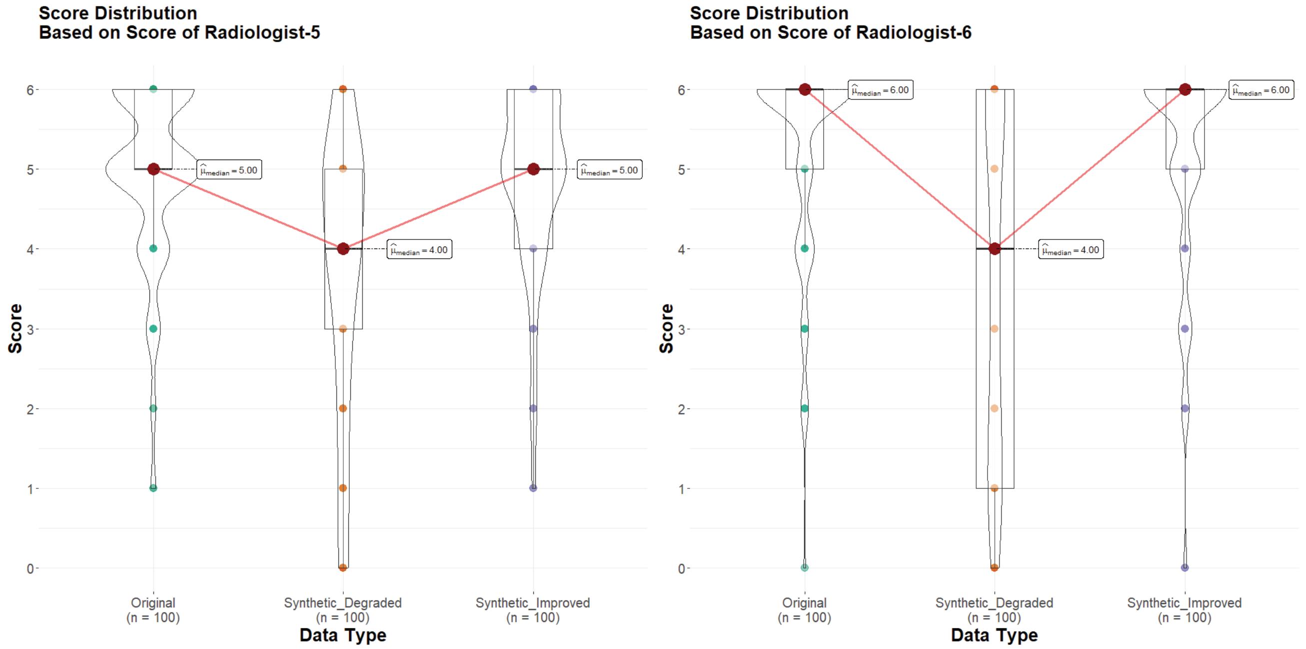
Comparison of radiologists' score for original, synthetic degraded, and synthetic improved prostate MRI images.



Score Distribution Based on Score of Radiologist-3

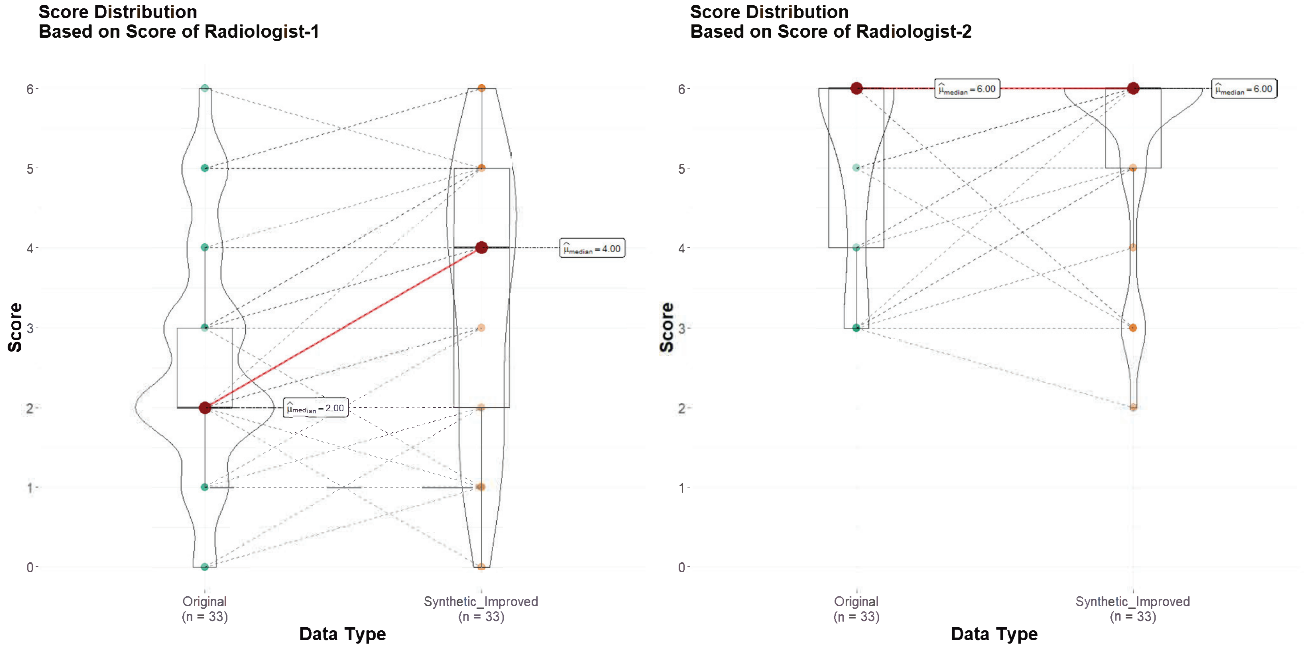
Score Distribution Based on Score of Radiologist-4

Comparison of radiologists' score for original, synthetic degraded, and synthetic improved prostate MRI images.

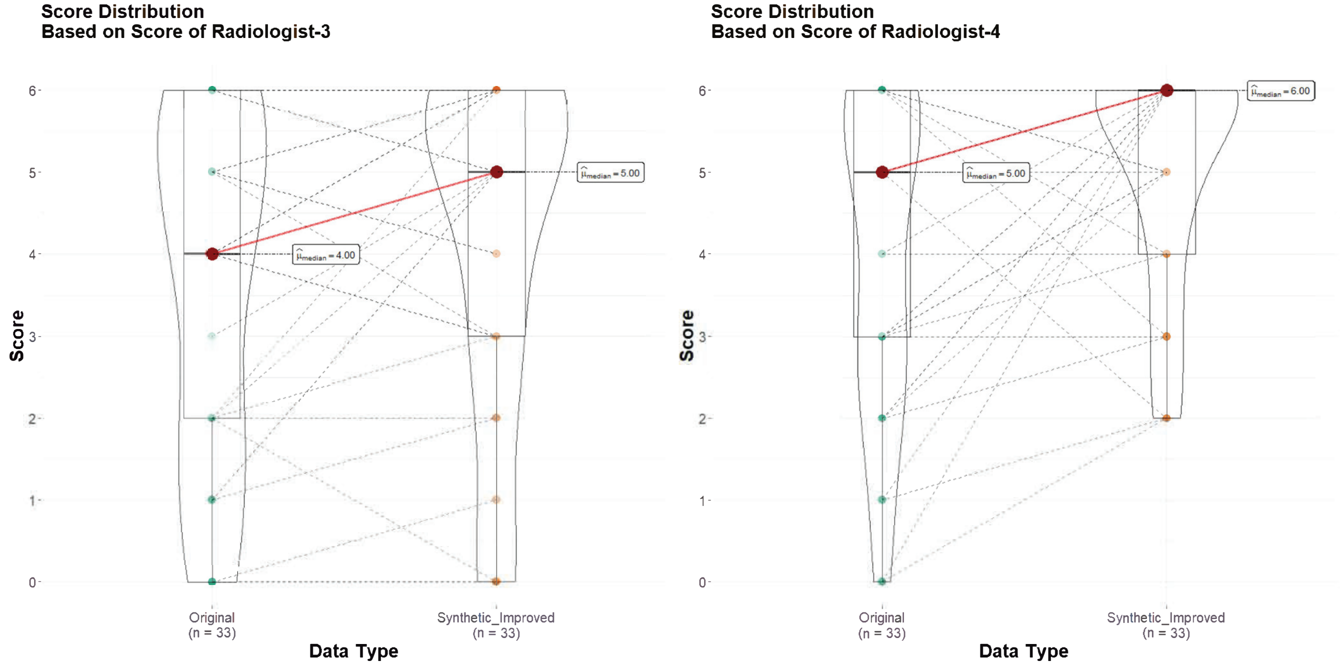


Comparison of radiologists' score for original, synthetic degraded, and synthetic improved prostate MRI images.

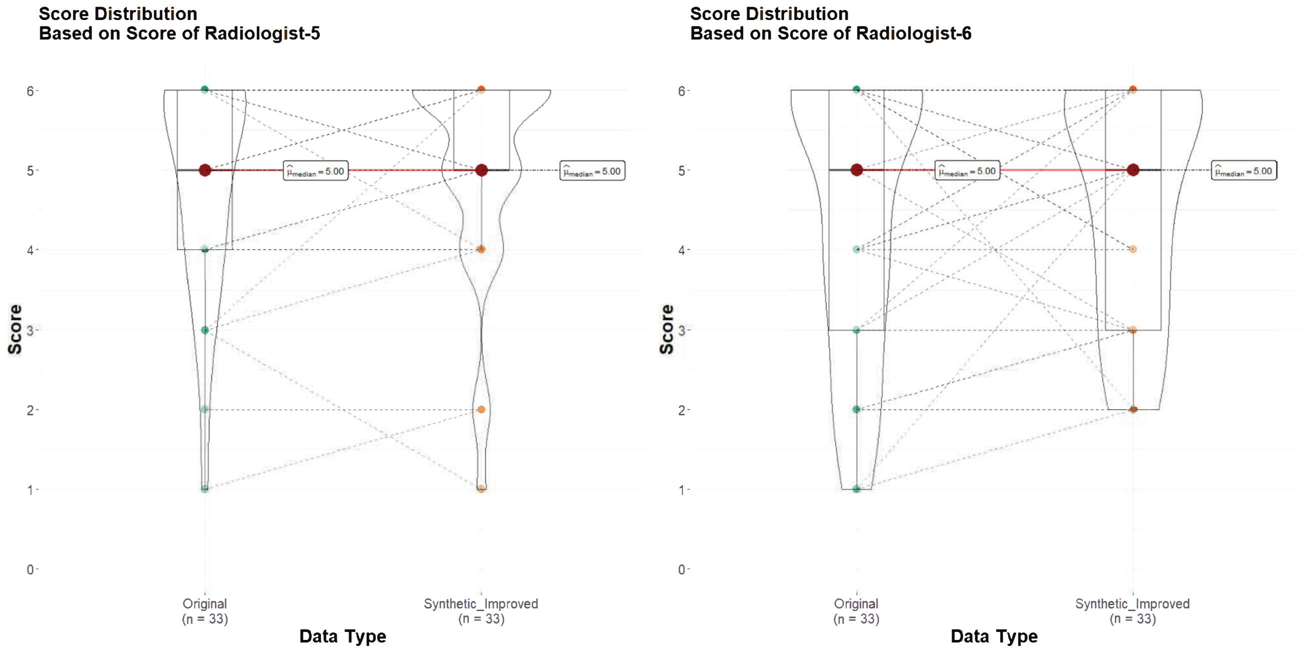
Comparison of radiologists' score for original and synthetic improved prostate MRI images from external test set.



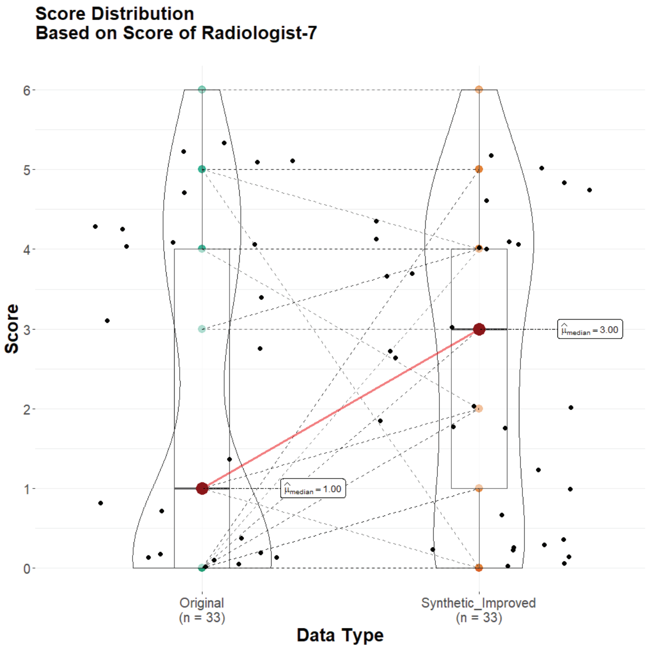
Comparison of radiologists' score for original and synthetic improved prostate MRI images from external test set.



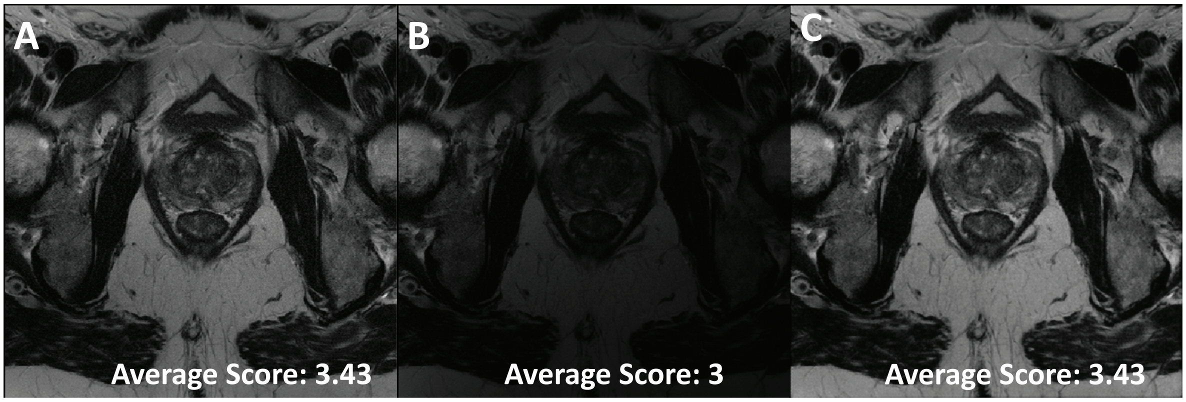
Comparison of radiologists' score for original and synthetic improved prostate MRI images from external test set.



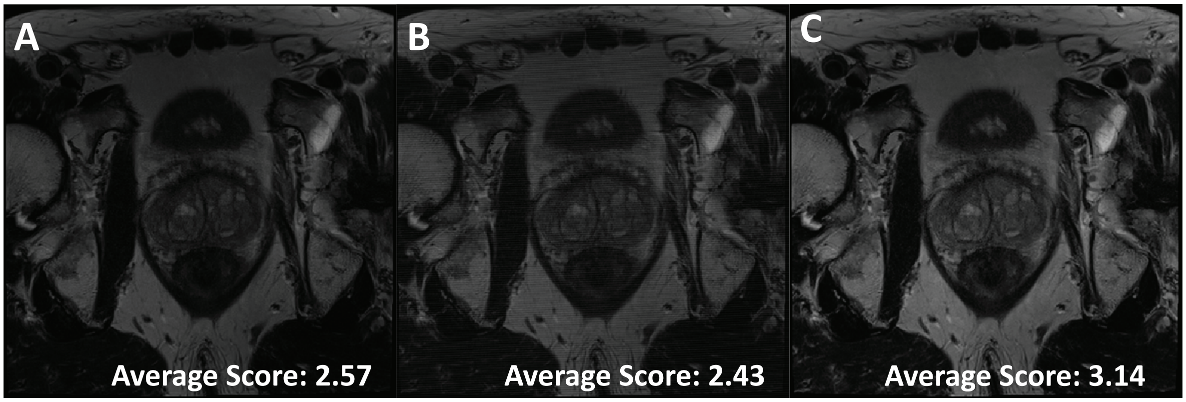
Comparison of radiologists' score for original and synthetic improved prostate MRI images from external test set.



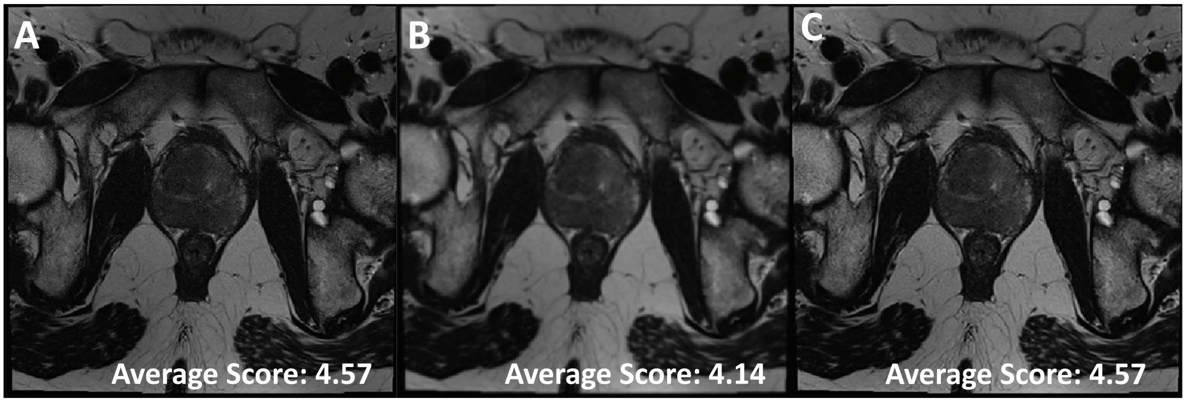
A representative prostate MRI scan (A) original image, (B) synthetic degraded image, (C) synthetic improved image.



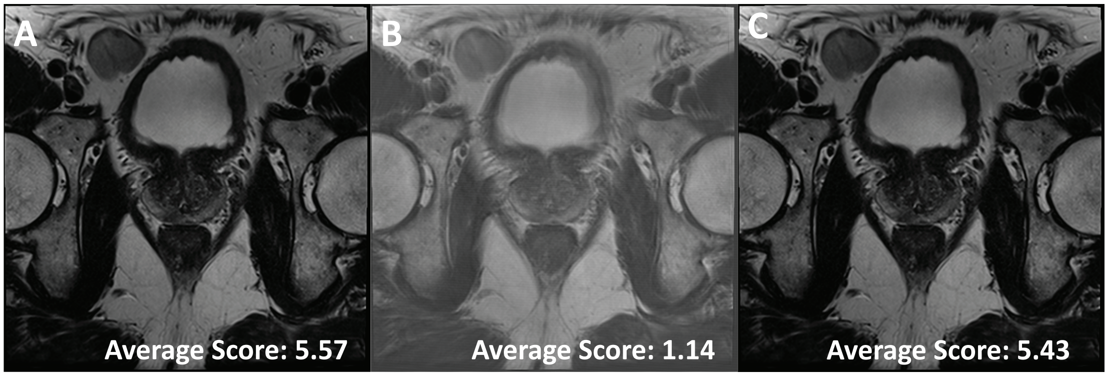
A representative prostate MRI scan (A) original image, (B) synthetic degraded image, (C) synthetic improved image.



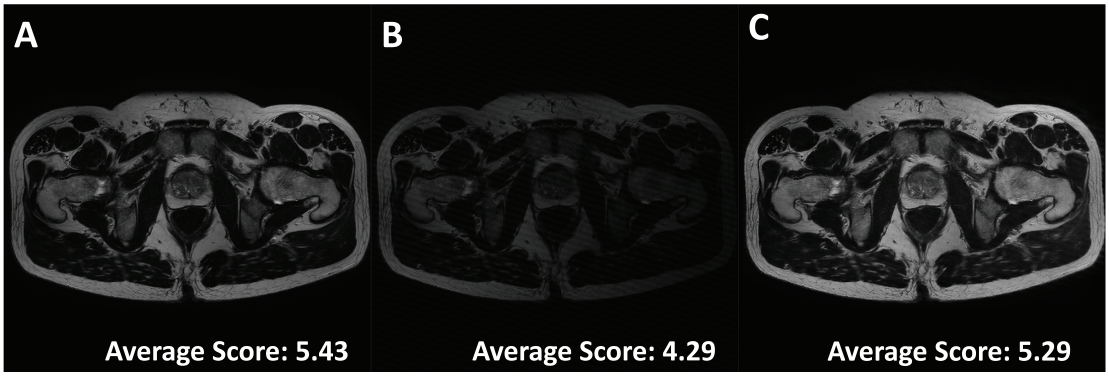
A representative prostate MRI scan (A) original image, (B) synthetic degraded image, (C) synthetic improved image.



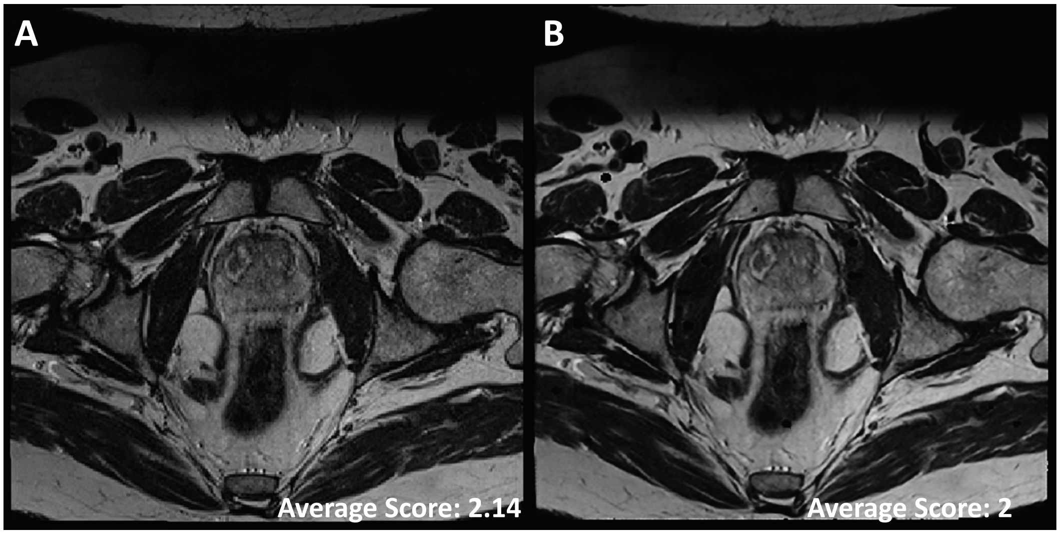
A representative prostate MRI scan (A) original image, (B) synthetic degraded image, (C) synthetic improved image.



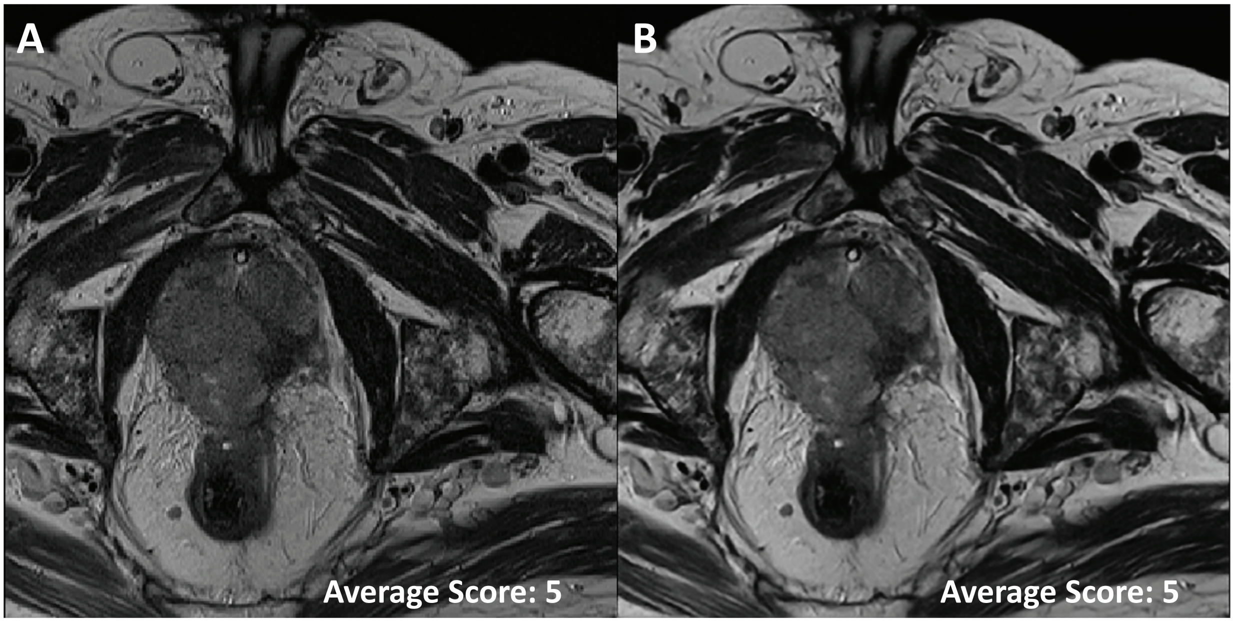
A representative prostate MRI scan (A) original image, (B) synthetic degraded image, (C) synthetic improved image.



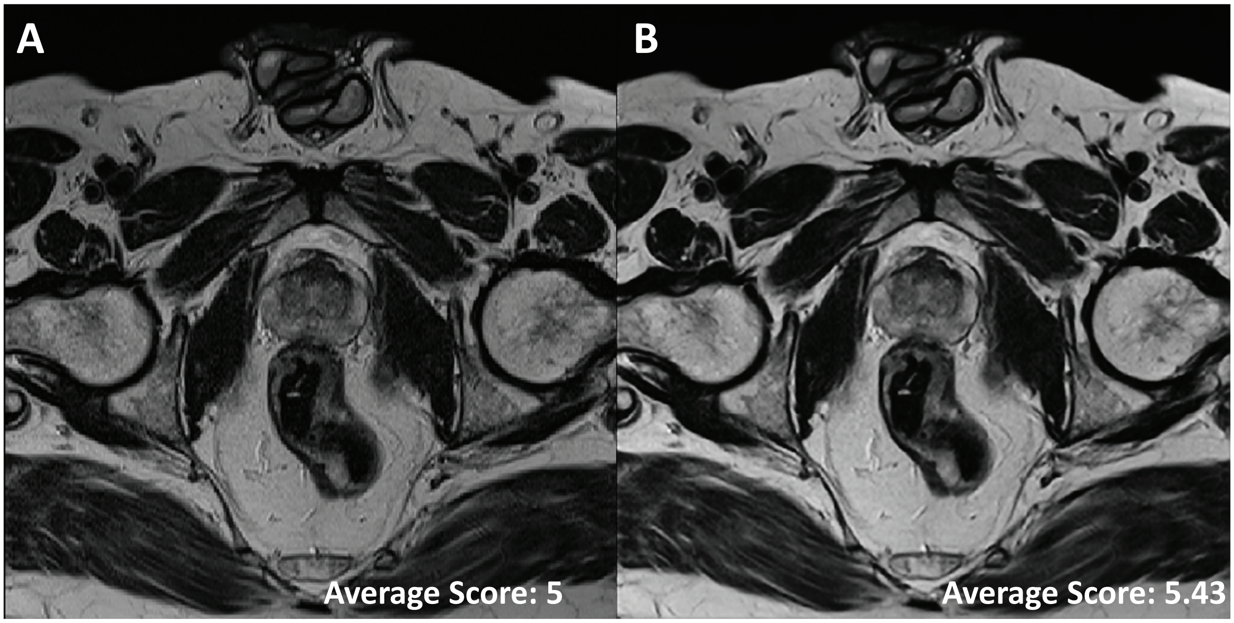
A representative prostate MRI scan from external test set (A) original image, (B) synthetic improved image.



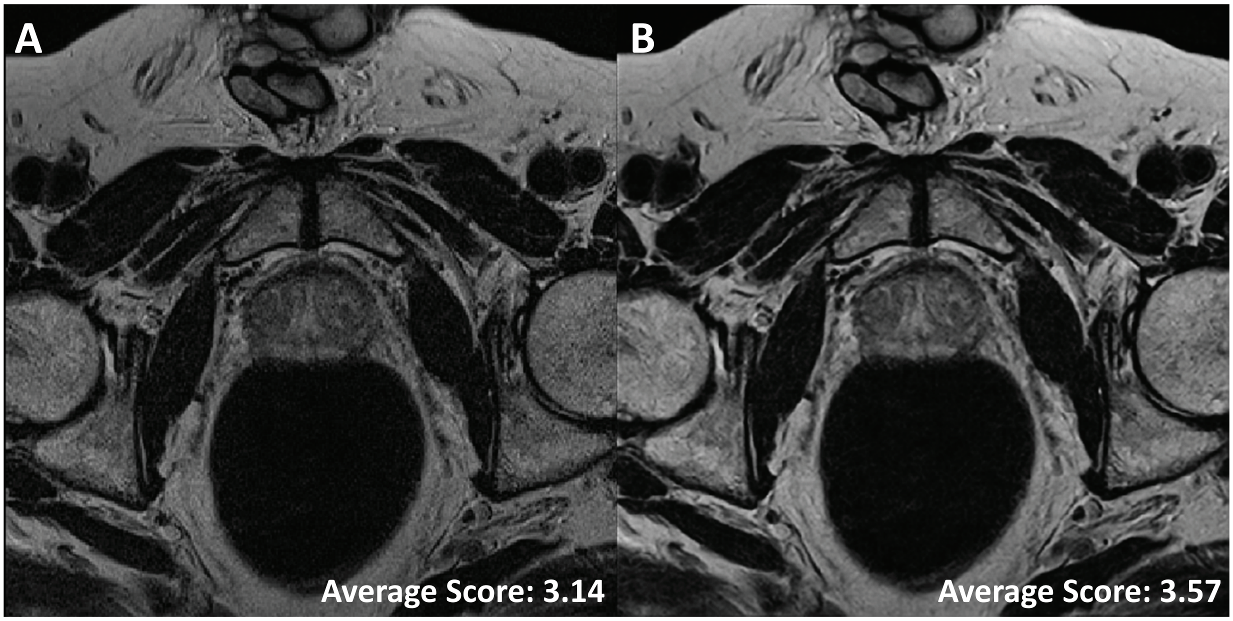
A representative prostate MRI scan from external test set (A) original image, (B) synthetic improved image.



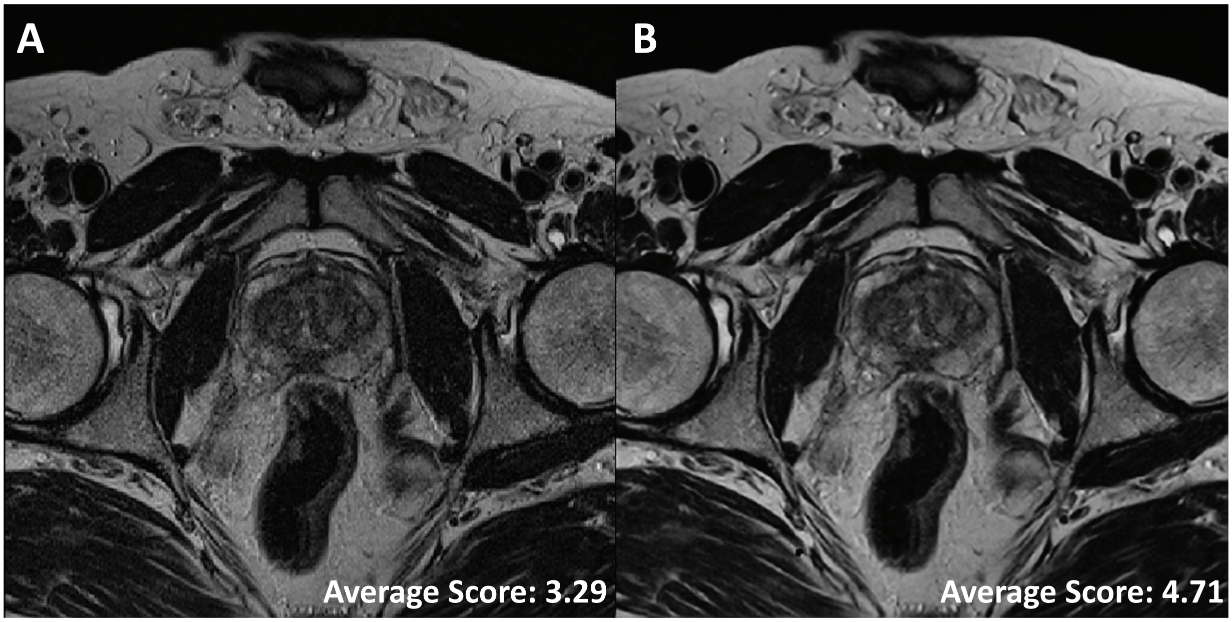
A representative prostate MRI scan from external test set (A) original image, (B) synthetic improved image.



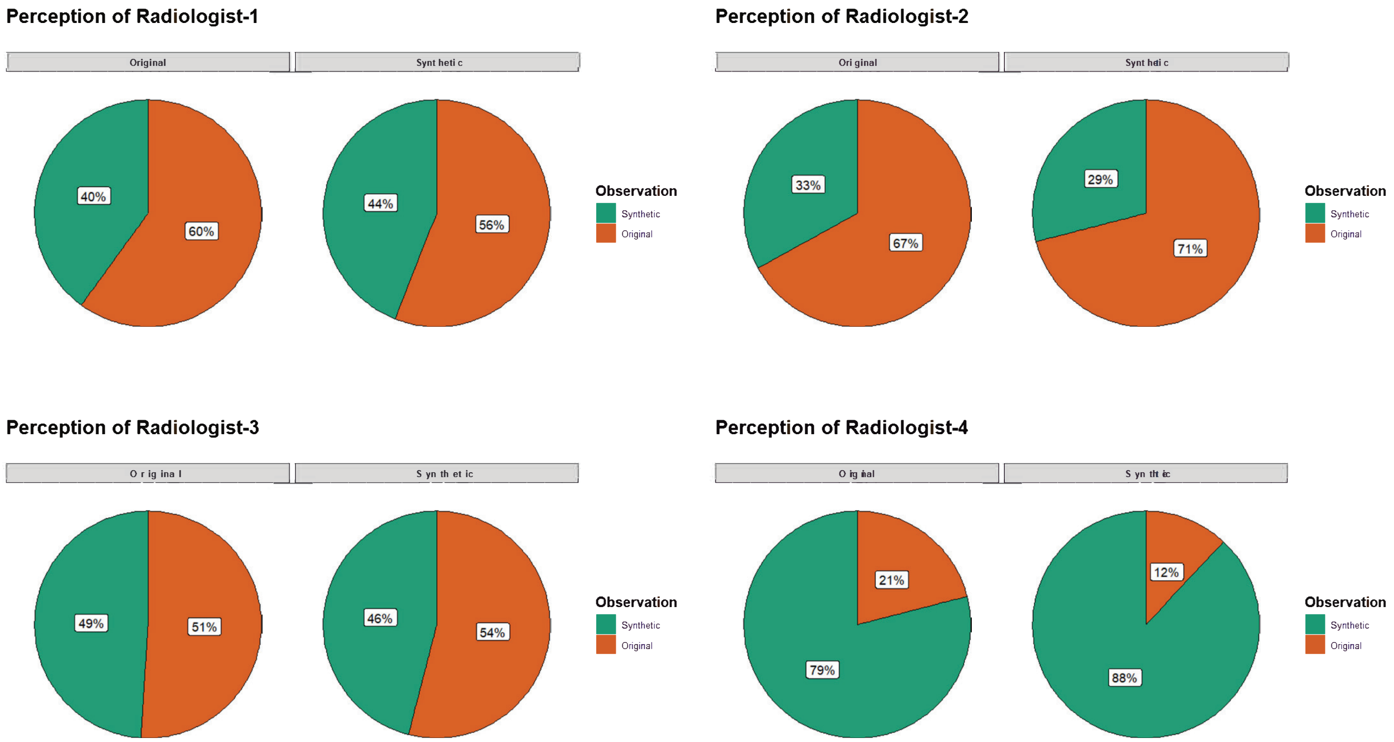
A representative prostate MRI scan from external test set (A) original image, (B) synthetic improved image.



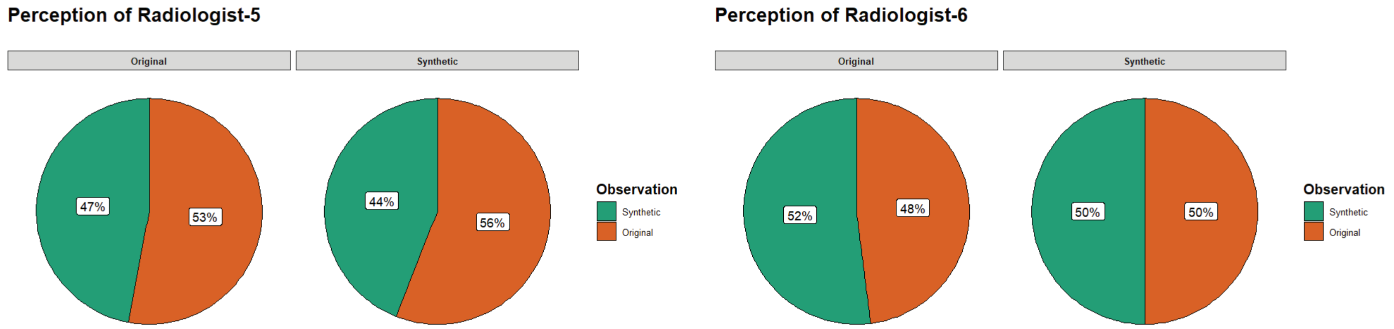
A representative prostate MRI scan from external test set (A) original image, (B) synthetic improved image.



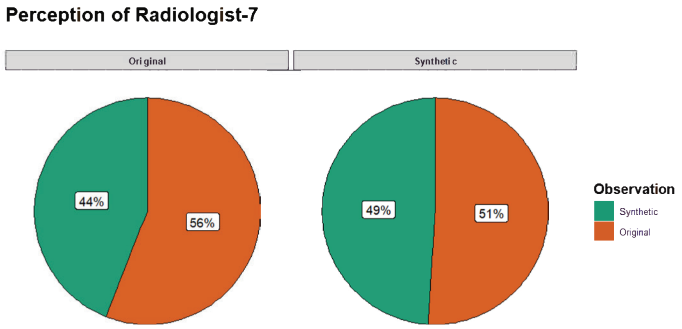
Comparison of radiologists' assessment on whether the image is original or synthetic.



Comparison of radiologists' assessment on whether the image is original or synthetic.



Comparison of radiologists' assessment on whether the image is original or synthetic.



Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253272

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR

Retrieval-augmented generation for answering Breast Imaging Reporting and Data System (BI-RADS)-related questions with large language models

Esat Kaba

Recep Tayyip Erdoğan University Training and Research Hospital, Department of Radiology, Rize, Türkiye

KEYWORDS
Artificial intelligence, BI-RADS, large language model, questions, retrieval-augmented generation

Dear Editor,

I read with great interest the article titled "Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition" published in Diagnostic and Interventional Radiology.[1] The study explores how large language models (LLMs) respond to multiple-choice and some image-based questions based on the Breast Imaging Reporting and Data System (BI-RADS) 5th edition and presents the impressive results achieved by these models. Research of this kind is crucial to understanding the growing potential role of artificial intelligence technologies, particularly LLMs, in radiology decision-making processes. As a contribution to the valuable findings of this study, I believe that considering the retrieval-augmented generation (RAG) approach could be beneficial for more effectively combining information retrieval and text generation in such scenarios.

Retrieval-augmented generation enables language models to address existing knowledge gaps by accessing external information sources, allowing them to generate more accurate, up-to-date, and contextually appropriate text.[2] It consists of two main components: retrieval and generation. In the retrieval phase, queries are converted into vector format (e.g., using OpenAI embeddings) to create text embeddings. These vectors are then compared with pre-indexed documents using similarity search algorithms to retrieve the most relevant content (top-k retrieval). In the generation phase, the retrieved information is added to the input of the LLM, which then generates text based on this context.[3,4] This method holds strong potential, especially in fields that require complex information processing, such as radiology and detailed analyses based on BI-RADS.

In a study highlighting the effectiveness of this method in radiology, Tozuka et al.[5] performed tumor, node, metastasis staging of lung cancer using LLMs with and without RAG. In this study, Google's NotebookLM, a system incorporating RAG, achieved the highest performance in lung cancer staging. GPT-4o was also tested with and without RAG, and the use of RAG resulted in more successful outcomes across all stages of staging.[5] Given the limitations of current static models–such as knowledge gaps and the risk of generating misleading content (hallucinations)–RAG offers a promising approach to mitigate these issues and provide a more practical solution in both radiology education and clinical practice.

I would like to express my gratitude once again for your study's contribution to the field. I believe that incorporating RAG in future research, particularly in studies evaluating the knowledge level of LLMs on radiology guidelines, as in this case, could further enhance model accuracy and reliability.

Corresponding author: Esat Kaba

E-mail: esatkaba04@gmail.com

Received 31 January 2025; accepted 17 February 2025.

Epub: 20.05.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253272

You may cite this article as: Kaba E. Retrieval-augmented generation for answering Breast Imaging Reporting and Data System (BI-RADS)-related questions with large language models. *Diagn Interv Radiol.* 2025;31(6):566-567.

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

References

1. Güneş YC, Cesur T, Çamur E, Günbey Karabekmez L. Evaluating text and visual diagnostic capabilities of large language models on questions related to the breast imaging reporting and data system atlas 5th edition. *Diagn Interv Radiol*. 2025;31(2):111-129. [Crossref]
2. Steybe D, Poxleitner P, Aljohani S, et al. Evaluation of a context-aware chatbot using retrieval-augmented generation for answering clinical questions on medication-related osteonecrosis of the jaw. *J Craniomaxillofac Surg*. 2025;53(4):355-360. [Crossref]
3. Toro S, Anagnostopoulos AV, Bello SM, et al. Dynamic retrieval augmented generation of ontologies using artificial ıntelligence (DRAGON-AI). *J Biomed Semantics*. 2024;15(1):19. [Crossref]
4. Zakka C, Shad R, Chaurasia A, et al. Almanac - Retrieval-Augmented language models for clinical medicine. *NEJM AI*. 2024;1(2):10.1056/aioa2300068. [Crossref]
5. Tozuka R, Johno H, Amakawa A, et al. Application of NotebookLM, a large language model with retrieval-augmented generation, for lung cancer staging. *Jpn J Radiol*. 2025;43(4):706-712. [Crossref]

DIR

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253361

BREAST IMAGING

ORIGINAL ARTICLE

Early prediction of neoadjuvant chemotherapy efficacy among patients with triple-negative breast cancer using an ultrasound-based radiomics nomogram

Min-Jia Lin[1]
Hai-Ling Zha[2]
Man-Qi Zhang[3]
Yu Du[4]
Min Zong[5]
Cui-Ying Li[2]

[1]Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Department of Ultrasound, Taizhou, China

[2]The First Affiliated Hospital of Nanjing Medical University, Department of Ultrasound, Nanjing, China

[3]Affiliated Drum Tower Hospital of Nanjing University Medical School, Department of Ultrasound Medicine, Nanjing, China

[4]Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Department of Ultrasound, Shanghai, China

[5]The First Affiliated Hospital of Nanjing Medical University, Department of Radiology, Nanjing, China

PURPOSE

To develop and validate a radiomics nomogram based on early ultrasound (US) imaging for predicting pathologic complete response (pCR) in patients with triple-negative breast cancer (TNBC) receiving neoadjuvant chemotherapy (NAC).

METHODS

This retrospective study included 328 patients with TNBC treated between September 2019 and January 2024, divided into a training cohort (n = 230) and a validation cohort (n = 98). Clinicopathologic data, US features before NAC, tumor volume reduction (TVR) after two cycles of NAC, and radiomics features were collected. Multiple logistic regression was applied to identify the potential predictors of pCR. The efficacy of the nomogram was evaluated through the receiver operating characteristic, calibration, and decision curve analyses. The study was approved by the ethics committee on February 28, 2024, with approval number 2023-SR-799, and the requirement for informed consent was waived.

RESULTS

Twelve features were selected to construct the radiomics signature (RS). The nomogram, incorporating tumor histologic grade, TVR, and RS, yielded an area under the curve of 0.856 [95% confidence interval (CI), 0.807–0.905] in the training cohort and 0.836 (95% CI, 0.749–0.923) in the validation cohort, outperforming both the clinico-ultrasonic and RS models. The calibration and decision curves confirmed the nomogram's excellent calibration and clinical utility.

CONCLUSION

The nomogram, which includes US characteristics, clinical variables, and radiomics features, exhibited satisfactory performance in predicting NAC efficacy in patients with TNBC.

CLINICAL SIGNIFICANCE

The US-based radiomics nomogram, incorporating histologic grade, TVR, and RS, shows preliminary clinical application potential for predicting NAC efficacy in patients with TNBC.

KEYWORDS

Triple-negative breast cancer, neoadjuvant chemotherapy, pathologic complete response, radiomics, ultrasonography

Corresponding authors: Min Zong, Cui-Ying Li

E-mails: mzong@njmu.edu.cn, lcy_njmu@163.com

Received 20 March 2025; revision requested 01 May 2025; last revision received 14 June 2025; accepted 06 August 2025.

Epub: 22.09.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253361

Breast cancer has become the most prevalent malignancy in women worldwide and the leading cause of cancer-related mortality, according to the latest global cancer statistics from 2022.[1] Triple-negative breast cancer (TNBC), which is characterized by the absence of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 (HER-2), is more likely to involve lymph nodes and has an early tendency toward recurrence.[2,3] Its heterogeneity and lack of specific molecular targets contribute to its higher likelihood of

You may cite this article as: Lin MJ, Zha HL, Zhang MQ, Du Y, Zong M, Li CY. Early prediction of neoadjuvant chemotherapy efficacy among patients with triple-negative breast cancer using an ultrasound-based radiomics nomogram. *Diagn Interv Radiol.* 2025;31(6):568-575.

early recurrence and poorer prognosis compared with other breast cancer subtypes.[4-7]

Neoadjuvant chemotherapy (NAC) has shown promising results for patients with TNBC by decreasing the tumor burden, increasing the breast-conserving rate, de-escalating axillary surgery, and allowing treatment reorientation in patients with tumor progression.[8] Pathologic complete response (pCR) after NAC has become a surrogate marker for both disease-free survival and long-term overall survival.[9,10] However, histological analysis of surgical specimens, the gold standard for response evaluation, is only accessible following the completion of NAC treatment.[11] Thus, evaluating the response to chemotherapy early in the treatment course to more accurately forecast the likelihood of achieving pCR in patients with TNBC could help clinicians modify ineffective treatments or avoid unnecessary treatment escalation.[12] Imaging methods for predicting the response to NAC include ultrasound (US), magnetic resonance imaging (MRI), and positron emission tomography (PET).[13,14] Compared with MRI and PET, US is more cost-effective and suitable for repeated evaluation during NAC, playing a pivotal role in detecting treatment response.[15] Nevertheless, few studies have examined the correlation between pCR in TNBC and US features, highlighting the need for further investigation into US as an early predictor of NAC response.

Radiomics is a new field of computer-aided technology that extracts and analyzes a large number of quantitative medical image features that are difficult for the human eye to detect, applying these features for clinical decision-making to enhance diagnostic accuracy.[16,17] Previous studies have shown that radiomics can be used to detect disease status, evaluate the response to NAC, and provide valuable information on cancer aggressiveness.[18-20] For the assessment and prediction of early NAC response, radiomics nomograms based on MRI are most widely used, with limited studies utilizing US for therapeutic evaluation.[21] We hypothesized that an early radiomics nomogram combining US features, clinical findings, and a radiomics signature (RS) could provide additional information to estimate the pCR in patients with TNBC.

Main points

- A clinico-ultrasonic model combining histologic grade and tumor volume reduction after two cycles of neoadjuvant chemotherapy (NAC) is associated with treatment efficacy in patients with triple-negative breast cancer (TNBC).
- A radiomics signature derived from ultrasound (US) features shows excellent potential for predicting pathologic complete response in patients with TNBC.
- An early US-based radiomics nomogram demonstrates favorable performance in predicting NAC efficacy and may assist clinicians in identifying potential poor pathological responders.

Therefore, this study aimed to develop an early US-based radiomics nomogram to predict pCR after NAC in patients with TNBC.

Methods

Patients in this study

The retrospective study was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University, and the requirement for informed consent was waived. The approval number was 2023-SR-799 and approval date was 28 February 2024. A total of 412 consecutive patients with pathologically confirmed TNBC who underwent NAC at our institution between September 2019 and January 2024 were considered, and 84 were excluded. All patients received eight cycles of NAC (taxane- and anthracycline-based chemotherapy regimens) followed by surgery. The inclusion criteria were (i) biopsy-confirmed TNBC without distant metastasis, (ii) US examination performed prior to NAC, (iii) US examination performed after two cycles of NAC, (iv) availability of complete clinicopathologic and US data, and (v) surgery performed after completion of NAC. The exclusion criteria were (i) a history of other tumors, (ii) incomplete NAC or surgery performed at an external institution, (iii) lack of pre-NAC or post-two-cycle NAC US data, (iv) poor-quality US images, and (v) presence of multiple or non-mass lesions. Ultimately, 328 eligible patients (age range: 20–82 years; mean age: 49.4 years) were included and randomly assigned to training and validation cohorts at a ratio of 7:3. The training cohort consisted of 230 patients (pCR: 82, non-pCR: 148), and the validation cohort consisted of 98 patients (pCR: 32, non-pCR: 66). The study flowchart is presented in Figure 1.

Ultrasound imaging acquisition and analysis

All breast US examinations were performed at two time points: before NAC and after two cycles of NAC, using either MyLab Twice (Esaote S.P.A., Italy) or Samsung RS80A (Samsung Medison Co. Ltd., Seoul, South Korea) equipment. The US features collected included the three largest tumor dimensions (length, width, and height) in transverse and

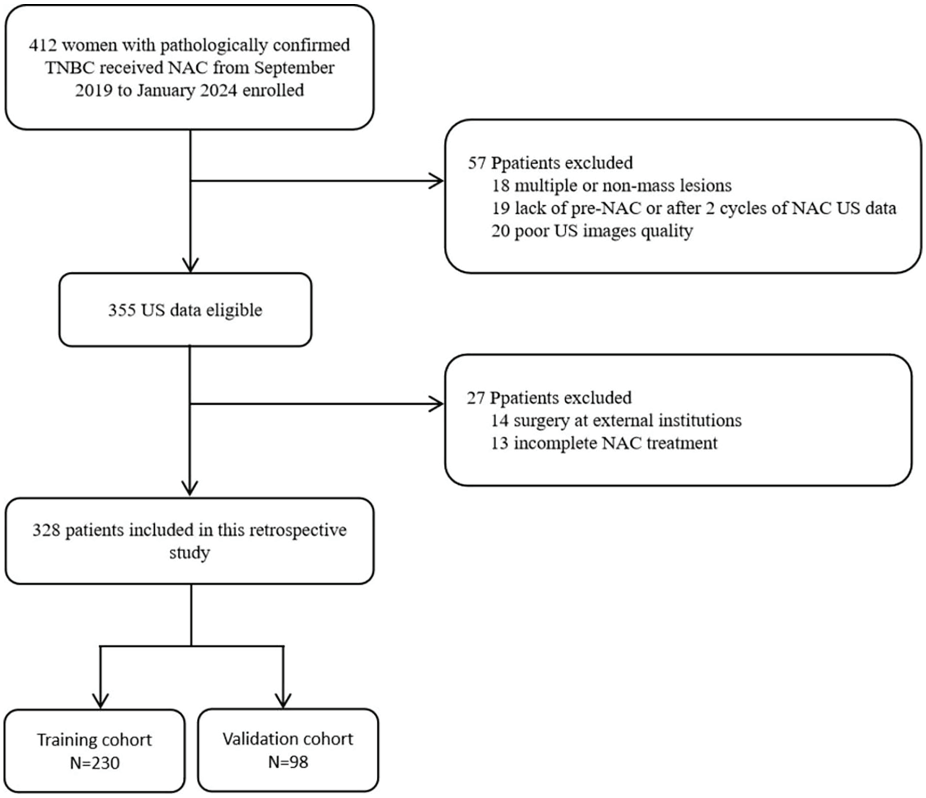


Figure 1. Flowchart showing the study exclusion criteria. TNBC, triple-negative breast cancer; NAC, neoadjuvant chemotherapy; US, ultrasound.

longitudinal views, calcification, posterior echo, blood flow, and axillary lymph node status. Measurements were obtained by two experienced breast sonographers (each with >5 years of work experience) who were blinded to the clinical information.

The tumor volume reduction (TVR) percentage between the pre-NAC and after two cycles of NAC was calculated using the formula:[12]

$$Volume = Length \times Width \times Height \times 0.52,$$

$$TVR = (V0 - V2)/V0$$

where V0 is the pre-NAC tumor volume and V2 is the tumor volume after two cycles of NAC.

Clinicopathologic data acquisition and analysis

Clinical and pathological information, including age, menstrual status, and histologic grade, was collected from institutional archives. TNBC was defined by negative expression of the estrogen receptor, progesterone receptor, and HER-2. For HER-2 assessment, immunohistochemistry scores of 0 or 1+ were classified as HER-2 negative, and tumors with immunohistochemical staining of 2+ that lacked HER-2 amplification by fluorescence in situ hybridization were also defined as HER-2 negative.[22]

Chemotherapy and pathologic response evaluation

All included patients received four cycles of epirubicin and cyclophosphamide once every 2 weeks, followed by four additional cycles of either the same regimen or taxane-based treatment. In this study, the US response was evaluated after the first two cycles of epirubicin and cyclophosphamide. Surgical specimens were assessed by pathologists. pCR was defined as the absence of invasive carcinoma in the breast and axillary lymph nodes, with ductal carcinoma in situ permitted (ypT0/is ypN0).

Radiomics signature construction

The radiomics workflow is illustrated in Figures 2a and 2b. Pre-treatment US data were collected for the region of interest (ROI) segmentation and feature extraction for all included patients. An experienced breast sonographer manually defined the ROI at the maximal diameter plane along the tumor contour on the US images using ITK-SNAP Version 3.6.0 (www.itksnap.org). Feature stability was assessed by calculating the intraclass correlation coefficient (ICC), and features with an ICC <0.80 were eliminated. To avoid overfitting in reducing the redundancy and dimensionality process, the least absolute shrinkage and selection operator (LASSO) regression algorithm was used to screen for features associated with pCR prediction. The development of the RS involved the linear combination of the corresponding coefficients from the validation and training cohorts.

Development and validation of the nomogram

Clinical data and US features that could indicate candidate risk factors for pCR were identified using univariate analysis, and those with $P < 0.05$ were subsequently incorporated into multiple logistic regression. Features with $P < 0.05$ were considered statistically significant. The clinical and US risk factors were then used to build a clinico-ultrasonic model.

The early US-based radiomic nomogram was created based on the RS and the clinico-ultrasonic characteristics. Therefore, three preoperative prediction models were fitted in the training cohort: (i) RS model, (ii) clinico-ultrasonic model (clinical factors plus US features), and (iii) early US-based radiomics nomogram (RS plus clinico-ultrasonic model).

Receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC) were used to evaluate the three models' diagnostic performance. Delong's validation was then applied to compare the AUCs between models and ascertain whether there were any notable variations in their diagnostic performance. The concordance index (C-index) was used to evaluate nomogram performance, where a C-index value of 0.5 denoted a random chance and 1.0 denoted an exact differentiation of the outcome. The Hosmer–Lemeshow test was used to evaluate differences between the anticipated and actual data. The calibration curve of the nomogram was plotted for the training and validation groups to assess the consistency of the predicted probabilities of pCR with the actual results. Ultimately, by calculating the net benefit at various threshold probabilities, a decision curve analysis was conducted to evaluate the clinical validity of the radiomics nomogram model.

Statistical analysis

All statistical tests were performed using SPSS software (version 26.0; IBM, Armonk, NY, USA) and R software (version 4.1.2), and a two-tailed $P < 0.05$ was deemed statistically significant. Quantitative variables were compared between the training and validation cohorts using the t-test. Categorical variables were compared using Fisher's exact test or the chi-square test. Continuous variables were presented as mean ± standard deviation. The LASSO logistic regression analysis was performed using the glmnet package. The ROC curve was plotted using the pROC

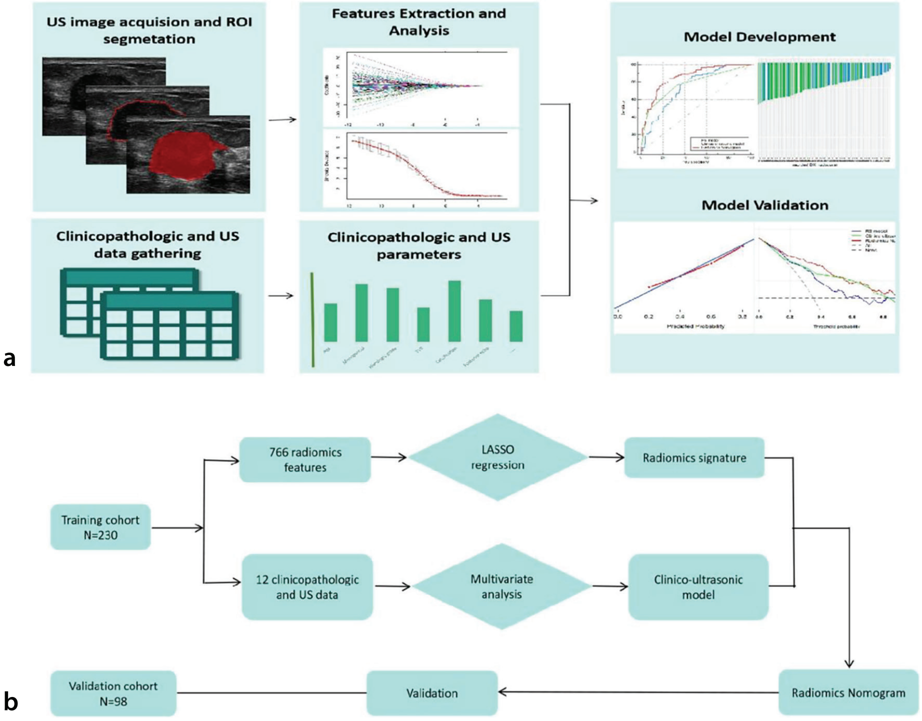


Figure 2. (a) Radiomics workflow and **(b)** study flowchart. US, ultrasound; ROI, region of interest.

package. The nomogram, calibration curves, and decision curve were created using the rms and rmda software packages. The Hmisc package was used to calculate the C-index.

Results

Patient characteristics and clinico-ultrasonic model

A total of 328 patients meeting the eligibility criteria between September 2019 and January 2024 were included. The overall pCR rate was 34.76% (114 of 328). There was no significant difference in pCR rates between the training and validation cohorts [35.7% (82/230) and 32.7% (32/98), respectively, $P = 0.602$; Table 1]. Histologic grade and TVR were significantly associated with pCR rate (both $P < 0.05$), and multiple logistic regression analysis revealed that both variables were independent clinical risk predictors for

Table 1. Clinical, histopathological, and US characteristics of the patients in the training and validation cohorts

Characteristics	Training cohort (n = 230)		*P*	Validation cohort (n = 98)		*P*
	pCR (n = 148)	non-pCR (n = 82)		pCR (n = 66)	non-pCR (n = 32)	
Age (years, mean ± SD)	49.2 ± 11.8	49.8 ± 9.7	0.692	49.9 ± 9.2	48.4 ± 10.0	0.471
Histologic grade						
I–II	42 (28.4%)	40 (48.8%)	0.002	15 (22.7%)	21 (65.6%)	<0.001
III	106 (71.6%)	42 (51.2%)		51 (77.3%)	11 (34.4%)	
Menopausal status						
Premenopausal	76 (51.4%)	55 (60.4%)	0.008	34 (51.5%)	12 (37.5%)	0.192
Postmenopausal	72 (48.6%)	36 (39.6%)		32 (48.5%)	20 (62.5%)	
US-reported tumor size (mm)						
≤20	18 (12.2%)	14 (17.1%)	0.284	3 (4.5%)	3 (9.4%)	0.646
20–50	106 (71.6%)	60 (73.2%)		52 (78.8%)	24 (75%)	
>50	24 (16.2%)	8 (9.7%)		11 (16.7%)	5 (15.6%)	
Calcification						
Non-calcification	86 (58.1%)	44 (53.7%)	0.514	33 (50%)	21 (65.6%)	0.145
Calcification	62 (41.9%)	38 (46.3%)		33 (50%)	11 (34.4%)	
Posterior echo						
No change	59 (39.9%)	36 (43.9%)	0.591	33 (50%)	17 (53.1%)	0.964
Enhancement	24 (16.2%)	17 (20.7%)		18 (27.3%)	9 (28.1%)	
Shadowing	41 (27.7%)	17 (20.7%)		9 (13.6%)	4 (12.5%)	
Mixed	24 (16.2%)	12 (14.7%)		6 (9.1%)	2 (6.3%)	
Blood flow grade						
0–I	25 (16.9%)	16 (19.5%)	0.619	7 (10.6%)	4 (12.5%)	0.781
II–III	123 (83.1%)	66 (80.5%)		59 (89.4%)	28 (87.5%)	
Resistant index						
≤0.7	53 (35.8%)	39 (47.6%)	0.081	22 (33.3%)	11 (34.4%)	0.918
>0.7	95 (64.2%)	43 (52.4%)		44 (66.7%)	21 (65.6%)	
The number of US-reported abnormal lymph nodes						
None	20 (13.5%)	13 (15.9%)	0.157	5 (7.6%)	2 (6.3%)	0.835
≤3	34 (23%)	27 (32.9%)		24 (36.4%)	9 (28.1%)	
3–9	71 (48%)	27 (32.9%)		29 (43.9%)	16 (50%)	
>9	23 (15.5%)	15 (18.3%)		8 (12.1%)	5 (15.6%)	
Lymph node short/long axis ratio						
≤0.5	82 (55.4%)	48 (58.5%)	0.646	34 (51.5%)	16 (50%)	0.888
>0.5	66 (44.6%)	34 (41.5%)		32 (48.5%)	16 (50%)	
Lymph node type						
Presence of lymphatic gates	85 (57.4%)	58 (70.7%)	0.046	43 (65.2%)	15 (46.9%)	0.084
Disappearance of the lymphatic portal	63 (42.6%)	24 (29.3%)		23 (34.8%)	17 (53.1%)	
TVR						
<80%	138 (93.2%)	43 (52.4%)	<0.001	56 (84.8%)	19 (59.4%)	0.007
≥80%	10 (6.8%)	39 (47.6%)		10 (15.2%)	13 (40.6%)	

SD, standard deviation; pCR, pathologic complete response; US, ultrasound; TVR, tumor volume reduction.

pCR (both $P < 0.05$; Table 2). The clinico-ultrasonic model developed from these two variables performed well in both the training cohort [AUC: 0.773; 95% confidence interval (CI): 0.711–0.835] and the validation cohort (AUC: 0.77; 95% CI: 0.674–0.866). Typical US images from patients with pCR and non-pCR are presented in Figure 3.

Development and validation of the radiomics signature model

From the pre-NAC US images, 766 radiomics features were initially extracted. Using LASSO regression, these were reduced to 12 features with non-zero coefficients. The ICC showed good reproducibility for selected radiomics feature extraction. A heatmap (Figure 4a) illustrates the pairwise correlations between the selected radiomics features. The formula for the final imaging omics score is as follows:

$$\text{Radscore} = -0.606 + \sum_{i=1}^{12} (a_i \times F_i)$$

The specific coefficient (α) and characteristics (*F*) are shown in Figure 4c. Patients achieving pCR had significantly higher Radscores than those without pCR in both cohorts (all $P < 0.01$, Figure 4b). The RS model exhibited good predictive performance with an AUC of 0.742 (95% CI: 0.678–0.806) in the training cohort and 0.758 (95% CI: 0.646–0.870) in the validation cohort.

Development and performance of the radiomics nomogram

Histologic grade, TVR, and RS were used to construct the early US-based radiomics nomogram (Figure 5), which showed excellent discriminatory power with an AUC of 0.836 (95% CI, 0.807–0.905) in the training cohort and 0.856 (95% CI, 0.749–0.923) in the validation cohort (Figures 6a and 6b). The radiomics nomogram outperformed both the clinico-ultrasonic model (AUC: 0.773, $P < 0.05$) and the RS model (AUC values of 0.742, $P < 0.05$) in the training cohort (Table 3). The nomogram also performed well in the validation cohort (AUC: 0.856 vs. 0.770, 0.758; $P < 0.05$). The C-index value for predicting pCR was 0.85 in the training cohort and 0.81 in the validation cohort. Calibration curves for both cohorts showed good alignment with the ideal curve (Figures 6c and 6d), and the

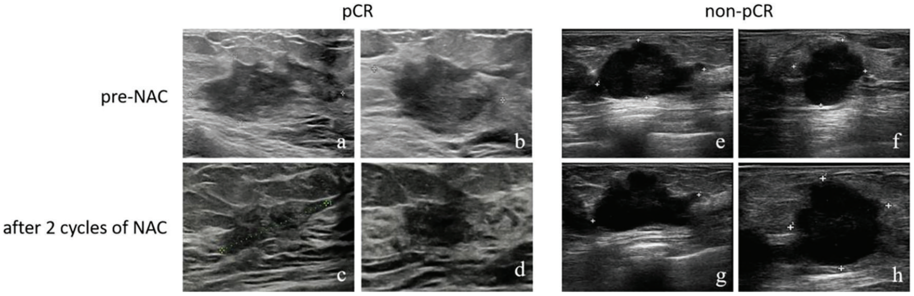


Figure 3. Grey-scale US images. **(a, b)** The three largest tumor dimensions in the transverse and longitudinal images before NAC. **(c, d)** The three largest tumor dimensions in the transverse and longitudinal images after two cycles of NAC. The patient completed neoadjuvant systemic therapy, and surgical pathology confirmed a pCR. **(e, f)** The three largest tumor dimensions in the transverse and longitudinal images before NAC. **(g, h)** The three largest tumor dimensions in the transverse and longitudinal images after two cycles of NAC. The patient completed neoadjuvant systemic therapy, and surgical pathology confirmed a non-pCR. US, ultrasound; NAC, neoadjuvant chemotherapy; pCR, pathologic complete response.

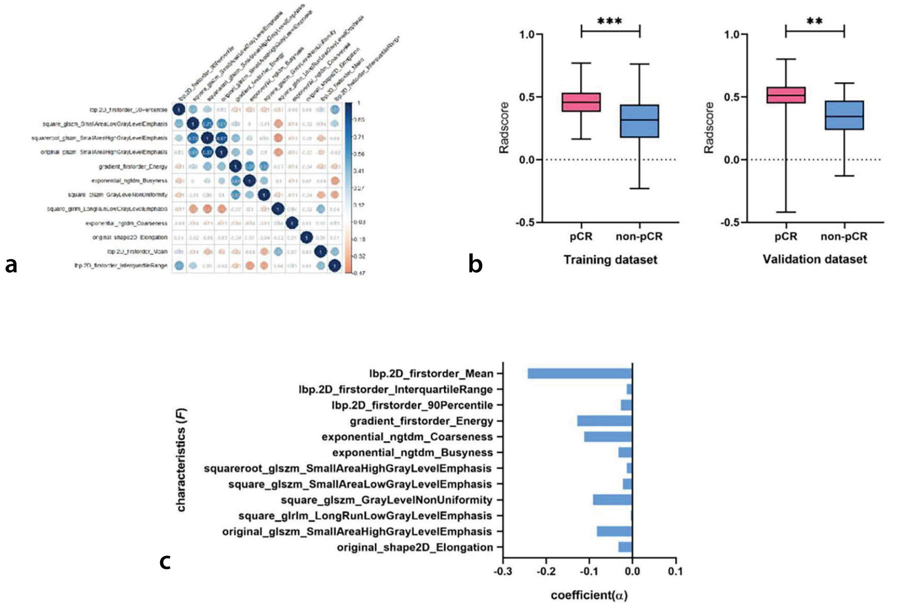


Figure 4. (a) Heatmap of selected radiomics features from the training cohort (GLSZM, grey level size zone matrix; NGTDM, neighboring gray tone difference matrix). **(b)** Comparison of Radscores between the pCR and non-pCR groups in the training and validation cohorts. **(c)** Features selected using LASSO regression. pCR, pathologic complete response; LASSO, least absolute shrinkage and selection operator.

Table 2. Multiple logistic regression analysis of risk factors for pCR

Characteristics	Training cohort (n = 230)			Validation cohort (n = 98)		
	β	Odds ratio (95% CI)	P	β	Odds ratio (95% CI)	P
Intercept	−0.546			−0.022		
Histologic grade						
I–II		1 (ref)			1 (ref)	
III	−1.050	0.350 (0.183 to 0.668)	0.001	−1.855	0.156 (0.059 to 0.411)	<0.001
TVR						
<80%		1 (ref)			1 (ref)	
≥80%	2.626	13.821 (6.195 to 30.837)	< 0.001	1.318	3.734 (1.270 to 10.977)	0.017

pCR, pathologic complete response; CI, confidence interval; TVR, tumor volume reduction.

Hosmer–Lemeshow test indicated no significant differences (P = 0.470 and 0.623). Decision curve analysis (Figure 6e) demonstrated that the early US-based radiomics nomogram enhanced the prediction of pCR in breast cancer across a wide range of risk threshold probabilities.

Discussion

In this study, we developed and validated an early US-based radiomics nomogram that integrates clinical data, US features, and RS derived from grayscale US images. The nomogram yielded satisfactory predictions, with an AUC of 0.856 in the training cohort and 0.836 in the validation cohort, outperforming both the clinico-ultrasonic and RS models.

Previous studies have shown that overestimating the extent of residual disease involvement can lead to unnecessary surgical expansion, whereas underestimating the residual cancerous area may result in positive margins and tumor recurrence after NAC in TNBC.[8,10,23,24] However, pathological findings from surgical specimens, considered the gold standard for evaluating NAC effectiveness, are only available after completing chemotherapy. Early prediction of NAC efficacy is therefore essential to allow appropriate therapy adjustments for potentially poor NAC responders. US, MRI, and PET have been investigated in several recent trials to track NAC efficacy, but US, being non-invasive and easily accessible, has an edge over other imaging modalities in predicting early tumor response.[25] A recent study confirmed that combining US features with clinicopathologic factors can accurately predict pCR preoperatively in breast cancer.[15,26] Therefore, we combined clinical prediction factors and US features to develop a clinico-ultrasonic model to predict pCR in patients with TNBC.

Our findings demonstrated a strong correlation between histologic grade and pCR rate in tumors, consistent with recent studies.[27,28] Jung et al.[29] demonstrated that tumors with high histologic grade are associated with improved pCR rates in patients with breast cancer due to their elevated mitotic index, supporting our findings. Ni et al.[30] showed that the changes in maximum tumor diameter after two cycles of NAC, four cycles of NAC, and six cycles of NAC were all independent predictors of pCR (P = 0.017, 0.005 and 0.009), though they did not evaluate TVR across the three largest tumor dimensions, which may be more accurate than predicting NAC efficacy from a single dimension. Adrada et al.[12] reported that a reduction of 80% or more in tumor volume after two NAC cycles predicted pCR in patients with TNBC. Gu et al.[31] also stated that the percentage reduction in tumor volume after four cycles was associated with early NAC response in breast cancer patients.

In our study, the optimum cut-off point on the ROC curve was based on the maximal value of the Youden index, and we set the TVR cut-off point to 80%. In the training cohort, the pCR rate was 69.2% (36 of 52) in patients with TVR ≥80% and 25.8% (46 of 178)

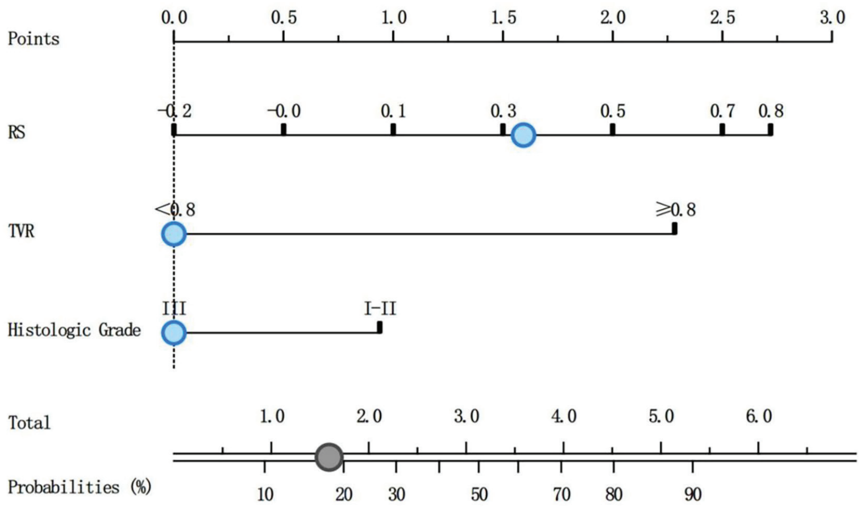


Figure 5. Development of the US-based radiomics nomogram for predicting pCR in TNBC. The nomogram integrates histologic grade, TVR, and RS. US, ultrasound; TNBC, triple-negative breast cancer; TVR, tumor volume reduction; RS, radiomics signature.

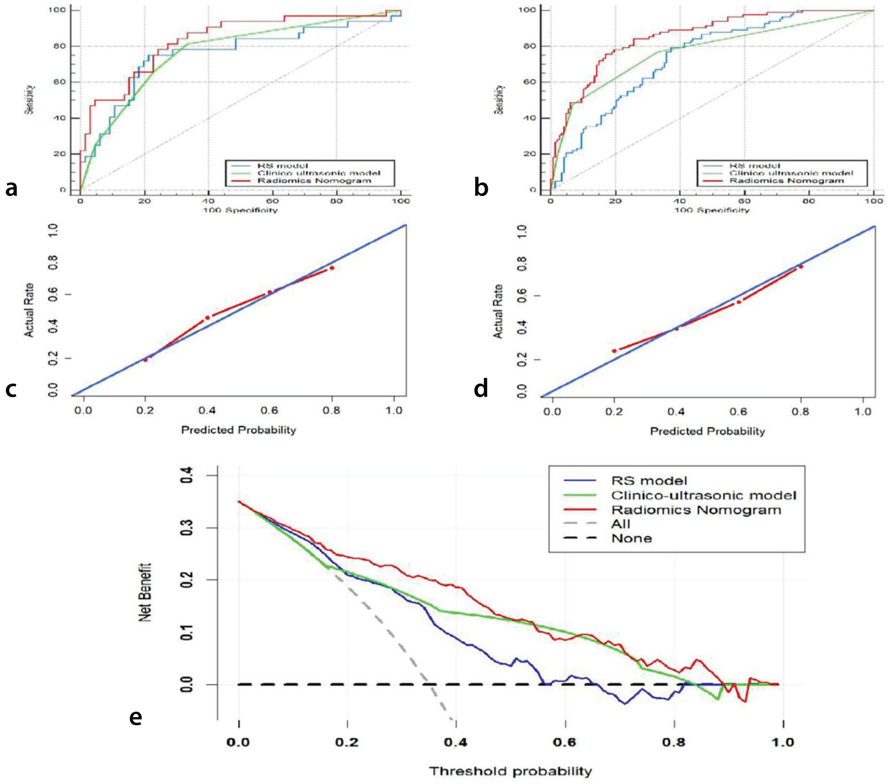


Figure 6. (a, b) ROC comparisons of the clinico-ultrasonic model, RS model, and radiomics nomogram in the training and validation cohorts. **(c, d)** Calibration curves of the radiomics nomogram in the training and validation cohorts. **(e)** Decision curve of the RS model, clinico-ultrasonic model, and radiomics nomogram. ROC, receiver operating characteristic; RS, radiomics signature; pCR, pathologic complete response.

Table 3. Diagnostic performance of the predictive models

Predictive model	Training cohort (n = 230)			Validation cohort (n = 98)		
	Clinico-ultrasonic model	RS model	Radiomics nomogram	Clinico-ultrasonic model	RS model	Radiomics nomogram
AUC (95% CI)	0.773 (95% CI: 0.711–0.835)	0.742 (95% CI: 0.678–0.806)	0.836 (95% CI: 0.807–0.905)	0.770 (95% CI: 0.674–0.866)	0.758 (95% CI: 0.646–0.870)	0.856 (95% CI: 0.749–0.923)
Sensitivity	0.768	0.793	0.756	0.813	0.750	0.667
Specificity	0.669	0.628	0.831	0.667	0.788	0.875
NPV	0.839	0.845	0.860	0.880	0.867	0.917
PPV	0.562	0.542	0.713	0.542	0.632	0.560
C-index	0.85			0.81		

RS, radiomics signature; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; C-index, concordance index.

in patients with TVR <80%. Multiple logistic regression analysis confirmed TVR as an independent risk factor (*P* < 0.001). Thus, clinical and US features can serve as excellent low-cost predictors in estimating early response in patients with TNBC. No other clinico-ultrasonic factors related to pCR in TNBC were found in this study, possibly due to the low number of patients enrolled.

Radiomics, a computer-aided technology, uses digital medical images to quantify tumor heterogeneity, converting those features into a series of mathematical data.[16,32] Recent studies have shown that radiomics analysis can be used for breast cancer subtype differentiation, therapeutic decision making, and axillary lymph node metastasis prediction.[33-36] Compared with MRI, US images are less expensive, simpler, and easier to obtain for preoperative pCR evaluation, offering substantial potential clinical and financial advantages. In this study, the final RS calculation formula was constructed based on twelve selected radiomics features to predict pCR before surgery in patients with TNBC, revealing excellent predictive performance with an AUC of 0.742 in the training cohort and 0.758 in the validation cohort.

The final twelve radiomics features included one shape-based, four first-order statistical, and seven texture-based features. Among them, the three most valuable were the neighboring gray tone difference matrix (NGTDM) coarseness and first-order mean and energy. First-order mean and energy measured the frequency distribution of the pixel intensity for the zones.[37] In our study, the mammary gland signal intensity correlated with the pCR rate among patients with TNBC. Higher mean and energy may be observed in TNBC non-pCR compared with pCR. The NGTDM coarseness is determined by the variations between voxels in adjacent image planes.[38] Current research indicates biological differences between responsive and non-responsive tumors, as evidenced by the decreased NGTDM coarseness in pCR compared with non-pCR. In addition, nine other features provided crucial insights into tumor physiology and the microenvironment of pCR, with its multiple tissue components.

Recent advances in radiomics nomograms emphasize the prognostic value of tumor and axillary lymph node status assessment following NAC, thereby informing treatment strategy decisions and avoiding unnecessary surgery.[39] Consequently, to provide an accurate individualized prediction of pCR, we developed a clinically applicable nomogram that integrates clinical data, US features, and RS. Our early US-based radiomics nomogram demonstrated appropriate calibration and excellent discrimination. It outperformed the RS and clinic-ultrasonic models, exhibiting superior predictive performance and net benefit in both the training and validation datasets. Therefore, the nomogram serves as a non-invasive preoperative predictive tool for pCR, helping clinicians identify patients with poor response to NAC and enabling consideration of alternative therapies for chemotherapy-insensitive patients with TNBC. These findings highlight the clinical value of the nomogram.

However, several limitations of this study should be acknowledged. First, this was a retrospective, single-center study, and future studies should involve larger sample sizes and multicenter data. Second, manual ROI segmentation is time-consuming; a fully automatic tool to encourage the clinical application of a nomogram should be explored. Third, elastography and contrast-enhanced US examinations were not included, potentially limiting lesion information and affecting the accuracy of the clinico-ultrasonic model for predicting pCR in TNBC.[40] Fourth, due to the exclusion criteria, patients with multiple cancers or non-mass lesions, and those who lacked pre-NAC or post-two-cycle NAC US data were excluded, potentially resulting in selection bias. Future studies should address these limitations to further refine the current model.

In conclusion, we developed and validated an early US-based radiomics nomogram to predict the likelihood of pCR in patients with TNBC. This tool may assist clinicians in formulating personalized treatment and identifying patients with a high probability of achieving pCR.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

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. Jiang YZ, Ma D, Suo C, et al. Genomic and Transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. *Cancer Cell*. 2019;35(3):428-440. [Crossref]
3. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? *Nat Rev Clin Oncol*. 2010;7(12):683-692. [Crossref]
4. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121(7):2750-2767. [Crossref]
5. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363(20):1938-1948. [Crossref]
6. Verma A, Singh A, Singh MP, et al. EZH2-H3K27me3 mediated KRT14 upregulation promotes TNBC peritoneal metastasis. *Nat Commun*. 2022;13(1):7344. [Crossref]

7. Zagami P, Carey LA. Triple negative breast cancer: pitfalls and progress. *NPJ Breast Cancer*. 2022;8(1):95. [Crossref]
8. Masuda N, Lee SJ, Ohtani S, et al. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. *N Engl J Med*. 2017;376(22):2147-2159. [Crossref]
9. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164-172. Erratum in: *Lancet*. 2019;393(10175):986. [Crossref]
10. Spring L, Greenup R, Niemierko A, et al. Pathologic complete response after neoadjuvant chemotherapy and long-term outcomes among young women with breast cancer. *J Natl Compr Cancer Netw*. 2017;15(10):1216-1223. [Crossref]
11. Pinder SE, Provenzano E, Earl H, Ellis IO. Laboratory handling and histology reporting of breast specimens from patients who have received neoadjuvant chemotherapy. *Histopathology*. 2007;50(4):409-417. [Crossref]
12. Adrada BE, Candelaria R, Moulder S, et al. Early ultrasound evaluation identifies excellent responders to neoadjuvant systemic therapy among patients with triple-negative breast cancer. *Cancer*. 2021;127(16):2880-2887. [Crossref]
13. Rauch GM, Adrada BE, Kuerer HM, van la Parra RF, Leung JW, Yang WT. Multimodality imaging for evaluating response to neoadjuvant chemotherapy in breast cancer. *AJR Am J Roentgenol*. 2017;208:290-299. [Crossref]
14. Fowler AM, Mankoff DA, Joe BN. Imaging neoadjuvant therapy response in breast cancer. *Radiology*. 2017;285(2):358-375. [Crossref]
15. Zhang MQ, Du Y, Zha HL, et al. Construction and validation of a personalized nomogram of ultrasound for pretreatment prediction of breast cancer patients sensitive to neoadjuvant chemotherapy. *Br J Radiol*. 2022;95(1140):20220626. [Crossref]
16. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med*. 2020;61(4):488-495. [Crossref]
17. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. [Crossref]
18. Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer*. 2021;147:95-105. [Crossref]
19. Wibmer A, Hricak H, Gondo T, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol*. 2015;25(10):2840-2850. [Crossref]
20. Du Y, Zha HL, Wang H, et al. Ultrasound-based radiomics nomogram for differentiation of triple-negative breast cancer from fibroadenoma. *Br J Radiol*. 2022;95(1133):20210598. [Crossref]
21. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res BCR*. 2017;19(1):57. Erratum in: *Breast Cancer Res*. 2017;19(1):80. [Crossref]
22. Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V. The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999-2004. *Cancer*. 2008;112(4):737-747. [Crossref]
23. Ma M, Gan L, Liu Y, et al. Radiomics features based on automatic segmented MRI images: prognostic biomarkers for triple-negative breast cancer treated with neoadjuvant chemotherapy. *Eur J Radiol*. 2022;146:110095. [Crossref]
24. Kuerer HM, Rauch GM, Krishnamurthy S, et al. A clinical feasibility trial for identification of exceptional responders in whom breast cancer surgery can be eliminated following neoadjuvant systemic therapy. *Ann Surg*. 2018;267(5):946-951. [Crossref]
25. Dobruch-Sobczak K, Piotrzkowska-Wróblewska H, Klimonda Z, et al. Multiparametric ultrasound examination for response assessment in breast cancer patients undergoing neoadjuvant therapy. *Sci Rep*. 2021;11(1):2501. [Crossref]
26. Zheng Q, Yan H, He Y, et al. An ultrasound-based nomogram for predicting axillary node pathologic complete response after neoadjuvant chemotherapy in breast cancer: Modeling and external validation. *Cancer*. 2024;130(S8):1513-1523. [Crossref]
27. Liedtke C, Hatzis C, Symmans WF, et al. Genomic grade index is associated with response to chemotherapy in patients with breast cancer. *J Clin Oncol*. 2009;27(19):3185-3191. [Crossref]
28. Hatzis C, Pusztai L, Valero V, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA*. 2011;305(18):1873-1881. [Crossref]
29. Jung YY, Hyun CL, Jin MS, et al. Histomorphological factors predicting the response to neoadjuvant chemotherapy in triple-negative breast cancer. *J Breast Cancer*. 2016;19(3):261-267. [Crossref]
30. Ni P, Li Y, Wang Y, et al. Construction of a nomogram prediction model for the pathological complete response after neoadjuvant chemotherapy in breast cancer: a study based on ultrasound and clinicopathological features. *Front Oncol*. 2025;15:1459914. [Crossref]
31. Gu J, Tong T, He C, et al. Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol*. 2022;32(3):2099-2109. [Crossref]
32. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [Crossref]
33. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;69(2):127-157. [Crossref]
34. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11(1):1236. [Crossref]
35. Granzier RWY, Ibrahim A, Primakov SP, et al. MRI-based radiomics analysis for the pretreatment prediction of pathologic complete tumor response to neoadjuvant systemic therapy in breast cancer patients: a multicenter study. *Cancers*. 2021;13(10):2447. [Crossref]
36. Yang M, Liu H, Dai Q, et al. Treatment response prediction using ultrasound-based pre-, post-early, and delta radiomics in neoadjuvant chemotherapy in breast cancer. *Front Oncol*. 2022;12:748008. [Crossref]
37. Nougaret S, Tibermacine H, Tardieu M, Sala E. Radiomics: an introductory guide to what it may foretell. *Curr Oncol Rep*. 2019;21(8):70. [Crossref]
38. Suga M, Nishii R, Miwa K, et al. Differentiation between non-small cell lung cancer and radiation pneumonitis after carbon-ion radiotherapy by 18F-FDG PET/CT texture analysis. *Sci Rep*. 2021;11(1):11509. [Crossref]
39. Gu J, Tong T, Xu D, et al. Deep learning radiomics of ultrasonography for comprehensively predicting tumor and axillary lymph node status after neoadjuvant chemotherapy in breast cancer patients: a multicenter study. *Cancer*. 2023;129(3):356-366. [Crossref]
40. Gu J, Polley EC, Denis M, et al. Early assessment of shear wave elastography parameters foresees the response to neoadjuvant chemotherapy in patients with invasive breast cancer. *Breast Cancer Res*. 2021;23(1):52. [Crossref]

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253354

CARDIOVASCULAR IMAGING

REVIEW ARTICLE

Non-pulmonary postoperative complications of cardiothoracic surgery

Furkan Ufuk
İclal Ocak
Lydia Chelala
Luis Landeras

The University of Chicago Medicine, Department of Radiology, Chicago, United States

ABSTRACT

Cardiothoracic surgery, including coronary artery bypass grafting, valve replacement, and transplantation, has considerably advanced, improving patient survival and outcomes. However, non-pulmonary postoperative complications remain a major concern, contributing to morbidity and mortality. These complications encompass cardiovascular events, vascular injuries, infections, and device-related issues that can severely impact recovery. Early diagnosis and timely intervention are crucial to mitigating risks and improving patient outcomes. Advanced imaging modalities such as computed tomography, magnetic resonance imaging, and echocardiography play a pivotal role in identifying and characterizing complications before clinical deterioration occurs. This review highlights the spectrum of acute non-pulmonary complications following cardiothoracic surgery, emphasizing the diagnostic value of imaging in guiding clinical decision-making. By improving the awareness of imaging findings associated with postoperative complications, radiologists and clinicians can facilitate early detection, enabling prompt surgical or medical interventions. A multidisciplinary approach that integrates imaging surveillance with clinical assessment is essential for optimizing patient care and reducing long-term morbidity.

KEYWORDS

Abscess, aorta, bypass, cardiac, cardiac catheterization, complications, computed tomography, dissection, endocarditis, hemorrhage, infection, mediastinum, surgery

Cardiothoracic surgery–such as coronary artery bypass grafting (CABG), valve replacement, and transplantation–has greatly evolved in both technique and perioperative care.[1,2] Despite being lifesaving, these complex procedures carry risks, especially when comorbidities are present.[2] Non-pulmonary postoperative complications–including vascular, infectious, and structural problems–can considerably affect recovery, morbidity, and mortality. Early detection and prompt management are critical. Imaging techniques such as chest radiography, computed tomography (CT), and magnetic resonance imaging (MRI) are vital for identifying complications, often before clinical symptoms arise.[3-6] Radiologists play a key role by enabling timely interventions such as surgical revisions or targeted therapies (Table 1). This review examines acute non-pulmonary complications following cardiothoracic surgery, highlights the diagnostic value of imaging, and explores strategies for improving outcomes through interdisciplinary collaboration.

2. Non-pulmonary complications

2.1. Cardiovascular complications

Cardiovascular complications considerably contribute to morbidity and mortality after cardiothoracic surgery. Key concerns include myocardial infarction (MI), cardiac tamponade, graft stenosis or occlusion, and pseudoaneurysms. Early detection with advanced imaging and prompt intervention are critical for optimal patient outcomes.

Presented as an educational exhibit at the RSNA 2024 Annual Meeting (CAEE-16).

Corresponding author: Furkan Ufuk

E-mail: furkan.ufuk@hotmail.com

Received 17 March 2025; revision requested 21 April 2025; accepted 10 June 2025.

Epub: 21.07.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253354

You may cite this article as: Ufuk F, Ocak İ, Chelala L, Landeras L. Non-pulmonary postoperative complications of cardiothoracic surgery. *Diagn Interv Radiol.* 2025;31(6):576-590.

a. Hemopericardium and cardiac tamponade

Cardiac tamponade is a life-threatening complication of cardiovascular surgery, commonly caused by surgical trauma, vascular injury, or anticoagulation.[3,5] Blood or fluid accumulates in the pericardial space, raising intrapericardial pressure and compressing the heart, thereby compromising cardiac output. Prompt recognition is essential to avert circulatory collapse, and imaging findings are crucial in detecting and assessing the severity of tamponade.[7,8] On chest radiography, a markedly enlarged, globular (water bottle) cardiac silhouette suggests considerable pericardial effusion.[8,9] Echocardiography is vital for identifying pericardial fluid and signs of cardiac chamber compression, such as right ventricular or right atrial collapse.[7] CT can delineate hyperattenuating or complex fluid collections indicative of hemopericardium, locate potential bleeding sources, and reveal secondary signs of tamponade, including inferior vena cava distension (Figure 1).[5] MRI further characterizes fluid composition and evaluates myocardial function.[10] Urgent pericardiocentesis is the primary intervention to evacuate the effusion and relieve cardiac compression. If substantial hemorrhage persists or the source of bleeding is unclear, surgical exploration may be needed.[7]

Main points

- Non-pulmonary complications after cardiothoracic surgery, including cardiovascular, vascular, infectious, and device-related issues, substantially impact patient recovery and outcomes. Early diagnosis and intervention are crucial to reducing morbidity and mortality.
- Cardiovascular issues such as myocardial infarction, cardiac tamponade, and thrombus formation, as well as vascular problems such as hemothorax and aortic dissection, require prompt detection through imaging techniques such as echocardiography, computed tomography (CT), and magnetic resonance imaging (MRI) to prevent severe outcomes.
- Postoperative infections, including mediastinitis, surgical site infections, and endocarditis, pose considerable risks. Sternotomy-related complications, such as sternal dehiscence and osteomyelitis, necessitate early recognition using CT and MRI for effective management.
- Prosthetic valves, pacemakers, and left ventricular assist devices (LVADs) can develop dysfunction, infections, or mechanical failures. Imaging plays a crucial role in detecting issues such as valve thrombosis, pacemaker lead displacement, and LVAD-associated thrombi.

b. Myocardial infarction

MI is an uncommon but serious complication of cardiothoracic surgery. Postoperative MI can complicate not only CABG but also other cardiac operations–including valve replacement/repair, resection of intracardiac masses, and complex congenital or aortic procedures–due to ischemia–reperfusion injury, graft or native-vessel occlusion, or embolization.[5] Perioperative MI occurs in about 2%–10% of patients undergoing CABG, underscoring the importance of early and accurate detection.[11] Recent diagnostic algorithms highlight postoperative troponin elevation as a crucial prognostic marker, even without overt ischemic symptoms.[12] Echocardiography is the first-line imaging modality, offering rapid assessment of cardiac function and wall motion. Coronary CT angiography (CTA) provides rapid, non-invasive evaluation of native coronary arteries, assesses bypass graft and stent patency, and demonstrates low-attenuation subendocardial perfusion defects or hypo-enhancement indicative of acute or evolving MI (Figure 2 and Supplementary Figure 1). Cardiac MRI quantifies infarct size, transmurality, and location; detects microvascular obstruction and hemorrhage; visualizes edema and acute injury; differentiates viable from non-viable myocardium for revascularization decisions; and flags early complications such as thrombus, papillary-muscle infarction, septal defect, and adverse remodeling when performed in hemodynamically stable patients.[13,14] Treatment focuses on prompt revascularization with percutaneous coronary intervention or reoperative CABG, supplemented by antiplatelet therapy, anticoagulants, beta-blockers, and other medical support.[12] A multidisciplinary approach that integrates advanced imaging with timely intervention is vital for optimal patient outcomes.

c. Cardiac thrombus

Cardiac thrombus is a serious postoperative complication, occurring in 1%–6% of cardiothoracic surgeries, particularly in valve replacements or procedures involving prosthetic devices.[15] Thrombi often form due to intraoperative endothelial injury, blood stasis, or hypercoagulability and carry a high risk of embolic events such as stroke or pulmonary embolism.[5,15] Echocardiography is the primary imaging method for detecting these thrombi, with transesophageal echocardiography (TEE) considered the gold standard, because it can reveal hyperechoic masses attached to cardiac structures or prosthetic valves.[16] Cardiac CT effectively identifies filling defects within heart chambers, especially in the left atrial appendage (Figure 3 and Supplementary Figure 2), whereas cardiac MRI uses late gadolinium enhancement (LGE) with extended inversion times (TI approximately 600 ms at 1.5T) to improve thrombus detection and assess its composition and surrounding flow dynamics.[14,16] Management typically involves anticoagulation therapy to prevent thrombus growth and embolization, although surgical removal may be necessary if there is considerable hemodynamic risk or embolic potential.[16] Timely imaging and diagnosis are essential for guiding intervention and reducing life-threatening complications.

d. Coronary artery bypass grafting complications

CABG complications include graft occlusion, which affects 10%–15% of patients within the first month (Figure 2). Early occlusion often arises from endothelial injury, platelet dysfunction, or hypercoagulability, whereas late occlusion is largely driven by atherosclerosis. Venous grafts are especially prone to failure, with more than half occluding within 15 years.[17] Recent studies show dual antiplatelet therapy can lower early occlusion risk, emphasizing the importance of optimal postoperative medical management.[12,17] Graft stenosis, often localized to anastomotic sites, compromises perfusion. However, refined surgical techniques and intraoperative imaging have reduced its incidence, prolonging graft durability.[18] Less frequently, pseudoaneurysms (occurring early from suture line disruption or infection) and true aneurysms (developing later from atherosclerosis) can arise (Figure 4). Both carry risks of rupture, thrombosis, or fistula formation and require urgent intervention.[19] Advances in surveillance imaging, including CTA and Doppler ultrasound, allow early detection and guide timely surgical or endovascular repair.

2.2. Vascular and hemorrhagic complications

Vascular and hemorrhagic complications are critical events that can have severe consequences in the postoperative period following cardiothoracic surgery. Prompt identification and management are essential to avoid catastrophic outcomes.

Table 1. Non-pulmonary postoperative complications following cardiothoracic surgery and imaging findings

Complication category	Complication	Imaging findings
Cardiovascular	Hemopericardium and cardiac tamponade	Radiograph: Globular (water bottle) silhouette Echo: Pericardial effusion with chamber collapse CT: Hyperattenuating pericardial fluid, active bleeding MRI: Hemorrhagic vs. serous characterization
Cardiovascular	Myocardial infarction	Echo: Wall motion abnormalities CT: Graft occlusion, myocardial hypoenhancement MRI: Infarct size, edema, hemorrhage, MVO
Cardiovascular	Cardiac thrombus	TEE: Hyperechoic mass on valve/chamber CT: In-chamber filling defect MRI: Thrombus vs. tumor with LGE
Cardiovascular	CABG complications	CT: Graft stenosis, pseudoaneurysm Doppler US: Turbulent or absent graft flow
Vascular and hemorrhagic	Hemothorax	Radiograph: Opacification, effusion US: Hyperechoic pleural fluid CT: Hyperdense pleural fluid, active extravasation
Vascular and hemorrhagic	Mediastinal hemorrhage	Radiograph: Mediastinal widening CT: High-density mediastinal fluid, active extravasation (multiphase imaging)
Vascular and hemorrhagic	Aortic dissection	CT: Intimal flap, double lumen MRI: Intimal flap, thrombus, late-stage complications
Vascular and hemorrhagic	Aortic graft complications	CT: Pseudoaneurysm, endoleak, thrombosis Doppler US: Graft stenosis
Infectious	Mediastinitis	Radiograph: Mediastinal widening, gas CT: Rim-enhancing abscess, wire displacement, osteomyelitis
Infectious	Surgical site infections	US: Fluid collection CT: Abscess, deep tissue gas
Infectious	Endocarditis and bloodstream infections	TEE: Vegetations, abscesses CT: Septic emboli, vegetations, pseudoaneurysm PET-CT: Prosthetic valve infection, emboli
Sternotomy	Sternal dehiscence	Radiograph: Wire displacement, sternal gap CT: Malalignment, hardware failure
Sternotomy	Sternal wound infection	CT: Bone destruction, abscess MRI: Marrow edema, early osteomyelitis
Sternotomy	Hardware failure	Radiograph: Wire fracture/migration CT: Displacement, associated infection
Device-related	Valve prosthesis dysfunction	TEE: Leak, dehiscence CT: Thrombosis, pannus, calcification MRI: Leak quantification, inflammation
Device-related	Pacemaker/ICD malfunction	Radiograph: Lead migration Echo: Pericardial effusion CT: Lead/device malposition, hemorrhage
Device-related	LVAD related complications	Echo: Thrombus, flow obstruction CT: Device thrombosis, debris, kinking, infection
Postoperative syndromes	Post-pericardiotomy syndrome	Radiograph: Enlarged silhouette, effusion Echo: Pericardial effusion MRI: Pericardial edema, effusion, inflammation
Postoperative syndromes	Post-sternotomy syndrome	CT: Chronic Bone changes MRI: Edema, infection, dehiscence
Neurological	Phrenic nerve injury	Radiograph: Elevated hemidiaphragm Fluoroscopy: Paradoxical motion MRI: Diaphragm atrophy
Neurological	Recurrent laryngeal nerve injury	Laryngoscopy: Vocal cord paralysis CT/MRI: Nerve compression, edema
Neurological	Brachial plexus injury	MRI: Plexus thickening, T2 hyperintensity, avulsion, pseudomeningocele

CT, computed tomography; MRI, magnetic resonance imaging; US, ultrasonography; TEE, transesophageal echocardiography; PET, positron emission tomography; ICD, implantable cardioverter defibrillator; LVAD, left ventricular assist devicess; MVO, microvascular obstruction.

a. Hemothorax (non-pulmonary origin)

Hemothorax is a serious postoperative complication of cardiothoracic surgery, often resulting from iatrogenic vessel injury, vascular graft failure, or bleeding from cardiac structures, with anticoagulation further heightening the hemorrhage risk.[1-3] Patients commonly present with acute hemoglobin decline, respiratory distress, hypotension, and diminished breath sounds, requiring prompt intervention to avoid hypovolemic shock or respiratory failure. Although most cases occur early in the postoperative period, delayed hemothorax can develop weeks later, emphasizing the importance of extended monitoring in high-risk patients.[3,4] Imaging is pivotal for diagnosis. A chest radiograph may show rapid, diffuse opacification of the hemithorax, whereas ultrasound offers a valuable bedside tool, revealing fluid collections of varying echogenicity. On non-contrast thoracic CT, hemothorax is characterized by hyperattenuating pleural collections–typically 30–60 Hounsfield units (HU)–that layer dependently and may demonstrate internal septations or fluid–fluid levels (Figure 5). When CTA is performed, active extravasation appears as a focal contrast blush within the pleural space, and injured thoracic aortic branches or intercostal vessels can often be localized to the site of vascular leak. In delayed cases, it can also characterize hematoma progression and identify complications such as infection or loculated collections.[3-6] Management typically involves thoracostomy drainage, surgical exploration if ongoing bleeding is suspected, and supportive measures such as blood transfusion and correction of coagulopathies.[20]

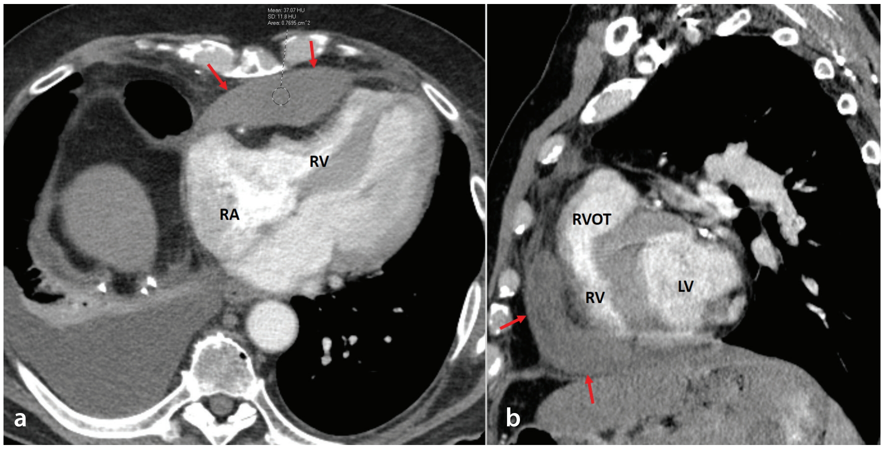


Figure 1. A 62-year-old male status post-robotic coronary artery bypass surgery. **(a)** Axial and **(b)** sagittal CT images demonstrate a loculated right pericardial hematoma (arrows; attenuation of 37 HU) causing considerable compression of the right ventricle, concerning for tamponade. LV, left ventricle; RVOT, right ventricular outflow tract; RA, right atrium; CT, computed tomography; HU, Hounsfield units.

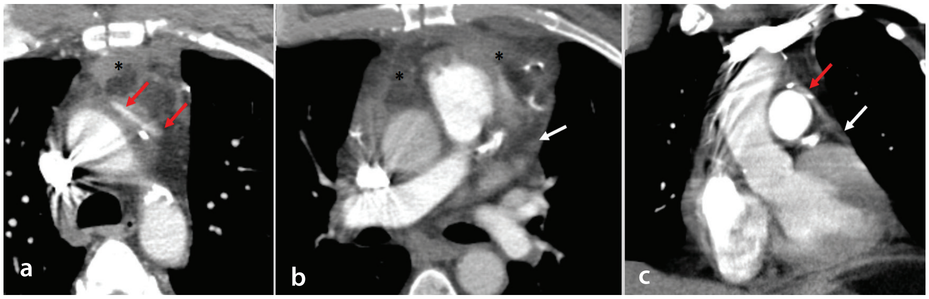


Figure 2. A 66-year-old male status post-coronary artery bypass surgery. **(a, b)** Axial and **(c)** coronal CT images demonstrate an anterior (prevascular) mediastinal postoperative changes with areas of fat stranding (*) and thrombotic occlusion (white arrows) of the aorto-obtuse marginal coronary artery bypass grafting (red arrows). CT, computed tomography.

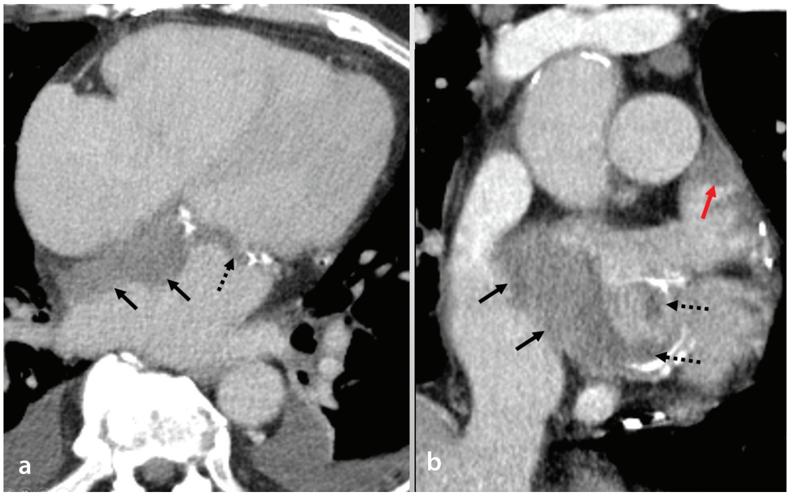


Figure 3. A 74-year-old female status post-mitral valve replacement. Axial and coronal CT images demonstrate an intermediate attenuated lesion most consistent with thrombus along the interatrial septum (solid arrows) and a prosthetic mitral valve, consistent with thrombus (dashed arrows) within the left atrium, as well as a left atrial appendage thrombus (red arrow). Follow-up echocardiography demonstrated complete resolution of the thrombus. CT, computed tomography.

b. Mediastinal hemorrhage

Mediastinal hemorrhage is a life-threatening complication of cardiac surgery, occurring in approximately 3.8% of cases requiring re-sternotomy for bleeding.[1,21] It typically results from intraoperative vascular injury involving the pulmonary arteries, internal thoracic arteries, or azygos vein, as well as postoperative coagulopathy.[3-5,21] Clinically, patients may present with hemodynamic instability, chest pain, and radiographic mediastinal widening, all of which necessitate urgent intervention to prevent progressive blood accumulation that can lead to cardiac tamponade, respiratory compromise, or fatal compression of mediastinal structures. Delayed diagnosis is associated with increased mortality, emphasizing the critical need for prompt imaging and treatment. Chest radiography may reveal mediastinal widening but lacks specificity. CT provides a more definitive evaluation, demonstrating high-density mediastinal fluid and identifying potential bleeding sources. Multiphase CT with non-contrast, arterial, and venous phase imaging is particularly useful for localizing active bleeding, assessing compression of adjacent structures, and guiding therapeutic decisions (Figure 6).[3-7] Management requires immediate surgical exploration to control hemorrhage, with interventional radiology techniques such as vessel embolization serving as a less invasive alternative in select cases. Hemodynamic stabilization through volume resuscitation and blood product administration is essential to support circulation and prevent further deterioration.[21]

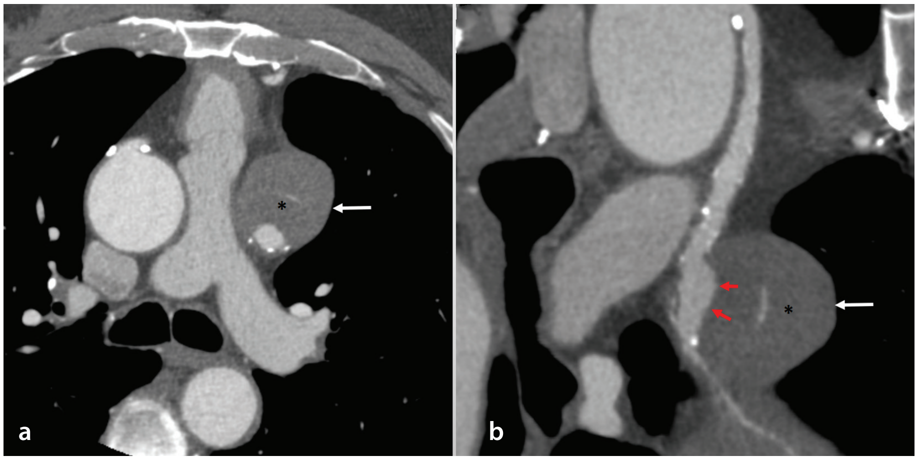


Figure 4. A 48-year-old female status post-robotic coronary artery bypass surgery. **(a)** Axial and **(b)** multiplanar reformat coronary CT angiography images demonstrate a partially thrombosed (*) pseudoaneurysm of the aorto-left circumflex coronary artery bypass grafting (white arrows). Wall irregularities of the bypass graft are also noted (red arrows). CT, computed tomography.

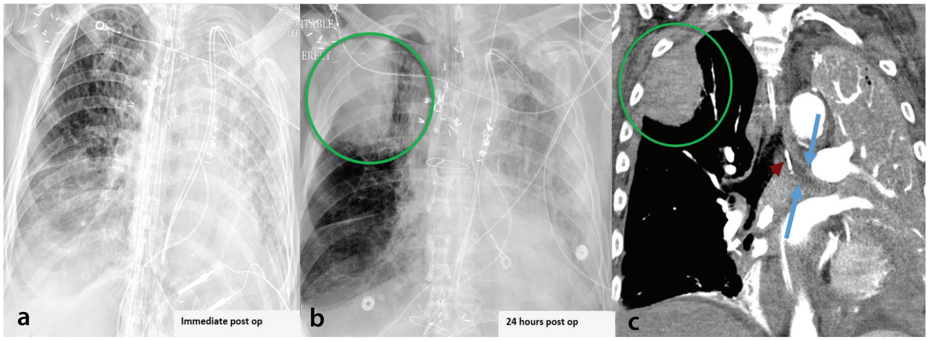


Figure 5. A 66-year-old male status post-heart transplant presenting with hemoptysis. **(a)** The immediate postoperative chest radiograph shows expected post-surgical findings. **(b)** Chest radiograph obtained 24 hours after surgery demonstrates a new broad-based density in the right upper hemithorax (circle). **(c)** Chest CT angiography reveals the opacity corresponding to a loculated hemothorax. A bronchial occluder extends into the left mainstem bronchus (arrowhead), which is filled with hemorrhagic debris (arrows). The left lung is entirely consolidated. CT, computed tomography.

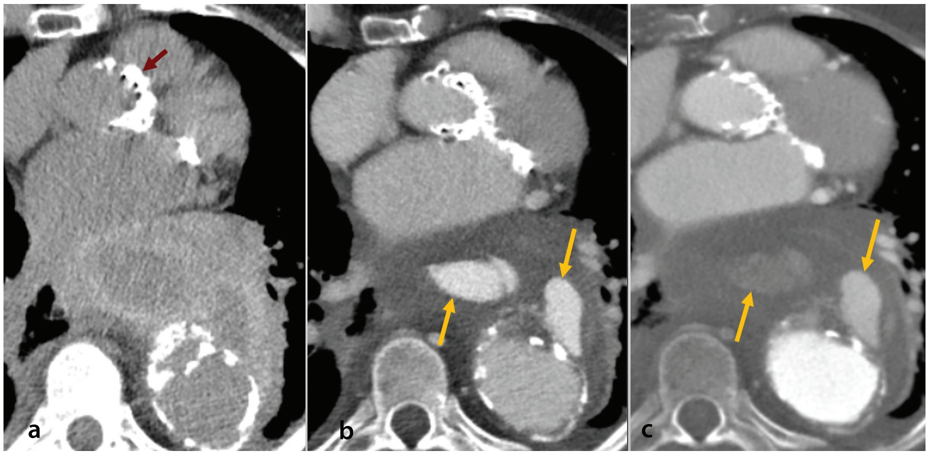


Figure 6. A 79-year-old female status post-transcatheter aortic valve replacement (red arrow). **(a)** Non-contrast, **(b)** arterial, and **(c)** venous phase chest CT images demonstrate arterial contrast extravasation within a periaortic collection, with contrast pooling on the venous phase (yellow arrows), consistent with aortic rupture. The collection contains areas of hyperdensity on the non-contrast image, suggestive of acute blood products/thrombus. CT, computed tomography.

c. Aortic dissection and pseudoaneurysms

Aortic dissection and pseudoaneurysms are rare but severe complications following cardiac or aortic surgery, each carrying high morbidity and mortality. Aortic dissection, occurring in 0.06%–0.23% of procedures, typically arises from intraoperative aortic wall trauma during cannulation, cross-clamping, or graft anastomosis (Figure 7 and Supplementary Figure 3). Chest pain is common but may be overshadowed by nonspecific symptoms such as hypertension or vague discomfort, making early detection challenging; intraoperative TEE can help identify dissections promptly.[22]

Pseudoaneurysms, affecting 2%–3% of aortic repairs, often form at anastomotic sites due to technical imperfections, infection, or graft dehiscence, posing risks of rupture, thrombosis, or compression. On CTA, true aneurysms are broad-necked dilations of all three vessel layers with smooth or scalloped walls and peripheral calcifications, whereas pseudoaneurysms are narrow-necked, irregular saccular outpouchings with wall disruption, periaortic hematoma, and lobulated margins. CTA is the preferred diagnostic tool for both conditions, distinguishing pseudoaneurysms from true aneurysms and delineating dissection flaps, lumina, and branch vessel involvement (Figure 8). Cardiac MRI improves the evaluation of chronic aortic disease by providing cine images for precise measurement of lumen and flap motion; phase-contrast flow quantification to detect false lumen thrombosis or branch compromise; T1/T2 black-blood and mapping sequences to distinguish thrombus or hematoma; LGE to identify wall inflammation, fibrosis, or pseudoaneurysm integrity; and simultaneous assessment of aortic regurgitation and ventricular function.[3-5,10] Management involves urgent surgical repair or endovascular stenting, combined with stringent blood pressure control to prevent progression. Early imaging and timely intervention are critical to avert hemodynamic instability and life-threatening hemorrhage.[23]

d. Aortic vascular graft complications

Aortic vascular graft complications–graft rupture, thrombosis, pseudoaneurysms, infection, and endoleaks–pose major risks after aortic aneurysm repair or dissection surgery.[3,5,24-26] Though rare, graft rupture can trigger sudden hemodynamic collapse, requiring emergency surgery. Thrombosis, presenting as limb pain, pulse deficits, or organ dysfunction from distal ischemia, is tempered by improved anticoagulation and endovascular thrombectomy. Pseudoaneurysms often arise from anastomotic leaks or graft degradation and carry a high rupture risk, but endovascular stent-graft placement offers a minimally invasive alternative to open repair.[26] Graft infections, frequently involving *Staphylococcus* species, require combined medical and surgical management, including antibiotics, debridement, and possible graft replacement (Figure 9).[24,25]

Endoleaks–a unique complication of endovascular aneurysm repair–persist in 20%–30% of cases and are classified into 5 types based on etiology: type I (inadequate proximal/distal seal), type II (collateral retrograde flow), type III (graft junction failure), type IV (graft material porosity), and type V (endotension with sac expansion).[27] Unaddressed endoleaks risk aneurysm sac enlargement and rupture, necessitating vigilant surveillance. CTA is pivotal for diagnosis, offering high-resolution visualization of graft integrity, thrombus formation, pseudoaneurysm morphology, and endoleak flow dynamics. Endoleak evaluation requires a three-phase CTA protocol: an unenhanced scan to distinguish calcifications or thrombus from contrast, an arterial-phase acquisition at 30–40 s post-injection to detect high-flow (type I/III) endoleaks and active contrast blush, and a delayed (venous) phase at 60–120 s to identify low-flow (type II) endoleaks that may only opacify on later images (Figure 10). Early imaging-guided intervention–whether endovascular coil embolization for type II endoleaks, stent relining for pseudoaneurysms, or antimicrobial therapy for infections–is critical to avert life-threatening sequelae. Structured postoperative surveillance protocols, including serial CTA, ensure timely identification of complications, underscoring the synergy between radiologists and surgeons in optimizing patient outcomes.[26,27]

2.3. Infectious complications

Non-pulmonary infectious complications pose critical threats to recovery after cardiothoracic surgery, driving morbidity and mortality through conditions such as mediastinitis, surgical site infections (SSIs), bloodstream infections (BSIs), and device-associated infections (e.g., prosthetic valve endocarditis). Risk factors span preoperative (e.g., diabetes, obesity), intraoperative (e.g., prolonged surgery duration, immunosuppression), and postoperative (e.g., inadequate wound care) variables.

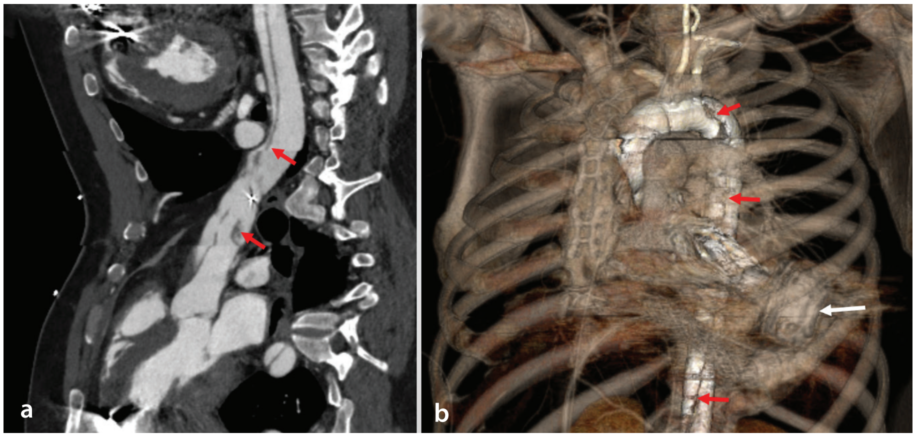


Figure 7. A 57-year-old male status post-left ventricular assist device placement (white arrow). **(a)** Aortic multiplanar reformatted chest CT angiography and **(b)** three-dimensional volume-rendered CT image demonstrate a dissection flap extending through the upper ascending aorta and descending thoracic aorta (red arrows). CT, computed tomography.

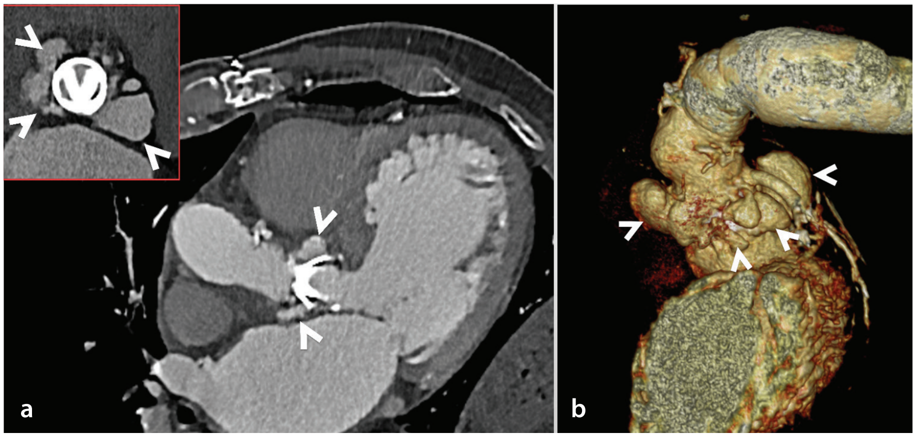


Figure 8. A 63-year-old man with a history of infective endocarditis and composite aortic valve graft with aortic root replacement (Bentall procedure). **(a)** Chest CT angiography images in a left ventricular outflow tract view and corresponding axial view through the aortic valve demonstrate multiple aortic pseudoaneurysms (arrowheads). **(b)** A three-dimensional volume-rendered thoracic CT angiography image also highlights the aortic pseudoaneurysms (arrowheads). CT, computed tomography.

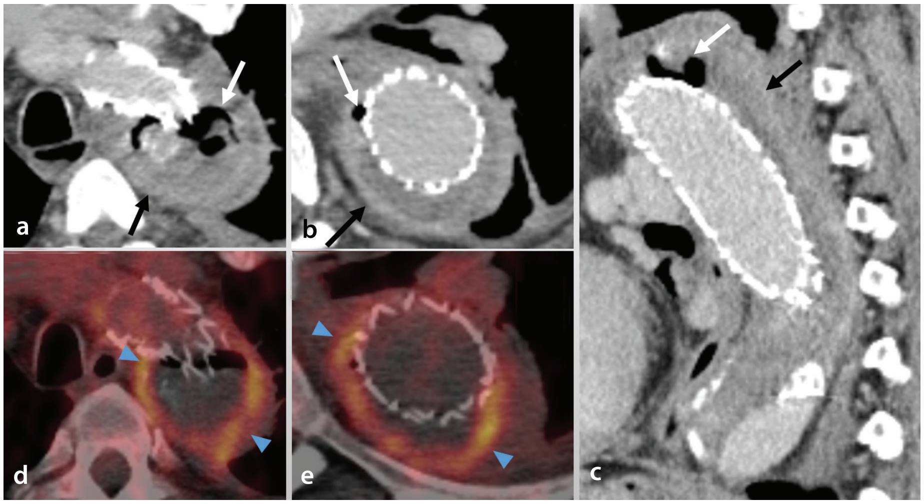


Figure 9. A 73-year-old male status post-aortic endovascular repair. **(a, b)** Axial and **(c)** sagittal CT angiography images demonstrate air surrounding the endograft (white arrows) within the excluded lumen, which exhibits an irregular thickened wall and pathological contrast-enhancement (black arrows). **(d, e)** Axial FDG-PET/CT images show corresponding FDG uptake (arrowheads), consistent with endovascular graft infection. CT, computed tomography; FDG-PET, fluorodeoxyglucose positron emission tomography.

a. Mediastinitis

Mediastinitis is a rare but life-threatening complication of cardiothoracic surgery, affecting 0.4%–5% of cases and occurring more frequently after median sternotomy, such as in CABG or valve replacement. Risk factors include diabetes, obesity, prolonged mechanical ventilation, and *Staphylococcus aureus* colonization.[28] Clinical signs often appear 2–4 weeks postoperatively, featuring fever, sternal instability, purulent drainage, and chest pain. Delayed diagnosis increases mortality beyond 20%.[29] Imaging is key for early detection, and chest radiography can show mediastinal widening, air–fluid levels, or new air collections, but it lacks specificity. CT is preferred, revealing mediastinal fluid collections, gas bubbles, soft tissue fat stranding, rim-enhancing abscesses, and possible sternal wire displacement or osteomyelitis (Figure 11).[3,5,28] Management requires prompt intravenous broad-spectrum antibiotics, surgical debridement, and drainage of any abscesses. Negative-pressure wound therapy aids in wound management, and early diagnosis and aggressive treatment are essential to lower mortality.

b. Surgical site infections

SSIs arise in 1%–8% of cardiothoracic surgeries, with rates influenced by procedure complexity and patient risk factors. They range from superficial infections–limited to skin and subcutaneous tissue–to deep infections involving muscle, fascia, or internal structures. Superficial SSIs typically present with localized redness, pain, swelling, and purulent discharge, whereas deep SSIs may exhibit systemic signs such as fever, leukocytosis, and hemodynamic instability.[29] Imaging is critical in identifying deep infections. Ultrasound effectively detects fluid collections suggestive of abscesses, whereas CT scans provide detailed views of deep-seated infections, including fluid collections, gas, and inflammatory changes (Supplementary Figure 4).[3,5] Treatment depends on infection severity: superficial infections often respond to local wound care and targeted antibiotics, whereas deep infections usually require surgical debridement, drainage, and systemic antimicrobial therapy.[29]

c. Endocarditis and bloodstream infections

Endocarditis is a serious complication of BSIs, especially in patients who undergo cardiothoracic surgery with extended hospital stays or invasive procedures.[30,31] The risk of infective endocarditis (IE) differs by pathogen, with certain *Streptococcus* species (e.g., *S. sanguinis, S. mutans*) linked to higher susceptibility. BSIs affect 2%–5% of postoperative patients, commonly stemming from catheter-related infections or mediastinitis, with *S. aureus*, coagulase-negative staphylococci, and gram-negative bacilli as frequent culprits. Clinically, IE may present with new murmurs, embolic events, or heart failure, requiring swift recognition and management.[30] TEE is the gold standard for diagnosing IE, providing high-resolution visualization of vegetations, abscesses, and prosthetic valve involvement. When echocardiography is inconclusive, cardiac MRI serves as the second-line test, particularly valuable in native valve IE or suspected myocardial/septic embolic extension. Guideline-endorsed fluorodeoxyglucose positron emission tomography/CT is preferred for prosthetic valve or device-related IE, culture-negative cases, and whole-body embolic surveys, with peri-annular hyperuptake beyond 3 months post-surgery confirming infection and extracardiac foci guiding antibiotic duration.[30,31] Cardiac CT–beyond identifying extracardiac septic emboli–directly visualizes perivalvular complications (abscesses, pseudoaneurysms, fistulae) and valve vegetations on both native and prosthetic valves, thereby enhancing diagnostic confidence when echocardiography is inconclusive (Figure 12).[31] Treatment involves a prolonged course of pathogen-targeted intravenous antibiotics. Surgical intervention may be necessary for prosthetic valve endocarditis, substantial valvular dysfunction, or abscess formation to avert further complications.[30,31]

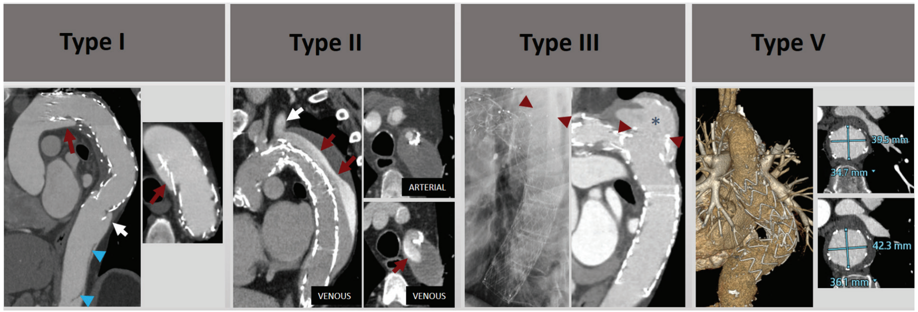


Figure 10. The images illustrate the common types of endoleaks and their corresponding CT findings.

- Type I (inadequate proximal/distal seal): Contrast extravasation into the excluded lumen at the proximal (red arrows) and distal (white arrow) stent margins. A residual dissection flap is noted in the distal descending segment (arrowheads).
- Type II (collateral retrograde flow): Retrograde contrast flow from the proximal left subclavian artery (white arrow) into the excluded false lumen (red arrows) with layering of bright contrast.
- Type III (graft junction failure): Outpouching of contrast (asterisk) extending into the excluded lumen across angulated and separated stent components (arrowhead).
- Type V (endotension with sac expansion): Mildly increased cross-sectional stent dimensions on follow-up. CTA (bottom right image), consistent with progressive sac expansion. CT, computed tomography; CTA, CT angiography.

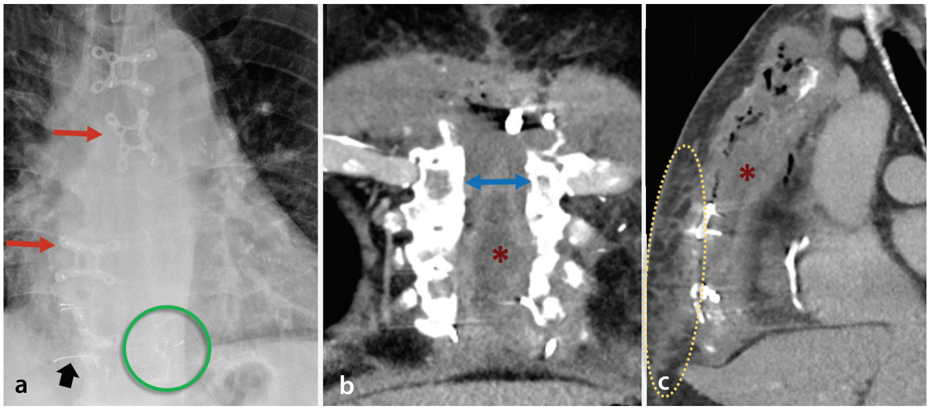


Figure 11. A 62-year-old male status post-coronary artery bypass grafting. **(a)** Chest radiograph demonstrates misaligned sternotomy hardware (arrows), fractured wires (short arrow), and dislodged screws (circle). **(b, c)** Coronal and sagittal chest CT images show a thick-walled air and fluid collection (asterisk) centered on the diastatic sternotomy (double arrow), consistent with an abscess. Associated inflammatory changes in the overlying anterior chest wall are also noted (dotted circle). CT, computed tomography.

2.4. Sternotomy complications

Sternotomy complications considerably affect patient recovery and pose a challenge due to their potential to cause severe morbidity and prolong hospital stay.

a. Sternal dehiscence

Sternal dehiscence, seen in 1%–3% of cardiothoracic surgeries, involves partial or complete separation of the sternum.[32] Common causes include inadequate fixation, infection (e.g., mediastinitis), or poor wound healing, and it often arises within weeks of surgery, presenting with instability, crepitus, pain, or swelling. Untreated cases can progress to mediastinitis, sepsis, or respiratory compromise. Imaging is crucial for early detection. Chest radiographs may show wire or plate displacement, sternal malalignment, or widening between sternal edges. CT offers a more detailed assessment of sternal alignment, wire integrity, and complications such as soft tissue edema, fluid collections, or abscesses (Figure 11 and Supplementary Figure 5).[32,33] Treatment depends on severity. Mild dehiscence may be managed conservatively with chest binders and physical therapy, whereas severe cases often require surgical re-closure with reinforced fixation, vacuum-assisted closure, and targeted antibiotic therapy.[32]

b. Sternal wound infection and osteomyelitis

Sternal osteomyelitis, a serious complication occurring in 0.5%–1% of median sternotomies, typically arises from deep sternal wound infections or mediastinitis.[29,32,34] Risk factors include prolonged operative times, diabetes, obesity, immunosuppression, and bilateral internal mammary artery grafting. Patients may present with persistent chest pain, localized erythema, swelling, purulent discharge, and systemic signs such as fever or leukocytosis. In advanced cases, sternal instability and fistula formation can lead to systemic sepsis or chronic osteomyelitis.[34] Early imaging is essential. Chest radiography may reveal sternal separation, and ultrasound can detect fluid collections or abscesses. CT provides detailed views of bone destruction, periosteal reactions, and soft tissue involvement, whereas MRI is highly sensitive for early marrow edema (Figure 11 and Supplementary Figure 6). Nuclear imaging, including bone scintigraphy, helps clarify ambiguous findings.[32,33] Management involves prolonged, culture-targeted intravenous antibiotics and extensive surgical debridement to remove necrotic tissue and infected bone. In cases of considerable bone loss or nonunion, reconstructive surgery with muscle or omental flaps may be necessary to restore stability.[35]

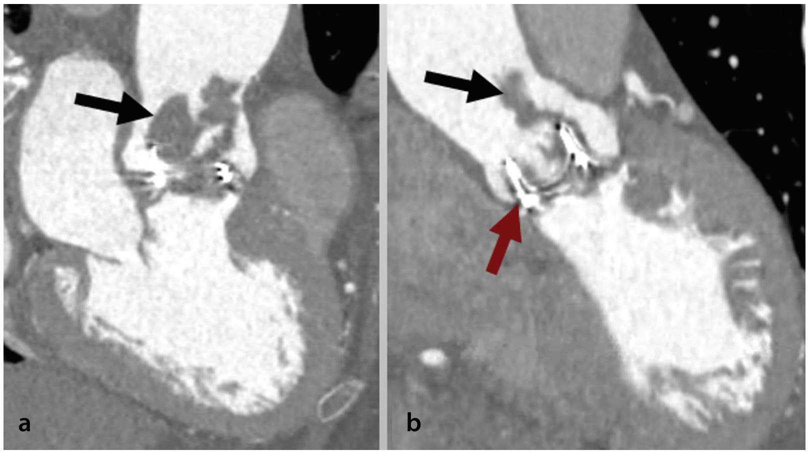


Figure 12. A 54-year-old female status post-surgical aortic valve repair. **(a, b)** Multiplanar reformatted CT angiography images demonstrate a sizeable filling defect (black arrows) adjacent to the metallic aortic valve (red arrow), consistent with a vegetation and perivalvular density and indicative of infective endocarditis. CT, computed tomography.

c. Hardware failure (wire, plates, and screws) and malposition

Hardware failure and malposition of sternal fixation devices occur in 1%–5% of post-sternotomy cases, often caused by mechanical stress, improper placement, or patient-related factors such as osteoporosis, obesity, infection, or inadequate surgical technique.[32,33] Fractured or displaced wires, loose screws, or misaligned plates can weaken sternal stability, increasing the risk of cardiac injury, sternal dehiscence, poor healing, chronic pain, or soft tissue irritation. Chest radiography helps identify broken or migrated wires, whereas CT offers a more detailed view of wire placement and related complications such as dehiscence or infection (Figure 11 and Supplementary Figure 5).[32,33] Management varies according to severity. Symptomatic broken or migrated wires may need removal, and substantial sternal instability may require re-wiring or alternative fixation to restore structural integrity and support healing.[35]

2.5. Device-related complications

Device-related complications can occur due to malfunction, migration, infection, or mechanical issues associated with medical devices utilized during or after cardiothoracic surgery. Early identification and appropriate management are crucial to ensuring patient safety and recovery.

a. Valve prosthesis complications

Prosthetic valve dysfunction, which may present as stenosis, paravalvular leaks, or dehiscence, can arise from mechanical failure, thrombosis, pannus formation, or IE. Mechanical valves are particularly prone to thrombosis if anticoagulation is suboptimal, whereas bioprosthetic valves are more susceptible to structural deterioration (e.g., leaflet tearing, calcification) and non-structural issues (e.g., paravalvular leaks from tissue ingrowth).[31,36] Distinguishing between structural and non-structural dysfunction via imaging is critical to prevent heart failure or thromboembolic events. Echocardiography is the primary diagnostic tool, with TEE offering enhanced visualization of small thrombi and leaks. CT cine reconstructions depict leaflet excursion, opening angle, and restricted mobility, enabling the detection of structural degeneration, flail segments, or stuck leaflets. Density analysis further differentiates pannus (mixed tissue >145 HU) from thrombus (low attenuation <90 HU), whereas high-resolution 3D maps reveal paravalvular abscesses, pseudoaneurysms, fistulae, and dehiscence and quantify cusp, leaflet, and annular calcification–an early marker of bioprosthetic failure (Figures 8, 9).[31] Cardiac MRI offers a radiation-free problem-solving tool that can accurately measure paravalvular leak regurgitant volume and jet direction when echocardiography is equivocal, and LGE with T1/T2 mapping characterizes peri-annular inflammation or abscess when CT findings are indeterminate.[31,36,37] Treatment depends on the underlying problem: anticoagulation for thrombosis, targeted antibiotics for IE, and surgical revision or valve

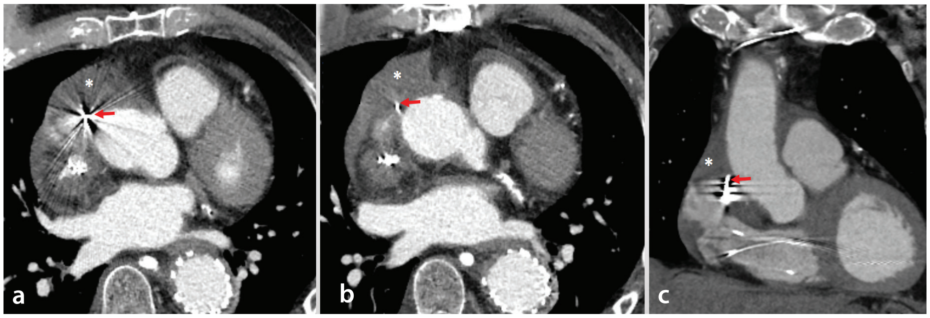


Figure 13. A 66-year-old man with a history of type B aortic dissection, status post-implantable cardioverter-defibrillator placement. **(a, b)** Axial and **(c)** coronal CT images demonstrate right atrial perforation by the pacer lead (red arrow) with associated peri-atrial hemorrhage and fat stranding (*). CT, computed tomography.

replacement for severe dysfunction, substantial leaks, or ongoing infection.[36]

b. Pacemaker or implantable cardioverter defibrillator complications

Pacemaker or implantable cardioverter defibrillator malfunction poses considerable risks for patients who rely on these devices for arrhythmia management. Common causes include lead displacement, battery depletion, infection, and pulse generator failure. Lead displacement may cause syncope, dizziness, or heart block, and infections typically present with local erythema, swelling, or fever. Battery depletion, often identified during routine device interrogation, can lead to sudden pacing or defibrillation failure if left unresolved.[37,38] Prompt recognition and intervention are vital to avert life-threatening arrhythmias, sepsis, or total device failure. Chest radiography is the first-line imaging modality to assess lead positioning, detect fractures, and identify generator or lead migration. Echocardiography helps evaluate intracardiac lead placement and potential complications such as pericardial effusion or hemopericardium. CT provides a more detailed view of device and lead placement, as well as related issues such as hematomas or infections (Figure 13).[37,38] Management depends on the underlying problem. Displaced or fractured leads often need repositioning or replacement, whereas infections require antibiotic therapy and, in severe cases, device removal.[38]

c. Left ventricular assist device complications

Acute complications, including bleeding, infection, right ventricular failure, or thrombosis, following left ventricular assist device (LVAD) implantation can affect up to 15%–30% of patients within the first month. Thrombosis is a key concern, often triggered by biodebris–microscopic particles from mechanical components–that cause microemboli, endothelial damage, and clot formation. These factors increase the risk of thromboembolic events, pump thrombosis, and device malfunction.[39] Clinically, patients may present with hemodynamic instability, worsening heart failure, or neurologic deficits. Sudden power surges or suction alarms often suggest pump thrombosis or biodebris accumulation; elevated D-dimer further supports the diagnosis. Although aortic root thrombi usually appear later, they can emerge early, warranting vigilant monitoring.[39,40]

Imaging is crucial for diagnosing LVAD-related complications. Echocardiography evaluates device flow, detects thrombus formation, and assesses ventricular function and pump integrity in real time. Although CT offers a detailed view of device positioning, structural connections, and associated issues such as biodebris, thrombosis, infection, or mechanical failure, CT images are often limited by beam-hardening and streak artifacts from the device components (Figure 14 and Supplementary Figures 7, 8).[39] Management depends on the underlying cause: anticoagulation therapy for pump thrombosis, targeted antibiotics (and possible device removal) for infection, and transfusion or surgical intervention for bleeding. Device exchange or repair may be necessary for mechanical failures. Current protocols emphasize a multidisciplinary approach–optimizing anticoagulation, employing endovascular interventions, and conducting structured imaging surveillance–ensuring close collaboration between radiologists and cardiologists for timely diagnosis and treatment.[39]

2.6. Postoperative syndromes

Postoperative syndromes encompass various symptoms and conditions following cardiothoracic surgery, affecting recovery and quality of life. Early recognition and management are crucial for optimizing outcomes.

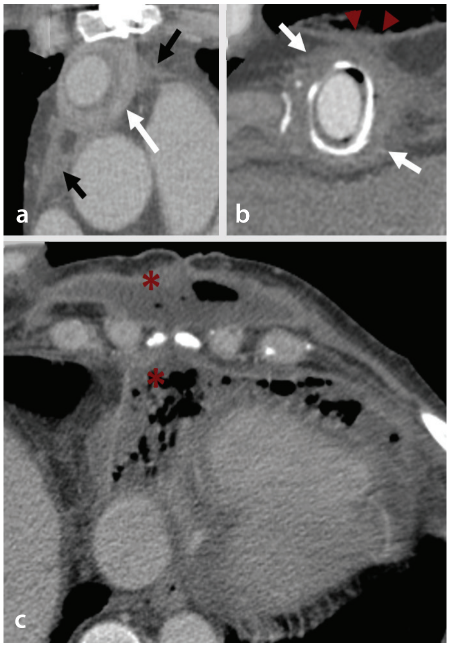


Figure 14. A 65-year-old man with left ventricular assist device (LVAD)-specific infection. **(a, b)** Axial chest CT images demonstrate irregular wall thickening of the LVAD outflow cannula with surrounding fat stranding (white arrows), ill-defined fluid collection (black arrows), and overlying skin thickening (arrowheads), consistent with an LVAD-related infection. **(c)** Status post-LVAD explant and heart transplant, with recurrent air and fluid collections in the mediastinum and chest wall (asterisk), compatible with an abscess. CT, computed tomography.

a. Post-pericardiotomy syndrome

Post-pericardiotomy syndrome (PPS) typically develops days to weeks after cardiothoracic surgery involving pericardial manipulation, triggered by an autoimmune or inflammatory response. Affecting 9%–29% of adult patients, PPS presents with fever, pleuritic chest pain, pericardial or pleural effusions, and elevated inflammatory markers.[41] Imaging is key, and chest radiographs may reveal an enlarged cardiac silhouette or pleural effusions, whereas echocardiography confirms effusions and rules out tamponade. CT and MRI can further detail pericardial edema and complex fluid collections (Figure 15).[42] Management centers on reducing inflammation and symptoms, with non-steroidal anti-inflammatory drugs (NSAIDs) as the first-line therapy and colchicine often added to prevent recurrence. Corticosteroids may be used in refractory cases, and pericardiocentesis is indicated for large or symptomatic effusions.[41]

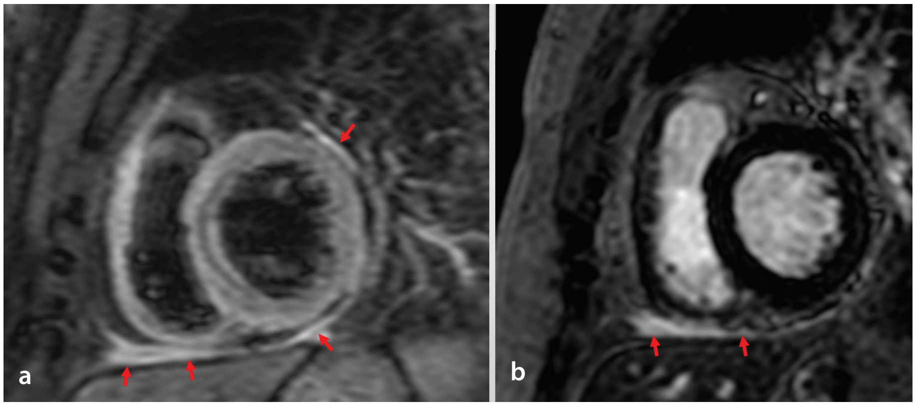


Figure 15. A 62-year-old male status post-coronary artery bypass grafting 2 months ago, presenting with prolonged pleuritic chest pain and mild fever. **(a)** Short-axis short tau inversion recovery image demonstrates patchy pericardial thickening with hyperintensity, suggestive of edema (arrows). **(b)** Late gadolinium enhancement images reveal patchy pericardial enhancement within the basilar inferior pericardium (arrows), consistent with pericardial inflammation.

b. Post-sternotomy syndrome

Post-sternotomy syndrome is a complex complication after median sternotomy, marked by chronic pain, inflammation, and delayed wound healing that impair recovery and quality of life. It affects about 30% of patients, with chronic pain often stemming from surgical trauma to the sternum, intercostal nerves, or soft tissues, whereas inflammation may result from tissue irritation, infection, or reactions to sternal hardware.[43] Imaging–using chest radiography, CT, bone scans, or MRI–is essential for assessing sternal integrity and detecting issues such as fractures, hardware loosening, dehiscence, osteomyelitis, or mediastinitis.[32] Treatment focuses on pain control and functional recovery through medications (NSAIDs, opioids, gabapentin, or pregabalin), physical therapy, rehabilitation, and psychological support.[43]

2.7. Neurological complications

Neurological complications after cardiothoracic surgery, including stroke, transient ischemic attacks, cognitive dysfunction, seizures, and nerve injuries, can severely affect recovery and quality of life. Early detection and prompt management are crucial for optimal outcomes.

a. Phrenic nerve injury

Phrenic nerve palsy complicates 1%–5% of cardiothoracic surgeries, especially those with extensive pericardial manipulation. It typically results from direct nerve injury, thermal damage, or traction, leading to diaphragmatic dysfunction. Patients may experience dyspnea, orthopnea, or reduced exercise tolerance, especially if they have underlying cardiopulmonary issues.[44] Chest radiographs usually show an elevated hemidiaphragm, and a fluoroscopic sniff test reveals paradoxical movement. Ultrasound provides real-time evaluation of diaphragm motion, whereas CT can detect secondary effects such as atelectasis or reduced lung volume, and MRI offers detailed soft tissue assessment in complex cases.[3] Most patients benefit from conservative measures such as breathing exercises and physical therapy, though non-invasive ventilation may be needed for considerable respiratory compromise. Surgical diaphragmatic plication is reserved for refractory cases to improve respiratory mechanics and quality of life.[44]

b. Recurrent laryngeal nerve injury

Recurrent laryngeal nerve injury complicates 3%–10% of neck and upper mediastinal surgeries, especially in oncologic and reoperative cases.[45] It most commonly occurs during thyroidectomy, parathyroidectomy, esophagectomy, or thoracic aortic procedures. Unilateral injury leads to hoarseness, vocal fatigue, and aspiration, whereas bilateral damage can cause severe airway obstruction requiring urgent care. Diagnosis is based on clinical evaluation and direct vocal cord visualization via flexible laryngoscopy, with CT and MRI used to detect compressive lesions. Management ranges from voice therapy and speech rehabilitation to injection laryngoplasty, and in persistent cases, advanced surgical options such as medialization thyroplasty, arytenoid adduction, or reinnervation may be necessary.[45]

c. Brachial plexus injury

Brachial plexus injury is a rare but serious complication of cardiothoracic surgery, often due to excessive arm traction, poor positioning, or direct trauma during sternal retraction. Prolonged hyperextension or aggressive retraction increases the risk.[46] Patients may experience weakness, paresthesia, or pain in the shoulder, arm, or hand, and severe cases can lead to paralysis, muscle atrophy, and limited motion. Diagnosis relies on electromyography, nerve conduction studies, and MRI–the gold standard for detecting nerve damage. CT myelography and ultrasound can also be useful. Preganglionic (root-level) injury manifests as root avulsion, empty nerve root sleeves, or pseudomeningocele; post-ganglionic injury shows nerve thickening, discontinuity, or neuroma with focal T2 hyperintensity. Chronic denervation is reflected by muscle edema progressing to fatty atrophy of the paraspinal and shoulder girdle muscles.[46,47] Treatment includes physical therapy, pain management, and surgery for severe cases, with orthotic devices providing additional support.[46]

In conclusion, non-pulmonary postoperative complications after cardiothoracic surgery can considerably affect recovery and morbidity if not identified and treated promptly. Imaging is essential for detecting and monitoring these conditions, including cardiovascular issues (e.g., graft failure, pericardial effusions), infections (e.g., mediastinitis, sternal osteomyelitis), and device-related complications (e.g., lead displacement, anastomotic leaks) (Table 1). Advanced techniques such as cardiac MRI and CTA have transformed non-invasive evaluation, enabling precise anatomical and functional assessments of vascular conduit stenosis and prosthetic valve dysfunction. These insights facilitate timely, targeted interventions.

Footnotes

Conflict of interest disclosure

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

References

1. Kindzelski BA, Bakaeen FG, Tong MZ, et al. Modern practice and outcomes of reoperative cardiac surgery. *J Thorac Cardiovasc Surg.* 2022;164(6):1755-1766.e16. [Crossref]
2. Ilcheva L, Risteski P, Tudorache I, et al. Beyond conventional operations: embracing the era of contemporary minimally invasive cardiac surgery. *J Clin Med.* 2023;12(23):7210. [Crossref]

3. Randhawa MK, Sultana S, Stib MT, Nagpal P, Michel E, Hedgire S. Role of radiology in assessment of postoperative complications of heart transplantation. *Radiol Clin North Am*. 2024;62(3):453-471. [Crossref]
4. Károlyi M, Eberhard M, Gloor T, et al. Routine early postoperative computed tomography angiography after coronary artery bypass surgery: clinical value and management implications. *Eur J Cardiothorac Surg*. 2022;61(2):459-466. [Crossref]
5. Goerne H, de la Fuente D, Cabrera M, et al. Imaging features of complications after coronary interventions and surgical procedures. *Radiographics*. 2021;41(3):699-719. [Crossref]
6. Rajiah PS, Sardá MJ, Ashwath R, Goerne H. Palliative procedures for congenital heart disease: imaging findings and complications. *Radiographics*. 2023;43(4):e220049. [Crossref]
7. Alerhand S, Adrian RJ, Long B, Avila J. Pericardial tamponade: a comprehensive emergency medicine and echocardiography review. *Am J Emerg Med*. 2022;58:159-174. [Crossref]
8. Ufuk F, Ocak I, Akpinar E, Chelala L, Landeras L, Montner S. Classic signs on chest radiographs: primer for residents. *Radiographics*. 2025;45(2):e240155. [Crossref]
9. Terry NLJ, Manapragada PP, Aziz MU, Singh SP. Review of pericardial disease on computed tomography. *J Med Imaging Radiat Sci*. 2021;52(3S):S65-S77. [Crossref]
10. Klein AL, Wang TKM, Cremer PC, et al. Pericardial diseases: international position statement on new concepts and advances in multimodality cardiac imaging. *JACC Cardiovasc Imaging*. 2024;17(8):937-988. [Crossref]
11. Smit M, Coetzee AR, Lochner A. The pathophysiology of myocardial ischemia and perioperative myocardial infarction. *J Cardiothorac Vasc Anesth*. 2020;34(9):2501-2512. [Crossref]
12. Gaudino M, Dangas GD, Angiolillo DJ, et al. Considerations on the management of acute postoperative ischemia after cardiac surgery: a scientific statement from the American Heart Association. *Circulation*. 2023;148(5):442-454. [Crossref]
13. Arslani K, Gualandro DM, Puelacher C, et al. Cardiovascular imaging following perioperative myocardial infarction/injury. *Sci Rep*. 2022;12(1):4447. [Crossref]
14. Beijnink CWH, van der Hoeven NW, Konijnenberg LSF, et al. Cardiac MRI to visualize myocardial damage after ST-segment elevation myocardial infarction: a review of its histologic validation. *Radiology*. 2021;301(1):4-18. [Crossref]
15. Roudaut R, Serri K, Lafitte S. Thrombosis of prosthetic heart valves: diagnosis and therapeutic considerations. *Heart*. 2007;93(1):137-42. [Crossref]
16. Camaj A, Fuster V, Giustino G, et al. Left ventricular thrombus following acute myocardial infarction: JACC state-of-the-art review. *J Am Coll Cardiol*. 2022;79(10):1010-1022. [Crossref]
17. Beerkens FJ, Claessen BE, Mahan M, et al. Contemporary coronary artery bypass graft surgery and subsequent percutaneous revascularization. *Nat Rev Cardiol*. 2022;19(3):195-208. [Crossref]
18. Botelho FE, Cacione DG, Leite JO, Baptista-Silva JC. Endoluminal interventions versus surgical interventions for stenosis in vein grafts following infrainguinal bypass. *Cochrane Database Syst Rev*. 2021;4(4):CD013702. [Crossref]
19. Vinciguerra M, Spadaccio C, Tennyson C, et al. Management of patients with aortocoronary saphenous vein graft aneurysms: JACC state-of-the-art review. *J Am Coll Cardiol*. 2021;77(17):2236-2253. [Crossref]
20. Demos D, McGough E. Postoperative management and complications in thoracic surgery. Thoracic Anesthesia Procedures. 2021;231. [Crossref]
21. LaPar DJ, Isbell JM, Mulloy DP, et al. Planned cardiac reexploration in the intensive care unit is a safe procedure. *Ann Thorac Surg*. 2014;98(5):1645-1651; discussion 1651-1652. [Crossref]
22. Williams ML, Sheng S, Gammie JS, Rankin JS, Smith PK, Hughes GC. Richard E. Clark Award. Aortic dissection as a complication of cardiac surgery: report from the society of thoracic surgeons database. *Ann Thorac Surg*. 2010;90(6):1812-1816; discussion 1816-1817. [Crossref]
23. Cai X, Shahandeh N, Ji J, et al. Ascending aortic pseudoaneurysm: a rare complication of transcatheter aortic valve replacement and thoracic surgery. *Circ Cardiovasc Imaging*. 2022;15(7):e014076. [Crossref]
24. Kim YW. Aortic endograft infection: diagnosis and management. *Vasc Specialist Int*. 2023;39:26. [Crossref]
25. Wilson WR, Bower TC, Creager MA, et al. Vascular graft infections, mycotic aneurysms, and endovascular infections: a scientific statement from the American Heart Association. *Circulation*. 2016;134(20):e412-e460. [Crossref]
26. Ufuk F, Hazirolan T. Cross-sectional imaging findings of cardiac outpouchings. *Diagn Interv Radiol*. 2023;29(1):68-79. [Crossref]
27. Moll FL, Powell JT, Fraedrich G, et al. Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery. *Eur J Vasc Endovasc Surg*. 2011;41(Suppl 1):S1-S58. [Crossref]
28. Buğra AK, Göde S, Buğra A, Eltutan S, Arafat Z, Şen O, Erentuğ V. Mediastinitis after cardiac surgery: risk factors and our vacuum-assisted closure results. *Kardiochir Torakochirurgia Pol*. 2021;18(4):195-202. [Crossref]
29. Jayakumar S, Khoynezhad A, Jahangiri M. Surgical site infections in cardiac surgery. *Crit Care Clin*. 2020;36(4):581-592. [Crossref]
30. McDonald EG, Aggrey G, Aslan AT, et al. Guidelines for diagnosis and management of infective endocarditis in adults: a wikiguidelines group consensus statement. Erratum in: *JAMA Netw Open*. 2023;6(10):e2341784. [Crossref]
31. Alexis SL, Malik AH, George I, et al. Infective endocarditis after surgical and transcatheter aortic valve replacement: a state of the art review. *J Am Heart Assoc*. 2020;9(16):e017347. [Crossref]
32. Hota P, Dass C, Erkmen C, Donuru A, Kumaran M. Poststernotomy complications: a multimodal review of normal and abnormal postoperative imaging findings. *AJR Am J Roentgenol*. 2018;211(6):1194-1205. [Crossref]
33. Semionov A, Kosiuk J, Ajlan A, Discepola F. Imaging of thoracic wall abnormalities. *Korean J Radiol*. 2019;20(10):1441-1453. [Crossref]
34. Cha YK, Choi MS, Bak SH, Kim JS, Kim CH, Chung MJ. Incidence and risk factors for sternal osteomyelitis after median sternotomy. *J Thorac Dis*. 2022;14(4):962-968. [Crossref]
35. Schimmer C, Sommer SP, Bensch M, Elert O, Leyh R. Management of poststernotomy mediastinitis: experience and results of different therapy modalities. *Thorac Cardiovasc Surg*. 2008;56(4):200-204. [Crossref]
36. Verma M, Pandey NN, Kumar S, Ramakrishnan S. Imaging spectrum of valvular and paravalvular complications of prosthetic heart valve at CT angiography. *Radiol Cardiothorac Imaging*. 2021;3(4):e210159. [Crossref]
37. Dipoce J, Bernheim A, Spindola-Franco H. Radiology of cardiac devices and their complications. *Br J Radiol*. 2015;88(1046):20140540. [Crossref]
38. Alandete Germán SP, Isarria Vidal S, Domingo Montañana ML, De la Vía Oraá E, Vilar Samper J. Pacemakers and implantable cardioverter defibrillators, unknown to chest radiography: review, complications and systematic reading. *Eur J Radiol*. 2015;84(3):499-508. [Crossref]
39. Gazda AJ, Kwak MJ, Akkanti B, et al. Complications of LVAD utilization in older adults. *Heart Lung*. 2021;50(1):75-79. [Crossref]
40. Mellnick VM, Raptis DA, Raptis C, Bhalla S. Imaging of left ventricular device complications. *J Thorac Imaging*. 2013;28(2):W35-41. [Crossref]

41. Lehto J, Kiviniemi T. Postpericardiotomy syndrome after cardiac surgery. *Ann Med.* 2020;52(6):243-264. [Crossref]

42. Alqahtani M, Mardigyan V, Chetrit M. The emerging role of multimodality imaging in the diagnosis and management of post pericardiotomy syndrome. *Curr Cardiol Rep.* 2025;27(1):1. [Crossref]

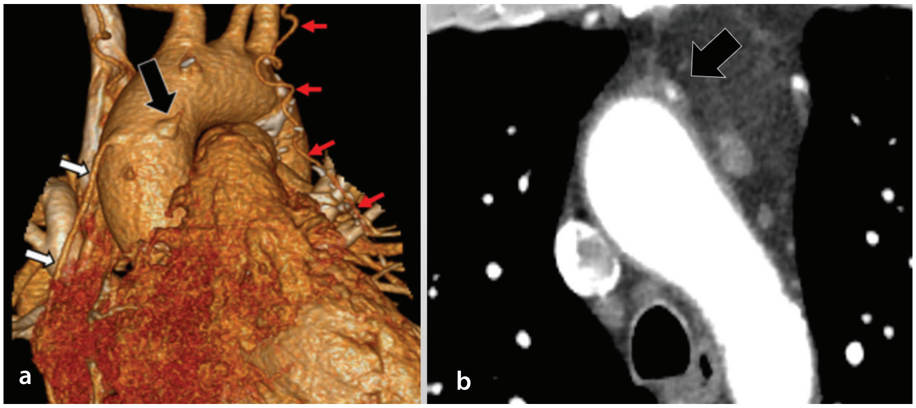
43. Sudin A, Chong C, Hassan R. Incidence and factors associated with post-sternotomy pain syndrome in the National Heart Institute, Malaysia. *J Cardiothorac Vasc Anesth.* 2024;38(2):466-474. [Crossref]

44. Nørskov J, Skaarup SH, Bendixen M, Tankisi H, Mørkved AL, Juhl-Olsen P. Diaphragmatic dysfunction is associated with postoperative pulmonary complications and phrenic nerve paresis in patients undergoing thoracic surgery. *J Anesth.* 2024;38(3):386-397. [Crossref]

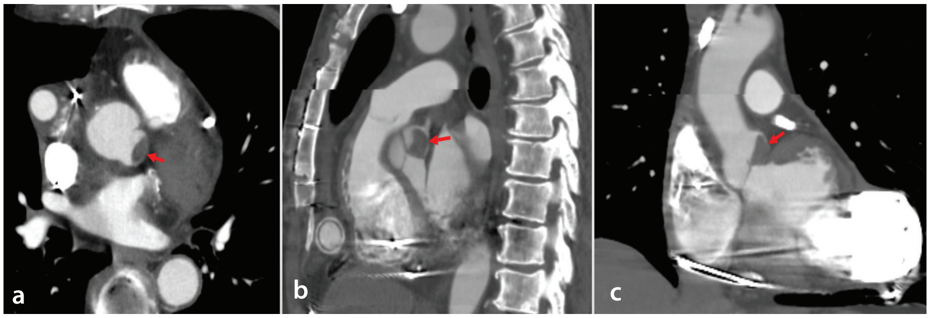
45. Taniyama Y, Okamoto H, Sato C, et al. Prevention and management of recurrent laryngeal nerve palsy in minimally invasive esophagectomy: current status and future perspectives. *J Clin Med.* 2024;13(24):7611. [Crossref]

46. Im YJ, Kang MS, Kim SW, Sung DH. Brachial plexus injury associated with median sternotomy during cardiac surgery: three cases of c8 radiculopathy due to the fracture of the first rib. *Diagnostics (Basel).* 2021;11(10):1896. [Crossref]

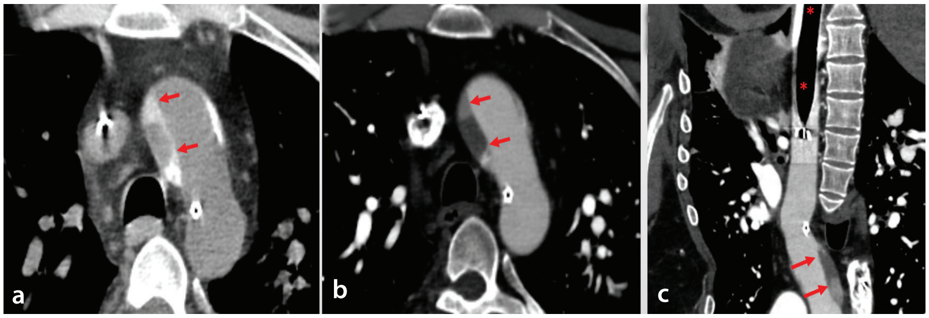
47. Zhang L, Xiao T, Yu Q, Li Y, Shen F, Li W. Clinical value and diagnostic accuracy of 3.0T multi-parameter magnetic resonance imaging in traumatic brachial plexus injury. *Med Sci Monit.* 2018;24:7199-7205. [Crossref]



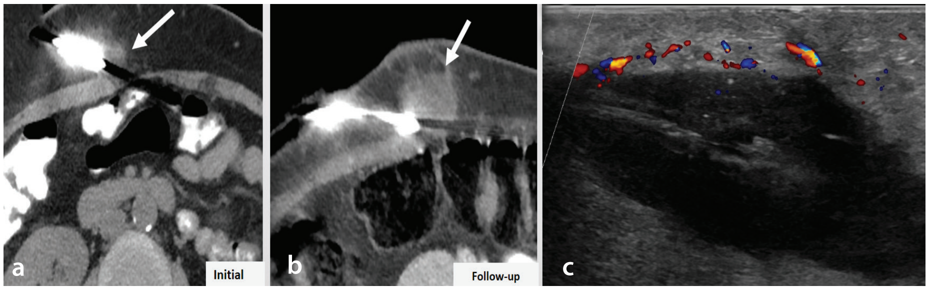
Supplementary Figure 1. A 73-year-old male status post-coronary artery bypass grafting. **(a)** Coronal three-dimensional (3D) and **(b)** axial computed tomography angiography images demonstrate a completely thrombosed saphenous vein graft (SVG) (white arrows), with only the residual most proximal portion of the graft remaining (black arrow). A patent SVG graft to the right coronary artery is seen arising from the right side of the ascending aorta (white arrows), along with a partially imaged left internal mammary artery graft (red arrows).



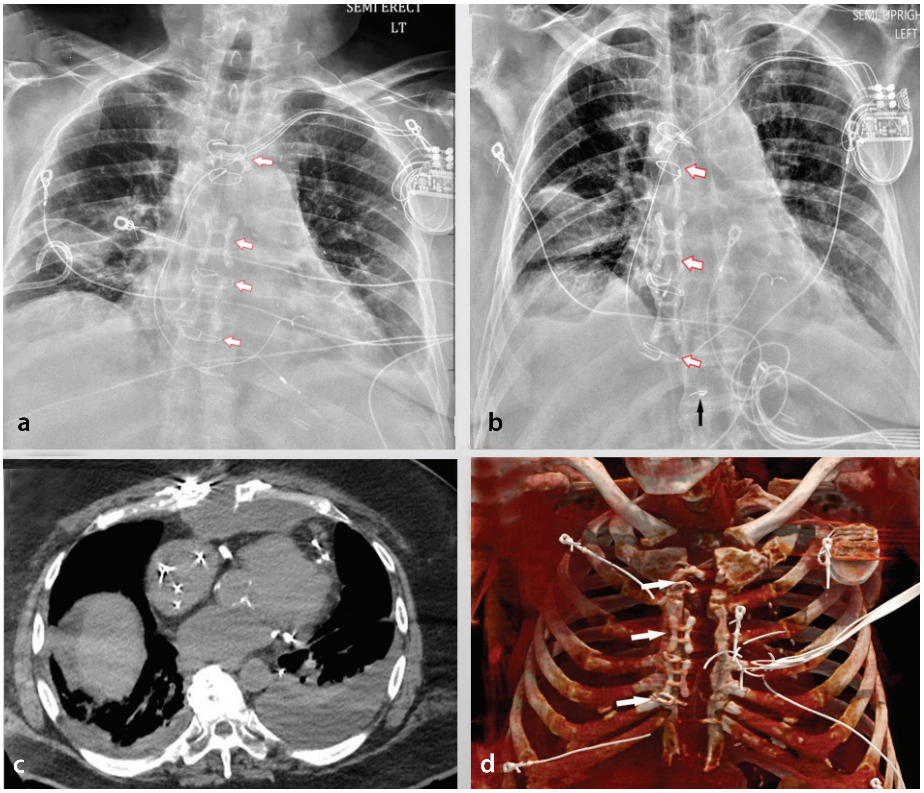
Supplementary Figure 2. A 55-year-old male status post-left ventricular assist device placement. **(a)** Axial, **(b)** sagittal, and **(c)** coronal CT angiography images demonstrate a hypodense lesion within the left sinus of Valsalva, consistent with thrombus (red arrows). CT, computed tomography.



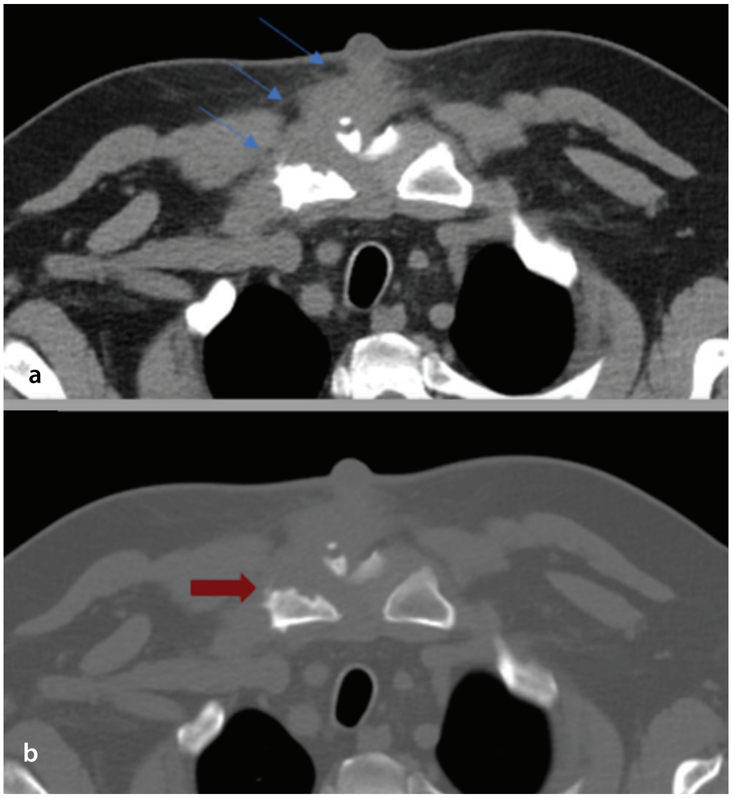
Supplementary Figure 3. A 65-year-old female status post-intra-aortic balloon pump placement. **(a)** Axial non-contrast, **(b)** axial arterial phase, and **(c)** aortic multiplanar computed tomography angiography images demonstrate an intimal defect within the left aortic arch, consistent with a mural hematoma (red arrows).



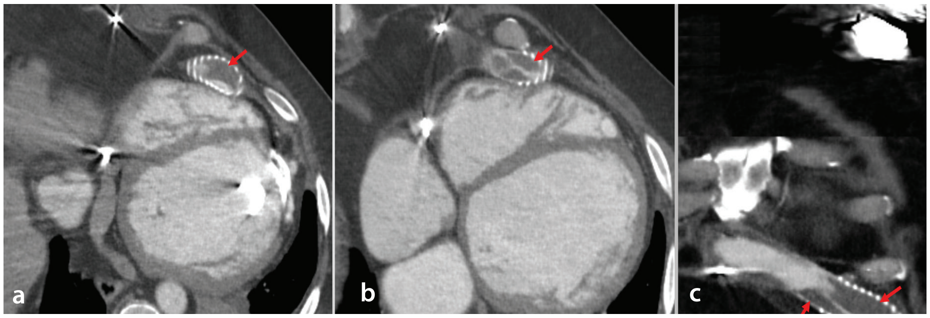
Supplementary Figure 4. A 65-year-old female status post-left ventricular assist device (LVAD) placement. **(a)** The initial axial chest computed tomography (CT) image demonstrates an unremarkable appearance of the LVAD driveline. **(b)** Follow-up axial chest CT image reveals increased soft tissue density (arrows) surrounding the LVAD driveline in the anterior abdominal wall, and **(c)** corresponding ultrasound image shows a fluid collection with surrounding echogenic fat and hyperemia, consistent with an abscess.



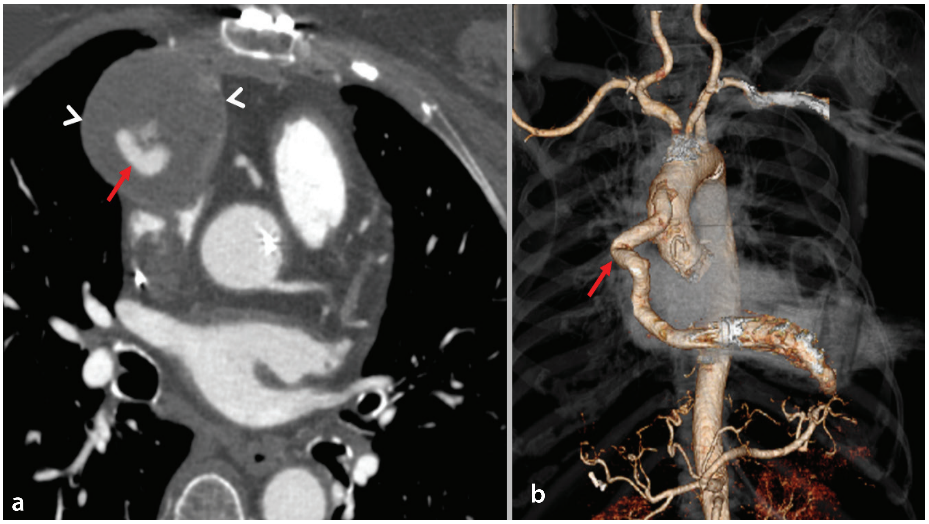
Supplementary Figure 5. A 79-year-old male status post-coronary artery bypass grafting. **(a)** Early postoperative chest radiograph demonstrates aligned sternotomy wires and plates (arrows) with expected postoperative changes and devices. **(b)** Chest radiograph obtained 1 day postoperatively shows marked rightward displacement of the sternal wires and plates (white arrows), representing a change from the initial postoperative radiograph. A fractured lower wire is also noted (black arrow). **(c)** Axial chest computed tomography (CT) image reveals bilateral pleural effusion, sternal dehiscence, and a peristernal fluid collection. **(d)** A three-dimensional (3D) volume-rendered chest CT image demonstrates sternal dehiscence with displaced sternal wires and plates (arrows).



Supplementary Figure 6. A 49-year-old male status post-heart transplantation. **(a)** Axial chest computed tomography image with mediastinal window and **(b)** bone window settings demonstrates a localized abscess surrounding the sternal manubrium, with an anterior sinus tract extending to the skin (thin blue arrows). Sclerosis and erosive changes involving the right clavicular head (red arrow), and to a lesser extent, the sternal manubrium, are better visualized on the bone window image, suggesting osteomyelitis.



Supplementary Figure 7. A 61-year-old man with left ventricular assist device (LVAD) outflow cannula stenosis due to biodebris. **(a, b)** Axial and **(c)** coronal computed tomography images demonstrate considerable luminal stenosis of the LVAD outflow cannula caused by large biodebris (arrows).



Supplementary Figure 8. A 63-year-old woman with left ventricular assist device (LVAD) outflow cannula stenosis due to kinking. **(a)** Axial computed tomography (CT) angiography and **(b)** three-dimensional volume-rendered CT angiography images demonstrate luminal stenosis of the LVAD outflow cannula caused by kinking (arrows). Surrounding fluid around the outflow cannula is also noted (arrowheads), which may suggest associated inflammation or serous collection.

DIR

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.243022

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Effectiveness of balloon angioplasty under intravascular ultrasound guidance in calcified below-the-knee tibial arteries

Muhammed Said Beşler[1]
Asiye Sözeri[2]
Murat Canyiğit[3]

[1]Necip Fazıl City Hospital, Clinic of Radiology, Kahramanmaraş, Türkiye

[2]University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Clinic of Radiology, Ankara, Türkiye

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

PURPOSE

This study aimed to assess the optimal balloon diameter for intravascular ultrasound (IVUS)-guided balloon angioplasty in calcified below-the-knee (BTK) tibial artery lesions.

METHODS

Between February 2024 and April 2024, a retrospective review was conducted on 17 patients with Rutherford category 4-6 severely calcified tibial arteries with >70% stenosis, treated with IVUS-guided balloon angioplasty. Sequentially, 3 mm and then 3.5 mm diameter balloons were inflated. The minimum lumen diameter and area were measured before and after the procedure in the proximal, mid, and distal segments of the tibial arteries. One- and three-month follow-ups were conducted using Doppler ultrasound.

RESULTS

Significant increases in lumen diameter ($P < 0.001$ for all) and lumen area ($P < 0.001$, $P = 0.003$, $P = 0.002$, respectively) were observed in the proximal, mid, and distal segments of the BTK arteries following IVUS-guided 3.5 mm balloon angioplasty. Ultra-low iodinated contrast media was used [median 2 mL (range, 1–4 mL)]. Lumen area increase ratios were similar among the proximal, mid, and distal segments ($P = 0.905$). No target vessel revascularization, major amputation, or mortality was observed during follow-up. Wound healing was seen in 62.5% of the cases with foot ulcers.

CONCLUSION

In the treatment of calcified BTK tibial arteries, a gradual balloon diameter increase up to 3.5 mm in IVUS-guided balloon angioplasty is safe and effective.

CLINICAL SIGNIFICANCE

Gradual balloon diameter increase up to 3.5 mm under IVUS guidance in calcified BTK lesions demonstrates significant potential. It enables ultra-low contrast usage, provides low complication rates, and achieves high patency and limb salvage, along with satisfactory wound healing in the short term.

KEYWORDS

Intravascular ultrasound, below-the-knee, tibial artery, calcified, ultra-low contrast, balloon angioplasty

Peripheral artery disease affects >200 million people globally.[1] The incidence of chronic limb-threatening ischemia, the severe stage of the disease, is approximately 0.3% and can result in high rates of limb loss and mortality.[2] In the treatment of below-the-knee (BTK) lesions, which are commonly seen in chronic limb-threatening ischemia cases, the primary goals are wound healing, pain relief, and limb salvage. The success of balloon angioplasty -the first-line standard treatment method for BTK lesions- is adversely affected by flow-limiting dissection and severe calcification. New methods continue to be investigated to improve the outcomes of balloon angioplasty.[3]

Corresponding author: Muhammed Said Beşler

E-mail: msbesler@gmail.com

Received 09 September 2024; revision requested 17 October 2024; accepted 23 November 2024.

Epub: 09.12.2024

Publication date: 06.11.2025

DOI: 10.4274/dir.2024.243022

You may cite this article as: Beşler MS, Sözeri A, Canyiğit M. Effectiveness of balloon angioplasty under intravascular ultrasound guidance in calcified below-the-knee tibial arteries. *Diagn Interv Radiol.* 2025;31(6):591-596.

The use of intravascular ultrasound (IVUS) has been deemed appropriate for all stages of endovascular procedures in tibial arteries, including vessel sizing, stenosis severity, and occlusion assessment.[4] Adding IVUS as an adjunct to conventional angiography in lower limb endovascular interventions improves 1-year patency rates[5] independently. Although the use of IVUS results in larger lumen measurements and consequently larger balloon sizes compared with angiography alone, balloons ranging from 2 to 3 mm in diameter are still preferred in tibial arteries to avoid vessel injury.[6,7] In BTK lesions, calcification surrounding the vessel wall is common. Although adjunctive treatments, such as atherectomy, intravascular lithotripsy, and specialty balloons, in addition to balloon angioplasty, can reduce complications and improve outcomes, they also result in increased costs and heterogeneous efficacy.[8] Although some adjunctive methods to balloon angioplasty in BTK lesions may increase limb salvage, 1-year patency rates remain similar.[9] It has been noted that increasing balloon lumen size in circumferentially calcified lesions may not carry a high risk of vessel injury.[10] The use of IVUS in lower extremity endovascular interventions significantly reduces the amount of contrast material required.[11] In tibial arteries, the use of IVUS is recommended to minimize contrast material usage.[4]

The optimal treatment for calcified BTK lesions remains unclear. In the present study, we aim to achieve effective treatment of challenging calcified BTK lesions by using IVUS to facilitate safe and incremental appropriate balloon sizing, elucidate angioplasty efficacy through changes in lumen area, and minimize contrast material usage. Based on the hypothesis that larger balloons can be used safely and effectively for the endovascular treatment of severely calcified BTK lesions with stand-alone balloon angioplasty, the study aims to determine the optimal balloon diameter.

Main points

- This is the first report demonstrating the effectiveness of gradually increasing balloon size up to 3.5 mm in calcified tibial arteries by measuring the vessel lumen area with intravascular ultrasound (IVUS) before and after the procedure.
- The use of IVUS guidance in this approach almost eliminates the need for contrast material, increasing the safety of the procedure.
- Balloon diameters larger than standard were successfully applied under IVUS guidance, treating tibial arteries effectively from proximal to distal.

Methods

Study design

This single-center, retrospective study was conducted in accordance with the Declaration of Helsinki after obtaining approval from the Ankara Bilkent City Hospital Institutional Review Board (decision number: TABED 2-24-224, date of approval: May 29, 2024), and informed consent was waived due to the study's retrospective nature. All consecutive patients who underwent IVUS-guided conventional balloon angioplasty for the treatment of severely calcified anterior tibial artery and posterior tibial artery lesions between February 2024 and April 2024 were reviewed. The study population comprised patients presenting with Rutherford category 4–6 and severely calcified lesions causing >70% stenosis on digital subtraction angiography (n = 17) (Figure 1).[12]

Intravascular ultrasound-guided balloon angioplasty procedure

After placing a 5F sheath in the ipsilateral common femoral artery, 70 IU/kg of heparin was injected intra-arterially. A CO_2 angiography from the groin to the ankle was performed. Lesions were crossed using a 0.018-inch guidewire (Gladius; Asahi Intecc, Aichi, Japan). The nominal pressure of the 3 and 3.5 mm diameter semi-compliant balloons used over a 0.014-inch guidewire is 6 atm (Minerva, Guangdong, China). The balloons were selected to be 150 or 200 mm in length to cover the target lesions. A 3 mm diameter balloon was inflated at 6–8 atm pressure for 30 seconds in the target tibial artery. In all patients, including those with chronic total occlusion, the first IVUS procedure was performed after 3 mm balloon dilation. Intraluminal imaging was conducted using a 20 MHz Eagle Eye Platinum ST catheter (Philips Volcano, San Diego, USA). Subsequently, a 3.5 mm diameter balloon was inflated at 4–6 atm pressure for 30 seconds. Afterward, intraluminal imaging was repeated with IVUS. At the end of the procedure, pedal arch patency was visualized using iodinated contrast material diluted with saline in a 1:4 ratio. No additional endovascular treatment methods were applied. Patients with wounds continued their medical treatment with a wound care specialist team. In IVUS frames, the minimum lumen area was defined as the area measured in the section with the smallest lumen area in the proximal, mid, and distal segments separately, and the mean diameter in this section was defined as the minimum lumen diameter. IVUS measurements were manually performed on the IVUS device by two radiologists with 20 years and 5 years of experience in consensus. The minimum lumen area and minimum lumen

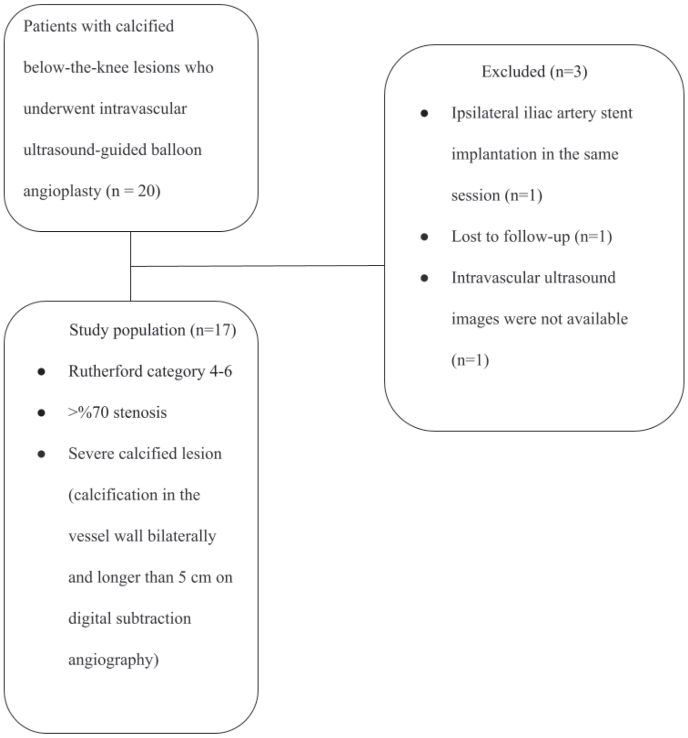


Figure 1. Flowchart of study population selection.

diameter were measured with IVUS before and after 3.5 mm balloon angioplasty in the proximal, mid, and distal segments. Due to the posterior acoustic shadowing created by severe calcifications surrounding the vessel wall, the lumen area was measured by tracing the inner boundary of the calcifications. When manual boundary drawing was performed, the area and diameter values were automatically provided as output (Figures 2-4). The severe calcification prevented the assessment of the internal and external elastic lamina; therefore, the lumen area and diameters were measured based on the inner border of the calcification. The lumen area measured before and after the procedure for the proximal, mid, and distal segments were compared to determine the lumen increase ratio. Thus, the efficacy of the 3.5 mm balloon angioplasty was assessed. In the IVUS images, vessel wall calcification was classified as ≥180°–<270° or ≥270° degrees circumferential. In the one- and three-month follow-up imaging, a peak systolic velocity of <200 cm/s in Doppler ultrasound was considered indicative of no significant restenosis.

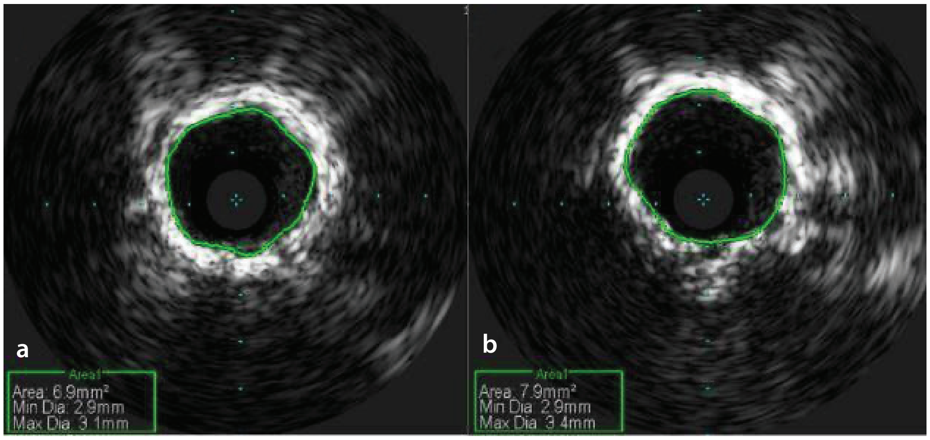


Figure 2. Measurement of minimum lumen diameter and area in intravascular ultrasound analysis of the calcified lesion in the proximal segment of the posterior tibial artery before **(a)** and after **(b)** 3.5 mm balloon angioplasty.

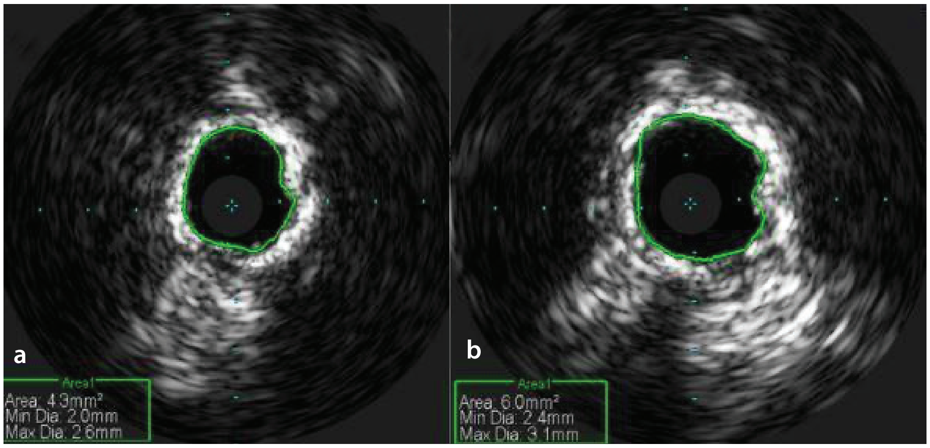


Figure 3. Measurement of minimum lumen diameter and area in intravascular ultrasound analysis of the calcified lesion in the mid segment of the posterior tibial artery before **(a)** and after **(b)** 3.5 mm balloon angioplasty.

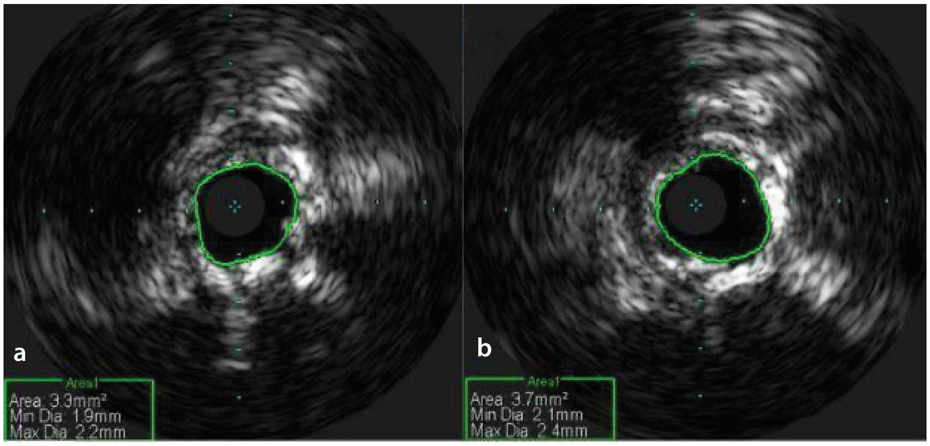


Figure 4. Measurement of minimum lumen diameter and area in intravascular ultrasound analysis of the calcified lesion in the distal segment of the posterior tibial artery before **(a)** and after **(b)** 3.5 mm balloon angioplasty.

Statistical analysis

Statistical analyses were performed using SPSS version 26.0 (IBM, IL, USA). The Shapiro–Wilk test was used to determine whether the data followed a normal distribution. Continuous variables were expressed as mean ± standard deviation and median (range). Comparisons of pre- and post-procedure minimum lumen area and minimum lumen diameter in the proximal, mid, and distal tibial artery segments were performed using Student's t-test. The Kruskal–Wallis test was used to compare the increase in minimum lumen area post-procedure among these three segments. A P value of <0.05 was considered statistically significant.

Results

Baseline characteristics

In the study population of 17 patients (88.2% male, mean age 67.9 ± 8 years), the mean follow-up duration was 134.1 ± 12 days (Table 1). All lesions were classified as category D according to the Trans-Atlantic Inter-Society Consensus classification.

Procedural details

No complications, such as flow-limiting dissection, distal embolization, or vascular perforation, were observed during the procedure. IVUS revealed that 61.9% of the arteries had circumferential calcification of ≥270°. Significant increases in lumen diameter ($P < 0.001$ for all) and lumen area ($P < 0.001$, $P = 0.003$, $P = 0.002$, respectively) were observed in the proximal, mid, and distal segments of the BTK arteries following IVUS-guided 3.5 mm balloon angioplasty (Table 2). Mean lumen diameters of the BTK arteries in the proximal, mid, and distal segments following IVUS-guided 3.5 mm balloon angioplasty increased by 0.47 mm, 0.53 mm, and 0.45 mm, respectively. However, there were no signif-

Table 1. Basic characteristics of the study population (17 patients and 21 lesions)

Age (years)	67.9 ± 8
Gender (male)	15 (88.2)
Hypertension	9 (52.9)
Diabetes mellitus	13 (76.5)
Chronic kidney disease	2 (11.8)
Coronary artery disease	9 (52.9)
Cerebrovascular ischemic event	2 (11.8)
Dyslipidemia	9 (52.9)
Smoking history (pack-years)	30 (range, 0–70)
Current smoker	5 (29.4)
Rutherford category 5/6	8 (47.1)
Chronic total occlusion	9 (42.9)
Lesion length (mm)	209 ± 44
Lesion location	
Posterior tibial artery	9 (42.9)
Anterior tibial artery	12 (57.1)
Procedure time (min)	44.1 ± 11.5

Variables are presented as mean ± standard deviation and number (percentage).

Table 2. Pre- and post-procedure intravascular ultrasound measurements

	Pre-procedure	Post-procedure	*P* value
Proximal			
Minimum lumen diameter (mm)	2.78 ± 0.42	3.25 ± 0.34	<0.001
Minimum lumen area (mm^2)	5.99 ± 1.79	8.16 ± 1.59	<0.001
Mid			
Minimum lumen diameter (mm)	2.32 ± 0.29	2.85 ± 0.40	<0.001
Minimum lumen area (mm^2)	4.21 ± 1.06	6.39 ± 1.87	0.003
Distal			
Minimum lumen diameter (mm)	2.16 ± 0.24	2.61 ± 0.37	<0.001
Minimum lumen area (mm^2)	3.56 ± 0.83	5.26 ± 1.46	0.002

Values are presented as mean ± standard deviation.

icant differences in the lumen area increase ratios among the proximal, mid, and distal segments ($P = 0.905$). Iodinated contrast material was only used for visualizing the pedal arch at the end of the procedure (median 2 mL; range 1–4 mL). The pedal arch was patent in all patients.

Follow-up outcomes

Foot ulcers were present in 47.1% of the patients, and 62.5% of these wounds healed completely during the short-term follow-up. All patients in Rutherford category 4 showed an improvement of ≥1 point. At the 4th month post-procedure, in a patient who experienced an increase in wound severity, the target vessel remained patent, but the occluded stents in the superficial femoral artery and external iliac artery were recanalized endovascularly. During the follow-up period, no restenosis, target vessel revascularization, or amputation was observed. Follow-up Doppler ultrasounds confirmed that all target vessels remained patent. No cases of contrast-induced nephropathy were observed.

Discussion

In the endovascular treatment of calcified BTK lesions with balloon angioplasty, determining the optimal balloon diameter to prevent vascular complications is a primary focus. In severely calcified lesions of the anterior tibial artery and posterior tibial artery, sequential balloon angioplasty with 3 mm and then 3.5 mm diameters under IVUS guidance achieved better lumen expansion without vascular complications. Short-term follow-up showed high patency rates, satisfactory wound healing, and effective limb salvage.

In studies investigating the efficacy of balloon angioplasty in BTK lesions under IVUS guidance, the rate of severe calcification in tibial arteries ranges from 60% to 90%.[6,7] Fujihara et al.[13] defined severe calcification as ≥180° in the superficial femoral artery using IVUS. In this study, all selected lesions were already severely calcified according to digital subtraction angiography, and all had ≥180° calcification on IVUS. Moreover, vessel wall calcification was more severe in three out of five patients, reaching ≥270°. The increased complexity of calcified lesions has led to the use of various methods, such as atherectomy, intravascular lithotripsy, and specialty balloons; however, study results are highly heterogeneous, and there is no demonstrated superiority over stand-alone balloon angioplasty.[9,14] Modifications to balloon angioplasty techniques, including incremental balloon oversizing, are also being explored. If balloon diameter oversizing is to be performed, a gradual approach is recommended to avoid flow-limiting dissection due to excessive stretching.[8] Balloon oversizing can lead to significant dissection in lesions with spotty calcification, whereas in circumferentially calcified lesions, it shows limited dissection and maximizes lumen gain.[10] Peripheral vascular procedures are safer with IVUS use and reduce the risk of vascular complications.[15] IVUS has shown that tibial artery diameters are larger compared with angiographic assessment, thus avoiding the use of undersized balloons and suboptimal lumen gain.[16] However, even in IVUS-guided studies, balloons typically in the 2–3 mm diameter range are preferred due to the risk of vessel injury.[6,7] Normal tibial artery diameters in men aged >60 years are approximately 3–3.5 mm and can exceed 4 mm.[17] In an IVUS study, the mean diameter of the proximal and distal segments of BTK arteries was found to be 4 ± 1 mm.[16] In light of these findings, a luminal diameter of ≥3 mm was targeted, aiming for resolution by increasing the balloon diameter. In the current study, the mean lumen diameter did not reach 3 mm with 3 mm balloon angioplasty, but significant lumen area expansion was achieved without complications after 3.5 mm balloon angioplasty. Following 3 mm balloon angioplasty, an IVUS assessment revealed that the minimum luminal diameters in all segments remained below even the targeted minimum value of 3 mm, leading to the use of a 3.5 mm balloon. Significant lumen enlargement was achieved in all segments of calcified tibial arteries with IVUS-guided incremental balloon oversizing. Effective and safe treatment was provided with balloon angioplasty as the primary treatment method without the need for additional cost-incurring methods. The rate of lumen enlargement provided by balloon angioplasty was similar in the proximal, mid, and distal segments. This indicates the applicability of the 3.5 mm oversized balloon to all segments and is beneficial for procedural simplicity by using the same balloon.

In the clinical research conducted by Brahmandam et al.[5], the 1-year patency rate in lower limb peripheral artery interventions, particularly in groups with chronic limb-threatening ischemia, long segment lesions, multiple stenoses/occlusions, or severe calcification, was positively impacted by IVUS assistance, exceeding 97%. In a similar current study population with severely calcified lesions, characterized by an average lesion length >20 cm and nearly half having chronic total occlusions, all target vessels remained patent at the short-term follow-up. A meta-analysis examining the mid-term effects of IVUS guidance in lower extremity peripheral artery disease endovascular interventions demonstrated a reduction in major amputation rates. However, it did not show a significant reduction in target lesion revascularization and mortality.[18] IVUS-guided endovascular intervention for BTK lesions positively influences wound healing and can reduce the healing period from 4–5 months to 3 months. The wound healing rate was observed to be approximately 60% by the 3rd–4th month of follow-up.[6,7] In this study, no mortality was observed in the short-term follow-up, and there was no need for target vessel revascularization or major amputation. Similarly, wound healing of >60% was observed during a mean follow-up of 4 months. Although short-term results are favorable, mid- and long-term follow-up is necessary.

Stenotic and occlusive disease of the tibial arteries is frequently seen in patients who are diabetic, and renal dysfunction is among the relative contraindications for endovascular procedures.[19] Although CO_2 angiography is successfully applied in above-the-knee arteries, it is not sufficient alone at the BTK level due to fragmentation and overestimation of stenosis degree.[20] Studies on BTK balloon angioplasty treatments have reported the use of an average of 60–130 mL of iodinated contrast material.[7,21] The use of IVUS can help mitigate these issues by reducing iodinated contrast material usage. In the current study population, where three out of four patients were diabetic, the use of IVUS-guided ultra-low contrast material (median 2 mL) completely avoided the risk of contrast nephropathy and preserved renal function, which is of great value.

The study has some limitations. It is a single-center, retrospective study with a small population size. Furthermore, there was no control group. Nominal pressure was used for inflation in all balloon angioplasties; therefore, the impact of high-pressure inflation on the outcomes could not be assessed. Due to the relatively short follow-up period, mid- and long-term follow-up results could not be presented; however, this does not affect the reporting of the primary outcomes of the study, such as complications and lumen gain. Selection bias is inevitable as the approach targets the severely calcified lesion group. Advancing the IVUS catheter in narrow lumens can be challenging; however, as the initial IVUS imaging in all cases was performed after 3 mm balloon angioplasty, no such issue was encountered. Although IVUS imaging inevitably extends the procedure time, no specific duration was measured for this in the study. To the best of our knowledge, this study is the first report to demonstrate the effectiveness of a gradual balloon size increase up to 3.5 mm in diameter by measuring the vessel lumen area before and after the procedure in calcified tibial arteries, almost eliminating the use of contrast material with IVUS guidance. Balloon diameters above the standard were used effectively and successfully with IVUS guidance from the proximal to the distal tibial arteries. In the future, randomized controlled multicenter prospective long-term follow-up studies will be beneficial for determining the optimal lumen diameter increase in IVUS-guided interventions for calcified BTK lesions and for standardizing the appropriate treatment approach.

In conclusion, the search for the optimal treatment of calcified BTK lesions is ongoing. Gradual balloon diameter increase up to 3.5 mm under IVUS guidance in calcified BTK lesions has demonstrated effectiveness and success by providing ultra-low contrast usage, low complication rates, high patency, and high limb salvage rates in short-term follow-up, along with satisfactory wound healing.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Shu J, Santulli G. Update on peripheral artery disease: epidemiology and evidence-based facts. *Atherosclerosis*. 2018;275:379-381. [CrossRef]
2. Kwong M, Rajasekar G, Utter GH, Nuño M, Mell MW. Updated estimates for the burden of chronic limb-threatening ischemia in the Medicare population. *J Vasc Surg*. 2023;77(6):1760-1775. [CrossRef]
3. Li J, Varcoe R, Manzi M, et al. Below-the-knee endovascular revascularization: a position statement. *JACC Cardiovasc Interv*. 2024;17(5):589-607. [CrossRef]
4. Secemsky EA, Aronow HD, Kwolek CJ, et al. Intravascular ultrasound use in peripheral arterial and deep venous interventions: multidisciplinary expert opinion from SCAI/AVF/AVLS/SIR/SVM/SVS. *J Vasc Interv Radiol*. 2024;35(3):335-348. [CrossRef]
5. Brahmandam A, Kim TI, Parziale S, et al. Intravascular ultrasound use is associated with improved patency in lower extremity peripheral arterial interventions. *Ann Vasc Surg*. 2024;106:410-418. [CrossRef]
6. Soga Y, Takahara M, Ito N, et al. Clinical impact of intravascular ultrasound-guided balloon angioplasty in patients with chronic limb threatening ischemia for isolated infrapopliteal lesion. *Catheter Cardiovasc Interv*. 2021;97(3):376-384. [CrossRef]
7. Fujihara M, Yazu Y, Takahara M. Intravascular ultrasound-guided interventions for below-the-knee disease in patients with chronic limb-threatening ischemia. *J Endovasc Ther*. 2020;27(4):565-574. [CrossRef]
8. Giannopoulos S, Varcoe RL, Lichtenberg M, et al. Balloon angioplasty of infrapopliteal arteries: a systematic review and proposed algorithm for optimal endovascular therapy. *J Endovasc Ther*. 2020;27(4):547-564. [CrossRef]
9. Nugteren MJ, Welling RHA, Bakker OJ, Ünlü Ç, Hazenberg CEVB. Vessel preparation in infrapopliteal arterial disease: a systematic review and meta-analysis. *J Endovasc Ther*. 2024;31(2):191-202. [CrossRef]
10. Welling RHA, de Borst GJ, van den Heuvel DAF, et al. Vessel calcification patterns should determine optimal balloon size strategy in below the knee angioplasty procedures. *Eur J Vasc Endovasc Surg*. 2020;60(4):636-637. [CrossRef]
11. Pesonen LO, Knol MC, Shelgikar CS, Battaglia MA, Heidenreich MJ, Aziz A. Efficacy and utilization of intravascular ultrasound in

office-based lower extremity endovascular interventions for peripheral arterial disease. *Vascular*. 2024;32(1):57-64. [CrossRef]

12. Rocha-Singh KJ, Zeller T, Jaff MR. Peripheral arterial calcification: prevalence, mechanism, detection, and clinical implications. *Catheter Cardiovasc Interv*. 2014;83(6):212-220. [CrossRef]
13. Fujihara M, Kozuki A, Tsubakimoto Y, et al. Lumen gain after endovascular therapy in calcified superficial femoral artery occlusive disease assessed by intravascular ultrasound (CODE Study). *J Endovasc Ther*. 2019;26(3):322-330. [CrossRef]
14. Benfor B, Sinha K, Lumsden AB, Roy TL. Scoping review of atherectomy and intravascular lithotripsy with or without balloon angioplasty in below-the-knee lesions. *J Vasc Surg Cases Innov Tech*. 2023;9(2):101185. [CrossRef]
15. Sheikh AB, Anantha-Narayanan M, Smolderen KG, et al. Utility of intravascular ultrasound in peripheral vascular interventions: systematic review and meta-analysis. *Vasc Endovascular Surg*. 2020;54(5):413-422. [CrossRef]
16. Shammas NW, Shammas WJ, Jones-Miller S, et al. Optimal vessel sizing and understanding dissections in infrapopliteal interventions: data from the idissection below the knee study. *J Endovasc Ther*. 2020;27(4):575-580. [CrossRef]
17. Lorbeer R, Grotz A, Dörr M, et al. Reference values of vessel diameters, stenosis prevalence, and arterial variations of the lower limb arteries in a male population sample using contrast-enhanced MR angiography. *PLoS One*. 2018;13(6):0197559. [CrossRef]
18. Tsukagoshi J, Shimoda T, Yokoyama Y, et al. The mid-term effect of intravascular ultrasound on endovascular interventions for lower extremity peripheral arterial disease: a systematic review and meta-analysis. *J Vasc Surg*. 2024;79(4):963-972. [CrossRef]
19. Spiliopoulos S, Del Giudice C, Manzi M, Reppas L, Rodt T, Uberoi R. CIRSE standards of practice on below-the-knee revascularisation. *Cardiovasc Intervent Radiol*. 2021;44(9):1309-1322. [CrossRef]
20. Taha AG, Saleh M, Ali H. Safety and efficacy of hybrid angiography in chronic lower extremity ischemia patients at risk of contrast-induced acute kidney injury. *Vasc Endovascular Surg*. 2022;56(2):180-189. [CrossRef]
21. Del Giudice C, Gandini R. Six-month angiographic and clinical outcomes of therapeutic ultrasound pretreatment associated with plain balloon angioplasty for below-the-knee lesions in patients with critical limb ischemia: a prospective, single-center pilot study. *J Endovasc Ther*. 2019;26(2):191-198. [CrossRef]

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253191

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Comparison of the novel simultaneous biplane versus in-plane imaging technique in ultrasound-guided biopsy: a prospective randomized multi-operator cross-over phantom study

Baki Akca[1,2]
Florian Vafai-Tabrizi[1,2]
Michel Bielecki[3]
Georg-Christian Funk[1,2]

[1]Klinik Ottakring - Wiener Gesundheitsverbund, Medical Department II, Division of Pneumology, Vienna, Austria

[2]Karl-Landsteiner-Institute, Department of Lung Research and Pulmonary Oncology, St. Pölten, Austria

[3]University of Zurich Faculty of Medicine, Department of Epidemiology, Biostatistics and Prevention, Zurich, Switzerland

Corresponding author: Baki Akca

E-mail: b.akca@gmx.net, baki.akca@gesundheitsverbund.at

Received 08 January 2025; revision requested 09 February 2025; accepted 25 March 2025.

Epub: 12.05.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253191

PURPOSE

To evaluate and compare the in-plane and novel biplane imaging techniques in ultrasound-guided biopsies (USBx). USBx are effective for obtaining tissue samples in suspected malignancy or infection. The in-plane technique is the gold standard, offering continuous needle visualization. The biplane technique enables simultaneous in-plane and out-of-plane visualization, potentially improving biopsy outcomes. A study was conducted using gel phantoms to simulate USBx, with the goal of determining whether one technique offers distinct advantages over the other.

METHODS

A total of 30 participants (mean age: 30 ± 7 years; 20 men) were recruited, primarily consisting of physicians in training with varying levels of experience. Each participant performed biopsies on gel phantoms using both the in-plane and biplane techniques in a randomized order after watching a standardized tutorial video. Procedure-related parameters were analyzed, and post-intervention questionnaires, including the NASA task load index (NASA-TLX), were collected to assess cognitive workload and personal preferences.

RESULTS

All participants achieved successful biopsies with both techniques. The first-puncture success rate was significantly higher with the biplane technique (83% vs. 63%; $P = 0.01$). The biplane technique required significantly fewer biopsy attempts than the in-plane approach (37 vs. 43; $P = 0.03$). Although the biplane technique had a longer "mean time to first successful biopsy" (120 seconds vs. 72 seconds), this difference was not statistically significant ($P = 0.09$), likely due to high variability. No significant differences were found in safety-related parameters, including the number of skin punctures, needle retractions, percentage of time the needle tip was visible, and the number of biopsy attempts without needle tip visualization. The NASA-TLX indicated higher mental demand with the biplane technique ($P = 0.013$), but other dimensions showed no significant differences. Overall, 83% of participants, including 88% of more experienced operators, preferred the biplane technique, citing enhanced visualization and perceived safety.

CONCLUSION

In this study, the biplane technique in USBx was substantially superior in terms of total biopsy attempts and first-puncture success rate compared with the in-plane approach. It may offer safety and efficiency advantages, particularly for less-experienced operators. Further studies with larger sample sizes and experienced operators, especially in clinical settings, are needed to determine clear superiority.

CLINICAL SIGNIFICANCE

These findings suggest that biplane imaging may be especially beneficial for training less-experienced operators and in cases with elevated complication risk.

KEYWORDS

Biplane imaging, handheld ultrasound device, in-plane vs biplane, phantom study, ultrasound, ultrasound guided biopsy

You may cite this article as: Akca B, Vafai-Tabrizi F, Bielecki M, Funk GC. Comparison of the novel simultaneous biplane versus in-plane imaging technique in ultrasound-guided biopsy: a prospective randomized multi-operator cross-over phantom study. *Diagn Interv Radiol.* 2025;31(6):597-604.

Ultrasound is an easily available and safe imaging modality, offering a wide range of applications in daily clinical practice. While depicting the anatomical structures of the examined area for diagnostic purposes, it can also provide reliable guidance in procedures such as vascular punctures or biopsies.[1-4]

Ultrasound-guided biopsy (USBx) allows for minimally invasive diagnosis with high success rates and a favorable safety profile. It can be used to diagnose thoracic, abdominal, neck, and musculoskeletal pathologies.[2-5] In cases of malignancy, molecular profiling can also be easily achieved, supporting the modern era of targeted therapy.[4,6,7] The diagnostic yields for tissue identification and molecular profiling are comparable to those reported for computed tomography-guided biopsy methods.[7] In addition to its high success rate and safety, USBx offers several other advantages: it does not require sedation, involves no radiation exposure for patients and staff, can be safely performed in patients with poor performance status, and presents economic benefits over other diagnostic procedures.[7-10]

In standard ultrasound devices, depending on the specific case and preferred approach, the operator can choose between the in-plane and out-of-plane techniques. In the context of vascular access, the in-plane approach aligns with the long axis of the vessel, while the out-of-plane approach corresponds to the short axis. The probe is parallel to the needle in the in-plane technique and perpendicular to it in the out-of-plane technique.[11] Despite its widespread use, the in-plane approach requires precise alignment, which can be challenging in anatomically complex regions or for less-experienced operators.

Main points

- While the in-plane approach remains the gold standard in ultrasound-guided biopsies (USBx), the novel biplane technique—which provides simultaneous in-plane and out-of-plane views in real time—was substantially superior in terms of total attempts and first-time success rate, potentially offering safety and efficiency advantages.
- According to the post-intervention questionnaire, the majority of participants expressed a preference for the biplane technique over the in-plane approach.
- Based on these results, biplane imaging appears to be the method of choice not only for less-experienced operators but also for USBx procedures with a higher expected risk of complications.

A novel option is biplane imaging, which allows for simultaneous visualization of both axes in real-time, combining the advantages of both imaging techniques without the need to rotate the probe. By simultaneously displaying the needle trajectory and the lesion with its surrounding structures, the biplane approach appears to offer clear advantages over other well-established imaging techniques. The ability to visualize both in-plane and out-of-plane views at once may reduce complications, improve accuracy, and enhance operator confidence.

Although previous studies have explored biplane imaging for vascular catheterization and other specific interventions, its utility in USBx remains underexplored.[12,13] Research into vascular access suggests that the multiplane—or biplane—approach may be a safer and more reliable technique, particularly for less-experienced operators.

This study compares the biplane and in-plane techniques in simulated USBx performed on phantoms, evaluating differences and potential superiority between the two techniques.

Methods

This randomized, multi-operator, crossover study was conducted in the Department of Pulmonology at the [Klinik Ottakring, Vienna]. Operators from different departments with varying levels of experience in ultrasound-guided diagnostics and interventions were included in the study: medical students, trainee doctors, physicians, radiologists, and specialists. Each participant went through four steps.

Pre-intervention phase (steps 1 and 2)

First, participants completed a pre-intervention questionnaire regarding their basic characteristics and level of experience. Operators with prior experience using the Butterfly ultrasound device or the biplane technique were excluded to ensure unbiased skill evaluation.

In step 2, participants were asked to watch a standardized 10-minute educational tutorial video. The video introduced the objectives of the study, presented relevant background information, and lastly, explained the key steps for performing successful USBx using either technique. The video could be watched more than once, and there was no time limit for this phase. Afterward, participants were allowed to ask questions; however, study assistants were only permitted to answer questions addressed in the tutorial video.

Intervention phase (step 3)

In this phase, the actual procedures were performed. It consisted of two intervention sessions, in which the operator had to successfully perform a biopsy using one of the two techniques—biplane or in-plane—at a time. Both interventions were performed in succession. To minimize learning effects, the sequence of techniques was randomized for each participant using an online randomization tool.

Each intervention phase ended when the operator successfully obtained a biopsy using the assigned technique. After completing the first intervention, the operator performed the procedure using the other technique. The entire intervention phase was recorded on camera, allowing for precise analysis and measurement of the relevant parameters using a video editor.

Post-intervention phase (step 4)

After both interventions, participants were asked to complete a post-intervention questionnaire, including the modified NASA task load index (NASA-TLX) protocol[14], to assess personal preferences and experiences. Using the NASA-TLX, the workload for both intervention phases was quantified on a scale from 0 (low) to 20 (high) across six categories: mental, physical, and temporal demand, performance, effort, and level of frustration.

Ultrasound device, biopsy needle system and phantom

The novel Butterfly iQ3 (Butterfly Network, Inc., Burlington, MA, USA) device with its standard probe was used for both imaging techniques: in-plane and biplane. In biplane mode, the needle is visible in-plane on one half of the screen, while the altitude can be adjusted simultaneously in the perpendicular (out-of-plane) view on the other half of the screen (Figures 1 and 2). The biopsy itself was performed using a semi-automatic firing core biopsy needle (BARD Mission®, Disposable Core Biopsy Instrument, 18 G × 10 cm, adjustable throw of 10 or 20 mm). All participants were instructed to use a 20 mm throw.

The phantom models used in this study were self-made, composed of gelatine, and constructed similarly to those used in several other studies.[15-17] The gelatine solution was standardized across batches to ensure

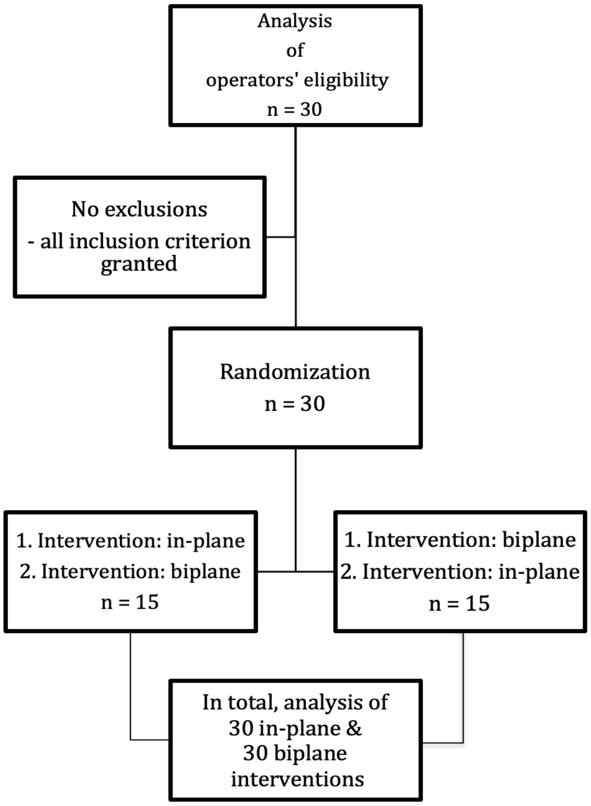


Figure 1. Operator recruitment and randomization.

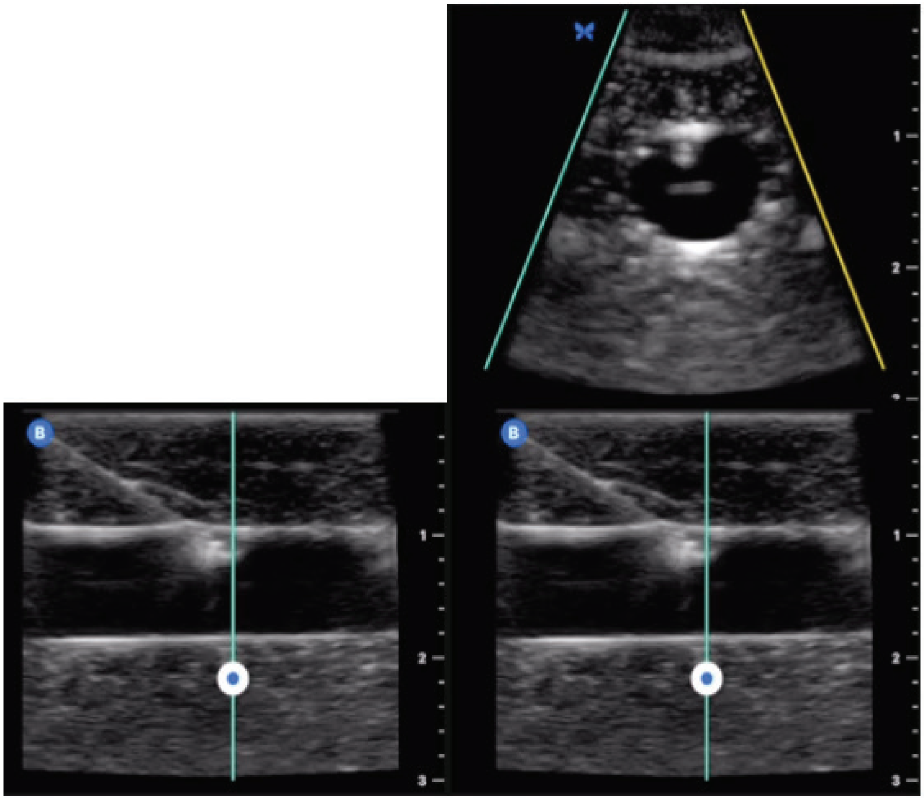


Figure 2. Both techniques illustrated. Left: in-plane mode. Right: biplane mode. The probe and needle are simultaneously displayed in both axes in real time.

uniform consistency, and lesions were uniformly embedded at a depth of 5 cm. The lesions within the phantom were made from Play-Doh, a modeling compound for children (Figure 2). The diameter of each lesion was approximately 1 cm. Soluble food coloring was added to the gelatine solution to prevent target visualization without ultrasound (Figures 2 and 3).

The transducer of the Butterfly iQ3 ultrasound device enables the combination of phased, curved, and linear arrays within a single probe, using Ultrasound-on-Chip™ technology to acquire images at a frequency range of 1–12 MHz. Unlike traditional ultrasound devices that rely on piezoelectric crystals, Ultrasound-on-Chip™ integrates thousands of transducer elements directly onto a semiconductor-based micro-electro-mechanical systems array, replacing bulky piezoelectric transducers with a more compact, software-driven solution.

Outcome variables

During the intervention phase, several parameters were measured and assessed to compare both techniques. The primary outcomes were: "time to first successful biopsy," defined as the time from when the probe was first placed on the phantom until the biopsy needle was triggered and a sample successfully obtained; "number of biopsy attempts"; and "first puncture success rate."

To evaluate the safety profile and potential complications, several surrogate parameters were also considered: "percentage of time with needle tip visualization," "number of biopsy attempts," "number of skin punctures," "number of biopsy attempts without needle tip visualization," and "number of times the cutting biopsy needle was retracted within each attempt."

The post-intervention questionnaire was used to assess the preferred technique for USBx and to evaluate workload differences across all six NASA-TLX categories: mental, physical, and temporal demand; performance; effort; and frustration.

Statistical analysis

All statistical analyses were performed using Stata (version 17.0). Data processing and analysis scripts (.do files), along with the dataset, are available at the GitHub repository: https://github.com/kushiel42/butterfly_paper

Figure 3. Intervention setting. The tablet screen displays the biplane mode. The top half of the screen shows the short-axis (out-of-plane) view, while the bottom half depicts the long-axis view. The operator uses the biplane mode to perform an ultrasound-guided biopsy in the presented gel phantom.

Descriptive statistics were used to summarize participant characteristics and procedural parameters. Categorical variables were presented as counts and percentages, as shown in Tables 1 and 2. Continuous variables were reported as medians with interquartile ranges (IQRs) or means with standard deviations (SDs), depending on data distribution assessed via visual inspection.

The primary exposure variable was the biopsy technique used—either in-plane or biplane. Outcome variables included time to first successful biopsy, total number of biopsy attempts, first puncture success rate, and NASA-TLX scores.

Given the paired nature of the data and the sample size, non-parametric statistical tests were applied. The Wilcoxon signed-rank test was used to compare paired continuous or ordinal variables between the two techniques. This test was appropriate due to the small sample size and the ordinal or non-normally distributed nature of several variables. Variables analyzed with the Wilcoxon signed-rank test included time to first successful biopsy, the total number of biopsy attempts, percentage of time the needle tip was visible, percentage of time the needle was partially visible, number of skin punctures, number of needle retractions and number of biopsy attempts without needle tip visualization.

McNemar's test was used to compare paired categorical data—specifically, first puncture success rates between the in-plane and biplane techniques. This test is suitable for analyzing dichotomous outcomes in paired samples.

Mixed-effects linear regression models were employed to account for repeated measurements and intra-participant variability in NASA-TLX scores. Each NASA-TLX dimension (mental demand, physical demand, temporal demand, performance, effort, and frustration) was modeled separately, with the biopsy technique as a fixed effect and participant ID as a random effect. No additional covariates were included, as the randomized crossover design inherently controlled for potential confounders. A two-tailed *P* value < 0.05 was considered statistically significant for all analyses. Effect sizes and 95% confidence intervals (CIs) were reported where applicable to enhance result interpretation.

Ethics approval

The institutional review board waived the need for formal approval. Informed consent was obtained from all participants.

Results

A total of 30 participants (Table 1) were included in the study, with a mean age of 30.3 years (± 7 years). The cohort comprised 20 men (67%) and 10 women (33%). The majority were trainees, including medical students and junior doctors, each group representing 20% of the participants. Only 13% (n = 4) were specialists; 80% (n = 24) of participants had advanced training in ultrasound diagnostics.

Analysis of procedural metrics (Table 2) revealed no significant differences between the in-plane and biplane techniques across several parameters. Although the time to first successful biopsy was longer with the biplane technique (120 seconds vs. 72 seconds, *P* = 0.096), this difference was not statistically significant, likely due to variability among participants. The percentage of time the needle tip was visible was similar between the two techniques (59% for in-plane vs. 61% for biplane; *P* = 0.909). Likewise, the percentage of time the needle was partially visible showed no significant difference (*P* = 0.885). Safety-related measures, including the number of skin punctures (*P* = 0.833), needle retractions (*P* = 0.563), and biopsy attempts without needle tip visualization (*P* = 0.433), were also comparable. These findings suggest that both techniques demonstrate similar profiles in terms of procedural and safety parameters.

The median time to first successful biopsy was comparable between the in-plane (69.5 seconds; IQR: 44–109 seconds) and biplane techniques (73.5 seconds; IQR: 50–128 seconds). Mean times were 84.07 seconds (SD: 50.62) for in-plane and 107.8 seconds (SD: 86.69) for biplane. Although the biplane technique exhibited a higher mean time, the difference was not statistically significant (*P* = 0.096; Wilcoxon signed-rank test), likely due to substantial variability in the data. The Bland–Altman plot (Figure 4) highlights this variability, showing considerable overlap in times between the two techniques.

Efficacy of biopsy techniques

Despite the absence of significant differences in the previously mentioned parameters, the biplane technique demonstrated superior efficacy in key outcome measures (Figure 5). The total number of biopsy attempts required was significantly lower with the biplane technique compared with the in-plane technique. Participants required fewer attempts to achieve a successful biopsy using the biplane method, indicating greater procedural efficiency (*P* = 0.030).

Moreover, the biplane technique achieved a significantly higher first puncture success rate (83%) compared with the in-plane technique (63%, *P* = 0.01). McNemar's test confirmed the statistical significance of this difference (*P* = 0.01), supporting the rejection of the null hypothesis that there is no difference in first puncture success rates between the techniques. These results highlight the enhanced efficacy of the biplane technique in achieving successful biopsies on the initial attempt (Figure 6).

Although the biplane technique proved more effective, it was associated with increased mental workload for practitioners.

Table 1. Baseline characteristics of all participants

Characteristics	Number of participants (n = 30)
Age	30.3 ± 7
Sex	
Men	20 (67%)
Women	10 (33%)
Field of practice	
Medical student	7 (23.3%)
Junior doctors	6 (20%)
Trainee doctors	13 (43.3%)
Specialists	4 (13.3%)
Field of expertise	
Students	7 (23.3%)
Physicians	6 (20%)
Radiologists	2 (6.7%)
Internal medicine (various specialties)	13 (43.3%)
Anesthesiologists	1 (3.3%)
Surgeons	1 (3.3%)
Advanced training in ultrasound diagnostics	24 (80%)
Number of diagnostic ultrasound examinations performed	
Never	2 (6.7%)
<25 times	7 (23.3%)
<50 times	5 (16.7%)
<100 times	5 (16.7%)
>100 times	11 (36.7%)
Subjectively estimated level of ultrasound diagnostic skills	
Not confident	9 (30%)
Little confident	10 (33.3%)
Confident	9 (30%)
Very confident	2 (6.7%)
Advanced training in USBx	4 (13.3%)
Number of USBx performed	
Never	17 (56.7%)
<25 times	9 (30%)
<50 times	3 (10%)
>50 times	1 (3.3%)
Subjectively estimated level of USBx	
Not confident	20 (66.7%)
Little confident	5 (16.7%)
Confident	5 (16.7%)
Very confident	0 (0 %)
Subjectively estimated level of ultrasound Overall ultrasound skill level (1 = high, 5 = low)	
1	2 (7%)
2	6 (20%)
3	6 (20%)
4	14 (47%)
5	2 (7%)

USBx, ultrasound-guided biopsy.

The NASA-TLX revealed that mental demand was significantly higher for the biplane technique than for the in-plane technique. Mean mental demand scores were 10.8 for the biplane and 9.1 for the in-plane. A mixed-effects linear regression model, accounting for repeated measures within participants, demonstrated that this difference was statistically significant (95% CI: 0.357–3.043, *P* = 0.013), indicating lower mental demand scores with the in-plane technique. The higher NASA-TLX mental demand scores for the biplane technique suggest a steeper learning curve, which may lessen with experience.

Discussion

The simultaneous display of both axes in real time may inherently suggest the superiority of the novel biplane approach over the gold standard in-plane technique in USBx. Although this study appears to be unique in investigating the role of biplane imaging in USBx, several studies have explored its application in other interventions. In ultrasound-guided regional anesthesia, biplane imaging has been reported to decrease procedure time and the number of attempts and needle passes, improve block success, and enhance safety by reducing the risk of unintended intraneural, intrapleural, or intravascular injection.[18] More commonly, studies examining the biplane technique in the context of vascular access have reported improved performance and feasibility, fewer puncture attempts and needle redirections, and a lower incidence of complications.[12,13,18-20] Similar findings have also been observed in other, more specific ultrasound-guided interventions.[21,22]

To confidently claim the superiority of one technique over the other, it is essential first to identify the factors that determine a technique's effectiveness. A technique is considered superior based on three key criteria: efficiency, safety, patient comfort and convenience, and the difficulty of execution. While all participants successfully conducted biopsies using both techniques, it is important to note that the biopsies were performed on phantoms with no pre-specified limit on attempts. Each operator was allowed to take their time until one biopsy was successfully secured in each intervention phase. Therefore, the total number of biopsy attempts stands out as a potential indicator of superiority. The biplane technique performed substantially better in this regard. This may also suggest a safer approach, as fewer biopsy attempts reduce the likelihood of complications. The first puncture success rate was also

Table 2. Procedural parameters and outcomes for in-plane and biplane techniques

Parameters	In-plane	Biplane	*P* value
Time to first successful biopsy	72 sec	120 sec	0.09
First puncture success rate	19/30 (63%)	25/30 (83%)	**0.01**
Percentage of time with needle tip (mean)	59%	61%	0.43
Number of biopsy attempts (in total)	43	37	**0.03**
Number of skin punctures	52	52	0.83
Number of biopsy attempts without needle tip visualization (in total)	13	10	0.89
Number of retractions of the cutting biopsy needle (total)	88	108	0.56
Preferred technique	5/30 (17%)	25/30 (83%)	
NASA-TLX (0 = low, 20 = high)			
Mental demand	9.1	10.8	**0.013**
Physical demand	4.0	4.2	>0.05
Temporal demand	5.7	5.9	>0.05
Performance	6.2	6.5	>0.05
Effort	6.6	7.2	>0.05
Frustration	5.6	5.6	>0.05

NASA-TLX, NASA task load index.

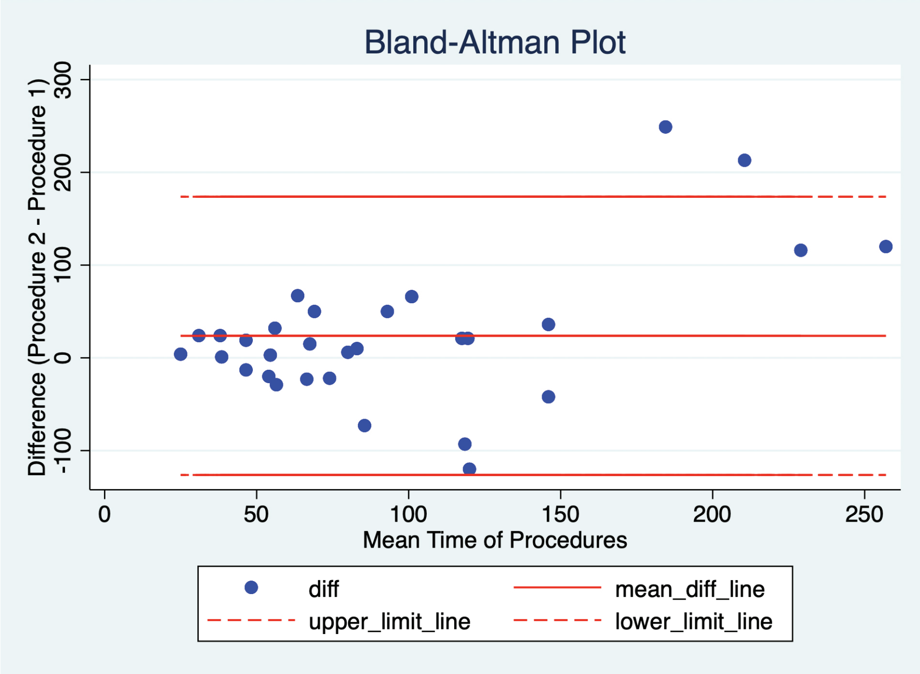


Figure 4. Bland-Altman plot of time differences between in-plane and biplane techniques.

statistically significant, indicating a higher safety profile, greater ease of execution, and improved patient comfort. This provides further support in favor of the biplane technique.

While there was no significant difference in the time to achieve the first successful biopsy individually, the mean total time notably favored the in-plane technique, possibly indicating greater efficiency and a less complex procedure overall. A faster method could be particularly relevant for unstable patients or those with poor performance status. However, as indicated by the post-intervention questionnaire, this result may be influenced by limited operator experience with the biplane technique. The additional out-of-plane axis in biplane mode introduces a mental challenge, requiring extra time for optimal needle placement and increasing procedural complexity compared with the in-plane approach. The post-intervention questionnaire further supports this, showing a statistically significant increase in mental demand for the biplane mode. These findings may, therefore, be misleading, as increased practice and familiarity with the biplane approach could yield different results.

In fact, another study comparing in-plane and biplane ultrasound-guided central venous catheterization found that participants who had considerably more training and were familiar with the biplane technique achieved shorter times to first successful catheterization using the biplane mode.[13] This suggests that the biplane mode does not, in fact, complicate USBx. Therefore, given the limitations of the study, the difference in "time to achieve the first successful biopsy" should be considered neither significant nor meaningful.

If taken into account, one could argue that a shorter time to first successful biopsy with the in-plane approach might imply a higher safety profile due to the reduced in situ duration of the biopsy needle. However, this, too, can be reasonably dismissed, as the additional axis view in the biplane mode arguably reduces the risk of complications—even with longer procedure times.

Based on the points made thus far, it can be concluded that the biplane mode not only seems to be superior in terms of safety but also appears to be at least as efficient as, if not more efficient than the in-plane mode—especially with more experience, considering the results of the other biplane study.[13] Interestingly, other surrogate parameters for patient safety, such as the percentage of time the needle tip is visualized or the number of retractions during biopsy, showed no significant differences. However, these results may vary in real-life settings.

The additional out-of-plane axis view, often cited as the main argument for a higher safety profile, can be disputed, as experienced operators are able to assess the perpendicular axis (out-of-plane) while remaining in-plane, using basic probe-tilting motions and adjusting the angle during the biopsy. Some of the more experienced participants cited this as the main reason why they saw no additional benefit in the biplane technique and, therefore, preferred the in-plane approach. However, there may be a status quo bias and a potential benefit even for highly experienced operators due to the continuous perpendicular axis view without requiring additional maneuvers—particularly in complicated biopsy cases. Nevertheless, this argument remains inconclusive in this study, as the number of very experienced operators was limited. It also remains unclear whether experienced radiologists with advanced manual skills would find the extra axis view distracting. However, it is likely that this factor would become increasingly irrelevant with more practice.

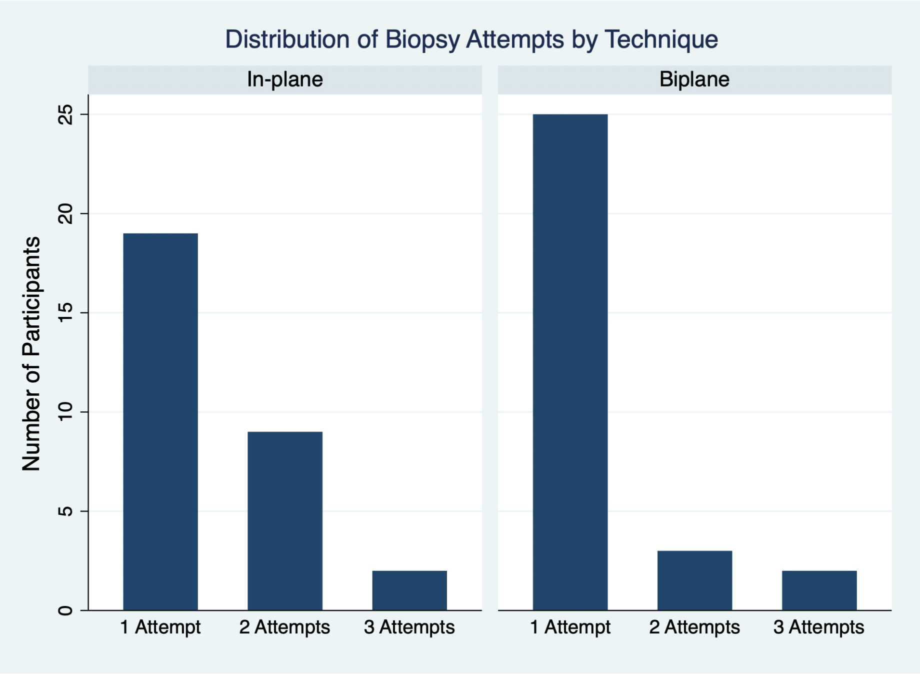


Figure 5. Distribution of biopsy attempts by technique. This figure shows the number of biopsy attempts required for both the in-plane and biplane techniques. The majority of participants in the biplane group (n = 21) required only one attempt, compared with the in-plane group, where 17 participants succeeded on the first attempt. A higher proportion of in-plane participants required two or more attempts, indicating greater variability in success rates with this method.

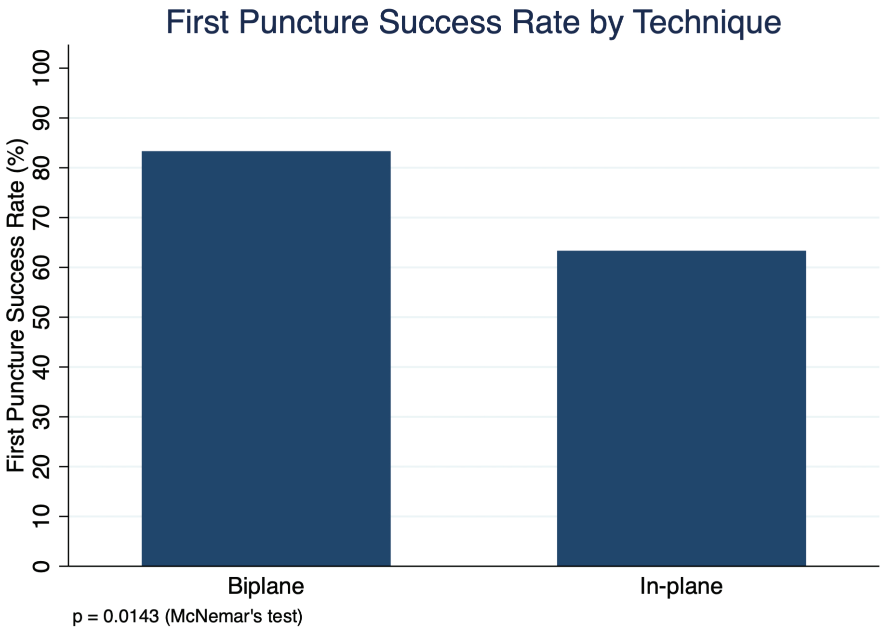


Figure 6. First puncture success rate by technique. This figure compares the first puncture success rates between the in-plane and biplane techniques. The biplane technique demonstrated a significantly higher first puncture success rate (83%) compared with the in-plane technique (63%), as confirmed by McNemar's test (*P* = 0.01).

One could argue that there might have been a learning effect or that the randomized order of execution methods— in-plane then biplane versus biplane then in-plane—may have influenced the outcome of this study. However, this assumption was disproven statistically. No significant differences based on the order of intervention were observed.

In terms of subjective impressions, the post-intervention questionnaire revealed that the majority favored the biplane approach for USBx. They reported that although the biplane technique felt more mentally demanding—due to lack of experience and unfamiliarity with three-dimensional thinking—they also felt safer and substantially more confident compared with the in-plane technique, due to the additional information provided. While very experienced operators may argue that the simultaneous additional axis is unnecessary, the biplane technique appears to be the method of choice for less experienced users. This conclusion was also drawn in the previously mentioned study comparing single-plane and biplane ultrasound-guided central venous catheterization.[13]

Another key criterion is patient comfort. Although the study was conducted on phantoms and definitive conclusions cannot be drawn, it is likely that with sufficient practice and experience using the biplane technique, patient comfort would be comparable to that of biopsies performed using the in-plane approach. In fact, given the expected lower complication rates and higher first-puncture success rate associated with the biplane approach, patient comfort may even be substantially improved by comparison.

In addition to the arguments above, it is also important to consider the limitations of the present study. First and foremost, the sample size and the lack of experienced ultrasound operators substantially limit the ability to determine clear superiority between the two techniques. The study's generalizability is restricted by the high proportion of trainee doctors, as only 13% of participants were specialists with extensive experience. The biopsies were conducted on phantoms and did not fully replicate real-life settings or actual patients. It should also be emphasized that only the biopsy of solid lesions was simulated; non-mass or partly solid lesions were not examined, which further limits the generalizability of the findings. However, it is reasonable to assume that real-time imaging of both short- and long-axis views may aid in navigating around and avoiding the puncture of critical structures such as nerves, vessels, bones, muscles, and fascial planes, potentially reducing complication rates even in non-mass biopsies.[18]

Future studies should consider incorporating tissue models or clinical trials to better validate these findings. Another limitation of the current study is that only the freehand technique was used. Consequently, the potential drawbacks or advantages of the biplane approach in procedures where ultrasound imaging is performed separately from the biopsy have not been explored.

Overall, considering the results of this study and those of the other referenced in-plane versus biplane study, there appears to be a clear trend favoring the biplane technique—at least in the hands of less-experienced operators. Nevertheless, due to the limitations and objections outlined above, absolute superiority cannot be asserted with certainty. However, it can be reasonably argued that while experienced operators may benefit only marginally from the additional axis, the biplane approach appears to be the method of choice for beginners and intermediate users. Incorporating biplane imaging into training programs for less-experienced operators could enhance procedural success rates and safety. Although the increased cognitive effort required for biplane imaging may initially discourage adoption, this challenge could be mitigated through targeted training and continued practice.

In conclusion, in this study, the biplane technique in USBx was substantially superior in terms of total biopsy attempts and first-puncture success rate when compared with the in-plane approach, potentially offering safety and efficiency advantages—particularly in the hands of less-experienced operators. Further research involving larger sample sizes, varying levels of operator experience, and real-world clinical settings is essential to confirm the potential superiority of biplane imaging in USBx.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Taylor RW, Palagiri AV. Central venous catheterization. *Crit Care Med*. 2007;35(5):1390-1396. [Crossref]
2. Schmidt GA, Blaivas M, Conrad SA, et al. Ultrasound-guided vascular access in critical illness. *Intensive Care Med*. 2019;45(4):434-446. [Crossref]
3. Bourgouin PP, Rodriguez KJ, Fintelmann FJ. Image-guided percutaneous lung needle biopsy: how we do it. *Tech Vasc Interv Radiol*. 2021;24(3):100770. [Crossref]
4. Kiranantawat N, McDermott S, Fintelmann FJ, et al. Clinical role, safety and diagnostic accuracy of percutaneous transthoracic needle biopsy in the evaluation of pulmonary consolidation. *Respir Res*. 2019;20(1):23. [Crossref]
5. Otto R. Interventional ultrasound. *Eur Radiol*. 2002;12(2):283-287. [Crossref]
6. Sundaralingam A, Aujayeb A, Akca B, et al. Achieving molecular profiling in pleural biopsies: a multicenter, retrospective cohort study. *Chest*. 2023;163(5):1328-1339. [Crossref]
7. Livi V, Sotgiu G, Cancellieri A, et al. Ultrasound-guided needle aspiration biopsy of superficial metastasis of lung cancer with and without rapid on-site evaluation: a randomized trial. *Cancers (Basel)*. 2022;14(20):5156. [Crossref]
8. Koegelenberg CF, von Groote-Bidlingmaier F, Bolliger CT. Transthoracic ultrasonography for the respiratory physician. *Respiration*. 2012;84(4):337-350. [Crossref]
9. Stigt JA, Groen HJ. Percutaneous ultrasonography as imaging modality and sampling guide for pulmonologists. *Respiration*. 2014;87(6):441-451. [Crossref]
10. Diacon AH, Schuurmans MM, Theron J, Schubert PT, Wright CA, Bolliger CT. Safety and yield of ultrasound-assisted transthoracic biopsy performed by pulmonologists. *Respiration*. 2004;71(5):519-522. [Crossref]
11. Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med*. 2011;364(8):749-757. [Crossref]
12. Panidapu N, Babu S, Koshy T, Sukesan S, Dash PK, Panicker VT. Internal jugular vein cannulation using a 3-dimensional ultrasound probe in patients undergoing cardiac surgery: comparison between biplane view and short-axis view. *J Cardiothorac Vasc Anesth*. 2021;35(1):91-97. [Crossref]
13. Li YY, Liu YH, Yan L, et al. Single-plane versus real-time biplane approaches for ultrasound-guided central venous catheterization in critical care patients: a randomized controlled trial. *Crit Care*. 2023;27(1):366. [Crossref]
14. Backhaus T, von Cranach M, Brich J. Ultrasound-guided lumbar puncture with a needle-guidance system: A prospective and controlled study to evaluate the learnability and feasibility of a newly developed approach. *PLoS One*. 2018;13(4):e0195317. [Crossref]
15. Bude RO, Adler RS. An easily made, low-cost, tissue-like ultrasound phantom material. *J Clin Ultrasound*. 1995;23(4):271-273. [Crossref]
16. Silver B, Metzger TS, Matalon TA. A simple phantom for learning needle placement for sonographically guided biopsy. *AJR Am J Roentgenol*. 1990;154(4):847-848. [Crossref]
17. Ng SY, Kuo YL, Lin CL. Low-cost and easily fabricated ultrasound-guided breast phantom for breast biopsy training. *Appl Sci*. 2021;11(16):7728. [Crossref]
18. Hernandez N, Sen S, de Haan JB, Haskins S, Pawa A. The novel use of biplane imaging for ultrasound-guided regional anesthesia. *Korean J Anesthesiol*. 2022;75(3):286-289. [Crossref]
19. Convissar D, Bittner EA, Chang MG. Biplane imaging using portable ultrasound devices for vascular access. *Cureus*. 2021;13(1):e12561. [Crossref]
20. Scholten HJ, Meesters MI, Montenij LJ, Korsten EHM, Bouwman RA; 3DUI Study group. 3D biplane versus conventional 2D ultrasound imaging for internal jugular vein cannulation. *Intensive Care Med*. 2022;48(2):236-237. [Crossref]
21. Chalouhi GE, Guenuec A, Rameh G, Hamze H, Salomon LJ, Ville Y. Biplane mode for more precise intrauterine procedures. *Am J Obstet Gynecol*. 2022;226(2):215-219. [Crossref]
22. Siebert FA, Hirt M, Niehoff P, Kovács G. Imaging of implant needles for real-time HDR-brachytherapy prostate treatment using biplane ultrasound transducers. *Med Phys*. 2009;36(8):3406-3412. [Crossref]

DIR

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253263

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Efficacy and safety of percutaneous thermal ablation in Bosniak III and IV cystic renal masses: a systematic review and meta-analysis

Hans-Jonas Meyer[1]
Timo Christian Meine[2]
Manuel Florian Struck[3]
Silke Zimmermann[4]

[1]University of Leipzig, Department of Diagnostic and Interventional Radiology, Leipzig, Germany

[2]Hannover Medical School, Department of Diagnostic and Interventional Radiology, Hannover, Germany

[3]University Hospital Leipzig, Department of Anesthesiology and Intensive Care Medicine, Leipzig, Germany

[4]University Hospital Leipzig, Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, Leipzig, Germany.

PURPOSE

Local thermal ablation is considered a standard treatment for small kidney masses. However, few studies have investigated the efficacy and safety of thermal ablation for cystic kidney masses.

METHODS

The MEDLINE library, Cochrane, and SCOPUS databases were screened for studies investigating the efficacy of thermal ablation for cystic renal masses, comprising studies between 1995 and February 2024. In total, seven studies were deemed suitable and included in the present analysis.

RESULTS

The studies included a total of 113 participants with 134 cystic renal masses. The sample sizes ranged from 5 to 38 participants. There were 76 men (67.2%) and 37 women (32.8%), with a mean age of 64.7 years (range: 50 to 75.4 years). Overall, 55 cystic masses were classified as Bosniak III (41%) and 79 as Bosniak IV (59%). Technical success of local thermal ablation was reported in 133 cystic masses (99.2%). The pooled meta-analytic technical success rate was 100% [95% confidence interval (CI): 96%–100%, $I^2 = 0.0\%$]. Complications were reported in 9 cases (6.7%). According to the Society of Interventional Radiology classification system, there were 3 major complications (2.6%) and 6 minor complications (5.3%). The pooled meta-analytic complication rate was 10% (95% CI: 5%–20%, $I^2 = 40\%$). No tumor recurrence was reported during follow-up.

CONCLUSION

Local thermal ablation can be considered a highly effective and safe procedure for cystic kidney masses. Most studies were performed using radiofrequency ablation, underscoring the need for further studies on alternative ablation techniques such as microwave ablation and cryoablation.

CLINICAL SIGNIFICANCE

Local thermal ablation is an effective and safe procedure for treating cystic kidney masses.

KEYWORDS

Meta-analysis, systematic review, percutaneous thermal ablation, renal cell carcinoma

Renal cell carcinoma (RCC) is the most common malignant renal tumor, representing over 30,000 cases every year in the United States of America (USA).[1] Besides surgical resection, a promising treatment option for small RCCs (stage 1) is percutaneous thermal ablation, most commonly performed with radiofrequency ablation (RFA) and microwave ablation (MWA).[2-6]

Percutaneous thermal ablation is a widely accepted minimally invasive treatment for surgically unresectable RCC, with good published results in small solid renal tumors and a high reported efficacy, with complete ablation rates ranging from 90% to 100%.[4,7] However, not all malignant renal tumors are solid masses, as cystic renal masses can also harbor malignancy.[8,9]

Corresponding author: Hans-Jonas Meyer

E-mail: hans-jonas.meyer@medizin.uni-leipzig.de

Received 29 January 2025; revision requested 25 February 2025; accepted 25 March 2025.

Epub: 18.08.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253263

You may cite this article as: Meyer HJ, Meine TC, Struck MF, Zimmermann S. Efficacy and safety of percutaneous thermal ablation in Bosniak III and IV cystic renal masses: a systematic review and meta-analysis. *Diagn Interv Radiol.* 2025;31(6):605-611.

According to the renal cyst classification, Bosniak III or IV cystic lesions may carry a particular risk for malignancy, even if some of these lesions (particularly Bosniak III) are proven to be benign after biopsy or surgery.[8-10] Based on the Bosniak category, the risks of malignancy for Bosniak III and IV lesions were 16%–100% and 90%–100%, respectively.[8,11]

Due to the high risk of malignant transformation in Bosniak III and IV cysts, thermal ablation has been utilized as a possible treatment option instead of surgery.[12] Partial nephrectomy remains the most commonly used treatment for Bosniak III and IV cysts. In some cases with high perioperative morbidity, active imaging-based surveillance can be justified.

Yet, the cyst composition, particularly the presence of a solid tumor component, could influence the outcome of thermal ablation.[13] One can assume that the differing tissue composition of cystic renal masses induces variable heating effects during local thermal ablation and may also impact the outcome of this treatment modality. However, reliable systematic data comparing these aspects between cystic and solid renal masses have not yet been published. Moreover, although thermal ablation is included in guidelines for the treatment of solid renal masses, these recommendations cannot yet be translated to cystic renal masses.[3]

It is noteworthy that only a few retrospective studies have investigated the efficacy and safety of thermal ablation in cystic renal masses, and a recent overview of the published literature is needed. Notably, no previous systematic review and meta-analysis have been conducted to investigate the outcomes of thermal ablation in cystic renal masses. Therefore, the purpose of the present systematic review and meta-analysis was to elucidate the efficacy and safety of local thermal ablation in participants with cystic kidney masses.

Main points

- Local thermal ablation of cystic renal masses is an effective and safe procedure.
- The data is mainly comprised for radiofrequency ablation and few microwave ablation ablations, whereas there is no data for cryoablation.
- Prospective evaluations are needed.

Methods

The institutional review board approved the meta-analysis.

The present analysis is an analysis of published results for which no ethics approval and informed consent is required

Data acquisition

The MEDLINE library, Cochrane, and SCOPUS databases were screened for studies investigating local thermal ablation of cystic renal masses. No other sources were used for data acquisition. The timeframe of the study search included studies between 1995 and February 2024. The paper acquisition process is summarized in Figure 1.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (2020) was used for the analysis.[14] The following search words were used: "radiofrequency ablation" OR "microwave ablation" OR "local thermal ablation" AND "renal mass" OR "cystic renal mass" OR "renal cyst."

The primary endpoints of the systematic review were the technical success rate of thermal ablation and the complication rate. Studies (or subsets of studies) were included if they satisfied the following criteria: (1) cystic renal mass, (2) treatment by local thermal ablation with RFA or MWA, and (3) reported technical success rate defined by complete coverage of the ablation zone of the cystic renal mass. The exclusion criteria were as follows: (1) systematic reviews, (2) case reports, (3) non-English language, and (4) solid renal mass. In total, seven studies were included in this analysis.[12,15-20]

Data extraction

Data extraction was performed by two authors (HJM and SZ), followed by an independent evaluation of the extracted data for correctness (MFS). For each study, details regarding study design, year of publication, country of origin, participant number, participant characteristics (age and sex), histopathological diagnosis, ablation type, ablation time, and Bosniak category[8] were extracted.

Quality assessment

The quality of the included studies was assessed using the Newcastle–Ottawa Scale (NOS).[21] Study quality assessment was conducted by two authors (SZ and HJM). In cases

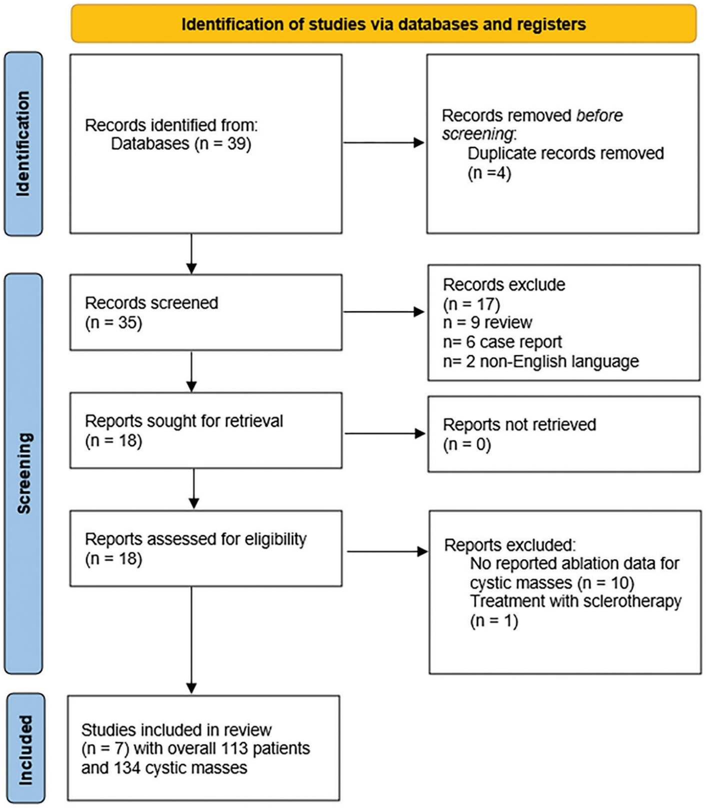


Figure 1. PRISMA flow chart (2020) providing an overview of the paper acquisition process. Overall, seven studies with 113 participants and 134 treated cystic masses were included in the analysis. PRISMA, The Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

of disagreement, consensus was reached by a third author. The NOS assesses the quality of studies based on the selection of cases, comparability of the cohort, and outcome assessment of exposure to risks. A score of 0–9 was assigned to each study, and a score ≥6 was considered indicative of high quality.

Statistical analysis

The meta-analysis was performed using RevMan 5.3 (2014; Cochrane Collaboration, Copenhagen, Denmark). Heterogeneity was assessed using the inconsistency index I^2.[22,23] The proposed thresholds for I2 according to the Cochrane Handbook are as follows: 0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; 75% to 100%: considerable heterogeneity.[24] DerSimonian and Laird[25] random-effects models with inverse variance weights were applied without further correction.

Results

Quality of the included studies

All included studies had a retrospective design. Table 1 provides an overview of the included studies. The overall risk of bias for the current research question can be considered low, indicated by high NOS values among the studies, with scores of 8 points (Table 2). However, there is a lack of information regarding long-term oncological outcomes in the included studies.

Participants

The included studies comprised a total of 113 participants with 134 cystic masses. The sample sizes ranged from 5 to 38 participants. There were 76 men (67.2%) and 37 women (32.8%), with a mean age of 64.7 years (range: 50 to 75.4 years). Three studies (42.7%) were conducted in North America (USA), two studies (28.5%) in Asia (South Korea), one study (14.4%) in South America (Brazil), and one study (14.4%) in Europe (Italy).

In total, 55 cystic masses (41.0%) were classified as Bosniak III and 79 cystic masses (59.0%) as Bosniak IV. The mean lesion size was 2.5 cm, ranging from 1.1 to 10.1 cm. In five studies (71.4%), RFA was used to treat 122 cystic masses (91.0%), whereas in two studies (28.6%), MWA was used to treat 12 cystic masses (9.0%).

Only one study,[20] involving 5 cystic masses, employed an aspiration technique prior to ablation; all other studies inserted the needle directly into the cystic mass. Three studies did not perform a pre-ablation biopsy.[15,18,19] Regarding histopathological confirmation, 60 of the 134 masses (44.4%) had a definitive diagnosis. Clear-cell carcinoma was diagnosed in 44 cystic masses, papillary carcinoma in 6 masses, and 10 lesions were diagnosed as undifferentiated carcinoma. The remaining 75 cystic masses (55.6%) had no definitive tumor diagnosis.

Technical success

Technical success of local thermal ablation was reported in 133 of 134 cystic masses (99.2%). The pooled meta-analytic technical success rate was 100% [95% confidence interval (CI): 96%–100%, $I^2 = 0\%$; Figure 2]. No tumor recurrence was reported after ablation in the included studies. A potential selection bias should be acknowledged when interpreting the technical success rate.

Complication rate

Complications were reported in 9 cases (6.7% of ablations). The pooled meta-analytic complication rate was 10% (95% CI: 0.05%–20%, $I^2 = 40\%$; Figure 3). Park et al.[19] reported two iatrogenic pneumothoraces requiring chest tube placement. In a previous study by Park et al.[18], four complications were reported: one arteriovenous fistula, one case of inguinal paresthesia, and two pneumothoraces. Allen et al.[12] reported three complications: one major case of flash pulmonary edema requiring emergency department transfer, and two minor complications–dysuria and mild hydronephrosis due to a blood clot in the ureter, both managed conservatively.

Using the classification system of the Society of Interventional Radiology,[26] there were 3 major complications (2.6% of all cases) and 6 minor complications (5.3% of all cases).

Discussion

This study is the first systematic review and meta-analysis on the success and complication rates of treating cystic renal masses with thermal ablation techniques. As shown, the technical success rate can be considered very high, with a reported rate of 100%. Moreover, the procedure can be considered safe, with a complication rate of 10% and no major events. In two cases, iatrogenic chest tube placement was needed due to pneumothoraces. No major bleeding events or cyst ruptures were reported in the included studies. The present results are therefore well comparable to published results in solid renal masses.[4,6]

However, a possible selection bias should be acknowledged for the very high success rate, as potential negative cases may not have been published in the literature. There will be cases in clinical routine with cystic components that are too large, and without full ablation coverage of the lesion. This should be kept in mind when interpreting the present results.

Notably, the studies included in this analysis did not report the size limits for which local ablation can still be considered treatable, raising further concerns about selection bias.

In a large recent meta-analysis, outcomes of RFA and MWA were reported for 2,258 ablations.[27] As in the present analysis, all studies were retrospective in nature. The primary technical efficacy rate of MWA was comparable to RFA, with a reported odds ratio = 0.89 (95% CI: 0.52%–1.51; $I^2 = 0\%$).[26] The complication rate was also not substantially different between the two methods.

In a meta-analysis by Choi et al.[7], the efficacy and safety of MWA for malignant renal tumors were analyzed. Overall, 13 articles with 616 renal masses were included. A very high efficacy rate of 97.3% (95% CI: 94.3%–99.4%; $I^2 = 0\%$) was reported, in line with our present results. The reported major complication rate was lower, at 1.8% (95% CI: 0.6%–3.3%; $I^2 = 0\%$).[7] Notably, the differences in complication rates should be discussed, as the reported complications of bleeding, hematoma, and pseudoaneurysm in the analysis by Choi et al.[7] were of a more severe quality compared with those reported in the present analysis. This likely explains the differences. Similar safety and efficacy results were reported in RFA analyses,[28] supporting the conclusion that both ablation methods can be considered equally effective and safe.

One important aspect of the current meta-analysis is that there were not enough cases to compare the treatment outcomes between RFA and MWA specifically for cystic renal masses, as only a few cases were treated with MWA. There is a definite need to further investigate the differences between the methods in cystic renal masses.

Another important technique for local thermal ablation is cryoablation, which has shown promising results for small renal masses.[29] However, no studies to date have reported on its use for cystic renal masses, which needs to be investigated in the future.

Notably, there are also no reports regarding long-term oncological outcomes for local thermal ablation.[30] Moreover, the cooling dynamics of cystic areas in the kidney may differ substantially from the heating mechanisms of RFA and MWA.

Table 1. Demographics and ablation characteristics of the included studies

Authors	Country	Study design	Included patients (n)	Cystic masses (n)	Mean age (years)	Technical success, n (%)	Gender female, n (%)	Performed biopsy	Bosniak III (n)	Bosniak IV (n)	Ablation technique	Complication, n (%)	Mean lesion diameter (cm) ± SD	Ablation time (min)	Localization	Follow-up
Allen et al.[12]	USA	Retrospective	38	40	71	40 (100)	19 (47.5)	During the ablation	25	15	RFA	3 (7.5)	2.3 (1.0–4.2)	n.a.	n.a.	33.8 months (range 12–78)
Carrafiello et al.[15]	Italy	Retrospective	6	7	74	7 (100)	1 (16.7)	No biopsy	1	6	MWA	0	1.7 (1.4–2.7)	10	7 exophytic	24 months
Menezes et al.[16]	Brasil	Retrospective	9	10	63.5	10 (100)	4 (44.4)	4 during ablation; 5 without biopsy	0	10	RFA	0	2.5 (1.5–4.1)	12	8 exophytic, 2 central	27 months (interquartile range, 23–38)
Felker et al.[17]	USA	Retrospective	16	23	62	21 (91.3)	5 (31.3)	During the ablation	7	16	RFA	0	3.1 (1.1–10.1)	n.a.	16 exophytic, 7 central	24 months (range 2–110)
Park et al.[18]	South Korea	Retrospective	9	14	50	15 (100)	1 (11.1)	No biopsy	5	9	RFA	4 (26.7)	2.5 (1.6–3.9)	6	10 exophytic, 5 central	8 months (range 1–19)
Park et al.[19]	South Korea	Retrospective	30	35	57	35 (100)	5 (16.7)	No biopsy	15	20	RFA	2 (5.7)	2 (1.1–4.3)	18	n.a.	24 months (range 6–70)
Zhou et al.[20]	USA	Retrospective	5	5	75.4	5 (100)	2 (40)	During the ablation	2	3	MWA	0	3.3 (2.3–5.2)	8.6	3 exophytic, 2 central	18 months (range, 6–36)

SD, standard deviation.

Table 2. Quality of the included studies based on the Newcastle-Ottawa Scale

Study	Case definition adequate	Representativeness of cases	Selection of controls	Definition of controls	Comparability of cases and controls	Ascertainment of exposure	Same method of ascertainment	Non-response rate	Quality score
Allen et al.[12]	*	*	*	*	*	*	*	*	8
Carrafiello et al.[15]	*	*	*	*	*	*	*	*	8
Menezes et al.[16]	*	*	*	*	*	*	*	*	8
Felker et al.[17]	*	*	*	*	*	*	*	*	8
Park et al.[18]	*	*	*	*	*	*	*	*	8
Park et al.[19]	*	*	*	*	*	*	*	*	8
Zhou et al.[20]	*	*	*	*	*	*	*	*	8

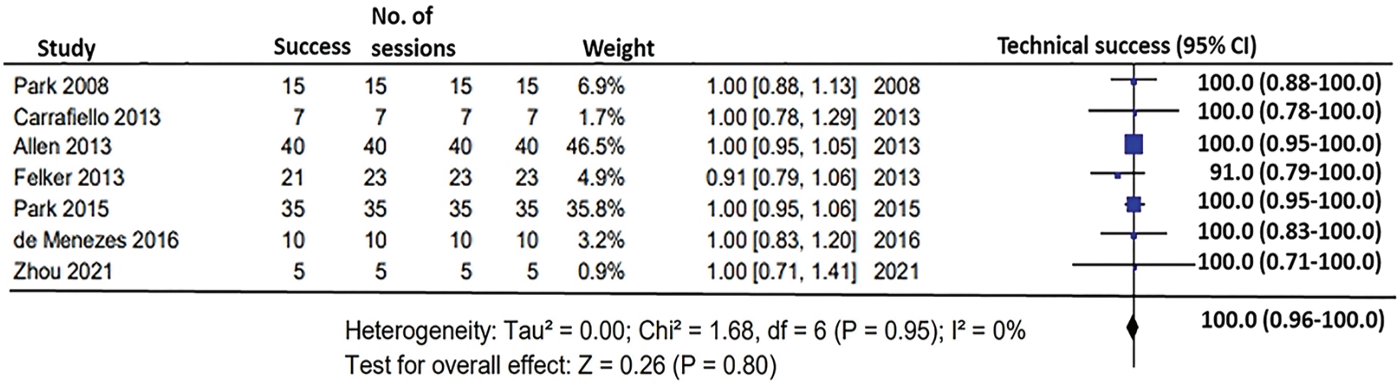


Figure 2. Forest plot of the technical success rate of thermal ablation. The pooled technical success rate was 100% (95% CI: 96%–100%, I^2 = 0%). CI, confidence interval.

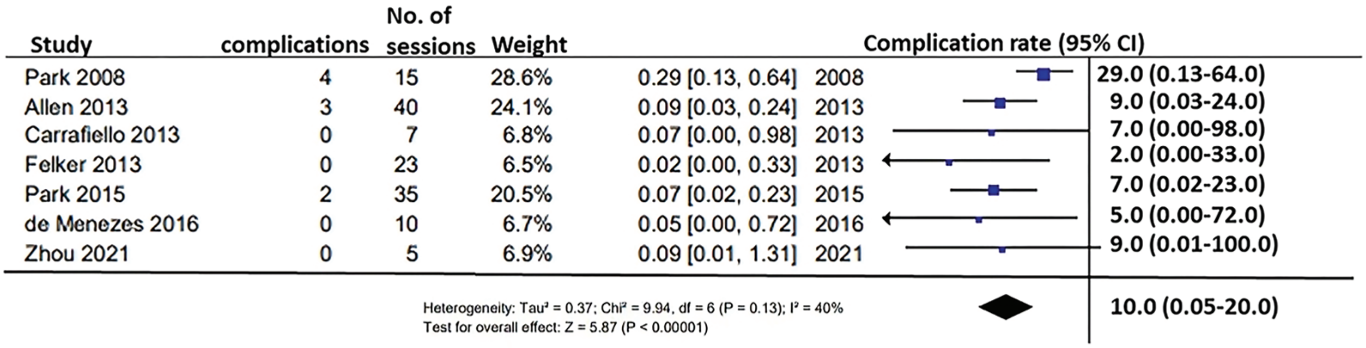


Figure 3. Forest plot of the complication rate of thermal ablation. The pooled complication rate was 10% (95% CI: 0.05%–20%, I^2 = 40%). CI, confidence interval.

Not every included study performed a biopsy prior to ablation. It therefore remains unclear whether every Bosniak III cyst treated was indeed a malignant mass. In a large meta-analysis, the malignancy rate of Bosniak III cysts was 55.1% (95% CI: 45.7%–64.5%), and for Bosniak IV cysts, 91% (95% CI: 87.7%–94.2%).[8] This raises the question of whether some Bosniak III cysts might be better suited for imaging surveillance rather than immediate definitive treatment with surgery or ablation.[31] However, treatment for both Bosniak III and IV cysts is currently recommended in the guidelines.[3]

The outcome of local thermal ablation is affected by factors such as lesion location, lesion size, ablation time, tissue impedance, and electrode surface area.[17,31] There may also be differences between MWA and RFA, especially regarding the heat sink effect, to which RFA is more susceptible.[5,6] Hypothetically, this is particularly relevant in cystic masses, as their fluid content may lead to a substantial heat sink effect. This could favor MWA in achieving full ablation of the mass. However, no direct comparison between RFA and MWA has been published to date, and such a comparison was not possible in the current meta-analysis.

One important aspect is also that the lesion composition and complexity–due to multilobulations and septae–could have an important impact on the treatment result. However, there is no reliable information in the included studies regarding the complexity of the cystic masses, and it remains unclear which imaging features of cystic renal masses have a clinically relevant impact on the outcome of ablation.

There is no systematic data or guideline on how to follow up with patients after local ablation treatment for cystic renal masses. One can only assume that follow-up for patients with cystic renal masses should be similar to that for solid renal masses. According to this, follow-up is recommended during the first year at 1, 3, 6, and 12 months, and thereafter every 12 months with contrast-enhanced cross-sectional imaging.[32,33] Presumably, regrowth of the cystic component alone should be considered tumor recurrence. However, as mentioned above, there is no data or published experience regarding tumor recurrence after ablation for cystic renal masses.

Notably, there are also no reports regarding long-term oncological outcomes after local ablation of cystic renal masses. One can only assume that outcomes should not differ from those of solid renal masses, provided full ablation of the lesion is achieved.

The present meta-analysis has several limitations to address. First, it is based on small published participant cohorts. Moreover, the included studies were retrospective case series without control groups, resulting in a low level of evidence. In addition, there was no direct comparison with surgical procedures in a randomized controlled trial. The overall level of evidence for these studies is considered level 5. Second, the analysis was restricted to English-language publications. Third, no long-term oncological outcomes were reported in the included studies. However, it seems plausible that no substantial recurrence occurred when the ablation zone covered the entire cystic mass with a safety margin. Fourth, comparisons between RFA and MWA were not possible, as only two studies investigated MWA. Fifth, the included studies used the older version of the Bosniak cyst classification, which was updated in 2019 to include more sophisticated assessments of septa and solid components.[10] However, there may be no substantial changes to the classification of Bosniak III and IV cysts.[32] Sixth, only one study reported aspiration of the cyst content prior to ablation. It remains unclear whether aspiration was performed in the other studies but not reported, or whether it was not performed at all. This should be addressed in future research. Seventh, there may be a publication bias in the analysis; however, we could not assess this, as tests for publication bias are recommended only when more than 10 studies are included.[24]

In conclusion, local thermal ablation can be considered a highly effective and safe procedure for cystic kidney masses. Most studies were performed using RFA, highlighting the need for new studies investigating MWA and cryoablation.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Chow WH, Devesa SS, Warren JL, Fraumeni JF Jr. Rising incidence of renal cell cancer in the United States. *JAMA.* 1999;281(17):1628-1631. [Crossref]
2. Lucas SM, Stern JM, Adibi M, Zeltser IS, Cadeddu JA, Raj GV. Renal function outcomes in patients treated for renal masses smaller than 4 cm by ablative and extirpative techniques. *J Urol.* 2008;179(1):79-80. [Crossref]
3. Campbell SC, Clark PE, Chang SS, Karam JA, Souter L, Uzzo RG. Renal mass and localized renal cancer: evaluation, management, and follow-up: AUA Guideline: Part I. *J Urol*. 2021;206(2):199-208. [Crossref]
4. Katsanos K, Mailli L, Krokidis M, McGrath A, Sabharwal T, Adam A. Systematic review and meta-analysis of thermal ablation versus surgical nephrectomy for small renal tumours. *Cardiovasc Intervent Radiol.* 2014;37(2):427-437. [Crossref]
5. Abdelsalam ME, Ahrar K. Ablation of small renal masses. *Tech Vasc Interv Radiol.* 2020;23(2):100674. [Crossref]
6. Mershon JP, Tuong MN, Schenkman NS. Thermal ablation of the small renal mass: a critical analysis of current literature. *Minerva Urol Nefrol.* 2020;72(2):123-134. [Crossref]
7. Choi SH, Kim JW, Kim JH, Kim KW. Efficacy and safety of microwave ablation for malignant renal tumors: an updated systematic review and meta-analysis of the literature since 2012. *Korean J Radiol.* 2018;19(5):938-949. [Crossref]
8. Sevcenco S, Spick C, Helbich TH, et al. Malignancy rates and diagnostic performance of the Bosniak classification for the diagnosis of cystic renal lesions in computed tomography - a systematic review and meta-analysis. *Eur Radiol.* 2017;27:2239-2247. [Crossref]
9. Hindman NM. Cystic renal masses. *Abdom Radiol (NY).* 2016;41(6):1020-1034. [Crossref]
10. Silverman SG, Pedrosa I, Ellis JH, et al. Bosniak classification of cystic renal masses, version 2019: an update proposal and needs assessment. *Radiology.* 2019;292(2):475-488. [Crossref]
11. McGrath TA, Bai X, Kamaya A, et al. Proportion of malignancy in Bosniak classification of cystic renal masses version 2019 (v2019) classes: systematic review and meta-analysis. *Eur Radiol.* 2023;33(2):1307-1317. [Crossref]
12. Allen BC, Chen MY, Childs DD, Zagoria RJ. Imaging-guided radiofrequency ablation of cystic renal neoplasms. *AJR Am J Roentgenol.* 2013;200(6):1365-1369. [Crossref]
13. Ziglioli F, De Filippo M, Cavalieri DM, et al. Percutaneous radiofrequency ablation (RFA) in renal cancer. How to manage challenging masses. A narrative review. *Acta Biomed.* 2022;93(5):2022220. [Crossref]
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:71. [Crossref]
15. Carrafiello G, Dionigi G, Ierardi AM, et al. Efficacy, safety and effectiveness of image-guided percutaneous microwave ablation in cystic renal lesions Bosniak III or IV after 24 months follow up. *Int J Surg.* 2013;11 Suppl 1:30-35. [Crossref]
16. Menezes MR, Viana PC, Yamanari TR, Arap MA, Reis LO, Nahas W. Safety and feasibility of radiofrequency ablation for treatment of Bosniak IV renal cysts. *Int Braz J Urol.* 2016;42(3):456-463. [Crossref]
17. Felker ER, Lee-Felker SA, Alpern L, Lu D, Raman SS. Efficacy of imaging-guided percutaneous radiofrequency ablation for the treatment of biopsy-proven malignant cystic renal masses. *AJR Am J Roentgenol*. 2013;201(5):1029-1035. [Crossref]
18. Park BK, Kim CK, Lee HM. Image-guided radiofrequency ablation of Bosniak category III or IV cystic renal tumors: initial clinical experience. *Eur Radiol*. 2008;18(7):1519-1525 [Crossref]
19. Park JJ, Park BK, Park SY, Kim CK. Percutaneous radiofrequency ablation of sporadic Bosniak III or IV lesions: treatment techniques and short-term outcomes. *J Vasc Interv Radiol.* 2015; 26:46-54. [Crossref]
20. Zhou W, Herwald SE, Arellano RS. Computed tomography-guided microwave ablation of cystic renal cell carcinoma: assessment of technique and complications. *J Vasc Interv Radiol*. 2021;32(4):544-547. [Crossref]
21. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses - 2008. [Crossref]
22. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008; 149:889-897. [Crossref]
23. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol.* 2006;6:31. [Crossref]
24. Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki AA (editors). Chapter 10: Chapter 10: Analysing data and undertaking meta-

analyses [last updated November 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024. [Crossref]

25. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. [Crossref]
26. Khalilzadeh O, Baerlocher MO, Shyn PB, et al. Proposal of a New Adverse Event Classification by the Society of Interventional Radiology Standards of Practice Committee. *J Vasc Interv Radiol*. 2017;28(10):1432-1437. [Crossref]
27. Castellana R, Natrella M, Fanelli G, et al. Efficacy and safety of MWA versus RFA and CA for renal tumors: a systematic review and meta-analysis of comparison studies. *Eur J Radiol*. 2023;165:110943. [Crossref]
28. Li L, Zhu J, Shao H, et al. Long-term outcomes of radiofrequency ablation vs. partial nephrectomy for cT1 renal cancer: a meta-analysis and systematic review. *Front Surg*. 2023;9:1012897. [Crossref]
29. Liu Y, Wang L, Bao EH, et al. Perioperative, functional, and oncological outcomes after cryoablation or partial nephrectomy for small renal masses in solitary kidneys: a systematic review and meta-analysis. *BMC Urol*. 2024;24(1):19. [Crossref]
30. Seager M, Kumar S, Lim E, Munneke G, Bandula S, Walkden M. Renal cryoablation - a practical guide for interventional radiologists. *Br J Radiol*. 2021;94(1118):20200854. [Crossref]
31. Kashan M, Ghanaat M, Hötker AM, et al. Cystic renal cell carcinoma: a report on outcomes of surgery and active surveillance in patients retrospectively identified on pretreatment imaging. *J Urol*. 2018;200(2):275-282. [Crossref]
32. Dai JC, Morgan TN, Moody D, McLaughlin J, Cadeddu JA. Radiofrequency ablation of small renal 343 masses. *J Endourol*. 2021;35:38-45. [Crossref]
33. Maas M, Beets-Tan R, Gaubert JY, et al. Follow-up after radiological intervention in oncology: ECIO-ESOI evidence and consensus-based recommendations for clinical practice. *Insights Imaging*. 2020;11(1):83. [Crossref]

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.242749

MUSKOLOSKELETAL IMAGING

ORIGINAL ARTICLE

Detection of synovial inflammation in the sacroiliac joint space through intravoxel incoherent motion imaging: an alternative to contrast agents

Murat Ağırlar[1]
Barış Genç[2]
Aysu Başak Özbalcı[2]

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

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

PURPOSE

We investigated the diagnostic accuracy of simplified intravoxel incoherent motion (IVIM) imaging for detecting synovial inflammation in the sacroiliac joint (SIJ) in a population with active sacroiliitis.

METHODS

In accordance with the Assessment of Spondyloarthritis International Society criteria, 86 SIJs of 46 patients with active sacroiliitis were included in this retrospective study conducted between November 2020 and January 2022. Based on T1-weighted post-gadolinium images, the SIJs were divided into two groups: synovial inflammation positive (SIP) (n = 28) and synovial inflammation negative (SIN) (n = 58). Synovial areas in the SIJ space were independently and blindly reviewed for the presence of inflammation by two radiologists with differing levels of expertise in radiology. Using four *b* values, apparent diffusion coefficient (ADC)= ADC (0, 800) and the simplified 3T IVIM method parameters true diffusion coefficient (D_1)= ADC (50, 800), D= ADC (400, 800), f_1= f (0, 50, 800), f_2= f (0, 400, 800), pseudodiffusion coefficient (D*)= D* (0, 50, 400, 800), ADC_{low} = ADC (0, 50), and ADC_{diff}= ADC_{low} – D were generated voxel by voxel for each patient. The IVIM and ADC parameters at the SIN and SIP joints were compared.

RESULTS

The D parameter was significantly increased in SIP areas ($1.23 \pm 0.34 \times 10^{-3}$ mm²/s) compared with SIN areas ($1.02 \pm 0.16 \times 10^{-3}$ mm²/s) ($P = 0.004$). Conversely, the D* parameter was significantly decreased in SIP areas ($21.78 \pm 3.77 \times 10^{-3}$ mm²/s) compared with SIN areas ($16.19 \pm 4.58 \times 10^{-3}$ mm²/s) ($P < 0.001$). When the optimal cut-off value of 1.11×10^{-3} mm²/s was selected, the sensitivity for the D value was 71% and the specificity was 72% [area under the curve (AUC): 0.716)]. When the optimal cut-off value of 21.06×10^{-3} mm²/s was selected, the sensitivity for the D* value was 78.6%, and the specificity was 79.3% (AUC: 0.829). The interclass correlation coefficient was excellent for f_1, f_2 D*, D, and ADC_{diff}, good for ADC_{low} and D_1, but reasonable for ADC.

CONCLUSION

The presence of synovial inflammation in the SIJ can be evaluated with high sensitivity and specificity using only four b values through the simplified IVIM method without the need for a contrast agent.

CLINICAL SIGNIFICANCE

IVIM imaging is a technique that allows us to gain insights into tissue perfusion without the administration of contrast agents, utilizing diffusion-weighted images. In this study, for the first time, we demonstrated the potential of detecting synovial inflammation in the SIJ using IVIM, specifically through the pseudodiffusion (D*) parameter, without the need for contrast agents.

KEYWORDS

Diffusion magnetic resonance imaging, intravoxel incoherent motion, perfusion fraction, pseudodiffusion, sacroilitiis, synovial inflammation

Corresponding author: Murat Ağırlar

E-mail: dr.muratagirlar@gmail.com

Received 06 March 2024; revision requested 16 April 2024; last revision received 03 June 2024; accepted 02 July 2024.

Epub: 19.08.2024

Publication date: 06.11.2025

DOI: 10.4274/dir.2024.242749

You may cite this article as: Ağırlar M, Genç B, Özbalcı AB. Detection of synovial inflammation in the sacroiliac joint space through intravoxel incoherent motion imaging: an alternative to contrast agents. *Diagn Interv Radiol.* 2025;31(6):612-618.

Synovial inflammation in the sacroiliac joint (SIJ) space can be evaluated using T1-weighted (T1w) post-gadolinium (Gd) images through magnetic resonance imaging (MRI). An assessment of the SIJ is commonly performed in cases of spondyloarthritis (SpA), which is a chronic inflammatory arthritis that can affect both axial and peripheral joints.[1] The most common clinical finding in patients affected by axial SpA (axSpA) is inflammatory low back pain (ILBP), and the frequency of radiological sacroiliitis is increasing.[2] In this context, the clinical value of MRI of the SIJ has increased since the presence of sacroiliitis in MRI was accepted as a diagnostic criterion by the Assessment of Spondyloarthritis International Society (ASAS).[3]

In studies conducted with patients with ankylosing spondylitis (AS), synovial inflammation has been demonstrated to predict disease activity and is associated with the degree of pain.[4] In 2019, the ASAS-MRI working group stated that increased contrast enhancement in the cartilaginous component of the SIJ reflects inflammation at the osteochondral interface, consistent with the early histopathological features of AS. Synovial inflammation may be associated with microenvironmental permeability changes in the joint space before bone marrow edema (BME) becomes apparent.[5] In recent years, concerns have emerged regarding the accumulation of Gd in the brain and the risk of nephrogenic systemic fibrosis associated with its use.[6,7] In this context, we believe that the simplified intravoxel incoherent motion (IVIM) technique could be a useful and effective method for quantitatively monitoring early inflammatory progression in axSpA cases without the need for Gd administration. The concept of IVIM, as defined by LeBihan, allows the separation of molecular diffusion and capillary perfusion using a biexponential model and helps calculate tissue perfusion without contrast agent administration.[8] The biexponential model calculates the true diffusion coefficient (D), pseudodiffusion coefficient (D*), and perfusion fraction (f). Notably, D* refers to blood velocity and the length of microvessel segments, whereas f signifies the microvascular blood flow.[8] The simplified IVIM model, developed recently using fewer b values, yielded more consistent and accurate results compared with the biexponential IVIM model.[9,10]

Main points

- Intravoxel incoherent motion (IVIM) imaging may serve as an alternative method to contrast agents for the detection of synovial inflammation in the sacroiliac joint.
- The most valuable parameter in the detection of synovial inflammation with simplified IVIM is pseudodiffusion coefficient.
- Using simplified IVIM with only four b values can shorten sequence time and yield results with high specificity and sensitivity.

The detection of synovial inflammation, which emerges in the early stages of axSpA, using quantitative methods such as IVIM, is likely to contribute to early diagnosis and treatment monitoring. In our literature review, we found that IVIM studies related to sacroiliac (SI) MRI have focused primarily on BME.[11-13] An IVIM study evaluating synovial inflammation in the SIJ is not yet available in the literature. The aim of this study is to detect the presence of synovial inflammation in SI MRI using the simplified IVIM model and compare its effectiveness with T1w post-Gd imaging in cases diagnosed with axSpA.

Methods

Study design and participants

The current study was designed retrospectively and approved by the Ondokuz Mayıs University Institutional Ethics Committee (decision no: 2021/206, decision date: 06.04.2021), along with a waiver for informed consent. This paper was drafted according to the guidelines laid down by the Standards for Reporting of Diagnostic Accuracy Studies.[14] Written consent was obtained from the patients to participate in the study.

Between November 2020 and January 2022, a total of 519 patients who underwent SIJ MRI examination on a 3Tesla (Philips, Ingenia, Best, the Netherlands) instead of 3T Philips Ingenia device with diffusion-weighted imaging at b values of 0, 50, 400, and 800 and T1w post-Gd images for ILBP were retrospectively screened. In total, 46 individuals aged between 18 and 45 years, without a history of malignancy, and with active sacroiliitis in accordance with the ASAS MRI criteria [the presence of BME larger than 1 cm in coronal oblique plane short tau inversion recovery (STIR) sequence], were included in the study (14 men and 32 women; mean age: 34.4 ± 7.9). Cases with chronic sacroiliitis were excluded, and all included patients had active sacroiliitis. The inclusion and exclusion criteria for this study are listed in Figure 1.

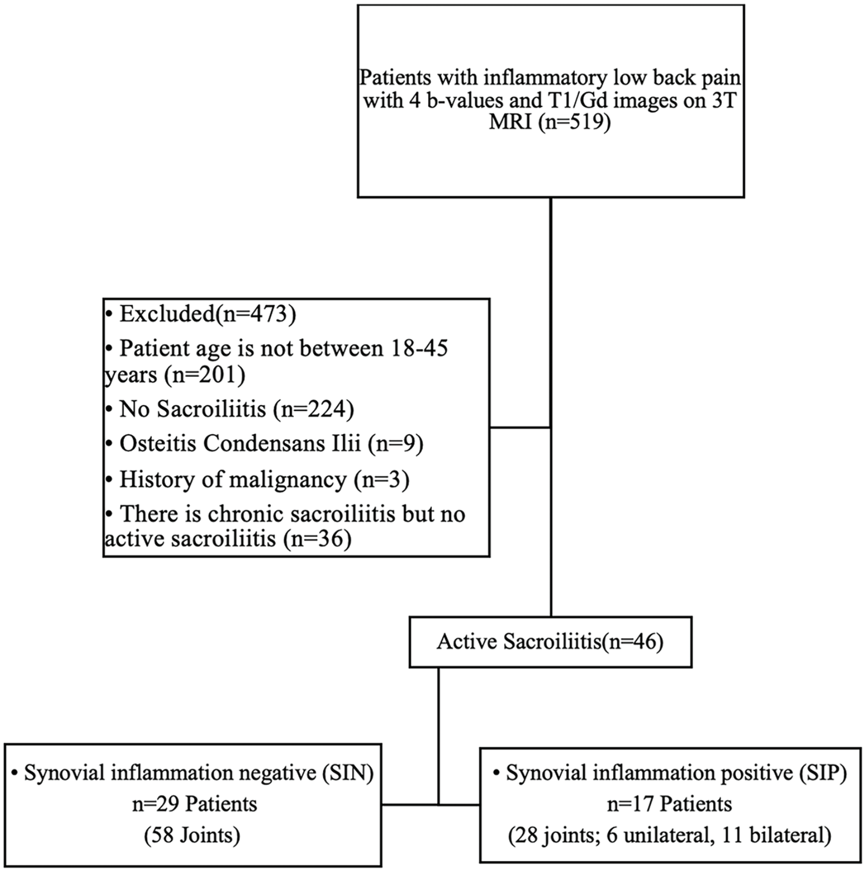


Figure 1. Study flowchart. Gd, gadolinium; MRI, magnetic resonance imaging.

To detect inflammation in the joint space of the selected patients, T1w post-Gd images were used as the gold standard test. The synovial areas in the lower third of the SIJ space were reviewed for the presence of contrast enhancement by two radiologists with 2 (M.A.) and 13 (A.B.Ö.) years of experience in musculoskeletal radiology following their radiology residency. Joints with contrast enhancement in the synovium were categorized as synovial inflammation positive (SIP), whereas those without contrast enhancement were categorized as synovial inflammation negative (SIN). Subsequently, a consensus was reached regarding the patients where a common opinion was not achieved to reach a final decision. Among the 17 patients with synovial enhancement, 11 displayed bilateral synovial involvement and 6 showed unilateral synovial involvement. The SIN group comprised 58 SIJs belonging to 29 patients for whom contrast enhancement in the joint space could not be detected, and 28 joints belonging to 17 patients exhibiting synovial enhancement formed the SIP group (Figure 1). Both radiologists were blinded to all clinical patient data and each other's findings. The clinical and laboratory parameters [sedimentation and C-reactive protein (CRP)] of the patients were obtained from the hospital information system 1 month before and after the MRI examination date. We confirmed from the hospital system that all patients included in the study received anti-inflammatory treatment. However, we were unable to standardize or compare their treatment protocols and follow-ups.

Magnetic resonance imaging technique

All examinations were performed on a 3.0 T Philips Ingenia MR (Philips, Netherlands) using a torso coil. The standard sequences included coronal oblique T1w, coronal oblique T2-weighted (T2w), coronal oblique T2w-fat suppressed (FS), coronal oblique STIR, axial oblique T2w-FS, axial oblique precontrast T1w-FS, coronal T1w-FS with IV contrast, and axial T1w-FS with IV contrast. The details of the MRI protocol are provided in Table 1. The diffusion MRI protocol is shown in Table 2.

Image analyses

Postprocessing image analysis

In the simplified IVIM method, IVIM parameters are calculated using the following equation, as stated in previous research:[8,9,15]

$$Apparent\ Diffusion\ Coefficient(ADC)(i,j) = \frac{\ln(S(b_i)) - \ln(S(b_j))}{j - i}$$

Using this equation, various ADC values were obtained from $b_0 = 0$, $b_1 = 50$, $b_2 = 400$, and $b_3 = 800$ s/mm², such as $D_1 = ADC$ (50, 800), $D_2 = D = ADC$ (400, 800), $ADC_{low} = ADC$ (0, 50), and ADC= ADC (0, 800). Using these ADC values, f_1 and f_2 values were obtained:

$$f_1 = f(0{,}50{,}800) = 1 - \frac{S(b_1)}{S(0)} \cdot exp^{D_1 \cdot b_1}$$
$$f_2 = f(0{,}400{,}800) = 1 - \frac{S(b_2)}{S(0)} \cdot exp^{D_2 \cdot b_2}$$

Furthermore, D* was obtained based on the above parameters using the following equation:

$$D^* = -\frac{1}{b_1} \cdot \ln\left(\frac{1}{f_2} \cdot \left(\frac{S(b_1)}{S(0)} - (1 - f_2) \cdot exp^{D_2 \cdot b_1}\right)\right)$$

Subsequently, the perfusion-sensitive diffusion parameter corresponding to the difference between D_2 and ADC_{low} was calculated.

$$ADC_{diff} = ADC_{low} - D_2$$

All images were generated in a Bash environment on macOS using a script developed in our radiology department. The Functional Magnetic Resonance Imaging of the Brain Software Library was employed as the image processing algorithm, and dcm2niix was used for DICOM-to-NIfTI conversion.[16,17] Furthermore, all voxels with a perfusion fraction below 0% and above 100% were removed from the image to avoid miscalculations.

Volume of interest definition

The entire synovium located in the lower third of the SIJ was segmented using ITK-SNAP (http://www.itksnap.org) with diffusion-weighted b = 0 images in the SIN group.[8] In the SIP group, subtracted images were created from the T1w-FS post-Gd and precontrast T1w-FS images. In these subtracted images, the entire contrast-enhanced synovial component was revealed. Subsequently, the subtracted images were fused with the diffusion-weighted b = 0 images on ITK-SNAP, and all areas exhibiting contrast enhancement in the synovial area were segmented. The segmented images in both groups were then extracted as mask images (Figure 2). Finally, the mean value of the intensities in the ADC, f_1, f_2, D, D_1, D*, ADC_{low}, and ADC_{diff} mask images was automatically calculated using the fslstats command.[16]

Statistical analysis

Statistical analyses were performed using SPSS software version 22.0. The variables were investigated by conducting a Shapiro–Wilk test to determine whether they exhibited a normal distribution. Descriptive analyses were performed using the means and standard deviations for the normally distributed variables. Additionally, the Student's t-test was employed to compare the normally distributed variables, and the Mann–Whitney U test was used to compare the non-normally distributed variables. Non-parametric variables were presented with standard deviation and median (min–max) or interquartile range (IQR) values. Furthermore, Pearson's χ^2 test was implemented to compare the genders of the two groups. The intraclass correlation coefficient (ICC) statistic was used to assess intraobserver agreement for the IVIM parameters. The diagnostic performance of

Table 1. Magnetic resonance imaging protocol

	T1W	T2W	T2W-FS	STIR	T2W-FS	T1W-FS (precontrast)	T1W-FS (post iv contrast)	T1W-FS (post iv contrast)
Plane	Coronal oblique	Coronal oblique	Coronal oblique	Coronal oblique	Axial oblique	Axial oblique	Coronal	Axial
Repetition time (ms)	500	4000	4000	shortest	4000	500	500	500
Echo time (ms)	8	80	80	80	80	8	8	8
FOV (RL × AP)	210 × 210	210 × 210	210 × 210	250 × 250	240 × 240	240 × 240	210 × 210	240 × 240
Number of excitations	1	1	1	1	1	1	1	1
Matrix size	264 × 279	212 × 199	212 × 199	256 × 204	268 × 264	344 × 294	264 × 279	344 × 294
Slice thickness (mm)	3	3	3	3	3.5	4	3	4
Gap (mm)	0.3	0.3	0.3	0.6	0.35	0.8	0.3	0.8

T1W, T1-weighted; T2W, T2-weighted; FS, fat-suppressed; STIR; short tau inversion recovery; iv, intravenous; FOV, field of view; RL, right-left; AP, anterior-posterior.

the IVIM parameters to detect synovial inflammation in the SIJ was calculated by conducting a receiver operating characteristic (ROC) curve analysis. The Youden index was used to determine the cut-off value in the ROC analysis.

Results

Patient and clinical characteristics

The results of the analyses revealed no significant difference between the age distributions of the two groups ($P = 0.11$). Notably, the number of women was significantly higher in the SIP group (22% vs. 78%) ($P = 0.007$). Although no statistically significant difference was observed between the CRP values of the groups, the sedimentation rate was significantly higher in the SIP group (SIN median: 13, IQR: 17; SIP median: 23, IQR: 20; $P = 0.027$). The clinical and demographic data of the patients are summarized in Table 3.

Intravoxel incoherent motion analyses of the sacroiliac joints in the study groups

The ADC, D_1, D, f_1, f_2, D*, ADC_{low}, and ADC_{diff} values of the joint spaces in the SIN and SIP groups are summarized in Table 4. Notably, although the means of ADC and D_1 were higher in the SIP areas than in the SIN areas, the differences were not statistically significant ($P = 0.057$, $P = 0.053$). However, D was significantly higher in the SIP group ($P = 0.004$). Furthermore, although f_1, ADC_{low}, and ADC_{diff} were higher in the SIP areas, the differences were not statistically significant ($P = 0.143$, $P = 0.131$, $P = 0.153$). In the SIP areas, D* was observed to be significantly decreased ($P < 0.001$).

Table 2. Diffusion magnetic resonance imaging protocol details

Name	Value
FOV (RL × AP)	250 × 250 mm
Slice thickness	4 mm
Spacing between slices	5
Slice number	39
Echo time	0.08 s
Repetition time	1.5 s
Diffusion gradients	3 orthogonal directions
b values	0, 50, 400, 800 (NEX: 4)
Acquisition time	4:30 min

NEX, number of excitation.

The diagnostic performance of D* and D in detecting synovial enhancement in patients with active sacroiliitis is shown in Figure 3. The best D cut-off value for diagnosing synovial enhancement in patients with active sacroiliitis was 1.11×10^{-3} mm^2/s, for which the sensitivity of D was 71% and specificity was 72% [area under the curve (AUC): 0.716, 95% confidence interval (CI): 0.574–0.857, $P = 0.001$]. Moreover, the best D* cut-off value for the diagnosis of synovial enhancement in patients with active sacroiliitis was 21.06 $\times 10^{-3}$ mm^2/s, with the sensitivity of D* being 78.6% and specificity being 79.3% (AUC: 0.829, 95% CI: 0.723–0.936, $P < 0.001$).

Interobserver agreement assessment

The interobserver agreement for f_1, f_2, D*, D, and ADC_{diff} were excellent, with the ICC being 0.793 for f_1, 0.802 for f_2, 0.978 for D*, 0.772 for D, and 0.774 for ADC_{diff} ($P < 0.001$). Furthermore, interobserver agreement for ADC_{low} and D_1 was good, with the ICC being 0.727 for ADC_{low} and 0.625 for D_1 ($P < 0.001$). Interobserver agreement for ADC was reliable, with an ICC of 0.535 ($P < 0.001$).

Discussion

In this study, simplified IVIM parameters were examined to quantitatively determine synovial inflammation in patients with axSpA

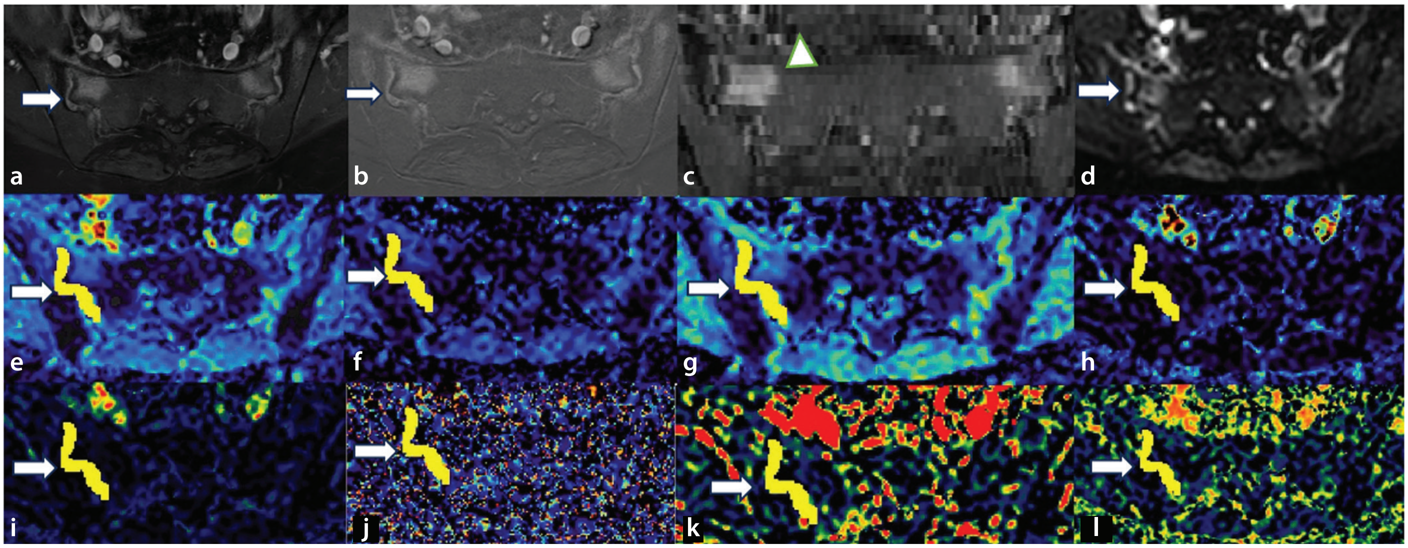


Figure 2. Sacroiliac magnetic resonance imaging (MRI) examination performed on a 40-year-old female patient diagnosed with axial spondyloarthritis reflecting active sacroiliitis in line with the Assessment of Spondyloarthritis International Society MRI criteria, with findings of bone marrow edema indicated by arrowheads on the short tau inversion recovery (STIR) series and synovial inflammation indicated by arrows on the fat-suppressed (FS) contrast-enhanced and diffusion-weighted imaging (DWI) series. **(a)** FS axial T1 postcontrast series and **(b)** subtracted images demonstrating enhancement in the joint space in the posteroinferior parts of both sacroiliac joints (SIJs). **(c)** Axial reformatted STIR images, **(d)** $b = 0$ DWI indicating fluid values in the joint space. **(e-l)** apparent diffusion coefficient (ADC), D, D_1, ADC_{low}, ADC_{diff}, D*, f_1, and f_2 maps with synovial enhanced areas marked with volume of interest in the right SIJ space indicated by arrows. At the marked volume, the following estimates were calculated: ADC = 1.9×10^{-3} mm^2/s, D = 1.22×10^{-3} mm^2/s, $D_1 = 1.43 \times 10^{-3}$ mm^2/s, $ADC_{low} = 1.76 \times 10^{-3}$ mm^2/s, $ADC_{diff} = 16.02 \times 10^{-3}$ mm^2/s, D* = 16.58×10^{-3} mm^2/s, f_1 = 7%, and f_2 = 20%.

Table 3. Demographic and clinical characteristics

Characteristic	SIP	SIN	*P* value
Female gender ratio (%)	78	22	0.007
Age (year) (mean ± SD)	30.88 ± 7.67	34.48 ± 8.10	0.85
CRP (mg/dL) (median, IQR)	3 (3)	3 (2)	0.659
ESR (mm/L) (median, IQR)	23 (20)	13 (17)	0.027

SIP, synovial inflammation positive; SIN, synovial inflammation negative; SD, standard deviation; CRP, C-reactive protein; IQR, interquartile range; ESR, erythrocyte sedimentation rate.

Table 4. Mean values of groups with and without synovial enhancement

Parameter	SIN (n = 58) Mean ± SD Median (min-max)	SIP (n = 28) Mean ± SD Median (min-max)	*P*
ADC (×10^{-3} mm^2/s)	1.39 ± 0.16 1.38 (0.99–1.90)	1.48 ± 0.21 1.52 (1.07–1.87)	0.057
D$_1$ (×10^{-3} mm^2/s)	1.25 ± 0.13 1.24 (0.92–1.71)	1.36 ± 0.26 1.38 (0.69–1.87)	0.053
D (×10^{-3} mm^2/s)	1.02 ± 0.16 1.01 (0.62–1.42)	1.23 ± 0.34 1.24 (0.46–1.87)	**0.004**
f$_1$%^{a}	16.85 ± 4.04 15.88 (9.30–25.98)	18.81 ± 5.52 16.71 (11.42–31.53)	0.143
f$_2$%	31.23 ± 6,34 30.90 (17.88–49.39)	31.68 ± 6.12 32.15 (20.47–43.74)	0.752
D* (×10^{-3} mm^2/s)	21.78 ± 3.77 21.86 (14.77–32.33)	16.19 ± 4.58 15.11 (8.16–26.22)	**<0.001**
ADC$_{\text{low}}$ (×10^{-3} mm^2/s)^a	4.59 ± 1.16 4.47 (2.35–7.69)	5.21 ± 1.55 4.54 (2.94–9.25)	0.131
ADC$_{\text{diff}}$ (×10^{-3} mm^2/s)^a	4.09 ± 1.17 3.77 (2.09–6.91)	4.62 ± 0.16 4.13 (2.40–8.92)	0.153

Mann–Whitney U test was performed for parameters denoted with ^{a} due to non-normal distribution according to the Shapiro–Wilk test. The Student's t-test was performed for the other parameters.
Significant test results are noted in bold. SIN, synovial inflammation negative; SD, standard deviation; min, minimum; max, maximum; ADC, apparent diffusion coefficient; D, true diffusion coefficient; D*, pseudodiffusion coefficient.

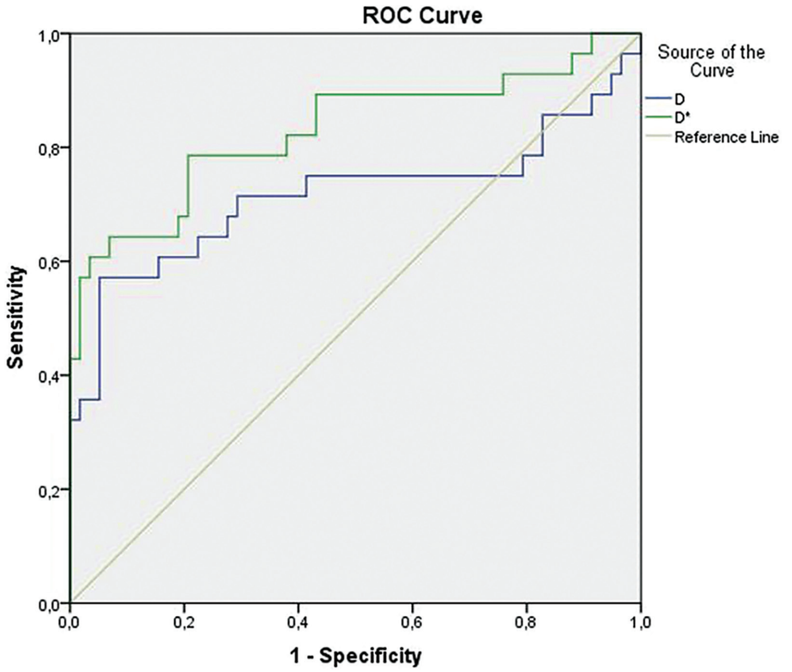


Figure 3. Receiver operating characteristic curve analysis of D* and D with regard to the diagnosis of synovial enhancement in active sacroiliitis cases. D, true diffusion coefficient; D*, pseudodiffusion coefficient.

accompanied by active sacroiliitis without using a contrast agent. The study results indicate that in areas where synovial contrast enhancement is detected in the SIJ and defined as SIP, the pseudodiffusion coefficient D* decreases, whereas the diffusion coefficient D, indicating pure diffusion in the extracellular space, increases. There was no significant difference between SIN and SIP areas in terms of ADC, f$_1$, f$_2$, D$_1$, ADC$_{\text{low}}$, or ADC$_{\text{diff}}$.

According to the study results, the decrease in the D* value in SIP areas indicates a reduction in intravascular blood flow and capillary network length. However, an increase in the D value suggests an increase in extracellular pure diffusion in SIP areas. Although f$_1$ and f$_2$ were affected by the amount of fluid in the extracellular space and intravascular compartment, ADC$_{\text{low}}$, D$_1$, ADC$_{\text{diff}}$, and ADC were impacted by both diffusion and perfusion. These findings indicate that, although an increase is observed in the amount of fluid in the extracellular space in areas of synovial inflammation, it is accompanied by a decrease in blood flow in the capillary bed in the SIP areas compared with the SIN areas. Furthermore, no significant change was observed in the value of f$_1$, which was affected by fluid areas in the extracellular space and the intravascular compartment. There was no significant difference in the ADC$_{\text{low}}$, D$_1$, and ADC values with regard to both diffusion and perfusion, or the ADC$_{\text{diff}}$ value, which represents the diffusion parameter associated with perfusion. These findings highlight that an increase in the amount of fluid in the extracellular space in areas of synovial inflammation is accompanied by a decrease in blood flow in the capillary bed.

Zhao et al.[13], who compared the parameters of the IVIM method with those of dynamic contrast-enhanced (DCE) MRI in the SIJ, found D to be moderately correlated with relative enhancement and maximum enhancement (ME). In the same study, D* was observed to have a moderate negative correlation with ME. The current study identified a decrease in D* and an increase in D in areas characterized by Gd uptake. Notably, a decrease in the D* of the areas marked by contrast enhancement is consistent with the results of the current study. In a similar study, Guo et al.[1] investigated the correlation between DCE-MRI and IVIM DWI parameters in patients with axSpA, finding that D correlated with perfusion parameters derived from DCE-MRI. This finding is consistent with the observed increase in D in the areas of synovial inflammation in the present study. Liu et al.[18] found a higher true diffusion coefficient

(D_{slow}) in the IVIM DWI study examining BME areas in cases of active sacroiliitis; however, this difference was not statistically significant. Guo et al.[19] correlated IVIM MRI parameters with the clinical activity index for BME areas in axSpA cases, identifying a correlation between D and D* with the clinical activity index. Although our study addressed a different topic, two of the most valuable IVIM parameters were found to be consistent with our study.

In our literature review, we identified two IVIM studies focusing on synovial inflammation that were also conducted on the knee joint. In the first study, an increase in f and D values was detected in inflamed synovial areas.[20] However, in the present study, although the f value registered an increase, it was not statistically significant. As noted by Andreou et al.[21], the decrease in D*, which was evident in the case of the patients participating in this study, may have prevented the f value from achieving a substantial increase. In their more recent pilot study, Huch et al.[22] compared the results obtained from the knee joint with detected synovitis in pediatric and young adult patients with data from healthy volunteers. This ongoing study has demonstrated that the D value is low and the f value is high based on data from eight patients.[22] Furthermore, the only study that compared the diffusion properties of inflamed and non-inflamed synovial fluid using contrast-enhanced MRI was conducted with a focus on the knee joints of children diagnosed with juvenile idiopathic arthritis, revealing that ADC values increase in the case of active disease.[23] In the present study, however, no significant change in ADC values was observed. Notably, the b values used to calculate the ADC in the aforementioned study were 50 and 600, as a result of which the tissue perfusion effect on the ADC was subtracted. This may be accepted as a parameter similar to the perfusion-free D value used in the present study. Therefore, in the present study, the D value exhibiting a minimal perfusion effect was significantly higher in the inflamed areas.

Notably, pathology studies have indicated that patients with AS experience extensive vascular congestion and obliteration, accompanied by an increase in lining cells, lymphocytes, and plasma cells in the SIJ.[24] Furthermore, histological studies have revealed increased mast cells, CD163 macrophages, and neutrophils in areas of synovial inflammation.[25] The mediators secreted from these cells can increase capillary permeability and the amount of extracellular fluid. This pathophysiological mechanism may explain the increase in D observed in this study, which reflects free diffusion in the extracellular space. In addition, vascular congestion and obliteration may have contributed to the decrease in D* values, which is associated with intravascular blood flow and overall decreased perfusion.

It should be noted that there is ongoing debate regarding the necessity of contrast application in SI MRI. According to the ASAS criteria, the administration of a contrast agent is primarily recommended to detect joint space enhancement rather than synovitis. Current guidelines suggest using the term "joint space enhancement" instead of synovitis.[26] In this context, this study is the first to investigate the performance of the simplified IVIM method in detecting synovial inflammation in SI MRI and its diagnostic superiority over using a contrast agent. In line with ASAS-EULAR recommendations, T1w post-Gd images are typically used to detect synovial inflammation.[27] However, the current study demonstrated the presence of synovial inflammation in patients with SpA with high specificity and sensitivity, without the need for T1w post-Gd images. Although previous IVIM studies on the SIJ have focused on BME areas detectable on STIR sequences, this pioneering study revealed synovial inflammation areas, traditionally requiring contrast agents, using the IVIM method. This approach not only demonstrated high specificity and sensitivity but also excellent interobserver agreement using only four *b* values.

Nonetheless, this study has several limitations. First, the patients' Bath Ankylosing Spondylitis Disease Activity Index scores and clinical follow-ups were not available for analysis; these data could enhance the reliability of results in future studies. Second, although pioneering, the study analyzed a relatively small number of patients. Third, although T1w post-Gd images were the gold standard for evaluating synovial inflammation, no evaluation based on histopathological data was conducted. Histopathological assessments were impractical because of the anatomical complexities of the SIJs and ethical considerations. Finally, as a cross-sectional study, it does not provide insights into longitudinal changes in the synovium over time. Future research should consider longitudinal studies to address these limitations comprehensively.

In conclusion, the results of this study indicate that in patients with axSpA and inflammation in the SIJ, there is an increase in fluid within the extracellular space accompanied by decreased blood flow. The simplified IVIM method, which involves a short sequence time and demonstrates high specificity and sensitivity using only four *b* values, allows for the detection of synovial inflammation without the need for any contrast material.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Guo C, Zheng K, Ye Q, et al. Intravoxel incoherent motion imaging on sacroiliitis in patients with axial spondyloarthritis: correlation with perfusion characteristics based on dynamic contrast-enhanced magnetic resonance imaging. *Front Med (Lausanne)*. 2022;8:798845. [CrossRef]
2. Navallas M, Ares J, Beltrán B, Lisbona MP, Maymó J, Solano A. Sacroiliitis associated with axial spondyloarthropathy: new concepts and latest trends. *Radiographics*. 2013;33(4):933-956. [CrossRef]
3. van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis*. 2017;76(6):978-991. [CrossRef]
4. Jee WH, McCauley TR, Lee SH, Kim SH, Im SA, Ha KY. Sacroiliitis in patients with ankylosing spondylitis: association of MR findings with disease activity. *Magn Reson Imaging*. 2004;22(2):245-250. [CrossRef]
5. Yang H, Jiang L, Li J, et al. Quantitative DCE-MRI: an efficient diagnostic technique for evaluating early micro-environment permeability changes in ankylosing spondylitis. *BMC Musculoskelet Disord*. 2020;21(1):774. [CrossRef]
6. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB; International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol*. 2017;16(7):564-570. [CrossRef]
7. Schlaudecker JD, Bernheisel CR. Gadolinium-associated nephrogenic systemic fibrosis. *Am Fam Physician*. 2009;80(7):711-714. [CrossRef]
8. Le Bihan D. What can we see with IVIM MRI? *Neuroimage*. 2019;187:56-67. [CrossRef]
9. Mürtz P, Pieper CC, Reick M, et al. Is liver lesion characterisation by simplified IVIM DWI also feasible at 3.0 T? *Eur Radiol*. 2019;29(11):5889-5900. [CrossRef]
10. Liu N, Yang X, Lei L, Pan K, Liu Q, Huang X. Intravoxel incoherent motion model in differentiating the pathological grades of esophageal carcinoma: comparison of mono-exponential and bi-exponential fit model. *Front Oncol*. 2021;11:625891. [CrossRef]

11. Zhao YH, Li SL, Liu ZY, et al. Detection of active sacroiliitis with ankylosing spondylitis through intravoxel incoherent motion diffusion-weighted MR imaging. *Eur Radiol.* 2015;25(9):2754-2763. [CrossRef]

12. Sun H, Liu K, Liu H, et al. Comparison of bi-exponential and mono-exponential models of diffusion-weighted imaging for detecting active sacroiliitis in ankylosing spondylitis. *Acta Radiol.* 2018;59(4):468-477. [CrossRef]

13. Zhao Y, Zhang Q, Li W, et al. Assessment of correlation between intravoxel incoherent motion diffusion weighted MR imaging and dynamic contrast-enhanced MR imaging of sacroiliitis with ankylosing spondylitis. *Biomed Res Int.* 2017;2017:8135863. [CrossRef]

14. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6(11):e012799. [CrossRef]

15. Pieper CC, Sprinkart AM, Meyer C, et al. Evaluation of a simplified intravoxel incoherent motion (IVIM) analysis of diffusion-weighted imaging for prediction of tumor size changes and imaging response in breast cancer liver metastases undergoing radioembolization: a retrospective single center analysis. *Medicine (Baltimore)*. 2016;95(14):e3275. [CrossRef]

16. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *Neuroimage.* 2012;62(2).782-790. [CrossRef]

17. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* 2006;31(3):1116-1128. [CrossRef]

18. Liu L, Zhou Z, Hua S, et al. Detection of the disease activity with ankylosing spondylitis through intravoxel incoherent motion diffusion-weighted MR imaging of sacroiliac joint. *Br J Radiol.* 2022;95(1133):20211074. [CrossRef]

19. Guo C, Zheng K, Xie Z, et al. Intravoxel incoherent motion diffusion-weighted imaging as a quantitative tool for evaluating disease activity in patients with axial spondyloarthritis. *Clin Radiol.* 2022;77(6):434-441. [CrossRef]

20. Hilbert F, Holl-Wieden A, Sauer A, Köstler H, Neubauer H. Intravoxel incoherent motion magnetic resonance imaging of the knee joint in children with juvenile idiopathic arthritis. *Pediatr Radiol.* 2017;47(6):681-690. [CrossRef]

21. Andreou A, Koh DM, Collins DJ, et al. Measurement reproducibility of perfusion fraction and pseudodiffusion coefficient derived by intravoxel incoherent motion diffusion-weighted MR imaging in normal liver and metastases. *Eur Radiol.* 2013;23(2):428-434. [CrossRef]

22. Huch B, Stumpf K, Bracher AK, et al. Intravoxel incoherent motion (IVIM) MRI in pediatric patients with synovitis of the knee joint: a prospective pilot study. *Pediatr Rheumatol Online J.* 2022;20(1):99. [CrossRef]

23. Barendregt AM, Nusman CM, Hemke R, et al. Feasibility of diffusion-weighted magnetic resonance imaging in patients with juvenile idiopathic arthritis on 1.0-T open-bore MRI. *Skeletal Radiol.* 2015;44(12):1805-1811. [CrossRef]

24. Chang CP, Schumacher HR Jr. Light and electron microscopic observations on the synovitis of ankylosing spondylitis. *Semin Arthritis Rheum.* 1992;22(1):54-65. [CrossRef]

25. van de Sande MG, Baeten DL. Immunopathology of synovitis: from histology to molecular pathways. *Rheumatology (Oxford)*. 2016;55(4):599-606. [CrossRef]

26. Maksymowych WP, Lambert RG, Østergaard M, et al. MRI lesions in the sacroiliac joints of patients with spondyloarthritis: an update of definitions and validation by the ASAS MRI working group. *Ann Rheum Dis.* 2019;78(11):1550-1558. [CrossRef]

27. Mandl P, Navarro-Compán V, Terslev L, et al. EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. *Ann Rheum Dis*. 2015;74(7):1327-1339. [CrossRef]

DIR

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.243148

NEURORADIOLOGY

PICTORIAL ESSAY

Clival and paraclival pathologies: imaging features and differential diagnosis

Ahmet Bozer[1]
Yeliz Pekçevik[1,2]

[1]İzmir City Hospital, Clinic of Radiology, İzmir, Türkiye

[2]University of Health Sciences Türkiye, İzmir University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

Corresponding author: Ahmet Bozer

E-mail: drahmetbozer@gmail.com

Received 02 December 2024; revision requested 01 January 2025; last revision received 29 January 2025; accepted 09 February 2025.

Epub: 27.03.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.243148

ABSTRACT

Clival and paraclival pathologies encompass a broad spectrum of benign and malignant lesions, necessitating accurate imaging for precise diagnosis and management. Magnetic resonance imaging and computed tomography are pivotal in evaluating these lesions, facilitating differentiation, and guiding therapeutic decisions. This study reviews the imaging characteristics, differential diagnoses, and clinical significance of clival and paraclival pathologies.

KEYWORDS

Clivus, computed tomography, chordoma, magnetic resonance imaging, skull base

The clivus, a sloping skull base bone, supports the brainstem and separates the posterior cranial fossa from the sphenoid sinus. It comprises the basisphenoid and basiocciput, which fuse in adulthood. In children, the spheno-occipital synchondrosis, a joint aiding skull base growth, separates these structures and fuses by the late teens to mid-twenties.[1] The central location of the clivus and its proximity to critical neurovascular structures make it vital in both normal and pathological conditions, especially in trauma or skull base lesions.

Clival and paraclival pathologies encompass a wide range of benign and malignant lesions (Table 1). Due to the proximity of the clivus to the brainstem, cranial nerves, and major vessels, a precise imaging diagnosis is essential. Magnetic resonance imaging (MRI) and computed tomography (CT) are essential for localizing, assessing, and differentiating pathologies such as neoplasms, inflammation, and congenital abnormalities, guiding targeted treatment. This essay highlights the imaging features and diagnostic approaches for clival and paraclival lesions. Informed consent was obtained from the patients for the use of the clinical and imaging data included in this review.

Table 1. Classification of clival and paraclival pathologies

Category	Subtype	Pathologies
Primary clival lesions	**Benign lesions**	Benign bone lesions, benign-appearing notochordal lesions
	Malignant lesions	Chordoma, chondrosarcoma, plasmacytoma/multiple myeloma, metastasis
Paraclival pathologies	**Infectious and inflammatory lesions**	Skull base osteomyelitis, IgG4-related disease
Secondary clival lesions: pathologies extending to the clivus	**Head and neck pathologies**	Nasopharyngeal carcinoma, sinonasal malignancies
	Intracranial pathologies	Meningioma, pituitary neuroendocrine tumor

You may cite this article as: Bozer A, Pekçevik Y. Clival and paraclival pathologies: imaging features and differential diagnosis. *Diagn Interv Radiol.* 2025;31(6):619-629.

Imaging techniques

CT and MRI serve as complementary techniques in evaluating clival and paraclival pathologies. CT is superior for the detailed analysis of bony structures, the identification of calcifications, and the evaluation of their patterns. Thin-slice axial and coronal reconstructions are particularly valuable for assessing bone erosion and destruction.

MRI offers superior soft tissue contrast and is essential for evaluating intracranial extensions and bone marrow signal abnormalities. It also plays a critical role in assessing the treatment response and follow-up. Standard multiplanar T1-weighted (T1W) and T2-weighted (T2W) sequences are used alongside post-contrast T1W sequences to evaluate lesion enhancement. Diffusion-weighted imaging and apparent diffusion coefficient (ADC) mapping assist in the differential diagnosis.

Advanced sequences such as fast imaging employing steady-state acquisition (FIESTA) and constructive interference in steady state (CISS) provide high spatial resolution with submillimeter section thickness. These sequences enhance the visualization of small lesions and structures, especially those with high T2 signal intensity, such as cerebrospinal fluid (CSF). Their multiplanar reconstruction capabilities are valuable for the detailed imaging of the brainstem, small cystic lesions, and the surrounding neural structures. These sequences may also be used with contrast, which is useful for the differential diagnosis and the further evaluation of the details and extension of the lesions. The integration of CT and MRI with these advanced imaging techniques enables a comprehensive evaluation of clival and paraclival pathologies. Additionally, fluorodeoxyglucose positron emission tomography (FDG PET)/CT can complement CT and MRI, particularly for staging purposes and the assessment of the treatment response.

Main points

- Clival and paraclival pathologies, which extend into the clivus, are categorized into primary clival lesions (benign and malignant), paraclival infectious/inflammatory conditions, and secondary clival lesions.
- High-resolution sequences, such as constructive interference in steady state/fast imaging employing steady-state acquisition, diffusion-weighted imaging, and post-contrast sequences, are valuable for the differential diagnosis of clival lesions.
- Although chordomas and chondrosarcomas have overlapping imaging features, higher apparent diffusion coefficient values in chondrosarcomas may aid in their differentiation.
- The radiological and pathological features of benign notochordal cell tumors overlap with those of clival chordomas, justifying the use of the term "benign-appearing notochordal lesions," which should be monitored with imaging every 6 months.

Normal anatomy of the clivus

The clivus is bordered superiorly by the dorsum sellae and sella, inferiorly by the foramen magnum, and laterally by the petroclival fissure, which includes the petro-occipital synchondrosis, petrous temporal bone, and internal carotid artery. Anteriorly, it adjoins the sphenoid sinus and nasopharynx, and posteriorly, the prepontine and premedullary cisterns, containing the brainstem (Figure 1).

The clivus is best visualized on a sagittal T1W MRI. It shows age-related changes in the bone marrow signal caused by the increasingly yellow (fatty) marrow, leading to T1W hyperintensity. A normal clival signal is typically isointense or hyperintense relative to the pons, and T1W hypointensity warrants suspicion. Clival marrow heterogeneity may manifest as hypointense foci. The T1W signal is categorized as follows: grade 1 indicates a low signal, occupying >50%; grade 2 covers 20%–50%; and grade 3 exhibits a high signal, with a low signal occupying <20% (Figure 2). Mild enhancement is often noted in grade 1 T1W signals.[2]

Primary clival lesions

Benign lesions

Fibrous dysplasia

Fibrous dysplasia is a benign fibro-osseous condition, in which normal bone is replaced with fibrous tissue. It affects all ages but is commonly diagnosed in childhood or early adulthood, with no gender predilection. Although usually sporadic, it can be linked to syndromes such as McCune–Albright and Mazabraud or isolated endocrinopathies.[3]

Most cases (70%–80%) are monostotic, commonly involving craniofacial bones, long bones, and the spine.[4] Clival involvement is rare, with only 44 cases reported in a 2023 review.[5] Diagnosis is primarily radiological, with histopathology reserved for symptomatic or atypical monostotic cases.

On imaging, CT typically reveals an expansile, ground-glass intramedullary lesion diagnostic of fibrous dysplasia. On MRI, it appears hypointense on T1W signals and exhibits a variable T2W signal with moderate-to-marked post-contrast enhancement (Figure 3).

Management typically involves observation and bone quality maintenance. Imaging should evaluate benign matrix transformation (e.g., aneurysmal bone cyst-like changes, myxoid degeneration) and malignant features, such as fractures, cortical destruction, aggressive periosteal reactions, marrow edema, solid masses, or soft tissue components.

Differential diagnoses include other fibro-osseous lesions, Paget's disease, and primary clivus lesions such as intraosseous

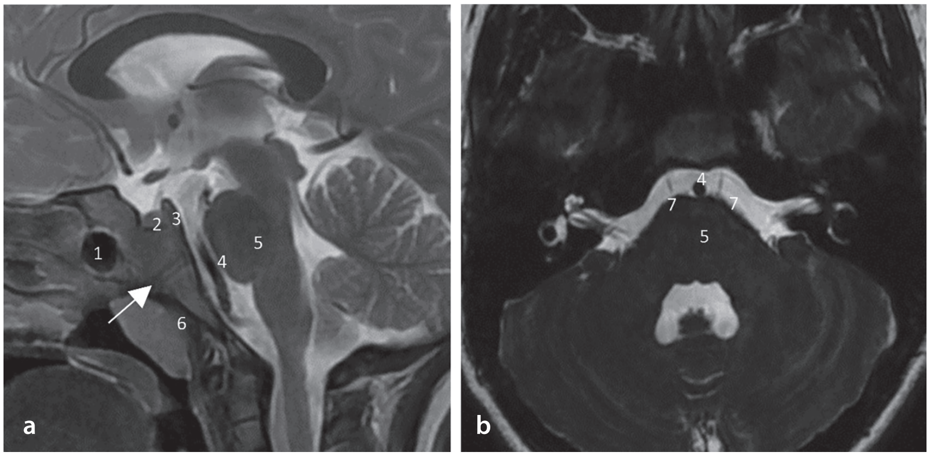


Figure 1. Magnetic resonance imaging anatomy of the clivus in a 4-year-old male pediatric patient. Sagittal T2-weighted (T2W) image **(a)** and axial heavily T2W image **(b)**, demonstrating the key anatomical structures. The spheno-occipital synchondrosis (arrow), located between the basisphenoid and basiocciput, is clearly visible. The key adjacent structures are labeled: 1. Sphenoid sinus, 2. pituitary gland, 3. dorsum sella, 4. basilar artery, 5. pons, 6. nasopharynx, 7. abducens nerve.

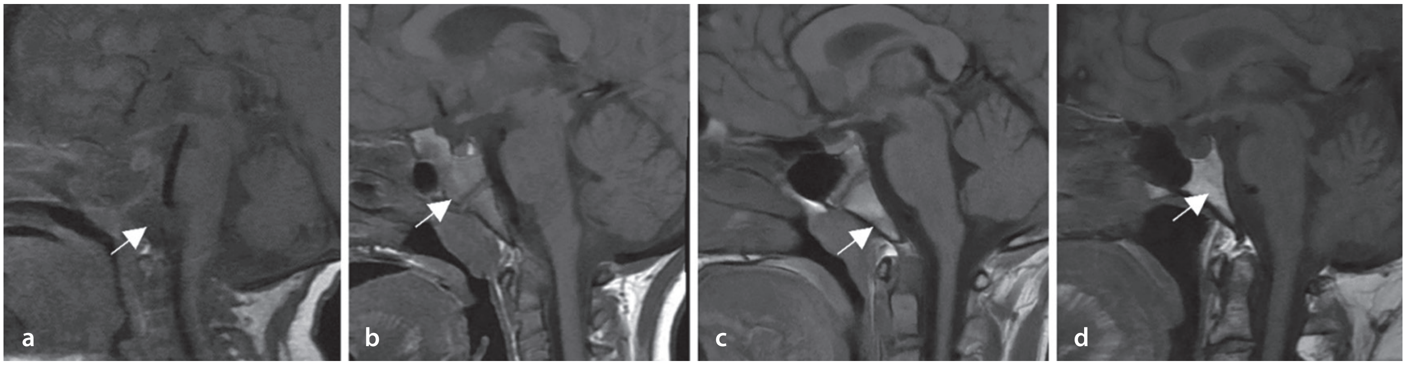


Figure 2. Signal changes in clival bone marrow on non-fat-suppressed T1-weighted (T1W) magnetic resonance imaging. Sagittal T1W images **(a-d)** without fat suppression from patients at 6 months **(a)**, 4 years **(b)**, 14 years **(c)**, and adulthood **(d)** are shown. The images demonstrate the gradual replacement of red bone marrow (hypointense) by fatty bone marrow. The signal intensity changes are graded as grade 1 **(a)**, grade 2 **(b)**, and grade 3 **(c, d)**, as indicated by the arrows.

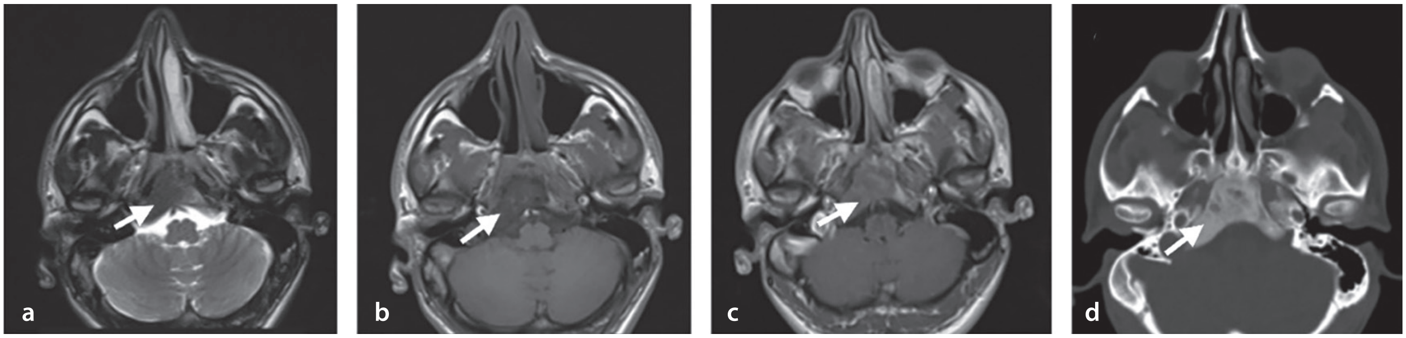


Figure 3. Imaging features of a lesion consistent with monostotic fibrous dysplasia in the clivus in a 31-year-old male patient. An incidental lesion in the clivus, appearing hypointense on T2-weighted **(a)** and T1-weighted (T1W) **(b)** sequences. The lesion is intramedullary in location and mildly expansile (arrow). On post-contrast T1W imaging **(c)**, the lesion exhibits moderate enhancement, raising suspicion for fibrous dysplasia, and a computed tomography (CT) scan was recommended. The CT scan **(d)** confirms a well-defined, mildly expansile lesion with ground-glass density, supporting the diagnosis of fibrous dysplasia (arrow).

meningioma and chordoma. Paget's disease may appear radiologically similar to fibrous dysplasia, but it primarily occurs in older individuals, whereas fibrous dysplasia can be seen across a broader range of age groups. Intraosseous meningiomas tend to be less extensive than fibrous dysplasia, with less bony remodeling. Chordoma can be easily distinguished from fibrous dysplasia because of its lytic destructive appearance, possible expansile solid component, and more pronounced contrast enhancement.

Benign-appearing notochordal lesions

Ecchordosis physaliphora and benign notochordal cell tumors

Ecchordosis physaliphora is a benign hamartomatous lesion from ectopic notochordal remnants. It is non-proliferative, non-invasive, and asymptomatic but can present with symptoms such as headache, CSF leaks, or diplopia. Ecchordosis physaliphora is classified as a benign notochordal cell tumor in current pathology systems.[6]

The radiological and pathological features of BNCT overlap with those of clival chordoma, making the term "benign-appearing notochordal lesions" preferable for EP/BNCT.[7] MRI reveals T1-hypointense and T2-hyperintense lesions, similar to chordomas, although chordomas are often enhanced with contrast. The FIESTA and CISS sequences are valuable for identifying small intracranial lesions, providing high-resolution images for multiplanar reconstruction. These sequences are recommended for diagnosing and classifying clival lesions.[8]

Benign notochordal cell tumors remain stable during follow-up, whereas low-grade chordomas exhibit slow growth, although differentiation remains challenging. These lesions can be defined as "benign-appearing notochordal lesions" and should be monitored through imaging at 6-month intervals.[7]

Surgical considerations include tumor size (<3 cm), growth, local invasion, bony erosion, patient age (<30 years), symptoms, and patient preference (Figure 4).

Malignant lesions

Chordomas

Chordomas, arising from notochord remnants, account for 1%–4% of bone malignancies and 0.5% of primary intracranial central nervous system tumors.[9] Cranial chordomas represent approximately 40% of cases, with sacral and spinal forms being less common. They predominantly affect adults, especially men, with peak incidence in the 70–80-years age range.[10]

Radiologically, CT shows a lytic, expansile mass with secondary calcifications caused by sequestration. On MRI, chordomas appear hyperintense on T2W images and hypointense on T1W images, occasionally with hemorrhagic foci. Post-contrast imaging often reveals a honeycomb pattern (Figures 5 and 6).

Despite their slow growth and low-grade histology, chordomas are locally invasive, recur frequently, and are classified as malignant. Treatment typically involves surgery with adjuvant radiotherapy. Surgical resection is critical for progression-free and overall survival.

Differential diagnoses include chondrosarcoma, metastasis, plasmacytoma, and, rarely, jugular paraganglioma. No conventional CT or MRI feature reliably distinguishes chordomas from chondrosarcomas; how-

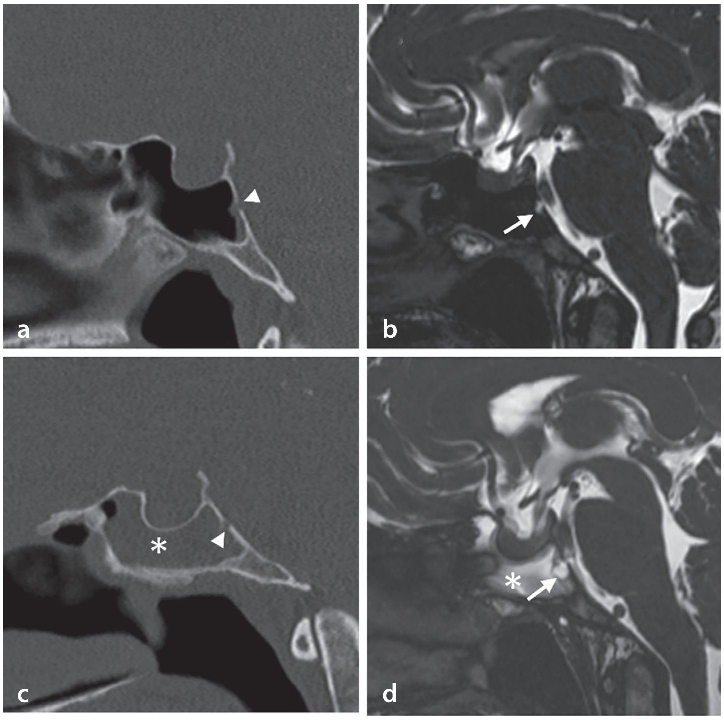


Figure 4. Benign-appearing notochordal lesion causing cerebrospinal fluid rhinorrhea and follow-up imaging in a 45-year-old female patient. Initial evaluation with computed tomography (CT) **(a)** reveals a clival defect (arrowhead) and a hyperintense focal clival lesion (arrow) on constructive interference in steady-state (CISS) imaging **(b)**. The patient declined surgical intervention and, after 4 years, presented with exacerbated symptoms. Follow-up CT **(c)** and CISS imaging **(d)** demonstrate fluid accumulation in the sphenoid sinus (asterisk) and an increase in the size of the clival lesion (arrow).

ever, ADC values, enhancement patterns, and bone changes may help differentiate the two, as they often share overlapping imaging features. Chondrosarcomas typically exhibit higher ADC values than chordomas.[11] The high T2 signal intensity of chondroid lesions helps differentiate them from metastases and plasmacytomas, whereas poorly differentiated chordomas may exhibit a low T2 signal. Paragangliomas have a distinct salt-and-pepper appearance, flow voids, and moth-eaten bone destruction.

Chondrosarcomas

Skull base chondrosarcomas are rare tumors that typically arise from the petro-occipital synchondrosis, which is believed to explain their usual off-midline location. However, 10%–30% may occur at the midline.[12] These slow-growing tumors often present with brainstem or cranial nerve compression symptoms.

On CT, these tumors may exhibit typical chondroid calcifications (popcorn-like or ring-and-arc patterns) and appear as destructive, heterogeneous masses. On MRI, they are hypointense on T1W images, hyperintense on T2W images, and demonstrate

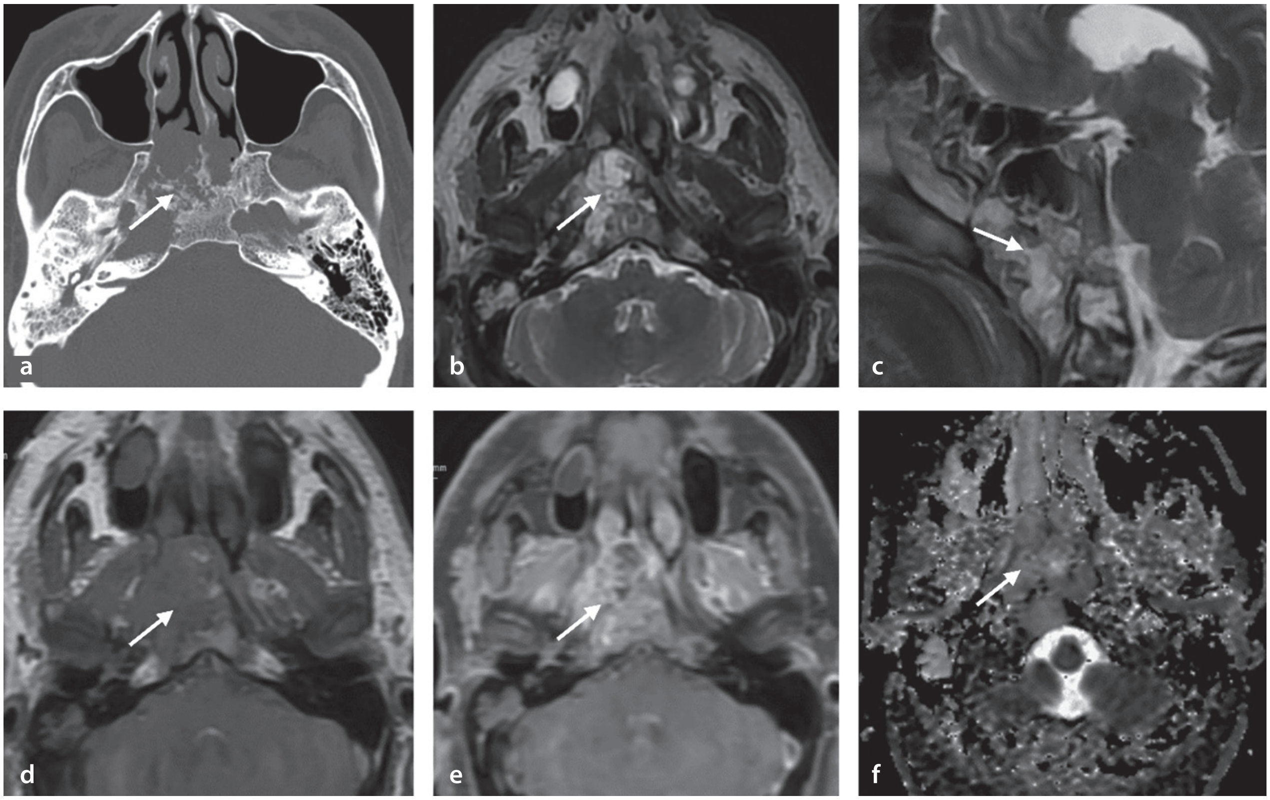


Figure 5. Imaging findings from a 60-year-old male patient who presented with hearing loss and sudden-onset diplopia, subsequently diagnosed with clival chordoma. Axial computed tomography image **(a)** reveals a lytic, destructive mass containing sequestered bone tissue (arrow). The T2-weighted magnetic resonance images **(b, c)** reveal a hyperintense lesion. The T1-weighted (T1W) image **(d)** demonstrates a hypointense lesion with hyperintense foci suggestive of hemorrhage. A post-contrast T1W image **(e)** displays moderate heterogeneous enhancement with a honeycomb-like pattern. On the apparent diffusion coefficient (ADC) map **(f)**, the ADC value is measured at 1.595 × 10^{-3} mm^2/sec. The patient received external beam radiotherapy following the surgical diagnosis of chordoma.

post-contrast heterogeneous enhancement with a ring-and-arc pattern (Figure 7). Chordomas should be considered in the differential diagnosis. Differentiation from chordomas is best made based on location, as chordomas arise from notochordal remnants and typically occur along the midline, whereas chondrosarcomas are typically centered on the petro-occipital fissure. Additionally, ADC values can help distinguish between the two, as chondrosarcomas generally have higher ADC values.

Treatment involves maximal safe resection followed by adjuvant radiotherapy.

Multiple myeloma

One of the diagnostic criteria for multiple myeloma is bone involvement. The most commonly affected region is the axial skeleton, particularly the vertebrae. Radiological evaluation plays a crucial role in supporting the diagnosis, excluding other causes, and identifying potential complications.

Typically, well-defined lytic bone lesions are observed in the commonly affected axial skeleton. However, one of the bone marrow involvement patterns may also present as entirely normal-appearing bone marrow.

On MRI, five active bone marrow involvement patterns have been described, including normal-appearing marrow, a focal pattern, a diffuse pattern, a salt-and-pepper (micronodular) pattern, and a combination of focal lesions within a diffuse pattern.

Active lesions are hyperintense on T2W images, hypointense on T1W images, and demonstrate enhancement on post-contrast T1W images. These lesions also demonstrate diffusion restriction (Figure 8).

The differential diagnosis includes lytic bone metastases.

Metastasis

Prostate, breast, and lung cancers account for most bone metastases. Lesions may appear as lytic, sclerotic, or mixed patterns on imaging and are often associated with bone marrow signal changes on MRI (Figure 9). Clival metastases can cause cranial nerve deficits or skull base compression symptoms, necessitating thorough radiological evaluation for an accurate diagnosis and treatment planning.

In elderly patients without a known malignancy, multiple lytic lesions raise suspicion for multiple myeloma as a key differential diagnosis. Features favoring metastases include the involvement of vertebral pedicles rather than vertebral bodies and distal ap-

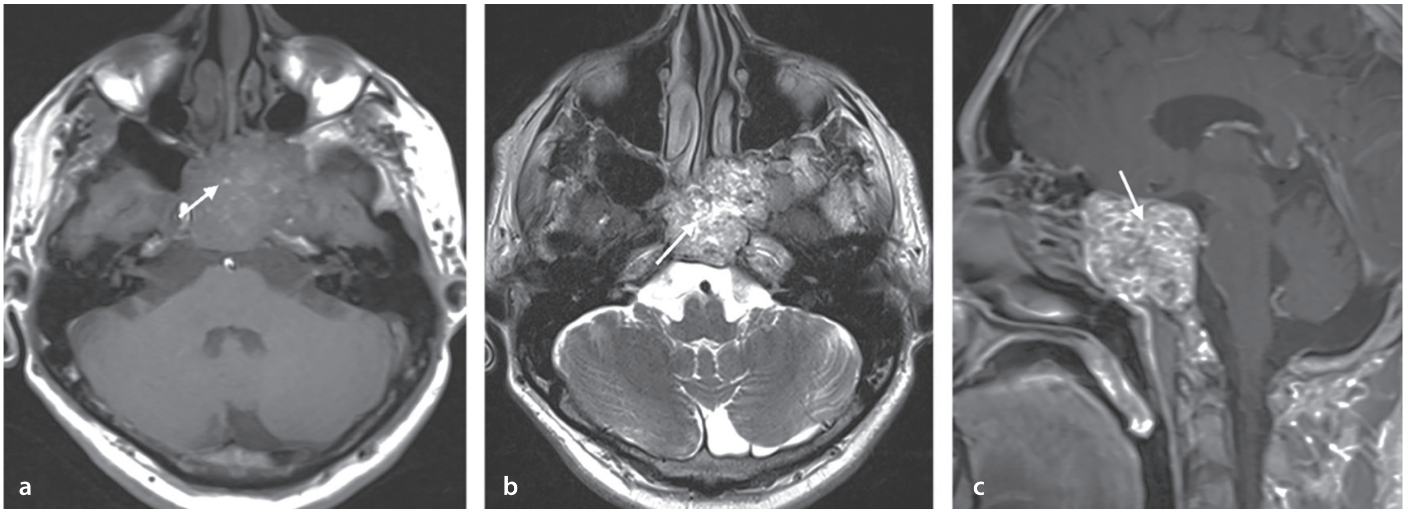


Figure 6. Magnetic resonance images of a 62-year-old male patient with a clival chordoma. The lesion appears hypointense on the T1-weighted (T1W) imaging **(a)**, with punctate hyperintense foci, hyperintense on T2-weighted imaging **(b)**, and demonstrates a honeycomb-like enhancement on post-contrast T1W imaging **(c)**, consistent with a diagnosis of clival chordoma (arrow).

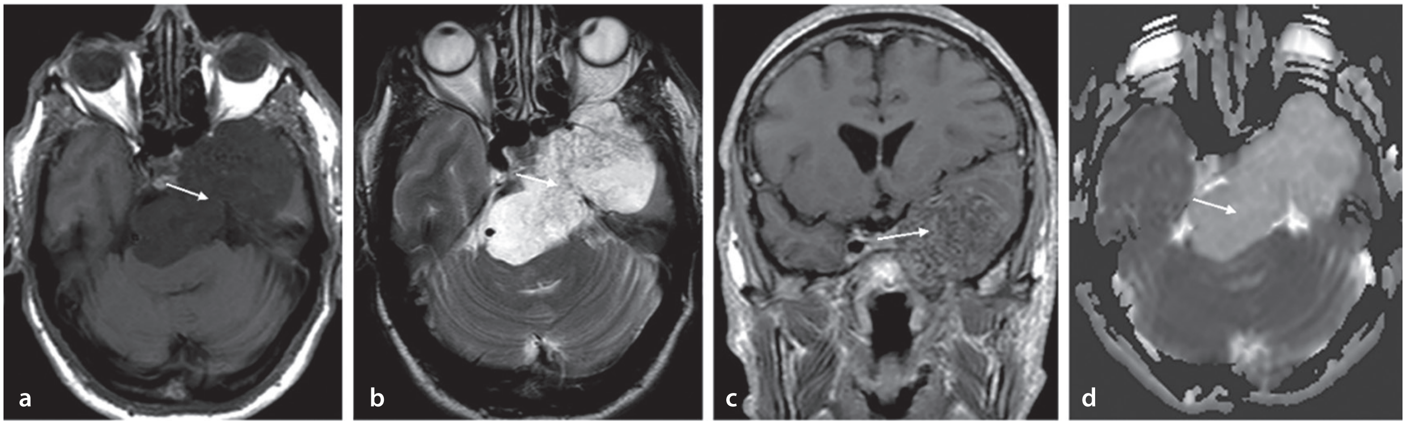


Figure 7. Magnetic resonance images of a 52-year-old male patient with chondrosarcoma. The lesion is hypointense on the T1-weighted (T1W) imaging **(a)**, hyperintense on the T2-weighted imaging **(b)**, and demonstrates septal enhancement on the post-contrast T1W imaging **(c)**. The apparent diffusion coefficient (ADC) map **(d)** reveals high ADC values (1.955 × 10^{-3} mm^2/sec) in this expansile mass arising from the petro-occipital synchondrosis, extending posteriorly toward the brainstem and anteriorly into the temporal region (arrow). The histopathological diagnosis confirmed chondrosarcoma.

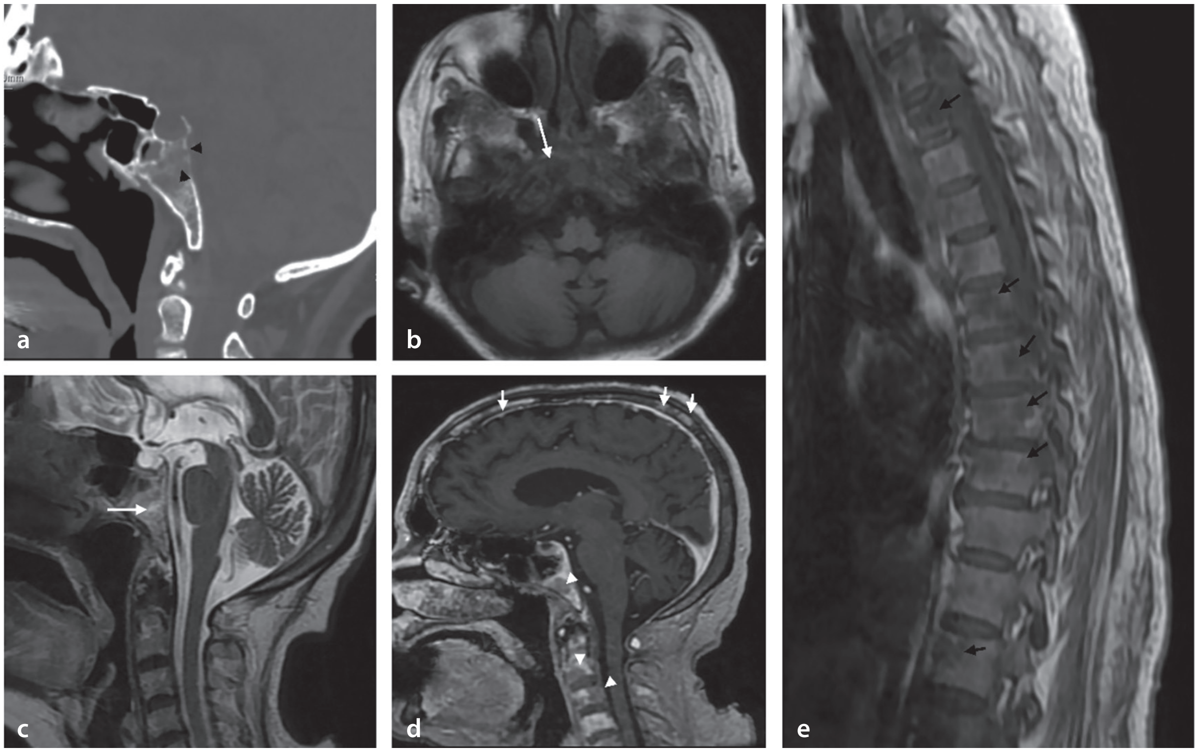


Figure 8. A 71-year-old female patient diagnosed with multiple myeloma. Computed tomography scan **(a)** reveals a lytic, destructive lesion in the clivus (black arrowhead). Axial T1-weighted (T1W) magnetic resonance imaging (MRI) **(b)** reveals the lesion as hypointense, whereas the sagittal T2-weighted MRI **(c)** demonstrates it as hyperintense (long white arrow). Post-contrast sagittal T1W MRI **(d)** reveals multiple lesions in the calvarium (short white arrows), clivus, and cervical vertebrae (white arrowheads). Sagittal T1W thoracic MRI **(e)** reveals diffuse bone marrow signal reduction with multiple focal hypointense lesions (black arrows). The combination of focal lesions within a diffuse pattern, along with bone marrow involvement, older age, and clinical-laboratory findings, supports the diagnosis of multiple myeloma.

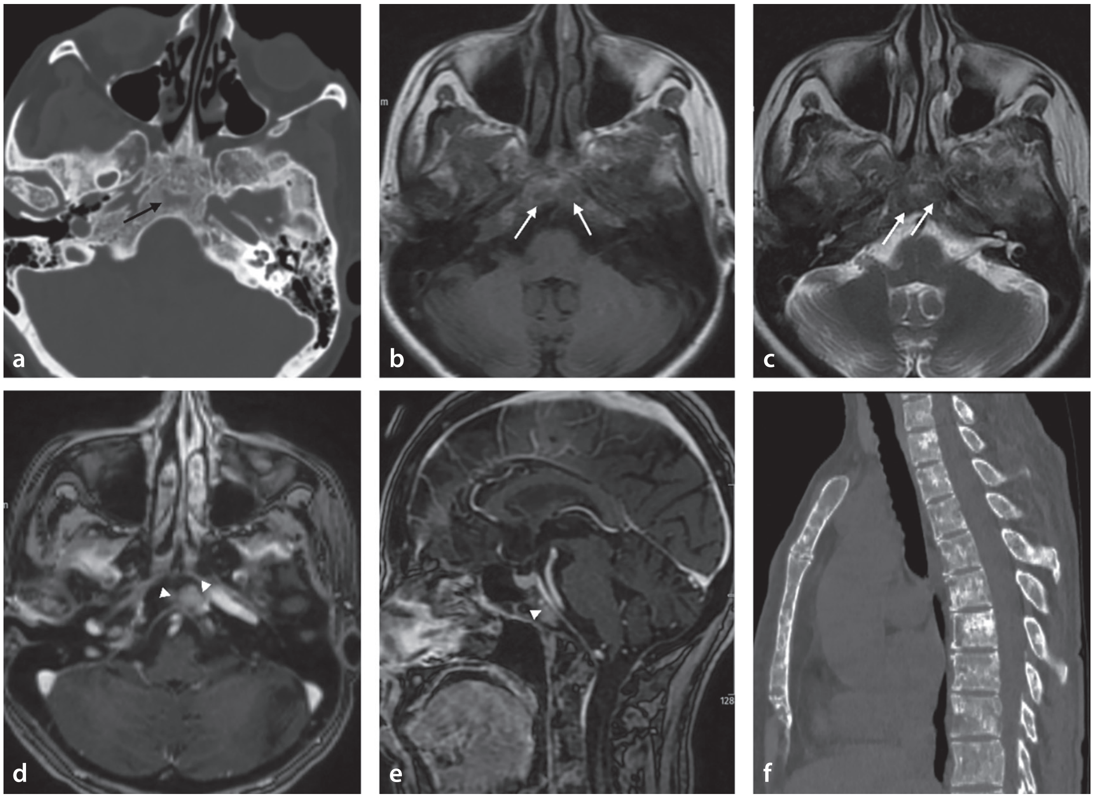


Figure 9. Images of a 60-year-old female patient diagnosed with breast cancer revealing clivus metastasis. Computed tomography (CT) scan **(a)** demonstrates a lytic lesion (black arrow). Axial T1-weighted (T1W) magnetic resonance imaging (MRI) **(b)** and T2-weighted MRI **(c)** reveal hypointense lesions (white arrows). Axial **(d)** and sagittal **(e)** post-contrast T1W images demonstrate homogeneously enhancing lesions (arrowhead). Sagittal CT image **(f)** revealing mixed lytic and sclerotic lesions consistent with mixed-type bone metastasis.

pendicular skeleton involvement. Although both conditions may exhibit variable bone scan findings on bone scintigraphy, including areas of increased (hot) or decreased (cold) uptake, extensive bone metastases rarely present with a normal scan appearance, unlike multiple myeloma.

Paraclival pathologies

Infectious and inflammatory lesions

Skull base osteomyelitis

Skull base osteomyelitis (SBO), often linked to necrotizing external otitis, is classified as lateral or central based on the infection source. Lateral SBO arises from otogenic or odontogenic infections, affecting the temporal bone and nearby structures, whereas central SBO originates from sinus infections, involving the clivus and sphenoid bones. It has an insidious onset, often presenting with delayed symptoms, such as persistent headaches and neurological deficits, after the apparent improvement of the initial otogenic infection.

CT and MRI are vital for diagnosing SBO, although imaging findings may lag behind clinical symptoms. Early bone destruction is often missed on CT, which primarily identifies skull base extension and temporomandibular joint involvement in chronic necrotizing otitis externa. MRI shows a decreased T1W signal, T2W hypointensity caused by necrosis, and diffuse post-contrast enhancement indicating extension of the infection (Figure 10). Diffusion-weighted imaging reveals higher ADC values than for malignancies (e.g., nasopharyngeal cancer, lymphoma, metastases).[13] Imaging is crucial for tracking infection spread, abscesses, and intracranial and vascular complications such as stroke, venous sinus thrombosis, or pseudoaneurysms.[14]

Differential diagnoses include nasopharyngeal carcinoma (NPC), which appears isointense to muscle on T1W images and iso- to hyperintense on T2W images, often revealing a mass with diffusion restriction, heterogeneous post-contrast enhancement, and potential perineural spread. Metastases typically exhibit variable, heterogeneous signal intensity on T1W and T2W images, with irregular post-contrast enhancement on T1W images. Lymphoma is characterized by mild hypointensity on T1W images, homogeneous hyperintensity on T2W images, and uniform post-contrast enhancement. Minor salivary gland tumors, such as in mucin-producing areas, display variable T2W signal intensity and are iso- to hypointense on T1W images with gradual post-contrast enhancement. Key MRI features, including a mass with low ADC values and cervical lymphadenopathy, assist in distinguishing malignancies; however, surgical biopsy remains essential for a definitive diagnosis.

SBO is most commonly caused by *Pseudomonas aeruginosa* in patients with diabetes and those who are immunocompromised, with *Aspergillus* as the leading fungal pathogen. Treatment involves prolonged intravenous antibiotics or antifungal therapy. Surgery is rarely performed because of the inaccessible location. Imaging may lag behind clinical recovery; therefore, treatment efficacy is primarily assessed by the resolution of the symptoms and inflammatory markers rather than imaging findings. Moreover, ^{18}F-FDG PET imaging may be used for the treatment response assessment and management guidance, especially in patients with SBO who experience treatment side effects.[15]

Secondary clival lesions: pathologies extending to the clivus

Head and neck pathologies

Nasopharyngeal carcinoma

Radiology is critical for NPC staging, treatment planning, and response monitoring, with MRI being the modality of choice because of its superior soft tissue resolution and sensitivity to perineural spread and intracranial extension. According to the American Joint Committee on Cancer tumor, node, and metastasis cancer staging system (8th edition), the involvement of bony structures (skull base, cervical vertebrae) and/or paranasal sinuses is classified as T3, with clival involvement also categorized as T3.[16]

On MRI, NPC appears isointense to muscle on T1W images and iso- to hyperintense on T2W images, with diffusion restriction, heterogeneous post-contrast enhancement, and perineural spread (Figure 11). In cases of skull base invasion, differential diagnoses include plasmacytoma, lymphoma, metastases, and other malignancies.

Sinonasal malignancies

CT is more effective for bone invasion, whereas MRI is more effective for tumor spread, dural invasion, and perineural spread, which is vital for pre-treatment staging. The T2W signal intensity varies—mucin-producing adenocarcinomas are hyperintense, whereas non-mucin tumors are iso-hypointense. On T1W images, tumors appear iso-hypointense with gradual post-contrast enhancement.

Differentiating sinonasal carcinoma from adenocarcinoma radiologically can be difficult. Sinonasal lymphoma and extramedullary plasmacytoma should be considered in the differential diagnosis. Lymphoma exhibits homogeneous enhancement, causes bone remodeling, lacks necrosis, and has a low ADC value. In our case, the differential diagnoses include sinonasal malignancy, lymphoma, and metastasis. The heterogeneous signal and enhancement, along with the lack of significant diffusion restriction, favor sinonasal malignancies over lymphoma (Figure 12).

Sinonasal carcinomas, which are locally invasive and prone to recurrence, are treated with surgery followed by adjuvant chemoradiotherapy. By contrast, lymphomas are primarily treated with chemotherapy and/or radiation therapy.

Intracranial pathologies

Pituitary neuroendocrine tumor/ pituitary adenoma

Pituitary adenomas are now classified as pituitary neuroendocrine tumors (PitNETs) in the revised fifth edition of the World Health Organization classification. Their behavior code has changed from "0" (benign) to "3" (malignant). Most PitNETs are benign, with a recurrence rate of less than 5% after surgery, requiring minimal further treatment. Aggressive PitNETs are rare, accounting for less than 1% of cases.[17]

Clival invasion by PitNETs occurs in 8.2% of cases, usually by direct extension, whereas extrasellar PitNETs, originating from the clivus, account for 7.2%.[18,19] PitNETs should be considered in the differential diagnosis of clival masses.

The MRI signal intensity of PitNETs varies depending on degeneration, hemorrhage, or infarction. Moreover, T1W images exhibit mild hypointensity or isointensity, whereas T2W images exhibit variable signal intensities. On contrast-enhanced MRI, PitNETs are typically hypointense or isointense compared with the pituitary gland (Figure 13). Imaging strategies depend on tumor size and hormonal activity.

Meningiomas

Bone involvement may result from hyperostosis, neoplastic infiltration, or prima-

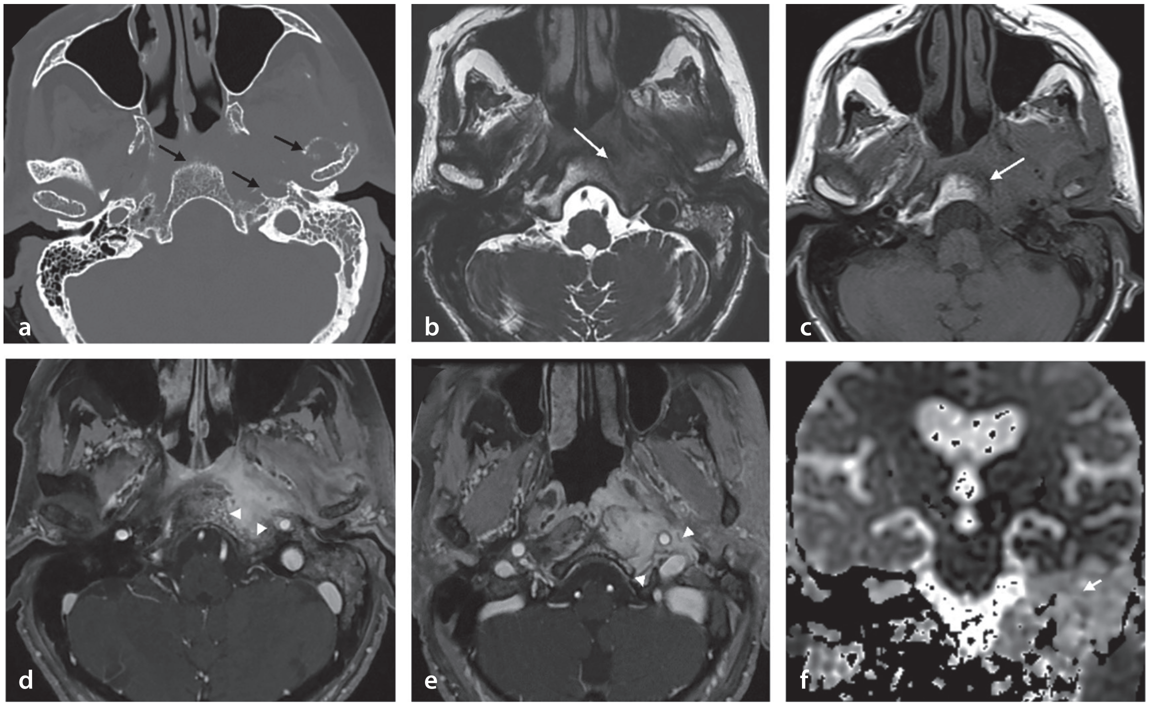


Figure 10. Imaging findings consistent with necrotizing otitis externa and skull base osteomyelitis in a 67-year-old female patient with a history of diabetes mellitus, presenting with prolonged severe ear pain. Computed tomography **(a)** reveals opacification in the mastoid and middle ear, along with erosion of the clivus, mandibular condyle, and temporal bone (black arrows). Axial heavily T2-weighted image **(b)** demonstrates slight hyperintensity, and non-fat-saturated T1-weighted (T1W) images **(c)** reveal reduced bone marrow signal (white arrow). Post-contrast fat-saturated T1W images **(d, e)** reveal clival invasion with extension into the pharyngeal mucosal area, parapharyngeal, and carotid space (arrowheads). The apparent diffusion coefficient (ADC) map **(f)** demonstrates hyperintensity and high ADC values (small arrow). The patient's ear culture grew *Pseudomonas aeruginosa*, and intravenous treatment is ongoing.

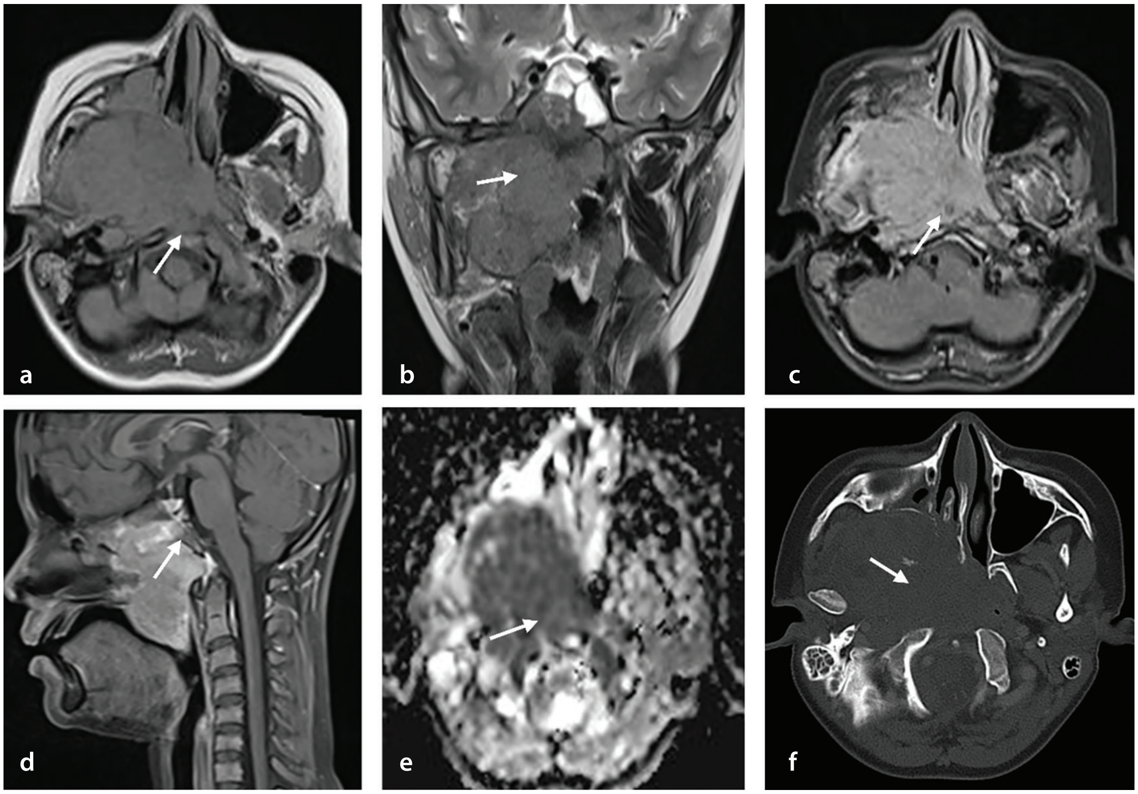


Figure 11. Imaging findings in a 14-year-old male patient diagnosed with T4 stage nasopharyngeal carcinoma. Magnetic resonance imaging reveals a mass that is isointense on T1-weighted images **(a)**, iso-hyperintense on T2-weighted images **(b)**, and has heterogeneous, marked contrast enhancement on post-contrast axial **(c)** and sagittal **(d)** images. It has low apparent diffusion coefficient values **(e)**. The histopathological diagnosis was nasopharyngeal carcinoma, classified as T4 stage nasopharyngeal carcinoma according to the American Joint Committee on Cancer tumor, node, metastasis classification system (8th edition). Computed tomography imaging **(f)** reveals extensive erosion of bony structures and skull base invasion.

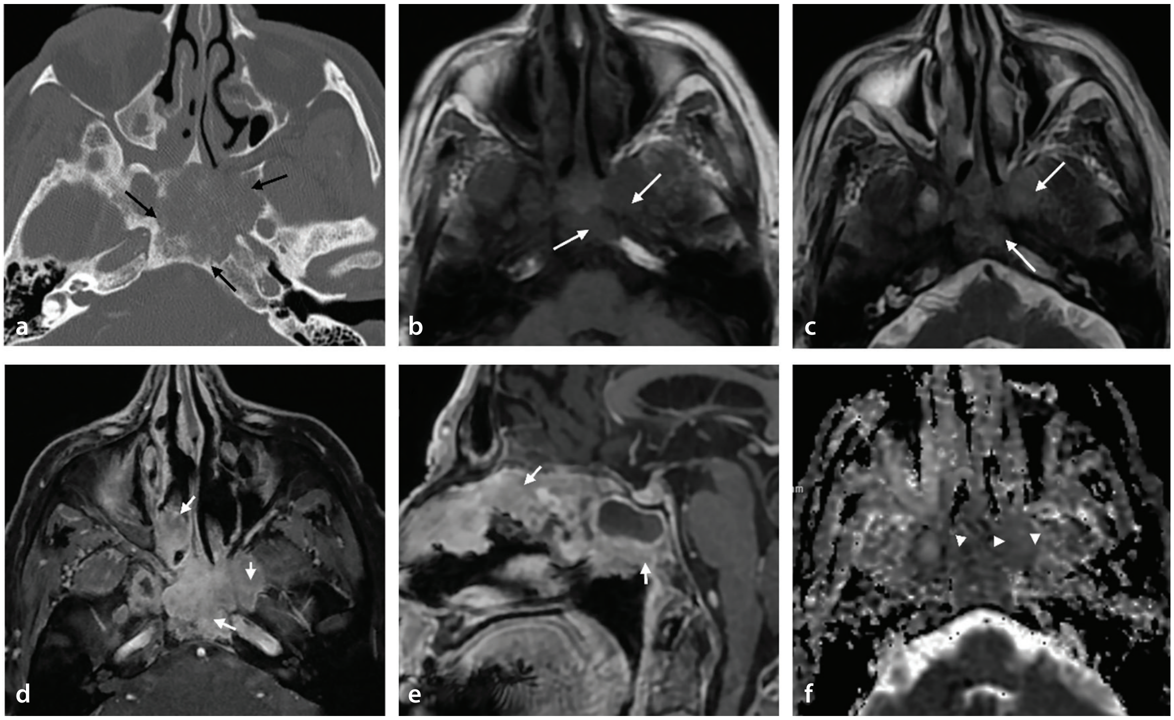


Figure 12. Imaging findings in a 72-year-old male patient presenting with strabismus and diplopia diagnosed with sinonasal mucoepidermoid carcinoma. Computed tomography images **(a)** reveal sclerosis and a lytic destructive appearance extending to the clivus in the sphenoid bone (black arrows). Magnetic resonance imaging reveals the lesion as iso-hypointense on T1-weighted (T1W) **(b)** and slightly hyperintense signal on T2-weighted **(c)** sequences (white arrows). Post-contrast T1W axial **(d)** and sagittal **(e)** images demonstrate invasion from the sphenoid sinus to the sphenoid wing, nasal cavity, cavernous sinuses, and skull base (short arrows). The apparent diffusion coefficient (ADC) map **(f)** reveals hypointensity and low ADC values (arrowheads). Histopathological examination confirmed high-grade mucoepidermoid carcinoma. The patient is undergoing concurrent chemoradiotherapy after surgery.

ry intraosseous meningioma, which is rare, comprising 2%–2.4% of all meningiomas.[20] It typically affects the frontal and parietal bones, with clival involvement being uncommon.

Most intraosseous meningiomas are osteoblastic, causing hyperostosis and occasionally mimicking fibrous dysplasia. In rare cases, they present as osteolytic lesions. CT typically reveals hyperostosis, with occasional lytic lesions. On MRI, the lesion is hypointense on T1W images, isointense with the cortex on T2W images, and demonstrates homogeneous enhancement on post-contrast images (Figures 14 and 15).

Differential diagnoses include Paget's disease, fibrous dysplasia, osteoma for osteoblastic forms, and metastasis and plasmacytoma for osteolytic forms. Skull base meningiomas should also be differentiated from perineural spread in head and neck malignancies.

Meningiomas grow slowly, with surgery as the primary treatment, using approaches such as endoscopic endonasal, middle fossa, or posterior fossa surgery. Radiation therapy is added for partial or subtotal resection.

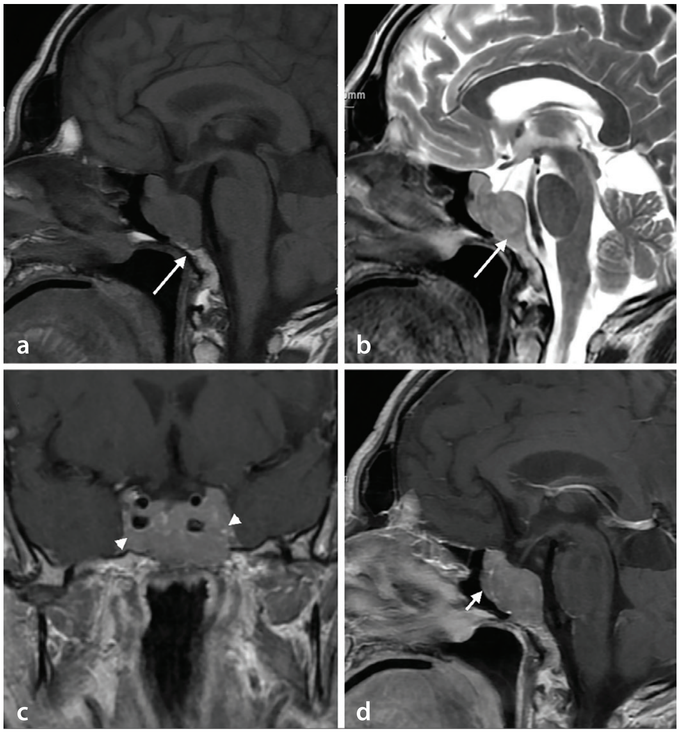


Figure 13. Magnetic resonance images (MRI) of a 43-year-old male patient presenting with headache and dizziness, revealing a pituitary neuroendocrine tumor (macroadenoma) invading the clivus. On MRIs, the lesion appears hypointense on T1-weighted (T1W) **(a)** and hyperintense on T2-weighted **(b)** images, invading the clivus (long arrow). The post-contrast T1W coronal image **(c)** demonstrates invasion of the cavernous sinuses (arrowhead), and the sagittal image **(d)** reveals indentation into the sphenoid sinus (short arrow) with mild contrast enhancement of the sellar mass. The lesion was histopathologically diagnosed as a pituitary neuroendocrine tumor.

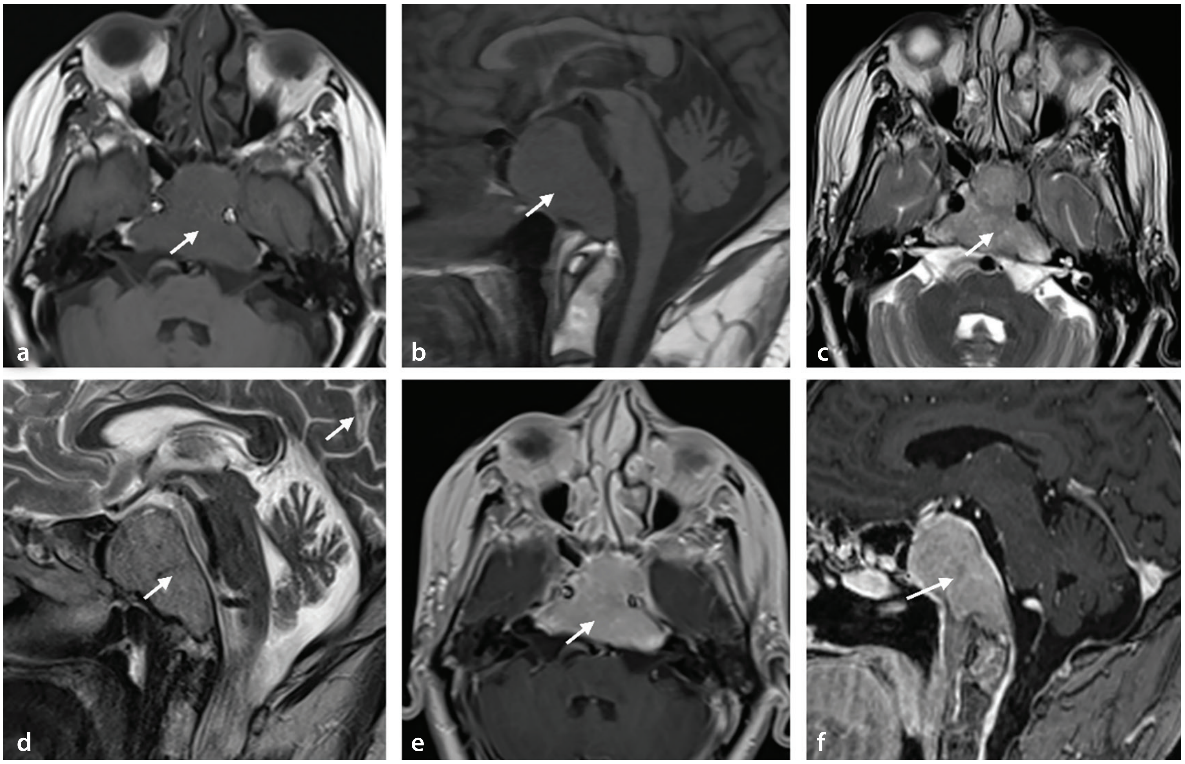


Figure 14. Magnetic resonance images of a 57-year-old male patient presenting with headache and restricted gaze, diagnosed with clival meningioma. The lesion (arrows) appears iso-hypointense on axial **(a)** and sagittal **(b)** T1-weighted (T1W) images and isointense on axial **(c)** and sagittal **(d)** T2-weighted images. Post-contrast T1W axial **(e)** and sagittal **(f)** images demonstrate homogeneous contrast enhancement of the clival mass, which extends laterally to the left cavernous sinus, anteriorly to the sphenoid sinus, and posteriorly to the prepontine cistern. Following the diagnosis of meningioma, the patient underwent radiotherapy.

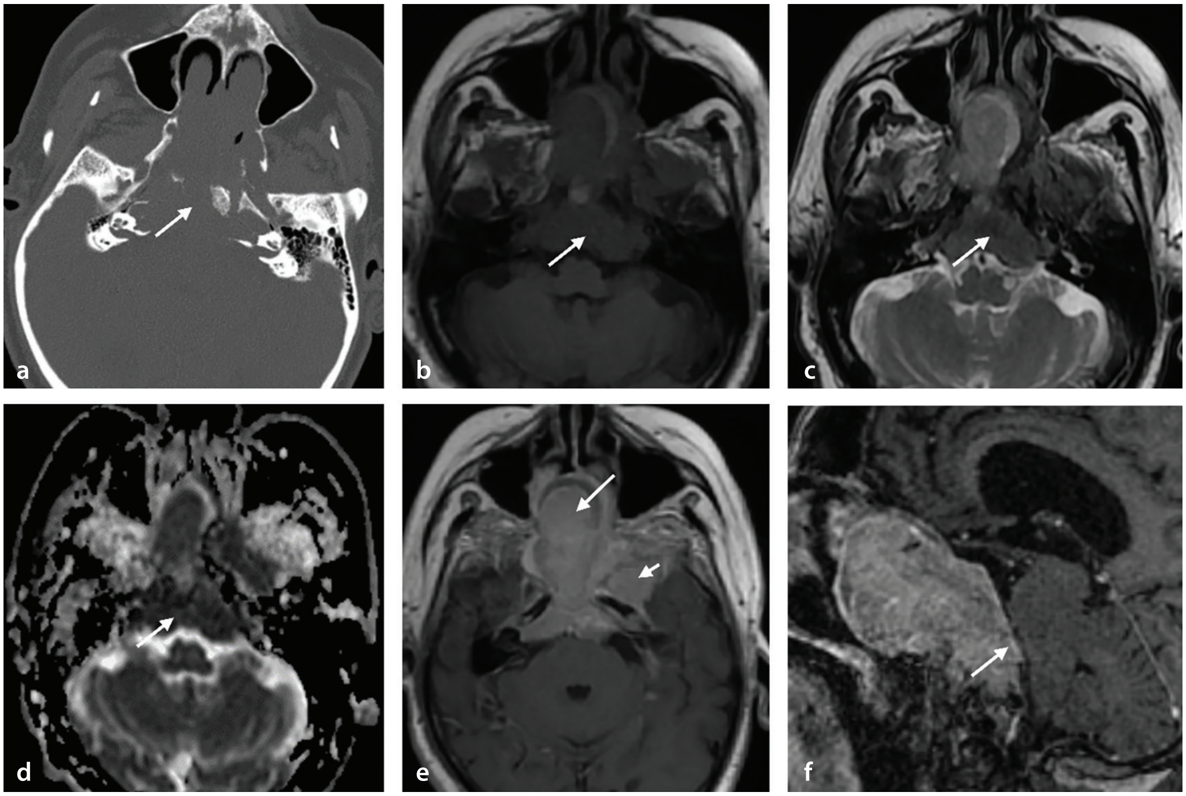


Figure 15. Radiological images of a 74-year-old male patient presenting with nausea, vomiting, and left eyelid ptosis, diagnosed with atypical meningioma (World Health Organization grade 2) invading the clivus. The computed tomography scan **(a)** reveals a lytic expansile lesion, whereas the lesion appears isointense on T1-weighted **(b)** and T2-weighted **(c)** images, with low apparent diffusion coefficient (ADC) values on the ADC map **(d)**. Post-contrast images **(e, f)** demonstrate homogeneous contrast enhancement of the expansile clival lesion (arrow). The lesion invades the right sphenoid sinus, extends bilaterally into the parasellar region, reaches the masticator space on the left, and posteriorly extends into the prepontine cistern.

Conclusion

This pictorial essay underscores the critical role of imaging in diagnosing and differentiating clival and paraclival pathologies. Comprehensive radiologic evaluation using advanced MRI and CT techniques facilitates accurate characterization, aiding in optimal management strategies. Understanding the imaging features and their clinical correlations is essential for diagnosing these complex conditions.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. McDonald SW, Miller J. When does the spheno-occipital synchondrosis close? *Clin Anat*. 2022;35(4):512-525. [CrossRef]
2. Kimura F, Kim KW, Friedman H, Russell EJ, Breit R. MR imaging of the normal and abnormal clivus. *Am J Roentgenol*. 1990;155(6):1285-1291. [CrossRef]
3. Adrien M, Nadège B, William K. Mazabraud's syndrome. *J Belgian Soc Radiol*. 2022;106(1):71. [CrossRef]
4. Kinnunen AR, Sironen R, Sipola P. Magnetic resonance imaging characteristics in patients with histopathologically proven fibrous dysplasia—a systematic review. *Skeletal Radiol*. 2020;49(6):837-845. [CrossRef]
5. Nkie VE, Martin S. Fibrous dysplasia of the clivus: case report and literature review. *Cureus*. 2023;15(9):45417. [CrossRef]
6. Choi JH, Ro JY. The 2020 WHO classification of tumors of bone: an updated review. *Adv Anat Pathol*. 2021;28(3):119-138. [CrossRef]
7. Stevens AR, Branstetter BF, Gardner P, Pearce TM, Zenonos GA, Arani K. Ecchordosis physaliphora: does it even exist? *AJNR Am J Neuroradiol*. 2023;44(8):889-893. [CrossRef]
8. Özgür A, Esen K, Kara E, et al. Ecchordosis physaliphora: evaluation with precontrast and contrast-enhanced fast imaging employing steady-state acquisition MR imaging based on proposed new classification. *Clin Neuroradiol*. 2016;26(3):347-353. [CrossRef]
9. Lee SH, Kwok KY, Wong SM, Chan CXJ, Wong YT, Tsang ML. Chordoma at the skull base, spine, and sacrum: a pictorial essay. *J Clin Imaging Sci*. 2022;12:44. [CrossRef]
10. Das P, Soni P, Jones J, et al. Descriptive epidemiology of chordomas in the United States. *J Neurooncol*. 2020;148(1):173-178. [CrossRef]
11. Welzel T, Meyerhof E, Uhl M, et al. Diagnostic accuracy of DW MR imaging in the differentiation of chordomas and chondrosarcomas of the skull base: a 3.0-T MRI study of 105 cases. *Eur J Radiol*. 2018;105:119-124. [CrossRef]
12. Kremenevski N, Schlaffer SM, Coras R, Kinfe TM, Graillon T, Buchfelder M. Skull base chordomas and chondrosarcomas. *Neuroendocrinology*. 2020;110(9-10):836-847. [CrossRef]
13. Ozgen B, Oguz KK, Cila A. Diffusion MR imaging features of skull base osteomyelitis compared with skull base malignancy. *Am J Neuroradiol*. 2011;32(1):179-184. [CrossRef]
14. Jain N, Jasper A, Vanjare HA, Mannam P, Mani SE. The role of imaging in skull base osteomyelitis - reviewed. *Clin Imaging*. 2020;67:62-67. [CrossRef]
15. Saxena A, Subramanyam P, Sarma M, Bhad B, Bhaskaran R, Palaniswamy SS. 18F-FDG PET imaging for treatment response assessment and management guidance in patients with skull base osteomyelitis. *Nucl Med Commun*. 2024;45(7):589-600. [CrossRef]
16. Amin MB, Greene FL, Edge SB, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. 2017;67(2):93-99. [CrossRef]
17. Tsukamoto T, Miki Y. Imaging of pituitary tumors: an update with the 5th WHO Classifications-part 1. Pituitary neuroendocrine tumor (PitNET)/pituitary adenoma. *Jpn J Radiol*. 2023;41(8):789-806. [CrossRef]
18. Tajudeen BA, Kuan EC, Adappa ND, et al. Ectopic pituitary adenomas presenting as sphenoid or clival lesions: case series and management recommendations. *J Neurol Surg B Skull Base*. 2017;78(2):120-124. [CrossRef]
19. Chen X, Dai J, Ai L, et al. Clival invasion on multidetector CT in 390 pituitary macroadenomas: Correlation with sex, subtype and rates of operative complication and recurrence. *AJNR Am J Neuroradiol*. 2011;32(4):785-789. [CrossRef]
20. Butscheidt S, Ernst M, Rolvien T, et al. Primary intraosseous meningioma: clinical, histological, and differential diagnostic aspects. *J Neurosurg*. 2020;133(2):281-290. [CrossRef]

DIR

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.242790

PEDIATRIC RADIOLOGY

ORIGINAL ARTICLE

A comparison of two artificial intelligence-based methods for assessing bone age in Turkish children: BoneXpert and VUNO Med-Bone Age

Evrim Özmen[1]
Hande Özen Atalay[1]
Evren Uzer[1]
Mert Veznikli[2]

[1]Koç University Hospital, Department of Radiology, İstanbul, Türkiye

[2]Koç University Faculty of Medicine, Department of Computational Biology and Biostatistics, İstanbul, Türkiye

PURPOSE

This study aimed to evaluate the validity of two artificial intelligence (AI)-based bone age assessment programs, BoneXpert and VUNO Med-Bone Age (VUNO), compared with manual assessments using the Greulich–Pyle method in Turkish children.

METHODS

This study included a cohort of 292 pediatric cases, ranging in age from 1 to 15 years with an equal gender and number distribution in each age group. Two radiologists, who were unaware of the bone age determined by AI, independently evaluated the bone age. The statistical study involved using the intraclass correlation coefficient (ICC) to measure the level of agreement between the manual and AI-based assessments.

RESULTS

The ICC coefficients for the agreement between the manual measurements of two radiologists indicate almost perfect agreement. When all cases, regardless of gender and age group, were analyzed, an almost perfect positive agreement was observed between the manual and software measurements. When bone age calculations were analyzed separately for boys and girls, no statistically significant differences were found between the two AI-based methods in any subgroup. For boys regardless of age, the ICCs were 0.995 for VUNO and 0.994 for BoneXpert ($z = 1.597$, $P = 0.110$), while for girls, the ICCs were 0.994 and 0.995, respectively ($z = -1.303$, $P = 0.193$). The overall agreement with manual measurements was high for both VUNO and BoneXpert. In both boys and girls, the agreement remained consistent across different age groups. These findings indicate that both AI-based bone age assessment tools have a high degree of agreement with manual measurements across all age and gender groups, with no significant superiority of one method over the other.

CONCLUSION

Both BoneXpert and VUNO demonstrated high validity in assessing bone age, with no statistically significant differences between the two methods across gender or pubertal status groups. Notably, this study represents the first evaluation of both BoneXpert and VUNO for bone age assessment in Turkish children, highlighting their potential as reliable and clinically relevant tools for this population.

CLINICAL SIGNIFICANCE

Investigating the most suitable AI program for the Turkish population could be clinically significant.

KEYWORDS

Bone age, BoneXpert, VUNO, artificial intelligence, deep learning

Corresponding author: Evrim Özmen

E-mail: evrimkilicdr@gmail.com

Received 06 April 2024; revision requested 02 May 2024; last revision received 17 July 2024; accepted 28 July 2024.

Epub: 02.09.2024

Publication date: 06.11.2025

DOI: 10.4274/dir.2024.242790

Bone age is a marker of skeletal maturation and is measured routinely by pediatricians, radiologists, and pediatric endocrinologists for the assessment of the maturation progress of children.[1] The most commonly used manual method for bone age measurement is the Greulich–Pyle (GP) method.[2] According to this method, the determination of bone age is based on the similarity between the image in the GP atlas and the patient's left-hand wrist radiography. Thus, the GP method is very subjective and has higher inter and intraobserver

You may cite this article as: Özmen E, Özen Atalay H, Uzer E, Veznikli M. A comparison of two artificial intelligence-based methods for assessing bone age in Turkish children: BoneXpert and VUNO Med-Bone Age. *Diagn Interv Radiol.* 2025;31(6):630-635.

variability in addition to inter and intrainstitutional variability.[3] Besides, there is no standardized protocol for assessing bones, and it is unclear which bones should be included in the assessment.[4] With the development of deep learning, which is a subclass of artificial intelligence (AI) that exploits artificial neural networks, several software programs have been developed to automate and standardize bone age assessment, thereby reducing interobserver variability. It has been reported previously that AI-based assessment methods have high accuracy, reproducibility, and time efficiency when compared with manual methods.[4] Although BoneXpert version 2.4.5.1 and 3.0.3 (Visiana, Denmark) is one of the most frequently used methods of these, there are other AI-based bone age calculation software packages, including VUNO Med-Bone Age version 1.0.3 (VUNO) (VUNO, Seoul, Korea). The Turkish population is composed of various ethnic groups. As far as we know, no data compares these software packages, and no published report compares the manual method with these AI-based bone age assessment methods in Turkish children. This study aims to analyze the accuracy of two AI-based bone age assessment programs, namely BoneXpert and VUNO, in comparison with manual assessments using the GP bone age atlas.

Methods

Study design and population

This retrospective cohort study was approved by the Ethics Committee of Koç University Faculty of Medicine (2024.050.IRB2.023) and conducted in accordance with the Declaration of Helsinki's ethical principles. Informed consent was not obtained from the participants due to retrospective design of the study.

Pediatric cases who underwent left-hand X-ray imaging between January 2016 and December 2023 in the hospital due to suspicion of an endocrinological pathology and whose left-hand X-ray evaluation revealed that their chronological age and bone age were compatible were determined. Patients whose bone age was compatible with chronological age but who had known endocrinologic genetic or orthopedic disorders were excluded from the study list. Cases were also excluded if the radiological images were of poor quality, as this could make bone age estimation difficult.

After that, these cases were anonymized and grouped according to their age and gender, and the groups were randomized within themselves. Due to the limited number of male and female cases in the 1-year age group (aged 1–2 years), 6 cases for each gender were selected from this group. In the evaluation made for the other age groups, it was determined that the group of 15-year-old girls had the fewest case numbers, and there were 10 cases in this group. For this reason, in the other groups, the first 10 cases from the randomized list for both genders were selected. The specific age distribution included 6 boys and 6 girls aged 1–2 years, and 10 boys and 10 girls were included for each subsequent age group (aged 2–16 years).

Main points

- Our study reveals that both VUNO Med-Bone Age (VUNO) and BoneXpert correlated well with the manual assessment and Greulich–Pyle atlas.
- Neither VUNO nor BoneXpert showed a statistically significant difference in performance across gender or pubertal status groups, indicating similar effectiveness for bone age assessment in Turkish children.
- The results of our study are particularly important as they represent the first evaluation of both VUNO and BoneXpert in the Turkish pediatric population, addressing the gap in research on the applicability of AI-based bone age calculations for this demographic.

Radiological assessment

Left-hand wrist posterior to anterior X-ray images were used for the evaluation of bone age. Two radiologists with 15 and 5 years of experience and unaware of the results determined by AI independently evaluated bone ages according to the GP bone age atlas. Bone age was determined to be the midpoint when a case exhibited some, but not all, of the typical bone characteristics of a particular age (e.g., aged 8 years) and had all the characteristics of the previous age (e.g., aged 7 years). This approach was adopted to provide a more detailed and precise assessment of bone maturity. A third radiologist, aware of the cases' clinical details but blind to the manual bone age assessments, documented the AI assessments using BoneXpert version 3.0.3 and VUNO version 1.0.3 (Figure 1).

Statistical analysis

Correlation analysis was performed using the Statistical Package for the Social Sciences, version 28.0 (IBM SPSS Statistics, Armonk, NY, USA).[5] Comparing correlation coefficients was done by the MedCalc Statistical Software version 12.7.7 (MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2013). The test used by MedCalc is a *z*-test on Fisher's *z*-transformed correlation coefficients.[6] The inter-reader agreement between the manual evaluations of two radiologists was measured to ensure consistency in the

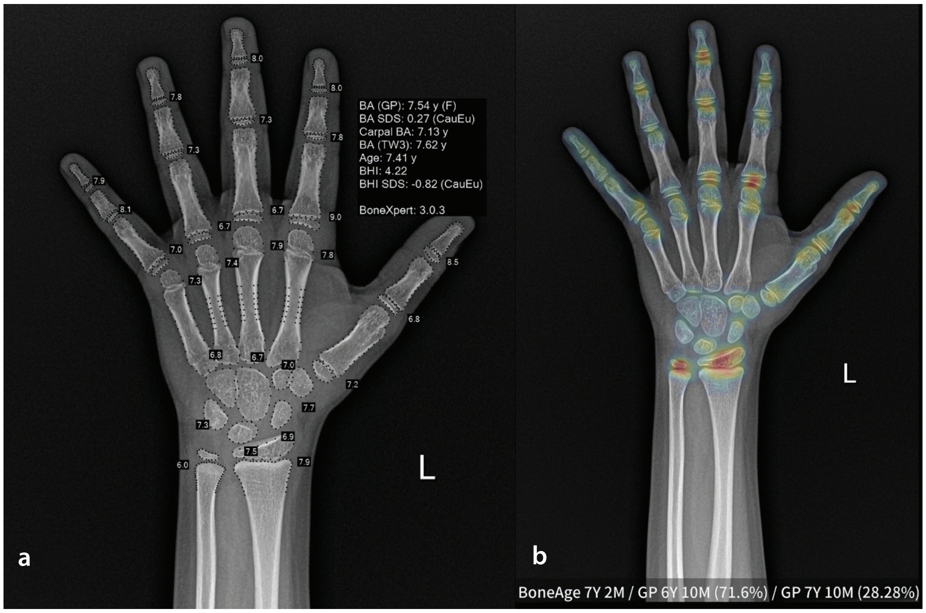


Figure 1. BoneXpert and VUNO images of a patient. In the 7.4-year-old female case, BoneXpert showed the bone age as 7.54 years **(a)** and VUNO indicated the bone age as 7 years 2 months **(b)**. As illustrated in the accompanying image, the BoneXpert version 3.0.3 system additionally evaluates carpal bones and provides separate carpal bone age, whereas the VUNO system assesses carpal bones as well as tubular bones and provides one bone age accordingly. Upon receipt of the image to be analysed, VUNO Med-Bone Age suggests three probable estimated bone ages (i.e., first- and second-rank artificial intelligence bone ages) based on probabilities, accompanied by similar images for comparison, while BoneXpert provides standard deviation information.

manual evaluation process. Intraclass correlation coefficients (ICC) were calculated for agreement between two radiologists using a two-way random-effects model, assessing absolute agreement. According to Shrout and Fleiss[7] (1979), this corresponds to ICC (2,1) for single measures and ICC (2,2) for average measures. Since the agreement was very high, manual evaluation was calculated with the arithmetic mean of these two measurements. The ICC values were used for assessing the agreement between software measurements and the mean radiologist measurements using a two-way random-effects model. According to Shrout and Fleiss[7] (1979), this corresponds to ICC (2,1) for single measures. To test the difference between two dependent correlations, the online tool "calculation for testing the difference between two dependent correlations" by Lee and Preacher (2013; https://quantpsy.org/corrtest/corrtest2.htm) was used. Bland–Altman analysis was used to further evaluate the agreement between manual and AI-based assessments. To also see the effect of gender and age on the measurements, all analyses were repeated for all combinations of subgroups: girls, boys, and different age groups. Boys over the age of 9 years and girls over the age of 8 years were considered to be pubescent.[8] The statistical significance level was accepted as 0.05.

Results

All pediatric patients aged 1–15 years with left-hand X-ray images generated in our institution were included in the study. Thirty-six patients with poor-quality radiological images and 54 patients with known endocrinologic genetic or orthopedic disorders were excluded from the study. The final study cohort included 292 cases with an equal distribution of genders across all age groups, ranging from 1 to 15 years (Figure 2). The ICC coefficients for the agreement between the manual measurements of two radiologists were calculated as 0.990 for ICC (2,1) and 0.995 for ICC (2,2) (Table 1). These values indicate almost perfect agreement. Based on these measurements, the average of the two observer values was taken and accepted as the manual measurement.

For the manual vs. software comparison, the ICC (2,1) values were calculated for single measurements. When all cases, regardless of gender and age group, were analyzed, an almost perfect agreement was observed between the manual and software measure-

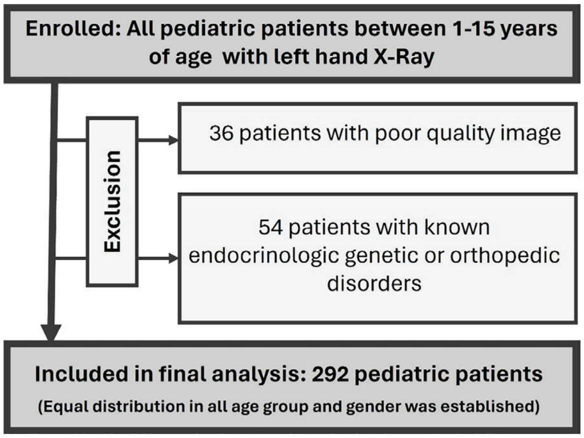


Figure 2. Flowchart of the study.

Table 1. Correlations for agreement between two radiologists

	Intraclass correlation[b]
Single measures ICC (2,1)	0.990[a]
Average measures ICC (2,2)	0.995

Two-way random effects model where both people effects and measures effects are random. [a], the estimator is the same, whether the interaction effect is present or not; [b], type A intraclass correlation coefficients using an absolute agreement definition; ICC, intraclass correlation coefficient.

Table 2. Intraclass Correlations for Agreement Between Software and Manual Measurements

	ICC (2,1)[a]			Difference between 2 correlation coefficients[b]	
Group		Vuno	BoneXpert	z	2-tailed *P*
Overall	**ICC**	0.995	0.995	0.000	1.000
	95% CI	0.993–0.996	0.991–0.997		
Males	**ICC**	0.995	0.994	1.597	0.110
	95% CI	0.989–0.997	0.970–0.998		
Females	**ICC**	0.994	0.995	−1.303	0.193
	95% CI	0.992–0.996	0.993–0.997		
≤9 years males & ≤8 years females	**ICC**	0.990	0.988	1.294	0.196
	95% CI	0.985–0.992	0.975–0.993		
>9 years males & >8 years females	**ICC**	0.977	0.978	−0.382	0.703
	95% CI	0.968–0.984	0.964–0.986		
≤9 years males	**ICC**	0.990	0.986	1.86	0.063
	95% CI	0.983–0.994	0.903–0.995		
>9 years males	**ICC**	0.977	0.976	0.262	0.793
	95% CI	0.902–0.991	0.906–0.99		
≤8 years females	**ICC**	0.988	0.992	−1.748	0.080
	95% CI	0.980–0.993	0.987–0.995		
>8 years females	**ICC**	0.977	0.980	−0.800	0.423
	95% CI	0.965–0.985	0.969–0.987		

[a]ICC estimates and their 95% confident intervals based on a single measures, absolute-agreement, 2-way random effects model. [b]Test of the difference between two dependent correlations with one variable in common. ICC, intraclass correlation coefficient; CI, confidence interval.

ments. When all cases, regardless of gender and age groups, were analyzed, an almost perfect positive agreement was observed between the manual and software measurements. The ICC was calculated as 0.995 for both VUNO and BoneXpert. No statistical difference was found between two AI-based methods.

When bone age calculations were analyzed separately for girls and boys, an ICC coefficient of 0.995 and 0.994 was calculated for VUNO and BoneXpert, respectively, for boys, and this difference was not significant (z = 1.597, *P* = 0.110). For girls, ICC coefficients of 0.994 and 0.995 were calculated for VUNO and BoneXpert, respectively, and this difference was not significant (z = -1.303, *P* = 0.193).

Upon categorization of all cases by age, a slight decrease in the software–manual agreement was observed for measurements of the older group. While the ICC coefficient was 0.990 for VUNO, it was calculated as 0.988 for BoneXpert in the younger age group (≤9 years for boys, ≤8 years for girls). Accordingly, it was evaluated that, in the measurements of prepubescent children, no significant difference was detected between two AI-based tools (z = 1,294, *P* = 0,196). After the age of 8 years for girls and 9 years for boys, the compliance of both software and manual measurements was calculated as 0.977 for VUNO and 0.978 for BoneXpert, and no significant difference was detected between the software (z = -0,382, *P* = 0,703) (Table 2).

Although there was no statistical significance between VUNO and BoneXpert, the difference between the agreements demonstrated by the two software packages with manual measurements in the prepubescent group was much more pronounced than older age group. The ICC values in prepubescent girls were calculated as 0.988 for VUNO and 0.992 for BoneXpert, and the difference was not significant (z =-1,748, *P* =0,080). In prepubescent boys, the ICC value was 0.990 for VUNO and 0.986 for BoneXpert; the difference was not statistically significant (z = 1.86, *P* = 0.063).

For girls aged >8 years and boys aged >9 years, the agreement between manual measurements and both AI software packages was similar. While the ICC values were 0.977 for VUNO, 0.976 for BoneXpert in boys aged >9 years, these values were 0.977 for VUNO and 0.980 for BoneXpert in girls aged >8 years (Table 2).

When examining Bland–Altman plot graphs, higher variability is observed on the left side of the graphs. Therefore, it can be seen that both AI-based bone age calculations tend to diverge more from manual measurements in the older group.

Discussion

This study represents the inaugural investigation into the comparative efficacy of AI-based systems, namely BoneXpert and VUNO, in the determination of bone age among a Turkish pediatric population. The results of our study indicate that both AI-based systems demonstrated a high level of agreement with each other and with manual methods in all our subgroups, including both genders and age groups. This is consistent with the findings of previous studies in the field. This highlights the potential for integrating AI-based bone age calculation into clinical practice, with the aim of enhancing the effectiveness of bone age assessment.

The GP method is the most widely used and well-known manual method, and according to Martin et al.[9], it is the method preferred by 76% of pediatric endocrinologists and radiologists.[10] The GP method is based

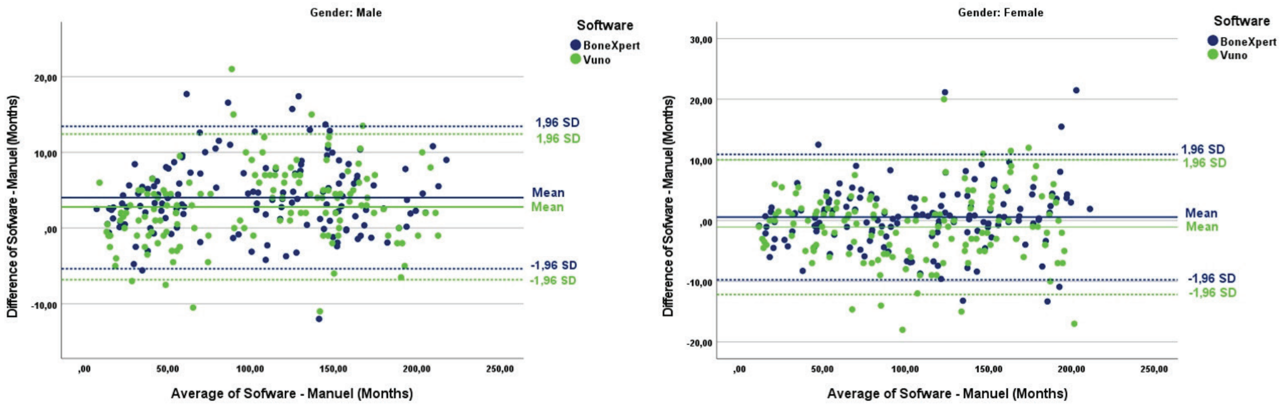


Figure 3. Bland–Altman plot. Difference between radiological bone age and automated bone age in boys and girls.

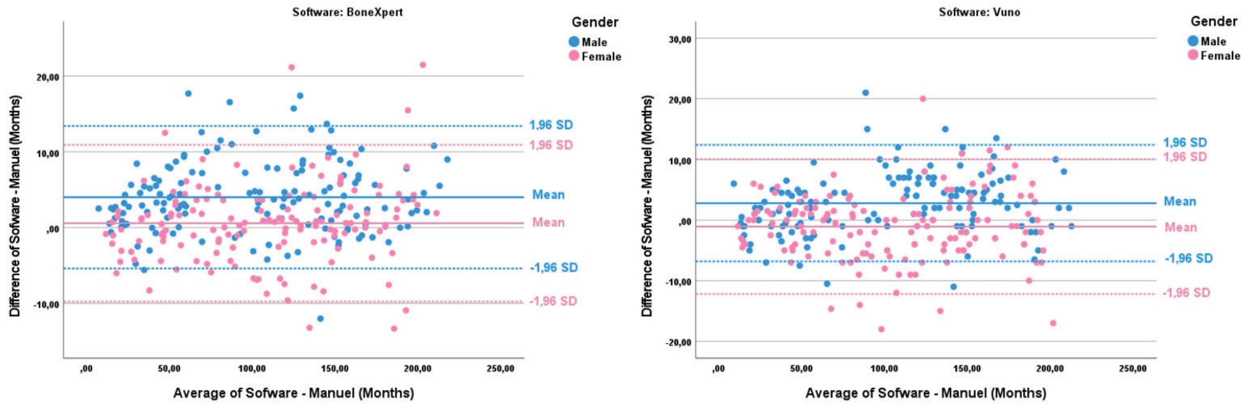


Figure 4. Bland–Altman plot. Difference between radiological bone age and automated bone age in BoneXpert and VUNO.

on the comparison of the cases' hand and wrist X-rays, with a standardized radiographic atlas compiled and standardized according to age and gender from birth to 18 years of age for girls and 19 years of age for boys.[10] However, bone age is influenced by ethnicity, gender, genetic factors, socioeconomic level, nutritional metabolic status, and bone disorders.[9-12] The standardized radiographic images of the atlas were derived from healthy North American and Western European-originated children.[13] They had good reliability in Australian and Middle Eastern ethnicity but were less reliable in Asian people. In addition to this, the evaluation of bone age with the GP method is also time-consuming; it takes a lot of time to evaluate the age of the bones individually with high accuracy when performed manually.[14] Furthermore, one of the major disadvantages of manual bone age assessment with the GP method is the possible risk of high inter and intraobserver error.[15] Therefore, before the comparison of manual bone age assessment with an AI-based system, the interobserver agreement between manual assessments performed by two radiologists was calculated and yielded an ICC of 0.990, thus establishing a solid basis for comparison of the AI-based measurements.

AI-based bone age calculation systems, developed to overcome all these disadvantages of manual calculation, can identify the morphological features of bone ossification automatically and provide rapid information about the patient's bone age. Therefore, this has resulted in a more objective and efficient method for assessing bone age.[16]

Numerous studies have demonstrated that newly developed AI technologies and software can accurately perform bone age assessments, surpassing the accuracy of the GP method.[1,4,9,15] Furthermore, these studies have shown that AI-based assessments exhibit excellent agreement with assessments made by experienced human observers.[1] In their study to compare deep learning systems, including AlexNet, GoogleNet and Vogg19, in performing age estimation with the Turkish population, Senel et al.[17] reported a success rate of 98.39%.

Similarly, we found a high level of agreement between manual assessments (using GP) and both AI-based systems, with an ICC of 0.995 for both VUNO and BoneXpert when the entire cohort was considered. This high correlation is particularly important given the lack of existing research on the applicability of AI-based bone age calculations in the Turkish pediatric population.

BoneXpert is an AI-based automated bone age assessment system and is known as the first AI radiology system.[13] This method, which is based on traditional machine learning methodology, predicts bone age by considering bone shape, density, and the degree of epiphyseal fusion.[18,19] Image analysis predicts bone age by measuring shape, density, and texture scores at specific locations.[14] If a bone's appearance falls outside the range covered by the machine learning process or if its bone age value deviates above the threshold value compared with the average of all tubular bones, it will not be included in the calculation. The final bone age is calculated using the evaluated bones. If fewer than eight bones are evaluated, the X-ray is not assessed due to possible inaccurate calculations, which is a major disadvantage of BoneXpert version 2.4.5.1[20] However, BoneXpert version 3.0 introduced several significant advancements over its predecessor. These features improves accuracy. Additionally, version 3.0 also provides carpal bone age determination which is typically determined for boys up to 11.5 years and girls up to 9.5 years for additional information about skeletal maturity in younger children. In addition to that, new version reduces image rejection rates by improving adaptability to variations in image post-processing and achieving more precise bone localization. Both versions of BoneXpert have been validated for bone age calculation in North American, Caucasian, African American, Hispanic, and Asian children and has also been reported to be applicable in various ethnic groups.[19,21,22] Many published reports show a notable distinction between bone ages determined by the GP method and chronological ages in Asian children.[23,24] Similarly, Ontell et al.[25] reported delayed bone age in preadolescence and increased bone age in adolescence in Asian boys. The process of skeletal maturation in Korean children is initiated at a later age and completed at an earlier age than in Caucasian children. The VUNO Korean bone age assessment method, which is based on deep learning, has demonstrated superior performance compared with the manual assessment from the GP atlas. Compared with the manual assessment with the GP atlas, the Korean model has a lower root mean square error and lower mean absolute error. VUNO is the first AI-based bone age assessment system approved by the Korean Food and Drug Administration. The system was developed by analyzing 18,940 left-hand wrist radiographs using the GP method.[25,26] VUNO provides the most likely estimated bone ages based on the examined wrist radiography.

A subgroup analysis of the data revealed subtle differences between the calculated bone ages by BoneXpert and VUNO, particularly when examining data based on gender and age subgroups. Both VUNO and BoneXpert demonstrated a high level of agreement with manual assessments in boys and girls, with no statistically significant differences observed between the two methods across any subgroup. This suggests that both tools are equally effective in bone age assessment regardless of gender or pubertal status. The analysis provided valuable insights into the applicability of AI-based bone age programs, showing that BoneXpert and VUNO maintain high reliability across different age and gender groups, even among prepubertal individuals in contrast to previous version of BoneXpert. In a comprehensive validation study comparing previous and latest versions of BoneXpert revealed that previous version had a tendency to underestimate bone age in girls aged 6–7 years and 12–15 years, but the latest version showed significant improvements in this regard, highlighting the importance of usage most updated version of bone age softwares.[27]

Our study had some limitations, including a small sample size and the fact that it focused on a single, heterogeneous ethnicity. Additionally, the study did not include participants aged <2 years or >15 years due to the unsuitability of the GP manual method for evaluating bone age in these age groups.

In conclusion, our study confirms that BoneXpert and VUNO are reliable AI-based systems for assessing bone age in the Turkish pediatric population. Both methods demonstrated comparable agreement with manual assessments across all gender and pubertal status groups, marking this study as a significant contribution to evaluating AI-based bone age assessment tools in this demographic.

Conflict of interest disclosure

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

References

1. Martin DD, Calder AD, Ranke MB, Binder G, Thodberg HH. Accuracy and self-validation of automated bone age determination. *Sci Rep*. 2022;12(1):6388. [CrossRef]
2. Greulich WW, Pyle IS. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford: Stanford University Press; 1959. [CrossRef]
3. Prokop-Piotrkowska M, Marszałek-Dziuba K, Moszczyńska E, Szalecki M, Jurkiewicz E. Traditional and new methods of bone age assessment-an overview. *J Clin Res Pediatr Endocrinol*. 2021;13(3):251-262. [CrossRef]
4. Kim PH, Yoon HM, Kim JR, et al. Bone age assessment using artificial intelligence in Korean pediatric population: a comparison of deep-learning models trained with healthy chronological and greulich-Pyle ages as labels. *Korean J Radiol*. 2023;24(11):1151-1163. [CrossRef]
5. IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp. [CrossRef]
6. Kinckle DE. Applied Statistics for the behavioral sciences. 2nd ed. Houghton Mifflin Harcourt; 1988. [CrossRef]
7. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Pshychol Bull*. 1979;82(2):420-428. [CrossRef]
8. National Pediatric Society of Turkey. Promed-Mail. Accessed July 21, 2024. [CrossRef]
9. Martin DD, Wit JM, Hochberg Z, et al. The use of bone age in clinical practice - part 1. *Horm Res Paediatr*. 2011;76(1):1-9. [CrossRef]
10. Artioli TO, Alvares MA, Carvalho Macedo VS, et al. Bone age determination in eutrophic, overweight and obese Brazilian children and adolescents: a comparison between computerized BoneXpert and Greulich-Pyle methods. *Pediatr Radiol*. 2019;49(9):1185-1191. [CrossRef]
11. Kaplowitz P, Srinivasan S, He J, McCarter R, Hayeri MR, Sze R. Comparison of bone age readings by pediatric endocrinologists and pediatric radiologists using two bone age atlases. *Pediatr Radiol*. 2011;41(6):690-693. [CrossRef]
12. Halabi SS, Prevedello LM, Kalpathy-Cramer J, et al. The RSNA pediatric bone age machine learning challenge. *Radiology*. 2019;290(2):498-503. [CrossRef]
13. Alshamrani K, Hewitt A, Offiah AC. Applicability of two bone age assessment methods to children from Saudi Arabia. *Clin Radiol*. 2020;75(2):156. [CrossRef]
14. Thodberg HH, Thodberg B, Ahlkvist J, Offiah AC. Autonomous artificial intelligence in pediatric radiology: the use and perception of BoneXpert for bone age assessment. *Pediatr Radiol*. 2022;52(7):1338-1346. [CrossRef]
15. Gräfe D, Beeskow AB, Pfäffle R, Rosolowski M, Chung TS, DiFranco MD. Automated bone age assessment in a German pediatric cohort: agreement between an artificial intelligence software and the manual Greulich and Pyle method. *Eur Radiol*. 2024;34(7):24407-24413. [CrossRef]
16. Dallora AL, Anderberg P, Kvist O, et al. Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS One*. 2019;14:e0220242. [CrossRef]
17. Senel FA, Dursun A, Ozturk K, Ayyildiz VA. Determination of bone age using deep convolutional neural networks. *Ann Med Res*. 2021;28(7):1381-1386. [CrossRef]
18. Acheson RM, Vicinus JH, Fowler GB. Studies in the reliability of assessing skeletal maturity from x-rays. 3. Greulich-Pyle Atlas and Tanner-Whitehouse method contrasted. *Hum Biol*. 1966;38(3):204-218. [CrossRef]
19. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging*. 2009;28(1):52-66. [CrossRef]
20. Booz C, Yel I, Wichmann JL, et al. Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. *Eur Radiol Exp*. 2020;4(1):6. [CrossRef]
21. Thodberg HH, Savendahl L. Validation and reference values of automated bone age determination for four ethnicities. *Acad Radiol*. 2010;17(11):1425-1432. [CrossRef]
22. Satoh M. Bone age: assessment methods and clinical applications. *Clin Pediatr Endocrinol*. 2015;24(4):143-152. [CrossRef]
23. Alshamrani K, Messina F, Offiah AC. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. *Eur Radiol*. 2019;29(6):2910-2923. [CrossRef]
24. Zhang A, Sayre JW, Vachon L, Liu BJ, Huang HK. Racial differences in growth patterns of children assessed on the basis of bone age. *Radiology*. 2009;250(1):228-235. [CrossRef]
25. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *AJR Am J Roentgenol*. 1996;167(6):1395-1398. [CrossRef]
26. Kim JR, Shim WH, Yoon HM, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. *AJR Am J Roentgenol*. 2017;209(6):1374-1380. [CrossRef]
27. Maratova K, Zemkova D, Sedlak P, et al. A comprehensive validation study of the latest version of BoneXpert on a large cohort of Caucasian children and adolescents. Front Endocrinol. 2023;14:1130580. [CrossRef]

Diagn Interv Radiol 2024 DOI: 10.4274/dir.2024.232551

RADIOLOGY PHYSICS

ORIGINAL ARTICLE

Institutional clinical indication-based typical dose values of multiphasic abdominopelvic computed tomography examinations

Süleyman Filiz[1,2]
Safiye Gürel[1,3]
Kamil Gürel[1,3]

[1]Bolu Abant Izzet Baysal University Faculty of Medicine, Department of Radiology, Bolu, Turkey

[2]Ankara Gazi Mustafa Kemal Occupational and Environmental Diseases Hospital, Department of Radiology, Ankara, Turkey

[3]University of Health Sciences, Gülhane Training and Research Hospital, Department of Radiology, Ankara, Turkey

PURPOSE

Our study aimed to obtain clinical indication-based typical dose values and size-specific dose estimates (SSDEs) for multiphasic abdominopelvic computed tomography (CT) examinations and to review our data with published diagnostic reference levels (DRLs).

METHODS

In this retrospective study, multiphasic liver, kidney, pancreas, and mesenteric ischemia protocol CT scans performed at our center between January 2018 and December 2021 were analyzed. The clinical indications were hepatocellular carcinoma, renal cell carcinoma, pancreas adenocarcinoma, and mesenteric ischemia. The computed tomography dose index volume ($\mathrm{CTDI}_{\mathrm{vol}}$) and dose-length product (DLP) values were recorded, and the SSDE and effective dose (ED) values were calculated. The water-equivalent diameter (Dw) value required for the SSDE calculation was measured using the automated calculation of the Dw program.

RESULTS

The total number of patients was 514, with 86 patients excluded from this study. The dose values were calculated for 426 patients (183 female and 243 male; 111 liver, 120 kidney, 85 pancreas, and 110 mesenteric). The median values for the $\mathrm{CTDI}_{\mathrm{vol}}$, DLP, SSDE, and ED were 6.86 mGy, 683.02 mGy.cm, 8.75 mGy, and 10.45 mSv for the liver CT; 8.37 mGy, 908.37 mGy.cm, 10.37 mGy, and 13.89 mSv for the kidney CT; 7.82 mGy, 517.98 mGy.cm, 10.01 mGy, and 7.92 mSv for the pancreas CT; and 9.48 mGy, 983.68 mGy.cm, 12.78 mGy, and 13.86 mSv for the mesenteric CT, respectively. All dose values were lower than the published DRLs.

CONCLUSION

The literature reveals large differences in the multiphasic abdominopelvic CT protocols, especially in the number of phases and scan length. This situation makes comparing dose values difficult. Dose studies revealing the protocol parameters in detail are needed so that institutions can compare and optimize their own protocols. Additionally, users should periodically check the dose values in their own institutions.

KEYWORDS

Clinical indication, computed tomography, diagnostic reference levels, multi-phase scan, size-specific dose estimate, water-equivalent diameter

Corresponding author: Süleyman Filiz

E-mail: suleymanfilizz@gmail.com

Received 12 October 2023; revision requested 14 November 2023; last revision received 17 January 2024; accepted 20 January 2024.

Epub: 20.02.2024

Publication date: 06.11.2025

DOI: 10.4274/dir.2024.232551

The frequency of computed tomography (CT) use and its contributions to diagnostic radiology have increased since the early 1970s. CT now constitutes a large part of the artificial radiation originating from medicine due to its increased prevalence and frequency of use.[1] This situation increases the cancer risk, and the optimization principle in radiation safety has become much more important. The diagnostic reference level (DRL) is used for diagnostic and interventional procedures to help optimize a patient's exposure to ionizing radiation. It is produced from radiation data collected locally, nationally, or regionally.[2] The use of CT scans should be reassessed and optimized when the patient's doses exceed the available DRLs.

You may cite this article as: Filiz S, Gürel S, Gürel K. Institutional clinical indication-based typical dose values of multiphasic abdominopelvic computed tomography examinations. *Diagn Interv Radiol.* 2025;31(6):636-640.

The computed tomography dose index volume (CTDI_{vol}) and dose-length product (DLP) are used to determine the DRL for CT examinations. These parameters are only an approximate estimate of the patient's dose. The CTDI_{vol} is a dose index specific to phantom sizes and does not consider the patient's size, thickness, and length of the scanned volume.

The size-specific dose estimate (SSDE) has been proposed by the American Association of Physicists in Medicine (AAPM) to give the CTDI_{vol} a more realistic dose value for the patient, considering the patient's size. In this method, the water-equivalent diameter (Dw) is calculated, and then the CTDI_{vol} value is multiplied by the corresponding conversion factor to the Dw in the AAPM Report 220.[3]

In 2015, Ataç et al.[4] reported the first Turkish national DRLs for single-phase head, chest, abdominal, and pelvic CT examinations of adults and children. In the following years, Atlı et al.[5] reported institutional typical dose values for single-phase head, neck, thorax, and abdomen CT examinations. In these dose studies in Turkey, data from single-phase CT examinations were collected, but there have been no national patient dose studies for multiphasic CTs so far. Recent DRL studies for multiphasic abdominopelvic CTs exist in other countries.[6-11] The dose values for liver CT were given in all of these studies, and the dose values for kidney and pancreas CTs were given in a few. However, there is no dose data for mesenteric ischemia protocol CTs. Additionally, the SSDE was not evaluated in any of these studies. Some studies did not include information such as the CTDI_{vol}, effective dose (ED), scan length, and phase number.

Most existing DRLs report dose values based on anatomical regions, such as head, chest, and abdomen CTs. However, the protocols to be selected in CT examinations are determined according to the clinical preliminary diagnosis or clinical indication. Different imaging protocols are used for varying clinical indications in the same anatomical region. For example, in our clinic, a non-contrast single-phase abdominopelvic CT protocol is used for a patient being investigated for kidney stones. However, if the patient is suspected of having renal cell carcinoma (RCC), the kidney is scanned four times (a precontrast phase followed by postcontrast corticomedullary, nephrogram, and urogram phases) for lesion characterization. This reveals that one of the most important things affecting dose values is clinical indication. The clinical indication-based approach to DRLs was mentioned by the International Commission on Radiological Protection in 2017.[2]

Main points

- The protocols used in multiphasic abdominopelvic computed tomography (CT) vary significantly between institutions. This makes it difficult to compare institutional dose values to diagnostic reference levels.
- The number of phases and scan length are the most important parameters that cause differences in multiphasic abdominopelvic CT protocols.
- Institutions must determine their own dose values and check them at regular intervals.

In our study, we aimed to evaluate the clinical indication-based typical dose values and SSDEs for multiphasic abdominopelvic CTs and review our data with published DRLs.

Methods

In this retrospective study, after obtaining approval from the Bolu Abant İzzet Baysal University Clinical Research Ethics Committee (decision number: 2022/81), multiphasic liver, kidney, pancreas, and mesenteric CT scans taken at the İzzet Baysal Training and Research Hospital between January 2018 and December 2021 were examined. Informed consent was waived by the ethics committee. The clinical indications were hepatocellular carcinoma (HCC), RCC, pancreatic adenocarcinoma, and mesenteric ischemia. The examinations were obtained with a 64-detector CT device (2017 GE Revolution EVO 128 slice, China). Table 1 summarizes the CT input parameters for each protocol. Automatic tube current modulation was used in all protocols.

The patient's age, gender, and indication for the CT examination were obtained from the hospital's information archive system. The CTDI_{vol} and DLP values were recorded from the picture archiving and communication system. The automated calculation of the Dw program was obtained from a free website (http://ctdose-iqurad.med.uoc.gr/) was used to calculate the Dw. For this, CT images of the patient were loaded into the program in the Digital Imaging and Communications in Medicine format, and then the program calculated the mean and median Dw values for each section (Figure 1). The Dw values were calculated from the median image, according to the AAPM Report 220, for each phase for each patient using this program. Afterward, the Dw value of that examination was calculated by taking the average of the Dw values obtained from each phase. For the SSDE calculation, the CTDI_{vol} values were multiplied by the Dw-appropriate conversion factors in the AAPM Report 220. While calculating the total DLP, the DLP values of all phases and the DLP value of bolus tracking were added. The scan lengths were calculated separately for each phase with the DLP/CTDI_{vol} ratio.

The ED was calculated by multiplying the DLP value with the conversion coefficients published in the International Commission on Radiological Protection (ICRP) 103. These coefficients were given as 0.0153 in an abdominal CT and 0.0141 in an abdominopelvic CT for a 120 kV tube current.[12] The average of the CTDI_{vol} and SSDE values of each phase and the sum of the DLP and ED values were taken.

Statistical analysis

The mean, standard deviation (SD), median, and the first, second, and third quartiles

Table 1. Input parameters for each CT protocol

CT protocol	Phase	Slice thickness (mm)	Tube current (min–max mAs)	Tube voltage (kV)	Gantry rotation time (sec)	Pitch
Liver	Late arterial	2.5	80–450	120	0.6	1.375
	Portal venous	2.5	80–450			
	Late	2.5	80–450			
Kidney	Non-contrast	5	100–350	120	0.6	1.375
	Corticomedullary	2.5	140–450			
	Nephrogram	2.5	140–450			
	Urogram	2.5	140–450			
Pancreas	Pancreatic	2.5	100–400	120	0.6	1.375
	Portal venous	2.5	100–400			
Mesenteric	Arterial	0.625	80–440	120	0.5	0.984
	Portal venous	0.625	80–440			

CT, computed tomography; kV, kilovoltage; mm, millimeter; mAs, milliampere-seconds; sec, second.

were calculated for the CTDI$_{vol}$, SSDE, DLP, and EDs using the Statistical Package for the Social Sciences (SPSS) for Windows (SPSS Inc., Chicago, Illinois, USA) version 26.0.

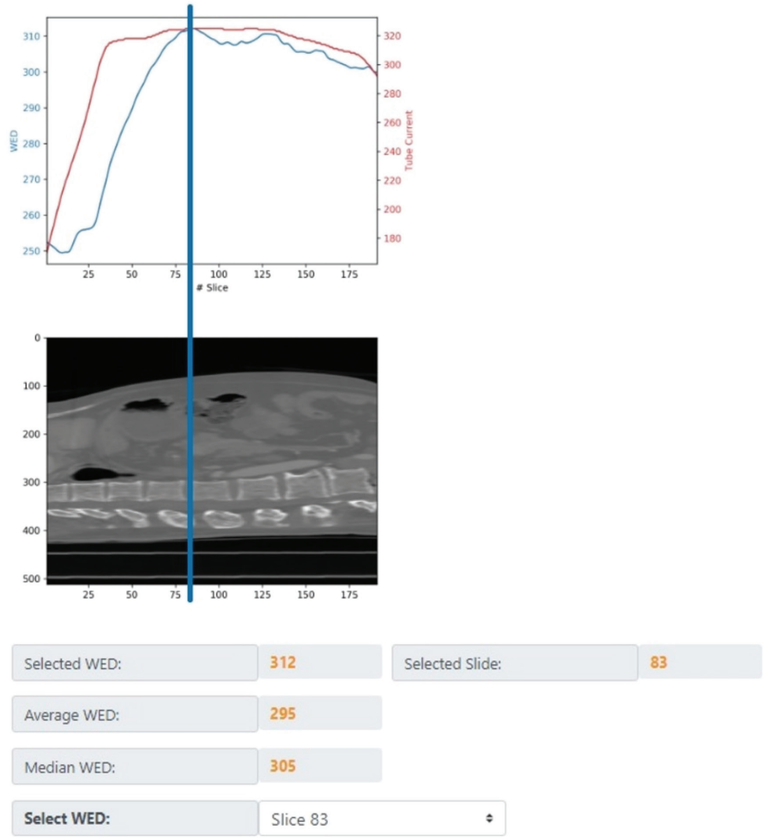
Results

The total number of patients was 514. Excluded from the study were 23 patients because their arms were in the imaging field, 16 patients who could not be positioned appropriately on the CT table, 6 patients who had metallic prostheses from lumbar stabilization surgery, and 1 patient who had a total hip prosthesis. Since the height and weight information of all the patients was not available, patients with the CTDI$_{vol}$ and DLP values between the minimum and maximum 5% for each protocol were not included in the calculation of the dose values, as recommended in the ICRP 135, to increase compliance with the standard patient definition.[2] As a result, 111 patients for liver CT [58 males (52%) and 53 females (48%)], 120 patients for kidney CT [81 males (67.5%) and 39 females (32.5%)], 85 patients for pancreas CT [46 males (54%) and 39 females (46%)], and 110 patients for mesenteric ischemia protocol CT [58 males (53%) and 52 females (47%)] were included in the study for the dose calculation. The mean ± SD age was 55.79 ± 14.76 years in liver CT patients, 62.35 ± 14.01 years in kidney CT patients, 62.74 ± 15.33 years in pancreatic CT patients, and 64.58 ± 14.14 years in mesenteric CT patients (Table 2).

The mean ± SD scan length was 32.3 ± 3.3 cm in liver CTs, 31.4 ± 4.8 cm in pancreas CTs, 51.8 ± 3.8 cm in mesenteric CTs, and equal for all phases in each protocol. The mean ± SD scan lengths in kidney CTs were 27.2 ± 4.4 cm in the non-contrast phase, 30.5 ± 2.8 cm in the corticomedullary and nephrogram phases, and 22.9 ± 3 cm in the urogram phase.

The mean ± SD and median values of the Dws were 29.63 ± 2.77 cm and 29 cm in the liver CT, 28.83 ± 2.44 cm and 29.51 cm in the kidney CT, 28.8 ± 2.25 cm and 28.85 cm in the pancreas CT, and 26.7 ± 2.32 cm and 26.75 cm in the mesenteric CT, respectively.

The median values for the CTDI$_{vol}$, DLP, SSDE, and ED were 6.86 mGy, 683.02 mGy.cm, 8.75 mGy, and 10.45 mSv for the liver CT; 8.37 mGy, 908.37 mGy.cm, 10.37 mGy, and 13.89 mSv for the kidney CT; 7.82 mGy, 517.98 mGy.cm, 10.01 mGy, and 7.92 mSv for the pancreas CT; 9.48 mGy, 983.68 mGy.cm, 12.78 mGy, and 13.86 mSv for the mesenteric CT, respectively. Tables 3, 4, and 5 detail the first, second, and third quartile values for the CTDI$_{vol}$, SSDE, DLP, and ED.



Selected WED: 312

Selected Slide: 83

Average WED: 295

Median WED: 305

Select WED: Slice 83

Figure 1. Water-equivalent diameter (Dw) measurement using the automated calculation of the Dw.

Table 2. Demographic data of the patients

	Patient numbers (percentage)		Age (mean years ± SD)
	Male	Female	
Liver CT	58 (52.3%)	53 (47.7%)	55.79 ± 14.76
Kidney CT	81 (67.5%)	39 (32.5%)	62.35 ± 14.01
Pancreas CT	46 (54.1%)	39 (45.9%)	62.74 ± 15.33
Mesenteric CT	58 (52.7%)	52 (47.3%)	64.58 ± 14.14

CT, computed tomography; SD, standard deviation.

Discussion

In our study, we found clinical indication-based typical dose values and SSDEs for multiphasic liver, kidney, pancreatic, and mesenteric CTs in 426 adult patients. Among the four indications we examined, the lowest ED value belonged to pancreatic adenocarcinoma. Our expectation was also in this direction because the scan length was shorter, and the number of phases was less compared with other protocols. We found that the clinical indication with the highest ED value was RCC. This was immediately followed by mesenteric ischemia. Although the highest DLP value was in mesenteric ischemia, the highest ED value was in RCC. This is because the conversion coefficient used in the ED calculation differs for these two indications (0.0153 in kidney CT and 0.0141 in mesenteric CT).[12]

Only a few dose studies have been conducted in Turkey.[4,5] In these studies, data for single-phase CT examinations were used. No studies in Turkey have been carried out with which we could compare our dose values for multiphasic abdominopelvic CTs.

Internationally, there are few DRL studies with which we could compare our data. van der Molen et al.[6] used data from 186 standard-sized patients for the DRLs of the 21 most frequently taken CT protocols in the Netherlands. The DRLs were only given for the DLP and ED, and the 75th percentile dose values of liver, kidney, and pancreatic CTs were higher than ours. The fact that the phase numbers and scan lengths of the CT protocols used in this study are higher than

Table 3. The first, second, and third quartile values for the CTDI_{vol} and SSDE (mGy)

Protocol	Phase	1st quartile	2nd quartile (median)	3rd quartile
Liver	Late arterial	4.82 (6.49)	7.00 (8.77)	11.09 (12.75)
	Portal venous	4.89 (6.52)	6.96 (8.80)	11.44 (13.22)
	Late	4.90 (6.44)	6.94 (8.74)	10.99 (13.04)
	Average	**4.88 (6.48)**	**6.86 (8.75)**	**11.19 (12.94)**
Kidney	Non-contrast	5.28 (7.25)	7.71 (9.48)	9.23 (11.12)
	Corticomedullary	6.25 (8.50)	8.33 (10.27)	10.25 (12.16)
	Nephrogram	6.25 (8.52)	8.47 (10.48)	10.16 (12.11)
	Urogram	6.24 (8.77)	8.84 (10.92)	10.67 (12.99)
	Average	**6.03 (8.25)**	**8.37 (10.37)**	**10.12 (12.14)**
Pancreas	Pancreatic	6.64 (9.16)	7.67 (10.01)	10.23 (12.09)
	Portal venous	6.77 (9.11)	7.71 (10.01)	10.00 (12.00)
	Average	**6.70 (9.17)**	**7.82 (10.01)**	**10.07 (12.09)**
Mesenteric	Arterial	6.56 (9.88)	9.49 (12.73)	11.26 (15.17)
	Portal venous	6.59 (9.86)	9.47 (12.75)	11.28 (15.09)
	Average	**6.57 (9.85)**	**9.48 (12.78)**	**11.27 (15.10)**

Values in the parentheses represent the SSDEs. CTDI_{vol}, computed tomography dose index volume; mGy, milligray; SSDE, size-specific dose estimate.

Table 4. The first, second, and third quartile values for the DLP (mGy.cm)

Protocol	1st quartile	2nd quartile (median)	3rd quartile
Liver	493.45	683.02	1074.31
Kidney	681.02	908.37	1163.16
Pancreas	418.11	517.98	701.26
Mesenteric	681.24	983.68	1212.62

DLP, dose length product; mGy, milligray.

Table 5. The first, second, and third quartile values for the ED (mSv)

Protocol	1st quartile	2nd quartile (median)	3rd quartile
Liver	7.54	10.45	16.43
Kidney	10.41	13.89	17.79
Pancreas	6.39	7.92	10.72
Mesenteric	9.60	13.86	17.09

ED, effective dose; mSv, millisievert.

ours may explain the higher dose values. In the national DRL study of Kim et al.[7] in South Korea, the data of 14,620 adult patients were used. In the study, the 75th percentile CTDI_{vol}, DLP, and ED values were higher than ours for liver, kidney, and pancreas CTs. In this study, these CTs were defined as "2–4 phase," and data such as phase and scan length of the protocols are unknown.[7] Tsapaki et al.[8] used data obtained from 14 European countries, and 10 clinical indications were given in the European Study on Clinical Diagnostic Reference Levels for X-ray Medical Imaging (EUCLID). The DRL values varied significantly between hospitals. This was mainly due to the technical protocol and variable phase number/scan lengths.[8] In the study of the DRL by Salama et al.[9] in Egypt, the CTDI_{vol} value was the highest in the literature. This may be due to the high body weights of the patients, the low pitch values, and the fact that the automatic tube current was not used in all patients.[9] In the DRL study by Aberle et al.[10] in Switzerland, the CT scans for HCC were taken in 2–4 phases, and our dose values were lower than in this study. In the study by Bos et al.[11], the DRLs were calculated for 10 clinical indications (EUCLID) from the CT scans of 3.7 million adult patients from seven countries. The dose values of CTs taken with the indication of HCC are higher than ours.

Our study has some limitations. The first is that the study was made from the data of a single CT device in a single center, and the number of patients was relatively small. Second, the height and weight information of all patients is not known. Third, multiphasic abdominopelvic CT protocols vary in different institutions due to specific parameters, such as phase numbers and the scan length on the z-axis. This situation causes difficulties in comparing the obtained data with the literature.

In our study, we reported the clinical indication-based typical dose values and SSDE's of multiphasic abdominopelvic CT protocols and compared our results with the published international data (Table 6). There are very few DRL studies of multiphasic abdominopelvic CTs in the literature, and none of these studies presented the SSDE data that would help us understand the impact of patient size on radiation dose.

In conclusion, additionally, until our study, no dose data for mesenteric ischemia protocol CTs were published. The DRL is a recommendation, and the purpose is to detect unusually high and low levels and to provide the necessary optimizations. Standard protocols are not used for multiphasic CTs, resulting in large differences in dose values between different devices, institutions, and countries. Studies that reveal the protocol parameters in detail are needed so that institutions can compare and optimize their protocols. Users should periodically evaluate dose values in their institutions to detect unforeseen deviations in doses in routine clinical practices and to take measures to correct them. The adequacy of the diagnostic image quality should be considered if the dose values are lower than the available DRLs. More studies are needed to evaluate clinical indication-based dose values in multiphasic abdominopelvic CTs. In our country, DRLs of single-phase CT examinations have been reported in pioneering studies, and similar DRL studies should be performed for multiphasic CTs.

Table 6. Comparison of the dose values (median) with the diagnostic reference values

Protocols	The authors' institution	van der Molen et al.[6]	Kim et al.[7]	Tsapaki et al.[8]	Salama et al.[9]	Aberle et al.[10]	Bos et al.[11]
Liver							
Number of phases	3	4	2–4*	4**	3	3.2***	-
Indication	HCC	Tx	-	HCC	Metastasis	HCC	HCC
CTDI$_{vol}$	6.86	-	14.70	9	31	11	14.60
DLP	683.02	1496.6	1693	1327	1425	1170	2032
ED	10.45	22.40	25.40	-	-	-	-
SSDE	8.75	-	-	-	-	-	-
SL (mean)	32.30	41.90	-	37	-	-	-
Kidney							
Number of phases	4	4	2–4*	-	-	-	-
Indication	RCC	RCC	-	-	-	-	-
CTDI$_{vol}$	8.37	-	14.20	-	-	-	-
DLP	908.37	1371.20	2100	-	-	-	-
ED	13.89	20.20	31.50	-	-	-	-
SSDE	10.37	-	-	-	-	-	-
SL (mean)	27.70	38.10	-	-	-	-	-
Pancreas							
Number of phases	2	3	2–4*	-	-	-	-
Indication	Adenoca	Adenoca	-	-	-	-	-
CTDI$_{vol}$	7.82	-	14	-	-	-	-
DLP	517.98	1000	1531	-	-	-	-
ED	7.92	14.70	23	-	-	-	-
SSDE	10.01	-	-	-	-	-	-
SL (mean)	31.4	40.90	-	-	-	-	-
Mesenteric							
Number of phases	2	-	-	-	-	-	-
Indication	Ischemia	-	-	-	-	-	-
CTDI$_{vol}$	9.48	-	-	-	-	-	-
DLP	983.68	-	-	-	-	-	-
ED	13.86	-	-	-	-	-	-
SSDE	12.78	-	-	-	-	-	-
SL (mean)	51.80	-	-	-	-	-	-

*Phase numbers not specified; **the number of phases was given as "4" most frequently; ***average number of phases. Adenoca, adenocarcinoma; CTDI$_{vol}$, computed tomography dose index volume; DLP, dose length product; ED, effective dose; HCC, hepatocellular carcinoma; RCC, renal cell carcinoma; SL, scan length (cm); SSDE, size-specific dose estimate; Tx, transplantation.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Hendee WR, O'Connor MK. Radiation risks of medical imaging: separating fact from fantasy. *Radiology*. 2012;264(2):312-321. [CrossRef]
2. Vañó E, Miller DL, Martin CJ, et al. ICRP publication 135: diagnostic reference levels in medical imaging. *Ann ICRP*. 2017;46(1):1-144. [CrossRef]
3. McCollough C, Bakalyar DM, Bostani M, et al. Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT: the report of AAPM task group 220. *AAPM Rep*. 2014;2014:6-23. [CrossRef]
4. Ataç GK, Parmaksız A, İnal T, Bulur E, et al. Patient doses from CT examinations in Turkey. *Diagn Interv Radiol*. 2015;21(5):428-434. [CrossRef]
5. Atlı E, Uyanık SA, Öğüşlü U, Çevik Cenkeri H, Yılmaz B, Gümüş B. Radiation doses from head, neck, chest and abdominal CT examinations: an institutional dose report. *Diagn Interv Radiol*. 2021;27(1):147-151. [CrossRef]
6. van der Molen AJ, Schilham A, Stoop P, Prokop M, Geleijns J. A national survey on radiation dose in CT in the Netherlands. *Insights Imaging*. 2013;4(3):383-390. [CrossRef]
7. Kim MC, Han DK, Nam YC, Kim YM, Yoon J. Patient dose for computed tomography examination: dose reference levels and effective doses based on a national survey of 2013 in Korea. *Radiat Prot Dosimetry*. 2015;164(3):383-391. [CrossRef]
8. Tsapaki V, Damilakis J, Paulo G, et al. CT diagnostic reference levels based on clinical indications: results of a large-scale European survey. *Eur Radiol*. 2021;31(7):4459-4469. [CrossRef]
9. Salama DH, Vassileva J, Mahdaly G, et al. Establishing national diagnostic reference levels (DRLs) for computed tomography in Egypt. *Phy Med*. 2017;39:16-24. [CrossRef]
10. Aberle C, Ryckx N, Treier R, Schindera S. Update of national diagnostic reference levels for adult CT in Switzerland and assessment of radiation dose reduction since 2010. *Eur Radiol*. 2020;30(3):1690-1700. [CrossRef]
11. Bos D, Yu S, Luong J, Chu P, et al. Diagnostic reference levels and median doses for common clinical indications of CT: findings from an international registry. *Eur Radiol*. 2022;32(3):1971-1982. [CrossRef]
12. Deak PD, Smal Y, Kalender WA. Multisection CT protocols: sex- and age-specific conversion factors used to determine effective dose from dose-length product. *Radiology*. 2010;257(1):158-166. [CrossRef]