

e-ISSN: 3023-6487

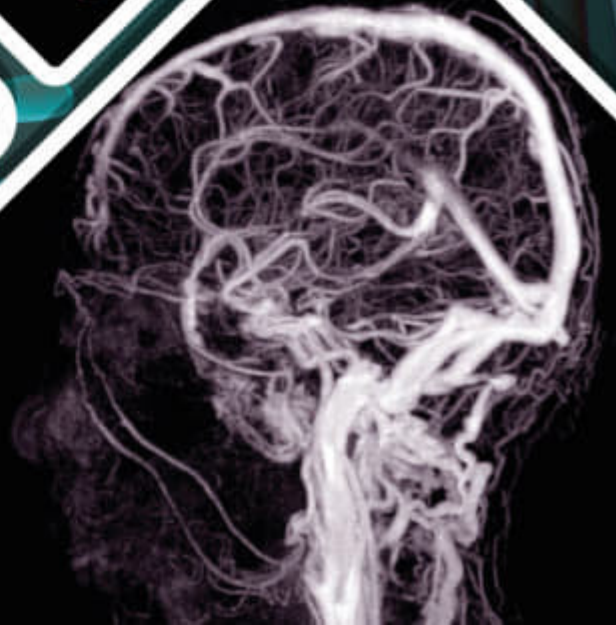
JRMI

Journal of
Radiology in **Medicine**

Volume: 3

Issue: 1

Year: 2026



EDITORS-IN-CHIEF

Assoc. Prof. Adnan ÖZDEMİR

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

Prof. Mehmet Hamdi ŞAHAN

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

EDITORIAL BOARD

Prof. Ayşegül ALTUNKESER

Department of Radiology, Konya City Hospital, Hamidiye Faculty of Medicine, University of Health Sciences, Konya, Türkiye

Prof. Veysel BURULDAY

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

Assoc. Prof. Çağrı DAMAR

Department of Pediatric Radiology, Ankara Bilkent City Hospital, Ankara, Türkiye

Assoc. Prof. Feride KURAL RAHATLI

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

Assoc. Prof. Hasan GÜNDOĞDU

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

Assoc. Prof. Mustafa KAYA

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

Assoc. Prof. Pelin Zeynep BEKİN SARIKAYA

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

Assoc. Prof. Serdar ARSLAN

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

Assoc. Prof. Yunus YILMAZSOY

Department of Radiology, Faculty of Medicine, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Asst. Prof. Melih AKŞAMOĞLU

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

Spec. Imad ALABDUL RAZZAK, MD

Department of Internal Medicine, Phoenixville Hospital, USA

Spec. Melih PEKCAN, MD

Department of Radiology, Dubai Hospital, Dubai, United Arab Emirates

ENGLISH LANGUAGE EDITOR

Assoc. Prof. Mehmet ZENGİN

Department of Medical Pathology, Ankara Training and Research Hospital, University of Health Sciences, Ankara, Turkiye

STATISTICS EDITOR

Asst. Prof. Emrah DOĞAN

Department of Radiology, Hôpital Temara Rabat Maroc, Fas

LAYOUT EDITOR

Kübra YÜRÜMEZ

Graphic/Design, MediHealth Academy Publishing, Ankara, Turkiye

Dear Readers,

Radiology plays a central role in diagnosis, treatment planning, and image-guided interventions. As technology continues to evolve including recent advances in artificial intelligence the importance of publishing reliable and high-quality scientific work remains essential. Our journal serves as a platform for original research, technical reports, and innovative imaging studies.

As we enter our third year, the “Journal of Radiology in Medicine” will continue to connect researchers and clinicians, supporting scientific collaboration and progress in radiology.

We sincerely thank our authors, reviewers, and readers for their invaluable support and dedication.

Kind regards,

Assoc. Prof. Adnan Özdemir

Prof. Mehmet Hamdi Şahan

Editors-in-Chief

Volume: 3

Issue: 1

Year: 2026

ORIGINAL ARTICLES

The relationship between anatomical measurements of the sella turcica and foramen magnum and their effect on nuchal fat thickness in patients with empty sella syndrome..... 1-4

Batur M, Akyol ME, Türkoğlu S.

Feasibility of a multimodal large language model in interpreting plain radiographs of bone tumors: a pilot study..... 5-8

Genç S, Özer H.

Functional Liver Imaging Score on gadoxetic acid-enhanced MRI: correlation with MELD, ALBI, Child-Pugh, FIB-4, and diagnostic performance for advanced hepatic dysfunction in chronic liver disease.....9-14

Akdal Dölek B.

REVIEW

Artificial intelligence and ethics in radiology15-21

Bozer A.

The role of ADC value and ADC ratio in the diagnosis and prognostic evaluation of prostate cancer 22-27

Berksoy AE, Karacaoğlu MC.

The relationship between anatomical measurements of the sella turcica and foramen magnum and their effect on nuchal fat thickness in patients with empty sella syndrome

 **Methiye Batur***¹,  **Mehmet Edip Akyol**²,  **Saim Türkoğlu**³

¹Department of Anatomy, Faculty of Medicine, Van Yüzüncü Yıl University, Van, Türkiye
²Department of Neurosurgery, Faculty of Medicine, Van Yüzüncü Yıl University, Van, Türkiye
³Department of Radiology, Faculty of Medicine, Van Yüzüncü Yıl University, Van, Türkiye

Received: 20.11.2025

Accepted: 27.12.2025

Published: 09.03.2026

Cite this article: Batur M, Akyol ME, Türkoğlu S. The relationship between anatomical measurements of the sella turcica and foramen magnum and their effect on nuchal fat thickness in patients with empty sella syndrome. *J Radiol Med.* 2026;3(1):1-4.

*Corresponding Author: Methiye Batur, methiyebatur@hotmail.com

ABSTRACT

Aims: Empty sella (ES) is a radiological finding characterized by flattening of the pituitary gland and filling of the sella turcica with cerebrospinal fluid (CSF). It is often associated with obesity and increased intracranial pressure (ICP). This study aimed to evaluate the relationship between anatomical measurements of the sella turcica and the foramen magnum (FM) in patients with ES and to investigate the potential association of these structures with nuchal subcutaneous fat thickness.

Methods: Thirty-one patients diagnosed with ES by magnetic resonance imaging (MRI) and 26 age- and sex-matched healthy controls were included in the study. Measurements of the sellar fossa included midsagittal narrow and wide diameters, craniocaudal depth, and transverse diameter. Additionally, FM anteroposterior and transverse diameters were measured. Nuchal subcutaneous fat thickness was also assessed at the FM level. Statistical analyses included group comparisons and correlation analyses.

Results: The ES group showed significantly larger narrow and wide diameters, as well as greater depth of the sellar fossa, compared to controls ($p < 0.05$). Although the anteroposterior diameter of the FM was similar between groups, the transverse diameter was significantly narrower in ES patients ($p = 0.019$). Nuchal subcutaneous fat thickness was significantly increased in the ES group compared to controls ($p = 0.028$). A strong positive correlation was observed between the wide diameter and the depth of the sellar fossa.

Conclusion: The significant enlargement of the sella turcica morphology and the narrowing of the FM transverse diameter in ES patients support the role of cranial bone structures and CSF dynamics in the pathogenesis of ES. Increased nuchal fat thickness may serve as a practical radiological indicator of obesity, which is commonly associated with ES.

Keywords: Anatomical measurements, empty sella syndrome, foramen magnum, obesity, sella turcica

INTRODUCTION

Empty sella (ES), also known as empty pituitary fossa, is a radiological finding characterized by a flattened pituitary gland within a sellar cavity filled with cerebrospinal fluid (CSF). It develops due to primary or secondary causes that lead to pituitary gland shrinkage or injury. The pathogenesis of primary ES involves mechanisms such as idiopathic benign intracranial hypertension (IIH), obesity, arterial hypertension, and insufficiency of the sellar diaphragm. Herniation of the arachnoid space into the pituitary fossa through a weak point in the diaphragma sellae results in the filling of the fossa with CSF, which compresses and shrinks the pituitary gland. Chronic CSF pulsations often cause bony enlargement and remodeling of

the sella turcica.¹ Secondary ES usually results from damage due to trauma, tumors, surgery, or radiation therapy, with the sella turcica typically appearing normal in size in these cases.²

On MRI, ES is characterized by varying degrees of flattening of the superior surface of the pituitary gland and CSF signal intensity within the sellar boundaries. This condition is often accompanied by bony enlargement and remodeling of the sella turcica. ES is observed in approximately 20% of the population, occurring more frequently in middle-aged obese women, with a female-to-male ratio of 5:1. It may be an incidental finding without clinical significance.



One factor contributing to the formation of ES may be a FM diameter that is narrower than normal. The FM is a key landmark of the skull base,³ and its morphometry has been linked to brain size, skull size, and intracranial volume.^{4,5} Physical examinations of ES patients often reveal obesity.⁶ In these patients, body-mass index (BMI), subcutaneous fat thickness measured at the arm and abdomen, and nuchal subcutaneous fat thickness have been correlated.⁷ Therefore, measuring nuchal fat thickness on MRI can serve as an indicator of obesity, even when the patient's weight is unknown.

FM stenosis may contribute to both IIH and non-IIH ES cases. This stenosis can result in increased intracranial pressure, altered CSF absorption, and expansion of CSF spaces. Our study aims to enhance understanding of ES pathogenesis by examining the relationship between anatomical measurements of the sella turcica and FM in ES patients. Additionally, given the high prevalence of obesity among ES patients, we investigate whether nuchal subcutaneous fat thickness correlates with these anatomical structures.

METHODS

Ethics

This retrospective study was conducted with the approval of the Non-interventional Clinical Researches Ethics Committee of Van Yüzüncü Yıl University (Date: 04.02.2025, Decision No: 2025/01-31). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

MRI images of 31 patients diagnosed with ES at the Van Yüzüncü Yıl University Medical Center Neurosurgery Department and 26 age- and sex-matched healthy controls were evaluated. All participants were aged 18 years or older. Patients with a history of intracranial or pituitary surgery or intracranial space-occupying lesions were excluded. Controls were healthy individuals without intracranial pathology who had undergone brain MRI. Age and sex data were recorded for all subjects. Retrospective analyses were performed on patients who were hospitalized between January 2023 and December 2024.

Radiological Diagnosis of an Empty Sella Syndrome

Pituitary MRI examinations were conducted using a 1.5 Tesla MRI scanner (Magnetom Amira, Siemens Medical Systems, Forchheim, Germany). Sagittal T1-weighted, coronal T2-weighted, and contrast-enhanced coronal T1-weighted images were obtained for all cases. A 0.5 molar gadolinium-based contrast agent was administered, and dynamic pituitary imaging was performed. All images were evaluated by a single radiologist. Pituitary thickness was measured on sagittal contrast-enhanced T1-weighted images. Patients were divided into two groups as follows: complete/total ES: More than 50% of the sella filled with CSF; pituitary thickness ≤ 2 mm, partial empty sella: Less than 50% of the sella filled with CSF; pituitary thickness ≥ 3 mm.

MRI Image Analysis

On T1 sequences, the widest and narrowest midsagittal diameters (Figure 1, red line), craniocaudal depth (Figure 1, yellow line), and transverse diameter of the sella fossa were measured

on axial T2 images at the level connecting the tuberculum sellae and dorsum sellae, where CSF signal was present (Figure 2, red line). For FM measurements, the basion and opisthion points were connected on midsagittal images; the anteroposterior diameter (Figure 1, blue line) and transverse diameter (Figure 3, red line) at this level were measured. Nuchal subcutaneous fat thickness was measured along the line drawn at the FM level (Figure 1, green line).



Figure 1. Magnetic resonance imaging (MRI) section showing the midsagittal widest diameter of the sella fossa (red line), craniocaudal depth (yellow line), anteroposterior diameter of the foramen magnum (blue line), and nuchal subcutaneous fat thickness (green line)



Figure 2. Magnetic resonance imaging (MRI) section showing the transverse diameter measurement of the sella fossa (red line)

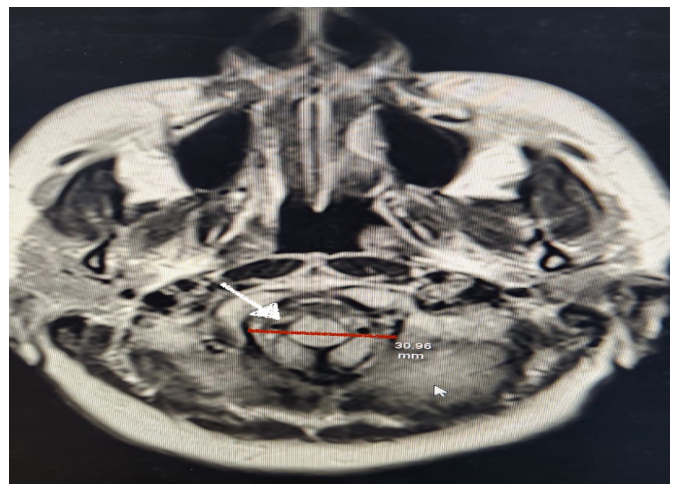


Figure 3. Magnetic resonance imaging (MRI) image showing the transverse diameter measurement of the foramen magnum (red line)

Statistical Analysis

The data analyses were conducted using SPSS software (version 21.0; Chicago, IL). Normality was assessed with the Kolmogorov-Smirnov test. Group comparisons were performed using Student's t-test or the Mann-Whitney U test, as appropriate. Correlation analyses employed Spearman's correlation coefficient. Results are presented as mean±standard deviation (SD), with $p < 0.05$ considered statistically significant.

RESULTS

The mean age of the ES group was 49.13 ± 10.26 years, while that of the control group was 46.31 ± 8.14 years. Seven patients (22.58%) in the ES group and six patients (23.07%) in the control group were male. The narrow and wide diameters of the sellar fossa were significantly larger in the ES group compared to the control group ($p = 0.041$ and $p = 0.040$, respectively). The transverse diameter of the sellar fossa was similar between groups ($p = 0.115$), whereas the craniocaudal depth was significantly greater in the ES group ($p = 0.0001$) (Table).

The anteroposterior diameter of the FM did not differ significantly between groups ($p = 0.159$); however, the transverse diameter was significantly narrower in the ES group compared to controls ($p = 0.019$) (Table). Nuchal subcutaneous fat thickness was 9.7 ± 2.5 mm in the ES group and 8.31 ± 2.15 mm in controls, with the ES group showing a significantly greater thickness ($p = 0.028$) (Table).

Table. Measurements of the SF, FM and nuchal subcutaneous fat thickness in patients with empty sella and in the control group

Measurement (mm)	Group	n	Mean±SD	p
SF narrowest diameter	Empty sella	31	12.10±2.23	0.041
	Control	26	11.96±1.47	
SF widest diameter	Empty sella	31	12.85±2.11	0.04
	Control	26	12.18±1.49	
SF transverse diameter	Empty sella	31	12.96±2.07	0.115
	Control	26	12.20±1.5	
SF depth	Empty sella	31	10.31±1.88	0.0001
	Control	26	7.37±1.87	
FM anteroposterior diameter	Empty sella	31	33.36±2.93	0.159
	Control	26	34.38±2.44	
FM transverse diameter	Empty sella	31	29.88±2.88	0.019
	Control	26	31.61±2.52	
Nuchal subcutaneous fat thickness	Empty sella	31	9.7±2.5	0.028
	Control	26	8.31±2.15	

SF: Sellar fossa, FM: Foramen magnum, mm: Milimetre, SD: Standard deviation

Correlation analysis between sellar fossa and FM measurements revealed a strong positive correlation between the width and depth of the sellar fossa in the ES group ($r = 0.556$, $p = 0.001$). In contrast, a negative correlation was observed between these parameters in the control group ($r = -0.443$, $p = 0.023$).

DISCUSSION

To our knowledge, no previous studies have directly examined the specific relationship or correlation between

the anatomical measurements of the sella turcica and the FM in ES patients. However, both anatomical structures are clinically significant in relation to IIH, which is closely associated with ES. Increased intracranial pressure (ICP) in IIH is believed to affect the morphology of the sella turcica. Chronic CSF pulsations can lead to bony expansion and remodeling of the sella turcica, resulting in compression of the pituitary gland.^{8,9}

Consistent with the literature, we observed enlargement of the sellar fossa dimensions in the ES group compared to controls. Quantitative analyses in IIH patients have demonstrated significantly larger sella turcica areas than those in healthy controls. The mean sella area in IIH patients was 200 ± 24 mm², compared to 124 ± 25 mm² in controls ($p < 0.0001$).¹⁰ Additionally, the pituitary gland to sella turcica area ratio (PG/S ratio) was significantly decreased in IIH patients ($p < 0.0001$).¹¹ This reduction is partly attributable to pituitary compression and sella turcica enlargement, particularly in severe or chronic cases, supporting the notion that the ES appearance primarily results from chronic ICP-induced sella expansion.^{10,12}

Elevated ICP can also affect adjacent skull base structures; for example, enlargement of the foramen ovale has been observed in IIH.¹³ This suggests a generalized remodeling effect of chronic ICP on central skull base anatomy. However, specific effects on FM anatomical measurements, similar to those observed in the sella turcica, have not been reported. In our study, although the anteroposterior diameter of the FM did not differ significantly, the transverse diameter was narrower in patients with ES. This finding suggests that FM stenosis may contribute to increased intracranial pressure, altered CSF absorption, and expansion of CSF spaces closely associated with ES.

No direct causal relationship has been established between specific anatomical measurements of the sella turcica or FM and nuchal fat thickness in ES patients. However, nuchal fat thickness is an important indirect marker related to the clinical status of IIH patients, a condition frequently associated with ES. Obesity, a common risk factor for both primary ES and IIH, underlies this association. In patients with IIH and ES, subcutaneous fat thickness measurements have been used to differentiate whether the ES is incidental or related to chronically elevated ICP. Patients diagnosed with IIH (with ES) had significantly higher scalp fat thickness (mean 9.0 mm) and neck soft tissue thickness (mean 19.5 mm) compared to controls with incidental ES ($p < 0.0001$).⁸ These thickness measurements suggest that obesity may serve as a potential imaging marker in IIH, which is prevalent among obese patients.¹⁴⁻¹⁶ Increased subcutaneous fat thickness in the neck and scalp, combined with clinical symptoms such as age, headache, or visual complaints, strongly suggests that the ES finding is associated with IIH.⁸

Anatomical measurements of the sella turcica, including maximum anteroposterior and craniocaudal dimensions as well as the position of the infundibulum, did not differ significantly between patients with IIH and incidental ES patients without IIH.⁸ This suggests that the appearance of the ES and sella turcica measurements alone are not specific

indicators; however, increased nuchal fat thickness and clinical symptoms contribute pathological significance to the ES findings associated with IIH.^{8,11} Therefore, rather than a direct effect of sella turcica or FM anatomical measurements on nuchal fat thickness, the latter serving as a marker of obesity strongly supports the presence of IIH in conjunction with sella turcica morphology as clinical and radiological indicators.¹⁴⁻¹⁶

FM measurements have not been directly correlated with nuchal fat thickness in the literature; however, the FM is clinically significant as a passageway for vital craniocervical junction structures in conditions of elevated ICP, such as IIH.^{7,17-19}

Limitations

Limitations of this study include the fact that the boundaries of the sellar fossa and FM are bony structures, making computed tomography (CT) more suitable than MRI for precise measurements. Additionally, the retrospective design and reliance on MRI images for measurements represent further limitations. Subgroup analyses comparing partial versus total ES and IIH status could provide more detailed insights into their relationships with sellar fossa and FM measurements; the absence of such subgroup data is another limitation. BMI data were not available in this study; therefore, only nuchal fat thickness was used as a measure of obesity. The lack of BMI data should be considered another limitation of this study. Finally, the relatively small sample sizes of the ES and control groups also limit the study.

CONCLUSION

Consistent with the literature, we observed increased sella turcica depth and anteroposterior diameter in ES patients. The narrower transverse diameter of the FM in ES patients compared to controls may contribute to ES pathogenesis; however, larger prospective studies are necessary to confirm this. Due to their hormonal profiles, ES patients often present with obesity, which is reflected by increased nuchal fat thickness compared to normal controls. Measuring nuchal fat thickness in ES patients can therefore be used as an indicator of obesity.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study was conducted with the approval of the Non-interventional Clinical Researches Ethics Committee of Van Yüzüncü Yıl University (Date: 04.02.2025, Decision No: 2025/01-31).

Informed Consent

As this was a retrospective study, formal written informed consent was not required and was therefore not obtained.

Peer Review Process

This manuscript was subject to external peer review.

Conflict of Interest

The authors declare no conflicts of interest related to this study.

Financial Disclosure

The authors received no financial support for the conduct or publication of this research.

Author Contributions

Concept: MB, MEA, ST; Design: MB, ST; Control: MB, MEA, ST; Data Collection and/or Processing: MB, MEA, ST; Analysis and/or Interpretation: MB, MEA; Literature Review: MB, MEA; Article Writing: MB, MEA; Critical Review: All authors.

REFERENCES

- Saundane AM, Lim PP, Aiken A, Chen Z, Hudgins PA. Factors determining the clinical significance of an "empty" sella turcica. *Am J Roentgenol*. 2013;200(5):1125-1131. doi:10.2214/AJR.12.9013
- Miljic D, Pekic S, Popovic V. Empty sella. *Endotext*. 2024.
- Ogolo DE, Ajare EC, Okwuoma O, et al. Morphometric evaluation of the foramen magnum in the West African population: implications for neurosurgical interventions. *Brain Disorders*. 2024;14:100140. doi:0.1016/j.dscb.2024.100140
- Zdilla MJ, Russell ML, Bliss KN, Mangus KR, Koons AW. The size and shape of the foramen magnum in man. *J Craniovertebr Junction Spine*. 2017;8(3):205-221. doi:10.4103/jcvjs.JCVJS_62_17
- Cirpan S, Yonguc GN, Mas NG, Aksu F, Orhan Magden A. Morphological and morphometric analysis of foramen magnum: an anatomical aspect. *J Craniofac Surg*. 2016;27(6):1576-1578. doi:10.1097/CS.0000000000002822
- Daniels AB, Liu GT, Volpe NJ, et al. Profiles of obesity, weight gain, and quality of life in idiopathic intracranial hypertension (pseudotumor cerebri). *Am J Ophthalmol*. 2007;143(4):635-641. doi:10.1016/j.ajo.2006.12.040
- Gaillard F, Weerakkody Y, Walizai T, et al. Empty sella. Reference article. *Radiopaedia.org*. 2024.
- Saundane AM, Lim PP, Aiken A, Chen Z, Hudgins PA. Factors determining the clinical significance of an "empty" sella turcica. *Am J Roentgenology*. 2013;200(5): 1125-1131. doi:10.2214/AJR.12.9013
- Agid R, Farb RI. Neuroimaging in the diagnosis of idiopathic intracranial hypertension. *Minerva Med*. 2006; 97(6):365-370
- Ranganathan S, Lee SH, Checkver A, et al. Magnetic resonance imaging finding of empty sella in obesity related idiopathic intracranial hypertension is associated with enlarged sella turcica. *Neuroradiology*. 2013;55(8):955-961. doi:10.1007/s00234-013-1207-0
- Yuh WT, Zhu M, Taoka T, et al. MR imaging of pituitary morphology in idiopathic intracranial hypertension. *J Magn Reson Imaging*. 2000;12(6):808-813. doi:10.1002/1522-2586(200012)12:6<808::aid-jmri3>3.0.co;2-n
- Alperin N, Ranganathan S, Bagci AM, et al. MRI evidence of impaired CSF homeostasis in obesity-associated idiopathic intracranial hypertension. *Am J Neuroradiol*. 2013; 34(1):29-34. doi:10.3174/ajnr.A3171
- Butros SR, Goncalves LF, Thompson D, Agarwal A, Lee HK. Imaging features of idiopathic intracranial hypertension, including a new finding: widening of the foramen ovale. *Acta Radiol*. 2012;53(6):682-688. doi:10.1258/ar.2012.110705
- Friedman DI, Rausch EA. Headache diagnoses in patients with treated idiopathic intracranial hypertension. *Neurology*. 2002;58(10):1551-1553. doi:10.1212/wnl.58.10.1551
- Silver RI, Moonis G, Schlosser RJ, Bolger WE, Loevner LA. Radiographic signs of elevated intracranial pressure in idiopathic cerebrospinal fluid leaks: a possible presentation of idiopathic intracranial hypertension. *Am J Rhinol*. 2007; 21(3):257-261. doi:10.2500/ajr.2007.21.3026
- Bruce BB, Kedar S, Van Stavern GP, Corbett JJ, Newman NJ, Biousse V. Atypical idiopathic intracranial hypertension: normal BMI and older patients. *Neurology*. 2010;74(22):1827-1832. doi:10.1212/WNL.0b013e3181e0f838
- Wang H, Rosenbaum AE, Reid CS, Zinreich S, Pyeritz R. Pediatric patients with achondroplasia: CT evaluation of the craniocervical junction. *Radiology*. 1987;164(2):515-519. doi:10.1148/radiology.164.2.3602395
- Bagley CA, Pindrik JA, Bookland MJ, Camara-Quintana JQ, Carson BS. Cervicomedullary decompression for foramen magnum stenosis in achondroplasia. *J Neurosurg Pediatrics*. 2006;104(3):166-172. doi:10.3171/ped.2006.104.3.166
- Yasuoka S, Takakura K. Proposal for the definition of "foramen magnum syndrome" foramen magnum tumor and abnormalities. *No To Shinkei*. 1983;35(10):1001-1007.

Feasibility of a multimodal large language model in interpreting plain radiographs of bone tumors: a pilot study

Samet Genez*, Hamza Özer

Department of Radiology, Faculty of Medicine, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Received: 05.02.2026

Accepted: 04.03.2026

Published: 09.03.2026

Cite this article: Genez S, Özer H. Feasibility of a multimodal large language model in interpreting plain radiographs of bone tumors: a pilot study. *J Radiol Med.* 2026;3(1):5-8.

*Corresponding Author: Samet Genez, samet.genez@ibu.edu.tr

ABSTRACT

Aims: To evaluate the diagnostic accuracy and clinical feasibility of a multimodal large language model (ChatGPT-5) in interpreting plain radiographs of bone tumors and differentiating between benign and malignant lesions.

Methods: This retrospective pilot study utilized 50 verified bone tumor cases (27 benign and 23 malignant) sourced from the Radiopaedia database. Anonymized radiographs were processed by ChatGPT-5 using a standardized zero-shot prompt in independent sessions to prevent contextual bias. Model performance was assessed based on the accuracy of the most likely diagnosis, the inclusion of correct diagnoses within the top three differentials, and benign–malignant classification metrics. Statistical analysis included the Clopper–Pearson binomial method for confidence intervals and McNemar’s exact test to evaluate improvements in diagnostic accuracy and potential systematic error asymmetry.

Results: The model achieved 100% accuracy in identifying the imaging modality and the affected bone. The accuracy for the single most likely diagnosis was 56.0% (95% CI: 41.3-70.0), which significantly increased to 70.0% (95% CI: 55.4-82.1) when two differential diagnoses were included ($p=0.016$). For benign–malignant classification, the model demonstrated an overall accuracy of 76.0%, with a high specificity of 96.3% but a notably limited sensitivity for malignancy at 52.2%. A statistically significant error asymmetry indicated a systematic tendency toward benign classification ($p=0.006$).

Conclusion: While ChatGPT-5 demonstrates proficiency in foundational radiographic identification, its low sensitivity for malignancy remains a critical limitation for independent clinical use. The results suggest that while multimodal LLMs may serve as promising educational or triage aids, they currently require rigorous human expert oversight to maintain diagnostic safety in image interpretation.

Keywords: Bone tumors, large language models, ChatGPT-5, artificial intelligence, radiology

INTRODUCTION

Multimodal large language models (LLMs) have rapidly moved from text-only assistants to general-purpose systems that can increasingly support clinical tasks.¹⁻³ With the emergence of LLMs capable of processing images, interest has grown regarding their potential role as assistive tools in image interpretation workflows.^{4,5}

Radiographs are the first-line modality for suspected bone tumors and can provide high-yield clues such as lesion location, matrix, pattern of bone destruction, periosteal reaction, and soft-tissue extension;⁶ however, in many cases the most clinically relevant output is a plausible ranked differential diagnosis and correct benign-malignant triage rather than a single definitive histologic label.⁷

Early evaluations of LLMs in radiology largely emphasized text-based applications, including report rewriting and exam-style question answering.⁸⁻¹⁰ As these models have begun to be tested on image-based tasks, performance has appeared more variable, particularly in image-only settings, highlighting the importance of rigorous validation designs and clinically meaningful endpoints.¹¹⁻¹⁴

Despite increasing interest in general-purpose Artificial Intelligence (AI) for radiology, evidence specifically addressing bone tumor differential diagnosis on plain radiographs remains limited. Therefore, this study aimed to evaluate the diagnostic performance of ChatGPT-5 on plain radiographs of bone tumors using open-access cases with reference diagnoses.



METHODS

Ethics

Ethics committee approval was waived as the study was based on anonymized, open-access data from a public database, involving no direct patient contact or identifiable information.

Study Design and Case Selection

This retrospective study was designed to evaluate the diagnostic performance of a large multimodal model in the interpretation of bone tumors on plain radiographs. A total of 50 cases of bone tumors with verified diagnoses were retrospectively identified and sourced from Radiopaedia (<https://radiopaedia.org>), a peer-reviewed open-source radiology database. Cases were selectively curated rather than consecutively sampled to form a pilot set of benign and malignant entities with diagnostic-quality radiographs and a clearly stated reference diagnosis. Rare tumors were not intentionally oversampled, although the educational nature of the source may favor more typical presentations. No formal case-difficulty grading was applied. The cohort was curated to include 27 benign and 23 malignant lesions. Inclusion criteria were: (1) availability of a diagnostic-quality conventional radiograph (X-ray) in at least one standard projection, (2) a definitive reference diagnosis confirmed by expert consensus or histopathology as provided by the database, and (3) clear visualization of the tumorous lesion. Cases with significant hardware, suboptimal positioning, or confounding secondary pathologies (e.g., advanced osteomyelitis mimicking a tumor) were excluded to ensure a focused assessment of tumor-specific features.

Image Acquisition and Reference Standard

For each case, the most representative plain radiograph was exported from the database as an anonymized 8-bit grayscale JPEG file. To ensure the model's assessment was purely image-based, all original annotations, markers, or clinical captions embedded in the images were removed. The reference standard consisted of the verified diagnosis provided by the Radiopaedia platform. For binary performance analysis, each diagnosis was further classified as either benign or malignant.

AI-Based Assessment Using ChatGPT-5

The diagnostic assessment was performed using ChatGPT-5.2 (Pro) (OpenAI, San Francisco, CA), a multimodal large language model, accessed via the web-based interface (<https://chatgpt.com>). All evaluations were conducted in February 2026 using the default vision-capable model settings without any task-specific fine-tuning or prior training.

For each case, the anonymized radiograph was uploaded into a new, independent chat session to prevent carryover effects or contextual bias from previous evaluations. A standardized zero-shot prompt was entered for every case:

“Interpret the provided plain X-ray image of a suspected bone tumor. Provide a brief radiographic assessment, then state the most likely diagnosis and 2 differential diagnoses”

The standardized prompting workflow is illustrated in **Figure**. The model was not provided with any patient demographics, clinical history, or laboratory findings; thus, this evaluation reflects an image-only scenario rather than real-world clinical

decision-making. The model's free-text responses, including its primary diagnosis and any mentioned differential diagnoses, were recorded for subsequent analysis.



Interpret the provided plain X-ray image of a suspected bone tumor. Provide a brief radiographic assessment, then state the most likely diagnosis and 2 differential diagnosis

Radiographic assessment (plain AP knee X-ray):

- Distal femoral metaphysis, eccentric cortical-based multilobulated lucent lesion along the lateral cortex
- Well-defined margins with a thin sclerotic rim, no cortical breakthrough
- No aggressive periosteal reaction, no soft-tissue mass, joint preserved

Most likely diagnosis:

Fibrous cortical defect / Non-ossifying fibroma (NOF spectrum lesion)

Differential diagnoses:

1. Osteofibrous dysplasia
2. Brodie abscess (subacute osteomyelitis)

Why NOF fits best:

Metaphyseal, eccentric, cortically based, geographic lytic lesion with sclerotic border and non-aggressive appearance in a long bone around the knee is classic — essentially a textbook benign developmental lesion.

📄 🗑️ 🗨️ 🔄 🔄 ...

Figure. Standardized prompt-and-response workflow used for ChatGPT-5 interpretation of bone tumor radiographs

Performance Metrics and Outcome Assessment

Model performance was assessed based on the accuracy of the most likely diagnosis, differential diagnosis (two differential diagnoses), and benign–malignant classification. Diagnostic correctness was assessed at the general diagnosis level; subtype-qualified labels (e.g., ‘conventional osteosarcoma’) were considered correct if they referred to the same underlying diagnostic entity (e.g., ‘osteosarcoma’). Additionally, the model's ability to correctly identify the imaging modality and the specific bone involved was recorded for each case.

Statistical Analysis

The data analyses were performed using R software (version 4.3.1). The primary endpoint was the accuracy of the initial most likely diagnosis. Proportions were reported as n/N with 95% confidence intervals (CIs) calculated using the exact (Clopper-Pearson) binomial method.

To assess the clinical value of differential diagnoses, the improvement from the initial diagnosis accuracy to the differential diagnosis was evaluated using McNemar's exact test. For benign–malignant classification, a 2×2 confusion matrix was used to calculate sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Balanced accuracy was reported as a point estimate (mean of sensitivity and specificity); 95% CIs were provided for sensitivity and specificity only. Potential systematic bias (e.g., a tendency toward benign classification) was evaluated by analyzing error asymmetry (false negatives vs. false positives) using McNemar's test. A p-value <0.05 was considered statistically significant.

RESULTS

A total of 50 plain radiographs of bone tumors were evaluated (reference labels: 23 malignant and 27 benign lesions). ChatGPT-5 provided responses for all cases, with no response refusals observed (0/50; 0.0% [95% CI: 0.0-7.1]). Additionally, the model correctly labeled the radiograph modality and identified the affected bone in 100% of cases (50/50; 95% CI: 92.9-100). Overall diagnostic accuracy—defined as an exact match between the model's single most likely diagnosis and the reference diagnosis—was 56.0% (28/50; 95% CI: 41.3-70.0). When the model was allowed to provide up to two differential diagnoses, the correct-diagnosis rate increased to 70.0% (35/50; 95% CI: 55.4-82.1), when the reference diagnosis was present in any of the three reported diagnoses (primary+two differentials) representing a statistically significant improvement over single-diagnosis performance (McNemar's exact test, $p=0.016$).

For benign–malignant classification, overall accuracy was 76.0% (38/50; 95% CI: 61.8-86.9; **Table 1**). Sensitivity for malignancy was 52.2% (12/23; 95% CI: 30.6-73.2) and specificity was 96.3% (26/27; 95% CI: 81.0-99.9). The positive predictive value was 92.3% (12/13; 95% CI: 64.0-99.8) and the negative predictive value was 70.3% (26/37; 95% CI: 53.0-84.1), yielding a balanced accuracy of 74.3% (**Table 2**). The model predicted malignancy less frequently than the reference prevalence (26.0% vs 46.0%), suggesting a tendency toward benign classification; this error asymmetry was supported by McNemar's exact test ($p=0.006$; **Table 2**).

Table 1. Confusion matrix summarizing the multimodal LLM's benign-malignant predictions for bone tumor radiographs

	Reference malignant (n=23)	Reference benign (n=27)
Predicted malignant (n=13)	TP=12	FP=1
Predicted benign (n=37)	FN=11	TN=26

LLM: Large language model, TP: True positive, FP: False positive, FN: False negative, TN: True negative

Table 2. Performance metrics for benign-malignant classification of bone tumors on plain radiographs by the multimodal LLM

Metric	Estimate	95% CI (exact)
Accuracy	76.0% (38/50)	61.8-86.9
Sensitivity	52.2% (12/23)	30.6-73.2
Specificity	96.3% (26/27)	81.0-99.9
PPV	92.3% (12/13)	64.0-99.8
NPV	70.3% (26/37)	53.0-84.1
Balanced accuracy	74.3%	—

LLM: Large language model, CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value

DISCUSSION

This pilot study provides critical insights into the diagnostic capabilities of ChatGPT-5 in the radiographic interpretation

of bone tumors. Our results demonstrate that while the model excels in basic descriptive tasks such as identifying imaging modalities and anatomical locations, its performance in definitive diagnostic reasoning and malignancy detection remains limited.

Consistent with prior literature, ChatGPT-5 achieved 100% accuracy in identifying the imaging modality and the affected bone. This aligns closely with the work of Hiredesai et al.,¹⁵ who reported a 98.9% accuracy rate for modality identification in upper extremity pathologies. Such high performance suggests that LLMs have successfully integrated foundational medical imaging characteristics into their visual processing frameworks. However, a significant gap exists between “recognizing” a radiograph and “interpreting” its complex pathological features.¹⁶

In our study, ChatGPT-5 reached a diagnostic accuracy of 56% for the most likely diagnosis, which increased to 70% when considering the top two differential diagnoses. This is notably higher than the 19.90% “image-only” accuracy reported by Atakır et al.¹⁷ in a general radiology dataset. This discrepancy may be attributed to the highly structured nature of bone tumor patterns—such as lesional margins and matrix mineralization—which may be more conducive to the model's pattern recognition than the heterogeneous cases used in broader studies. Nevertheless, a 56% diagnostic accuracy is unlikely to support unsupervised clinical use, suggesting that LLMs currently serve best as assistive tools under radiologist oversight.^{16,17}

A critical finding of our study is a tendency toward benign classification, with a sensitivity of only 52.2% for malignancy. This benign-leaning asymmetry ($p=0.006$) is clinically consequential, as missed malignant bone tumors may delay appropriate referral and definitive management. In an image-only workflow, these results suggest that a general-purpose LLMs may have limited sensitivity for malignancy, reinforcing the need for radiologist oversight and task-specific validation before clinical deployment.¹⁶ Furthermore, research by Atakır et al.¹⁷ suggests that LLMs are heavily text-biased; their diagnostic consistency and accuracy improve dramatically when textual clinical context is provided, yet the addition of the image itself often yields diminishing marginal returns.

The limitations observed in ChatGPT-5's visual reasoning are likely due to its architecture, which was primarily optimized for natural language processing rather than specialized medical computer vision.¹⁸ While the latest iterations show “numerical improvements” over previous versions in medical assessments, they still exhibit hallucinations and struggle with subtle findings that human experts easily identify.⁴ For instance, Hiredesai et al.¹⁵ noted that ChatGPT 4.0 often declined to provide a diagnosis or provided generalized information instead of specific findings.

Limitations

This study has several limitations. First, the use of static, single-view radiographs does not reflect the dynamic nature of clinical practice, where radiologists often utilize multiple views and longitudinal data. Second, cases were drawn from an open-access educational repository and

were curated primarily for teaching purposes, which may introduce selection/spectrum bias toward well-demonstrated archetypal examples and may not reflect the prevalence or full spectrum of presentations in the general clinical population. Third, our evaluation focused on a “zero-shot” prompting approach, prompt engineering and optimized instructions can significantly reduce information loss and improve the clarity of AI outputs. Future research should explore “few-shot” learning or the integration of LLMs with specialized Convolutional Neural Networks (CNNs), which have historically outperformed general-purpose LLMs in fracture and tumor detection.

CONCLUSION

ChatGPT-5 demonstrates a promising but still insufficient capacity for the independent interpretation of bone tumor radiographs. While it can serve as a valuable triage or educational aid, potentially enhancing the performance of residents in structured scenarios-its low sensitivity for malignancy underscores the indispensable role of human expert oversight. As these models evolve, their integration into radiology must be governed by rigorous validation to ensure they supplement, rather than compromise, diagnostic safety.

ETHICAL DECLARATIONS

Ethics Committee Approval

Since the study involved no human or animal subjects, clinical interventions, or identifiable patient data, ethics committee approval was not required.

Informed Consent

Since the study did not involve human or animal subjects, clinical interventions, or identifiable patient data, informed consent was not required.

Peer Review Process

This manuscript was subject to external peer review.

Conflict of Interest

The authors declare no conflicts of interest related to this study.

Financial Disclosure

The authors received no financial support for the conduct or publication of this research.

Author Contributions

Project Conceptualization and Study Design: SG; Data Collection: SG; Statistical Analysis: SG, HÖ; Manuscript Drafting: SG, HÖ; Review of the Final Manuscript Submitted for Publication: SG, HÖ.

REFERENCES

1. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. 2023;104(6):269-274. doi:10.1016/j.diii.2023.02.003
2. Genez S, Özer H, Buz Yaşar A, et al. Evaluation of ChatGPT-5 for automated ASPECTS assessment on non-contrast CT in acute ischemic stroke. *Diagnostics (Basel)*. 2025;15(24):3160. doi:10.3390/diagnostics15243160
3. Nguyen D, Rao A, Mazumder A, Succi MD. Exploring the accuracy of embedded ChatGPT-4 and ChatGPT-4o in generating BI-RADS scores: a pilot study in radiologic clinical support. *Clin Imaging*. 2025; 117:110335. doi:10.1016/j.clinimag.2024.110335
4. Wu Z, Li S, Zhao X. The application of ChatGPT in medical education: prospects and challenges. *Int J Surg*. 2025;111(1):1652-1653. doi:10.1097/JS9.0000000000001887
5. Liu M, Okuhara T, Chang X, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. *J Med Internet Res*. 2024; 26:e60807. doi:10.2196/60807
6. Costelloe CM, Madewell JE. Radiography in the initial diagnosis of primary bone tumors. *AJR Am J Roentgenol*. 2013;200(1):3-7. doi:10.2214/AJR.12.8488
7. Priolo F, Cerase A. The current role of radiography in the assessment of skeletal tumors and tumor-like lesions. *Eur J Radiol*. 1998;27 Suppl 1: S77-S85. doi:10.1016/s0720-048x(98)00047-3.
8. Kitamura FC. ChatGPT Is shaping the future of medical writing but still requires human judgment. *Radiology*. 2023;307(2):e230171. doi:10.1148/radiol.230171
9. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198
10. Liao W, Liu Z, Dai H, et al. Differentiating ChatGPT-generated and human-written medical texts: quantitative study. *JMIR Med Educ*. 2023;9:e48904. doi:10.2196/48904
11. Hayden N, Gilbert S, Poisson LM, Griffith B, Klochko C. Performance of GPT-4 with vision on text- and image-based ACR diagnostic radiology in-training examination questions. *Radiology*. 2024;312(3):e240153. doi:10.1148/radiol.240153
12. Handa P, Chhabra D, Goel N, Krishnan S. Exploring the role of ChatGPT in medical image analysis. *Biomed Signal Process Control*. 2023;86:105292. doi:10.1016/j.bspc.2023.105292.
13. Dehdab R, Brendlin A, Werner S, et al. Evaluating ChatGPT-4V in chest CT diagnostics: a critical image interpretation assessment. *Jpn J Radiol*. 2024;42(10):1168-1177. doi:10.1007/s11604-024-01606-3
14. Lacaita PG, Galijasevic M, Swoboda M, et al. The accuracy of ChatGPT-4o in interpreting chest and abdominal X-Ray images. *J Pers Med*. 2025; 15(5):194. doi:10.3390/jpm15050194
15. Hiredesai AN, Martinez CJ, Anderson ML, Howlett CP, Unadkat KD, Noland SS. Is artificial intelligence the future of radiology? Accuracy of ChatGPT in radiologic diagnosis of upper extremity bony pathology. *Hand (N Y)*. 2026;21(1):73-80. doi:10.1177/15589447241298982
16. Yang X, Chen W. The performance of ChatGPT on medical image-based assessments and implications for medical education. *BMC Med Educ*. 2025;25(1):1192. doi:10.1186/s12909-025-07752-0
17. Atakır K, Işın K, Taş A, Önder H. Diagnostic accuracy and consistency of ChatGPT-4o in radiology: influence of image, clinical data, and answer options on performance. *Diagn Interv Radiol*. 2025. doi:10.4274/dir.2025.253460
18. Lyu Q, Tan J, Zapadka ME, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Vis Comput Ind Biomed Art*. 2023; 6(1):9. doi:10.1186/s42492-023-00136-5

Functional Liver Imaging Score on gadoxetic acid-enhanced MRI: correlation with MELD, ALBI, Child-Pugh, FIB-4, and diagnostic performance for advanced hepatic dysfunction in chronic liver disease

 **Betül Akdal Dölek**

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

Received: 07.02.2026

Accepted: 04.03.2026

Published: 09.03.2026

Cite this article: Akdal Dölek B. Functional Liver Imaging Score on gadoxetic acid-enhanced MRI: correlation with MELD, ALBI, Child-Pugh, FIB-4, and diagnostic performance for advanced hepatic dysfunction in chronic liver disease. *J Radiol Med.* 2026;3(1):9-14.

Corresponding Author: Betül Akdal Dölek , b_akdal@yahoo.com

ABSTRACT

Aims: The objective was to evaluate the diagnostic performance of the Functional Liver Imaging Score (FLIS) for the identification of severe hepatic dysfunction and the relationship between the FLIS on gadoxetic acid-enhanced MRI and established liver function scores in patients with chronic liver disease (CLD).

Methods: 104 patients with chronic liver disease who had gadoxetic acid-enhanced MRIs between January 2021 and December 2024 were the subjects of this retrospective study. Twenty-minute hepatobiliary phase images were used to score FLIS (0-6) based on portal vein signal intensity, biliary excretion, and parenchymal enhancement. The MELD, ALBI, Child-Pugh, and FIB-4 scores were computed using laboratory data obtained within two weeks before MRI.

Results: FLIS correlated with key biochemical markers of hepatic reserve (higher albumin and sodium; lower bilirubin and cholestatic/hepatocellular enzymes; all $p < 0.001$) and showed strong correlations with MELD and ALBI ($r = -0.63$ each). Mean FLIS decreased from Child-Pugh A (5.66 ± 0.53) to B (3.85 ± 1.32) to C (1.57 ± 0.79) ($p < 0.001$). FLIS demonstrated high diagnostic accuracy for advanced dysfunction, with AUCs of 0.86 for MELD > 5 and 0.94 for Child-Pugh B/C.

Conclusion: FLIS has high correlations with biochemical and composite scores of liver function and can well predict severe hepatic dysfunction. The moderate correlation of FLIS with FIB-4 indicates that FLIS is more a measure of functional reserve than fibrosis and thus confirms its utility as an imaging biomarker.

Keywords: Functional Liver Imaging Score, chronic liver disease, MELD, ALBI, Child-Pugh, FIB-4

INTRODUCTION

Chronic liver disorders represent a major global health burden, with prevalence estimates reaching hundreds of millions internationally.^{1,2} To predict prognosis, guide treatment choices, and evaluate surgical or interventional risks in these patients, an accurate evaluation of hepatic functional reserve is crucial.³

Traditionally, liver function has been measured using laboratory-based tools such as the Child-Pugh, albumin-bilirubin (ALBI) scores and Model for End-Stage Liver Disease (MELD).⁴⁻⁶

Although carefully validated, these indices are influenced by acute biochemical fluctuation, systemic illness, and renal

function and are not regionally specific with respect to hepatic function.^{7,8} Consequently, interest in non-invasive, image-based biomarkers that directly reflect hepatocellular function is growing.

Magnetic resonance imaging (MRI) performed with gadoxetic acid allows evaluation of both liver morphology and contrast-related functional characteristics in the same examination.⁹ The uptake of gadoxetic acid depends mainly on organic anion transporting polypeptides (OATP)-mediated transport into hepatocytes, whereas its biliary elimination requires multidrug resistance-associated protein 2 (MRP2)-related excretion pathways.^{10,11}



The Functional Liver Imaging Score (FLIS), first described by Bastati et al.,¹² is a simple semi-quantitative scoring system that uses three variables from hepatobiliary phase MRI: parenchymal enhancement, biliary excretion, and portal vein signal intensity.

Several studies have demonstrated associations between FLIS and biochemical liver function tests, clinical scores, and postoperative outcomes.¹²⁻¹⁸ However, data directly comparing FLIS with a comprehensive panel of widely used liver function and fibrosis indices, including MELD, ALBI, Child-Pugh, and FIB-4, in a single mixed-etiology chronic liver disease (CLD) cohort remain limited.

Therefore, the primary aim of this study was to assess the correlation between FLIS and established liver function scores (MELD and ALBI) in patients with CLD. Secondary aims were to evaluate its relationship with Child-Pugh class and FIB-4, and to determine the diagnostic performance of FLIS for identifying advanced hepatic dysfunction.

METHODS

Ethics

The study was carried out with the permission of the Ankara Bilkent City Hospital Scientific Researches Evaluation and Ethics Committee (Date: 19.11.2025, Decision No: TABED 1-25-1842). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Population

All consecutive patients who underwent gadoteric acid-enhanced liver MRI between January 2021 and December 2024 were screened for eligibility. CLD was a mandatory inclusion criterion and was confirmed by at least one of the following: (1) a documented diagnosis of chronic hepatitis, fibrosis, or cirrhosis by a hepatologist or gastroenterologist; (2) characteristic imaging findings of CLD or cirrhosis, including nodular liver contour, segmental volume redistribution, caudate lobe hypertrophy, splenomegaly, or portosystemic collaterals; and/or (3) a documented chronic etiology (e.g., hepatitis B or C infection >6 months, alcohol-related liver disease, or metabolic-associated steatotic liver disease) with supportive clinical and laboratory findings. Etiology of CLD was determined from clinical records and categorized as viral, alcohol-related, or metabolic/cryptogenic.

Eligibility criteria included availability of hepatobiliary-phase images acquired ≥ 20 minutes after contrast administration, laboratory and clinical data obtained within two weeks of MRI, and absence of focal hepatic lesions involving more than one liver segment. During the study period, 312

consecutive gadoteric acid-enhanced MRI examinations were screened. Patients were excluded due to absence of a 20-minute hepatobiliary phase (n=15), laboratory or clinical data obtained more than two weeks from MRI (n=24), severe motion or breathing artefacts (n=51), portal vein thrombosis (n=11), prior hepatic surgery (n=32), or focal hepatic lesions involving more than one Couinaud segment (n=75). After applying these predefined exclusion criteria, 104 patients with CLD were included in the final analysis.

MRI Acquisition

All MR examinations were performed on a 3.0-T scanner (SIGNA, GE HealthCare, Waukesha, WI, USA) using a phased-array abdominal coil. Patients were imaged in the supine position. The protocol consisted of axial and coronal T2-weighted sequences (typical field of view 38-40 cm, TR/TE approximately 3750-4280/90 ms, slice thickness 5 mm, matrix 320-352×224-230) and T1-weighted sequences (FOV 40-44 cm, TR/TE 4.2-4.9/2.3-2.6 ms, slice thickness 4 mm, matrix 260-300×162-200).

Gadoxetic acid (Gd-EOB-DTPA, Primovist®, Bayer Healthcare) was delivered intravenously at a dosage of 0.025 mmol/kg, succeeded by a 30-ml saline flush. Hepatobiliary phase images were acquired 20 minutes after injection.

FLIS Assessment

FLIS was assessed on hepatobiliary phase images according to the criteria described by Bastati et al.¹² FLIS was derived by visually evaluating hepatobiliary-phase images in terms of (1) the relative brightness of the liver compared with the kidney, (2) the clarity of intrahepatic biliary ducts, and (3) the appearance of the portal vein compared with parenchyma. Each component contributes 0-2 points to the composite score, and the total FLIS was obtained by summation (range, 0-6) (Table 1, Figure 1).

An abdominal radiologist with 10 years of experience, blinded to laboratory values and clinical data, scored all examinations independently.

Laboratory and Clinical Data

All laboratory tests were performed within two weeks of the MRI examination. The following parameters were recorded: serum bilirubin, albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), international normalized ratio (INR), creatinine, serum sodium, and platelet count. Clinical liver function scores were calculated as follows: Child-Pugh score and class (A-C), MELD score, ALBI score and grade ($ALBI = [\log_{10} \text{bilirubin} \times 0.66] + [\text{albumin} \times 0.085]$), and FIB-4 index ($FIB-4 = [\text{age} \times \text{AST}] / [\text{platelet count} \times \sqrt{\text{ALT}}]$). All scores were calculated using the laboratory values closest to the MRI date.

Table 1. Functional Liver Imaging Score (FLIS) parameters evaluated on hepatobiliary-phase magnetic resonance imaging (MRI)

Parameter	Score 0	Score 1	Score 2
Parenchymal enhancement (compared with renal cortex)	Lower than renal cortex	Equal to renal cortex	Higher than renal cortex
Biliary excretion	No visible intrahepatic bile ducts	Faint visualization of bile ducts	Distinct visualization of intrahepatic bile ducts
Portal vein signal intensity (relative to liver parenchyma)	Hyperintense	Isointense	Hypointense



Figure 1. Gadoteric acid-enhanced hepatobiliary phase MRI in a 55-year old man with advanced chronic liver disease.

The Functional Liver Imaging Score is 1, based on absent intrahepatic biliary excretion (score 0), liver parenchymal signal intensity equal to the renal cortex (score 1), and hyperintense portal vein relative to the liver parenchyma (score 0). These imaging findings are concordant with severe hepatic dysfunction, as reflected by a MELD score of 18, Child-Pugh class C, and ALBI grade 3

MRI: Magnetic resonance imaging, MELD: Model for End-Stage Liver Disease, ALBI: Albumin-bilirubin

Statistical Analysis

All statistical computations were carried out using SPSS (version 26; IBM), and analyses were tailored to the distributional properties of each variable.

Quantitative measures are summarized as means \pm SD or medians and interquartile ranges, whichever was appropriate, given the distribution. Categorical data are described as absolute numbers and proportions. Distributional characteristics of numeric variables were assessed using the Shapiro-Wilk procedure. Between-group differences were investigated by parametric or non-parametric tests according to the distribution of the data: ANOVA or the Kruskal-Wallis test.

The main analysis examined the association of FLIS with MELD and ALBI scores using Spearman's rank correlation coefficient (r) with associated 95% confidence intervals (CIs). Secondary analyses explored associations between FLIS and other biochemical measures and FIB-4.

ROC analysis was performed to assess the diagnostic performance of FLIS for detecting advanced liver dysfunction defined as MELD >15 , Child-Pugh class B/C versus class A, and FIB-4 ≥ 3.25 . AUCs and 95% CIs were calculated for all endpoints investigated. Thresholds reflecting the best balance between sensitivity and specificity were selected based on Youden's method. A two-sided p -value <0.05 was considered statistically significant.

Additional analyses were conducted to assess whether the FLIS provided independent and incremental value beyond the established biochemical indices.

To assess incremental discrimination, the AUC was calculated for both the base model and the extended model.

RESULTS

The study included 104 participants with chronic liver disease. Fifty-two percent were male, and the average age was 54.1 \pm 11.1 years. **Table 2** summarizes the clinical, laboratory, and demographic features of each patient in the study. The cohort included a mixed etiologic spectrum, reflecting real-world clinical practice.

FLIS scores correlated significantly with both the biochemical and composite parameters of liver function.

Table 2. Baseline demographic, clinical and laboratory characteristics of the study population (n=104)

Parameter	Value
Age (years)	54.1 \pm 11.1
Male sex, n (%)	54 (52)
Etiology of chronic liver disease, n (%)	
Hepatitis B	49 (47)
Hepatitis C	21 (20)
Alcohol-related	11 (11)
NASH/cryptogenic	23 (22)
Child-Pugh score	7 \pm 2
Class A/B/C, n (%)	71 (68.3%)/26 (25.0%)/7 (6.7%)
MELD score	13 \pm 5
ALBI score	-1.8 \pm 0.6
Grade 1/2/3, n (%)	38 (36.5%)/47 (45.2%)/19 (18.3%)
FIB-4 index	4.8 \pm 2.9
Serum albumin (g/dl)	3.4 \pm 0.6
Total bilirubin (mg/dl)	1.9 \pm 1.7
INR	1.3 \pm 0.3
AST (U/L)	65 \pm 38
ALT (U/L)	52 \pm 44
ALP (U/L)	138 \pm 58
Platelet count ($\times 10^3/\mu\text{L}$)	122 \pm 61
Serum sodium (mmol/L)	138 \pm 4
Creatinine (mg/dl)	0.9 \pm 0.3
FLIS score	4.9 \pm 1.6
FLIS by Child-Pugh class (A/B/C)	5.7 \pm 0.5/3.8 \pm 1.3/1.6 \pm 0.8
Values are presented as mean \pm SD or n (%). AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, ALP: Alkaline phosphatase	

FLIS correlated positively with albumin ($r=0.56$, $p<0.001$) and serum sodium ($r=0.40$, $p<0.001$), and negatively with bilirubin ($r=-0.59$, $p<0.001$), AST ($r=-0.53$, $p<0.001$), ALP ($r=-0.51$, $p<0.001$), and INR ($r=-0.32$, $p<0.001$). Among composite indices, FLIS showed strong negative correlations with ALBI score ($r=-0.63$, $p<0.001$) and MELD score ($r=-0.63$, $p<0.001$), and a moderate negative correlation with FIB-4 ($r=-0.33$, $p<0.001$).

When categorized according to the Child-Pugh classification, FLIS decreased sequentially from Class A (5.66 \pm 0.53) to Class B (3.85 \pm 1.32) to Class C (1.57 \pm 0.79) ($p<0.001$, Kruskal-Wallis test) (**Figure 2**).

When the MELD score was categorized as ≤ 10 , 11-18 and >18 , the respective mean FLIS values were 5.8 \pm 1.1, 4.3 \pm 1.4 and 2.2 \pm 1.1 ($p<0.001$) (**Figure 2**).

Post hoc analysis also confirmed that FLIS decreased significantly in both the 11-18 and >18 groups compared with the ≤ 10 group (Bonferroni-corrected $p<0.01$).

Similarly, FLIS decreased stepwise along ALBI grades: Grade 1 (5.7 \pm 1.0), grade 2 (4.1 \pm 1.3) and grade 3 (2.4 \pm 1.0) ($p<0.001$) (**Figure 2**).

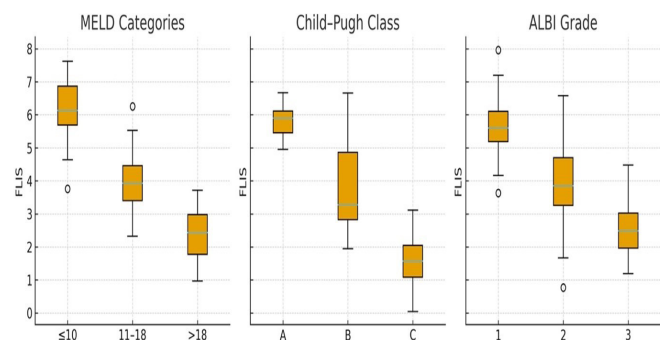


Figure 2. Boxplots showing the distribution of FLIS according to (A) MELD categories, (B) Child-Pugh class, and (C) ALBI grade
 FLIS: Functional Liver Imaging Score, MELD: Model for End-Stage Liver Disease, ALBI: Albumin-bilirubin

Thirty-three patients (32%) were classified as Child-Pugh class B/C. In a multivariable logistic regression model including age, sex, and ALBI grade, higher ALBI grade was independently associated with Child-Pugh class B/C (OR per 1-grade increase, 6.9; 95% CI, 1.8-18.2; $p=0.034$). When FLIS was added to this model, it remained a strong and independent predictor of advanced clinical stage.

The discriminative ability of the base model (age, sex, ALBI grade) for identifying Child-Pugh class B/C was high, with an AUC of 0.88 (95% CI, 0.81-0.95). Addition of FLIS further improved classification, increasing the AUC to 0.96 (95% CI, 0.92-0.99) (Figure 3). The incremental gain in AUC was 0.07 (95% CI, 0.03-0.13) based on bootstrap resampling, indicating significant added prognostic value of FLIS beyond biochemical indices alone.

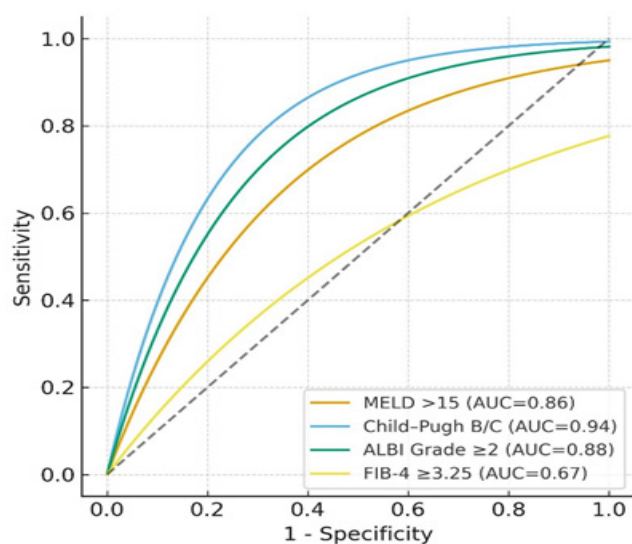


Figure 3. Receiver operating characteristic (ROC) curves demonstrating the diagnostic performance of the Functional Liver Imaging Score (FLIS) for identifying advanced hepatic dysfunction

Using the optimal FLIS threshold of ≤ 5 , the sensitivity for predicting MELD >15 was 100% (95% CI, 75-100%), and the negative predictive value was 100% (95% CI, 93-100%), while specificity was 55% and the positive predictive value was 24% (Table 3).

DISCUSSION

In this single-center cohort of patients with chronic liver disease, we found that FLIS derived from gadoxetic

Metric	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
FLIS ≤ 5	1.00 (0.75-1.00)	0.55 (0.44-0.65)	0.24	1.00
ALBI grade ≥ 2	0.85 (0.55-0.98)	0.51 (0.40-0.61)	0.20	0.96
Child-Pugh B/C	1.00 (0.75-1.00)	0.78 (0.68-0.86)	0.39	1.00

CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value, FLIS: Functional Liver Imaging Score, MELD: Model for End-Stage Liver Disease, ALBI: Albumin-bilirubin

acid-enhanced MRI correlated strongly with established biochemical and clinical measures of liver function, including MELD, ALBI, and Child-Pugh scores, while showing only moderate association with the fibrosis-based FIB-4 index. These findings support the concept that FLIS primarily reflects hepatocellular functional reserve and transporter activity, rather than structural fibrosis alone.

Our multivariable modeling shows that FLIS provides information that is independent from established biochemical markers of liver function. Specifically, after adjustment for age, sex, and ALBI grade, every 1-point decrease in FLIS increased the odds of being Child-Pugh class B/C by approximately 11-fold. Notably, the addition of FLIS to a model that contained ALBI grade increased the AUC for identifying Child-Pugh B/C from 0.88 to 0.96, reflecting a statistically and clinically meaningful improvement in discriminative capability. Such findings support the contention that FLIS reflects aspects of hepatocellular functional reserve that are not fully captured by laboratory-based scores in isolation. Moreover, the weaker relationship between FLIS and FIB-4 underscores its functional rather than structural biomarker status. Collectively, these results position FLIS as a complementary imaging tool that enhances risk stratification when combined with established biochemical indices. Our results complement and extend previous studies that have linked FLIS to Child-Pugh class, ALBI grade, and postoperative outcomes by providing a comprehensive comparison with MELD, ALBI, Child-Pugh, and FIB-4 in a mixed-etiology CLD population and by demonstrating high diagnostic performance of FLIS for advanced hepatic dysfunction. The mechanisms behind FLIS's diagnostic validity are biologically plausible.¹²⁻¹⁸

Gadoxetic acid uptake is governed by hepatocellular transport mechanisms involving OATP family carriers, while its excretion into bile depends on canalicular efflux pathways such as MRP2,8,10. The transport processes may be impaired due to hepatocellular dysfunction or cholestasis, resulting in diminished hepatobiliary-phase signal intensity and reduced FLIS scores. Thus, FLIS provides a direct measure of hepatocellular transporter function, which is a vital aspect of hepatic reserve that cannot be determined by structural fibrosis alone.

In line with this physiology, our study found strong inverse correlations between FLIS and bilirubin ($r=-0.59$), INR ($r=-0.32$), AST ($r=-0.53$) and ALP ($r=-0.51$) and positive correlations with albumin ($r=0.56$) and serum sodium ($r=0.40$). These biochemical correlations confirm that FLIS reflects both hepatocellular synthetic capacity and transporter-mediated excretory function.

The Child-Pugh score is still one of the most widely used for staging cirrhosis and predicting prognosis. However, it incorporates subjective criteria such as ascites and encephalopathy, which are susceptible to clinical heterogeneity and treatment variation, and which can lead to misclassification of functional status.^{19,20} Unlike the Child-Pugh system, ALBI relies exclusively on bilirubin and albumin values and avoids subjective clinical variables.^{4,21}

Various studies have demonstrated that the ALBI score is more effective than the Child-Pugh score at predicting prognosis, particularly in cases of hepatocellular carcinoma and in surgical populations.²²⁻²⁶

In our study population, FLIS decreased stepwise in both Child-Pugh classes and ALBI grades, reflecting the expected deterioration in liver function. This simultaneous behavior indicates that FLIS, ALBI and Child-Pugh all describe the same biological range of hepatocellular function, but in different ways. Child-Pugh scores the clinical aspects, ALBI scores the objective laboratory values, and FLIS scores the essential imaging findings of hepatobiliary transport capability. These scores are complementary, and this enhances the utility of FLIS as a non-invasive imaging surrogate for hepatic functional reserve that can be combined with existing scoring systems. FLIS scores were found to be monotonically decreasing with increasing MELD scores (≤ 10 , 11-18, and >18), suggesting a progressive reduction in hepatocellular functional reserve with worsening disease severity. The monotonic relationship suggests a strong correlation between imaging and biochemical parameters of liver dysfunction. Since MELD scores include bilirubin, INR, and creatinine, which can be affected by systemic factors, FLIS may reflect a more direct assessment of hepatocellular capability. A similar study by Sakai et al.¹⁴ found that FLIS is a better predictor of hepatectomy outcomes than MELD or ALBI scores in patients with normal liver function.

Our ROC analysis also confirmed that FLIS is highly discriminatory for severe hepatic dysfunction, with AUCs of 0.86 for MELD >5 and 0.94 for Child-Pugh B/C. These values are within the range reported for quantitative MR techniques such as T1 mapping and perfusion analysis in the literature, suggesting that a simple visual score like FLIS may offer comparable clinical utility with minimal additional post-processing.^{11,17}

The FIB-4 index, which was originally developed to evaluate the risk of advanced fibrosis in individuals with hepatitis C, is based on age, AST, ALT, and platelet count.^{27,28} It is also commonly used as an inexpensive, non-invasive method for ruling out or diagnosing significant fibrosis, according to threshold values of approximately 1.45 and 3.25. However, the FIB-4 index primarily reflects structural fibrotic burden rather than hepatocellular function. Therefore, we identified a moderate correlation between FLIS and FIB-4 (AUC=0.67 for FIB-4 ≥ 3.25), which is understandable on biological grounds given their differing pathophysiological foundations. The most recent large cohort population evidence appears to support the use of FIB-4 for predicting long-term liver-related events and mortality in clinical practice. However, it seems that for short-term functional markers such as portal hypertension and early outcomes, non-invasive fibrosis

scores such as FIB-4 are only associated to a limited extent, while ALBI is strongly associated with these functional outcomes.²⁹⁻³³ FLIS evaluates functional impairment, whereas FIB-4 reflects fibrotic burden two related but distinct processes in chronic liver disease.

Given its simplicity and reproducibility, FLIS may serve as an adjunct imaging biomarker that complements biochemical indices in clinical and surgical decision-making. Several recent studies have suggested that FLIS thresholds around $\leq 5-6$ may indicate clinically relevant impairment and postoperative risk.^{14,15,18} Our results confirm this range, confirming FLIS as a clinically effective measure in surgery and treatment planning.

Limitations

First, its retrospective, single-center design may introduce selection bias, and the relatively small number of patients with advanced disease (Child-Pugh C or very high MELD scores) limits the precision of estimates in these subgroups. Second, laboratory tests were allowed within a two-week window of MRI, during which liver function may fluctuate, particularly in decompensated patients. Third, FLIS assessment was performed by a single reader and interobserver variability was not assessed; although prior studies suggest good reproducibility, this should be confirmed in future multicenter work. Fourth, FIB-4 thresholds were originally derived in hepatitis C populations and were applied here to a mixed-etiology CLD cohort, which may partly explain the only moderate association between FLIS and FIB-4. Finally, we did not evaluate hard clinical outcomes such as decompensation, liver-related mortality, or post-hepatectomy liver failure, and future prospective studies are warranted to establish the prognostic value of FLIS in these settings.

CONCLUSION

FLIS values derived from gadoteric acid-enhanced MRI show strong correlations with MELD, ALBI, and Child-Pugh scores and accurately distinguish advanced hepatic dysfunction in patients with chronic liver disease. The only moderate association with FIB-4 suggests that FLIS predominantly reflects functional rather than structural impairment. In addition to strong univariable correlations, FLIS demonstrated independent and incremental prognostic value beyond ALBI for identifying advanced hepatic dysfunction, supporting its use as a practical imaging biomarker to complement existing liver function scores in risk stratification and treatment planning.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was carried out with the permission of the Ankara Bilkent City Hospital Scientific Researches Evaluation and Ethics Committee (Date: 19.11.2025, Decision No: TABED 1-25-1842).

Informed Consent

As this was a retrospective study, formal written informed consent was not required and was therefore not obtained.

Peer Review Process

This manuscript was subject to external peer review.

Conflict of Interest

The author declare no conflicts of interest related to this study.

Financial Disclosure

The author received no financial support for the conduct or publication of this research.

Author Contributions

The author is solely responsible for the conception, data collection, analysis, and writing of this manuscript.

REFERENCES

- Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *J Hepatol*. 2019;70(1):151-171. doi:10.1016/j.jhep.2018.09.014
- Devarbhavi H, Asrani SK, Arab JP, Nartey YA, Pose E, Kamath PS. Global burden of liver disease: 2023 update. *J Hepatol*. 2023;79(2):516-537. doi:10.1016/j.jhep.2023.03.017
- D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol*. 2006;44(1):217-231. doi:10.1016/j.jhep.2005.10.013
- Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *J Clin Oncol*. 2015;33(6):550-558. doi:10.1200/JCO.2014.57.9151
- Toyoda H, Johnson PJ. The ALBI score: from liver function in patients with HCC to a general measure of liver function. *JHEP Rep*. 2022;4(10):100557. doi:10.1016/j.jhepr.2022.100557
- Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: the model for end-stage liver disease updated for the modern era. *Gastroenterology*. 2021;161(6):1887-1895.e4. doi:10.1053/j.gastro.2021.08.050
- Río Bártulos C, Senk K, Schumacher M, et al. Assessment of liver function with MRI: where do we stand? *Front Med (Lausanne)*. 2022;9:839919. doi:10.3389/fmed.2022.839919
- Baleato-González S, Vilanova JC, Luna A, et al. Current and advanced applications of gadoteric acid-enhanced MRI in hepatobiliary disorders. *Radiographics*. 2023;43(4):e220087. doi:10.1148/rg.220087
- Van Beers BE, Pastor CM, Hussain HK. Primovist, Eovist: what to expect? *J Hepatol*. 2012;57(2):421-429. doi:10.1016/j.jhep.2012.01.031
- Tan CH, Chou SC, Inmutto N, et al. Gadoteric acid-enhanced MRI as a diagnostic tool in the management of hepatocellular carcinoma: report from a 2020 Asia-Pacific multidisciplinary expert meeting. *Korean J Radiol*. 2022;23(7):697-719. doi:10.3348/kjr.2021.0593
- Wang C, Yuan XD, Wu N, Sun WR, Tian Y. Optimization of hepatobiliary phase imaging in gadoteric acid-enhanced magnetic resonance imaging: a narrative review. *Quant Imaging Med Surg*. 2023;13(3):1972-1982. doi:10.21037/qims-22-916
- Bastati N, Beer L, Ba-Salamah A, et al. Gadoteric acid-enhanced MRI-derived functional liver imaging score (FLIS) and spleen diameter predict outcomes in ACLD. *J Hepatol*. 2022;77(4):1005-1013. doi:10.1016/j.jhep.2022.04.032
- Lee HJ, Hong SB, Lee NK, et al. Validation of functional liver imaging scores (FLIS) derived from gadoteric acid-enhanced MRI in patients with chronic liver disease and liver cirrhosis: the relationship between Child-Pugh score and FLIS. *Eur Radiol*. 2021;31(11):8606-8614. doi:10.1007/s00330-021-07955-1
- Sakai N, Takayashiki T, Takano S, Suzuki D, Ohtsuka M. Low functional liver imaging score is associated with poor prognosis following hepatectomy for hepatocellular carcinoma. *Sci Rep*. 2024;14(1):31290. doi:10.1038/s41598-024-82741-9
- Maino C, Romano F, Franco PN, et al. Functional liver imaging score (FLIS) can predict adverse events in HCC patients. *Eur J Radiol*. 2024;180:111695. doi:10.1016/j.ejrad.2024.111695
- Aslan S, Eryuruk U, Tasdemir MN, Cakir IM. Determining the efficacy of functional liver imaging score (FLIS) obtained from gadoteric acid-enhanced MRI in patients with chronic liver disease and liver cirrhosis: the relationship between Albumin-Bilirubin (ALBI) grade and FLIS. *Abdom Radiol (NY)*. 2022;47(7):2325-2334. doi:10.1007/s00261-022-03557-7
- Yu QJ, Luo YC, Zuo ZW, et al. Utility of gadoteric disodium-enhanced magnetic resonance imaging in evaluating liver failure risk after major hepatic resection. *Quant Imaging Med Surg*. 2024;14(5):3731-3743. doi:10.21037/qims-23-1504
- Ji YH, Son IW, Hong SB, et al. Functional Liver Imaging Score (FLIS) as imaging parameter for predicting post-hepatectomy complications in patients with liver cirrhosis. *Acta Radiol*. 2025;66(2):208-217. doi:10.1177/02841851241299088
- Tsoris A, Marlar CA. Use of the Child Pugh score in liver disease. In: *StatPearls*. Treasure Island (FL): StatPearls. 2019.
- Johnson PJ, Pinato DJ, Kalyuzhnyy A, Toyoda H. Breaking the Child-Pugh dogma in hepatocellular carcinoma. *J Clin Oncol*. 2022;40(19):2078-2082. doi:10.1200/JCO.21.02373
- Toyoda H, Johnson PJ. The ALBI score: from liver function in patients with HCC to a general measure of liver function. *JHEP Rep*. 2022;4(10):100557. doi:10.1016/j.jhepr.2022.100557
- Cho WR, Hung CH, Chen CH, et al. Ability of the post-operative ALBI grade to predict the outcomes of hepatocellular carcinoma after curative surgery. *Sci Rep*. 2020;10(1):7290. doi:10.1038/s41598-020-64354-0
- Pinato DJ, Sharma R, Allara E, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. *J Hepatol*. 2017;66(2):338-346. doi:10.1016/j.jhep.2016.09.008
- Zhao S, Wang M, Yang Z, et al. Comparison between Child-Pugh score and Albumin-Bilirubin grade in the prognosis of patients with HCC after liver resection using time-dependent ROC. *Ann Transl Med*. 2020;8(8):539. doi:10.21037/atm.2020.02.85
- Hiraoka A, Kumada T, Kudo M, et al. Albumin-bilirubin (ALBI) grade as part of the evidence-based clinical practice guideline for HCC of the Japan Society of Hepatology: a comparison with the Liver Damage and Child-Pugh classifications. *Liver Cancer*. 2017;6(3):204-215. doi:10.1159/000452846
- Zhang ZQ, Xiong L, Zhou JJ, et al. Ability of the ALBI grade to predict posthepatectomy liver failure and long-term survival after liver resection for different BCLC stages of HCC. *World J Surg Oncol*. 2018;16(1):208. doi:10.1186/s12957-018-1500-9
- Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317-1325. doi:10.1002/hep.21178
- Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology*. 2007;46(1):32-36. doi:10.1002/hep.21669
- Anstee QM, Berentzen TL, Nitze LM, et al. Prognostic utility of Fibrosis-4 Index for risk of subsequent liver and cardiovascular events, and all-cause mortality in individuals with obesity and/or type 2 diabetes: a longitudinal cohort study. *Lancet Reg Health Eur*. 2023;36:100780. doi:10.1016/j.lanepe.2023.100780
- Hsieh YC, Lee KC, Wang YW, et al. Correlation and prognostic accuracy between noninvasive liver fibrosis markers and portal pressure in cirrhosis: role of ALBI score. *PLoS One*. 2018;13(12):e0208903. doi:10.1371/journal.pone.0208903
- European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol*. 2021;75(3):659-689. doi:10.1016/j.jhep.2021.05.025
- Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. *Hepatology*. 2017;66(5):1486-1501. doi:10.1002/hep.29302
- Rabiee A, Deng Y, Ciarleglio M, et al. Noninvasive predictors of clinically significant portal hypertension in NASH cirrhosis: Validation of ANTICIPATE models and development of a lab-based model. *Hepatol Commun*. 2022;6(12):3324-3334. doi:10.1002/hep4.2091

Artificial Intelligence and ethics in radiology

 **Ahmet Bozer**

Department of Radiology, İzmir City Hospital, İzmir, Türkiye

Received: 09.08.2025

Accepted: 24.09.2025

Published: 09.03.2026

Cite this article: Bozer A. Artificial Intelligence and ethics in radiology. *J Radiol Med.* 2026;3(1):15-21.

Corresponding Author: Ahmet Bozer, draahmetbozer@gmail.com

ABSTRACT

Artificial Intelligence (AI) is reshaping radiology by enhancing diagnostic accuracy, optimizing workflows, and supporting clinical decision-making. Despite over 500 Food and Drug Administration (FDA)-approved algorithms, adoption remains limited due to ethical, legal, and operational challenges. Key concerns include data privacy, algorithmic bias, explainability, and accountability. Inadequate representation in training datasets can perpetuate healthcare disparities, while “black-box” decision-making undermines trust and complicates liability. Ethical governance must integrate transparency, fairness, and human oversight from system design through implementation. Data protection frameworks, such as GDPR and Türkiye’s KVKK, mandate anonymization, informed consent, and secure handling of imaging data. Privacy safeguards-metadata cleaning, pixel-level masking, and defacing-are essential to prevent re-identification. Commercial use of health data requires explicit consent, strict oversight, and equitable benefit-sharing. Explainable AI techniques and human-in-the-loop designs can improve trust and reliability. In Türkiye, AI-specific regulations are emerging, with the 2024 Draft AI Law introducing risk-based classification and governance requirements. Global frameworks, including FUTURE-AI and UNESCO guidelines, provide structured pathways for ethical integration. Ensuring AI in radiology is safe, fair, and socially responsible depends on embedding ethical principles into every stage of its lifecycle, fostering interdisciplinary collaboration, and maintaining active oversight to align technological progress with human dignity.

Keywords: Artificial intelligence, radiology, ethics, data privacy, algorithmic bias

INTRODUCTION

Artificial Intelligence (AI) has initiated a transformative shift in medical imaging, with the potential to enhance diagnostic accuracy, reduce processing times, and increase efficiency. Deep learning-based models can analyze modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and mammography with accuracy comparable to that of human experts, and in certain cases, even surpass their performance. However, despite more than 500 AI algorithms approved by the U.S. FDA, only approximately 2% have been widely integrated into routine clinical practice.¹ This indicates the existence of ethical, legal, and operational barriers that go beyond mere technical capability.

Recently, with the rise of technologies such as generative AI, tasks like automated reporting and patient communication have reduced the workload on clinicians, allowing radiologists to focus on more cognitive and interpretative processes.^{2,3} Nevertheless, these advances also introduce a range of ethical challenges. Issues such as embedded algorithmic biases,

opaque decision-making mechanisms (“black box”), lack of explainability, and questions regarding accountability are directly linked to fundamental medical ethics principles.^{4,5}

Imbalanced representation of variables such as gender, race, and age in training datasets risks exacerbating healthcare disparities, particularly among vulnerable patient groups.^{4,6} Moreover, despite high performance, decisions generated by non-explainable models can undermine clinical trust and create uncertainty in legal accountability.² In high-risk fields such as radiology, it is imperative that decision-making is not solely delegated to algorithms but is reinforced through human oversight and adherence to explainability principles.

Thus, ethical implementation should be grounded not only in technical excellence but also in a multilayered governance framework encompassing justice, accountability, transparency, and respect for human dignity. Initiatives such as the European Commission’s Ethics Guidelines for



Trustworthy AI and the FUTURE-AI framework offer guidance toward this goal.^{7,8} In addition, multinational ethics statements published by professional bodies such as the American College of Radiology (ACR), Radiological Society of North America (RSNA), and European Society of Radiology (ESR) have established comprehensive principles covering data usage, algorithm training, and shared accountability.^{2,9}

In conclusion, the safe and equitable integration of AI into radiology requires not only software accuracy but also the establishment of robust ethical governance principles. This book chapter aims to examine the ethical dimensions of AI in radiology, focusing on data governance, algorithm reliability, shared accountability, and social justice.

APPLICATIONS OF AI IN RADIOLOGY

AI and machine learning (ML) significantly enhance clinical efficiency in radiology by improving diagnostic accuracy and reducing workload. Systematic reviews have demonstrated that AI-assisted image analysis shortens diagnostic times, increases interpretation speed, and reduces false-positive and false-negative rates.¹⁰ In modalities such as MRI, AI has the potential to save radiologists' time through tasks such as automated segmentation, reduction of scan times, and worklist prioritization.¹¹ In the diagnostic domain, AI tools are already employed in routine clinical practice with high sensitivity and specificity for detecting urgent findings such as lung nodules, intracranial hemorrhage, and pulmonary embolism.⁴

In treatment evaluation and monitoring, radiomics enables the analysis of parameters such as tumor heterogeneity, treatment response, and disease progression.¹² AI-assisted analyses allow prediction of therapeutic response, thereby enriching clinical decision-making processes.

In the reporting field, studies have shown that the automatic transfer of AI-generated findings into structured reporting templates reduces reporting times by approximately 20% while significantly improving report quality.¹³ Furthermore, in Large Language Model (LLM)-based systems, such as those using GPT-4, draft reports refined by radiologists provide both time and quality advantages.¹⁴ AI also optimizes workflow in structured reporting systems through functionalities such as auto-populating fields, error checking, and quality assurance, leading to substantial time savings and a marked reduction in human errors.

In the context of education, AI models can analyze reports prepared by radiology residents, provide targeted feedback, and accelerate learning processes.¹⁵

UNIVERSAL ETHICAL PRINCIPLES FOR AI

The integration of AI into high-impact domains, particularly healthcare, has made it imperative to clearly define ethical principles and standards. In fields such as medicine, which directly affect human life, universal ethical principles for AI have been shaped by numerous international organizations, academic institutions, and policymakers.^{5,16} These principles establish a

framework of responsibility that extends not only to technology developers but also to healthcare professionals using AI systems, patients, decision-makers, and society as a whole.¹⁷

Respect for human dignity and the protection of individual autonomy form the foundation of AI ethics. Ensuring patients' control over their own data and adopting dynamic informed consent processes are essential in this context.⁴ The existence of continuously learning algorithms allows for repeated use of data, making it advisable that consent be ongoing and renewable over time. In addition, the confidentiality of personal data must be supported by legal frameworks such as the European Union's General Data Protection Regulation (GDPR) and Türkiye's Personal Data Protection Law (KVKK).¹⁶

The principles of non-maleficence and beneficence are directly linked to medical ethics. AI applications should not cause harm to individuals and should provide potential benefits. However, algorithms trained on biased datasets carry the risk of reinforcing systemic inequities. For example, a widely used healthcare algorithm in the United States was shown to assign disproportionately lower risk scores to Black patients.¹⁸ Preventing such inequities requires representative diversity in datasets and rigorous oversight during algorithm training.¹⁹

The principle of justice aims to ensure equitable access to healthcare services. Universal ethical principles call for preventing algorithmic decisions that could lead to disparities among different socioeconomic and demographic groups.³ In this regard, ethical design must be evaluated not only in terms of technological competence but also through a social justice perspective.

Transparency and explainability are prominent ethical concerns in modern AI systems. It is essential to provide external traceability regarding how algorithms reach decisions, along with understandable explanations for users and patients.²⁰ This facilitates both scientific auditability and the establishment of trust. However, the high parameter counts and complex architectures of contemporary deep learning models make explainability challenging.

According to the principle of accountability, ultimate responsibility for erroneous decisions made by AI systems still lies with humans. Clinicians and developers are expected to understand system risks, apply effective oversight during use, and maintain a role in the decision-making chain.^{2,17} Awareness of issues such as automation bias and balancing decision-support tools with human judgment are critical for ethical compliance.

In conclusion, universal AI ethical principles are built on respect for human rights, non-maleficence, justice, explainability, and accountability. Binding ethical frameworks, such as UNESCO's 2021 Recommendation on the Ethics of AI and the Council of Europe's legally binding instrument enacted in 2024, demonstrate that these principles have evolved from guidance into normative regulatory structures.^{16,17} Ensuring their applicability requires interdisciplinary collaboration, educational initiatives, and

the development of ethical oversight mechanisms. In high-risk domains such as radiology, regular review and updating of these principles is essential to ensure that technology advances with a human-centered and equitable approach (Figure 1).

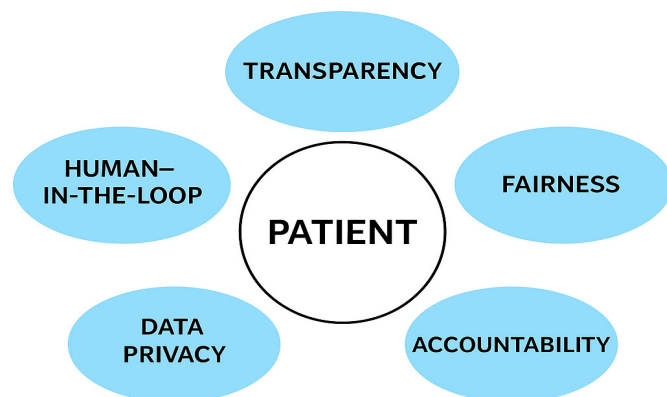


Figure 1. Core ethical principles of ai in radiology

DATA PROTECTION AND PRIVACY

Given that applications of AI in healthcare require the processing of highly sensitive medical data, data protection and privacy constitute one of the most fundamental components of the ethical framework. Because AI algorithms are typically trained on large volumes of patient data, such data must be appropriately anonymized or, at a minimum, processed in a purpose-limited and pseudonymized manner. The EU's GDPR incorporates core principles such as data subject consent, data minimization, accountability, and transparency.^{4,6}

In Türkiye, amendments to the KVKK in 2020 introduced certain exceptions for the use of health data in scientific research. Specifically, if data have been appropriately pseudonymized and the research serves the public interest, they may be used for scientific purposes without obtaining explicit consent from individuals, provided that approval from an ethics committee is obtained and adequate data protection measures are implemented.²¹

In radiology, the use of imaging data in AI projects necessitates more specific regulations. For instance, personal identifiers within the metadata of DICOM-format images must be removed, and any information such as patient names or identification numbers displayed on the images should be masked. Additionally, in three-dimensional imaging, body contours are often removed to prevent the possibility of personal identification.

Nevertheless, the contribution of data sharing to scientific advancement should not be overlooked. Organizations such as the ICMJE define the sharing of clinical research data as an ethical obligation, emphasizing that the accessibility of datasets used in AI research is essential for transparency and reproducibility.⁵

In conclusion, safeguarding personal data in AI research is not only a legal obligation but also an essential ethical responsibility to ensure research integrity, protect

participants' rights, and maintain public trust. Within this scope, data controllers must reinforce technical infrastructure, conduct regular audits of research processes, and comply with both national and international ethical guidelines.

PRIVACY OF IMAGING DATA AND MASKING METHODS

Medical imaging data not only contain clinical information but may also include elements that can directly or indirectly reveal an individual's identity. In DICOM-format images, header (metadata) fields may contain identifiers such as patient name, date, device, and protocol numbers; moreover, in scans of the facial region, re-identification can be achieved through pixel-level anatomical details.²²

Therefore, two levels of measures are necessary to ensure image privacy: metadata-level and pixel-level. At the metadata level, de-identification or pseudonymization techniques are used to either completely remove identifying fields or replace them with random identifiers.²³ At the pixel level, identifiers such as patient names, faces, or wristbands are rendered unrecognizable through blurring, masking, or regional deletion techniques.²⁴

However, in 3D images of the head and face, these methods may be insufficient, as facial contours can still be used for re-identification. To mitigate this risk, approaches such as defacing (face removal), algorithms that distort facial modeling, or segmentation-based automatic masking are recommended.²⁵ Recently developed machine learning-based systems aim to automatically conceal identity-related features while preserving clinical information.

Nonetheless, studies published in 2024 and 2025 have emphasized that despite the high accuracy rates of automated DICOM anonymization tools, human oversight remains indispensable in sensitive cases. According to the final report of the NIH's Medical Imaging De-Identification (MIDI) Project in June 2025, the ethical use of DICOM data should incorporate human-supervised processes alongside automated masking systems.²⁶

In conclusion, privacy in medical imaging data is not only a legal requirement but also an ethical necessity. Therefore, a combined approach involving metadata cleaning, pixel-level masking, 3D face removal, machine learning-based concealment systems, and human verification processes is recommended. Ethics committees should technically assess these procedures and document them in accordance with traceability principles, ensuring that AI-assisted image analysis is conducted both safely and within the legal framework.

DATA OWNERSHIP, SHARING, AND ETHICAL LIMITS OF COMMERCIAL USE

Data ownership is a central ethical issue, particularly in the context of health data used in AI applications, as it directly relates to the protection of individual rights. The ability of individuals to exercise control over their own

data is intrinsically linked to both privacy and the principle of autonomy. However, the replicable, reproducible, and shareable nature of digital data creates an ethical debate distinct from traditional concepts of property rights.²⁷

While the sharing of medical data for scientific research purposes generally offers a more ethically acceptable framework, data sharing with commercial entities requires far stricter oversight, transparency, and explicit consent mechanisms. The use of publicly available datasets for commercial purposes without the informed consent of data owners can lead to ethical concerns such as erosion of trust, risk of discrimination, and the commodification of personal information.²⁸ In particular, when large technology companies utilize patient data for financial gain in AI development processes, it raises questions about the boundary between public interest and market-driven benefits.

A prominent recent approach, the concept of data commons, advocates for the open sharing of data for societal benefit. However, even within this model, data ownership, intended use, access conditions, and sharing limitations must be clearly defined, ensuring adherence to principles of fairness and inclusivity.²⁹ Furthermore, in cases where data generate revenue, increasing attention is being paid to whether data owners are adequately informed and included in benefit-sharing arrangements.

In conclusion, the ethical boundaries of data ownership and sharing should be defined not only by legal frameworks but also by principles of societal trust, individual rights, and transparency. A careful balance must be struck between supporting scientific progress and safeguarding individual rights, with explicit consent, robust oversight, and strict adherence to ethical guidelines being essential—particularly in commercial use cases.

ALGORITHM ETHICS AND TRANSPARENCY IN DECISION-MAKING SYSTEMS

AI-assisted decision-making systems are increasingly used in healthcare, finance, and law, yet their “black box” nature and risk of bias raise significant ethical concerns. To address these, explainable AI (XAI) methods—such as feature importance ranking, decision tree surrogates, and SHAP—can enhance transparency and user understanding, while human-in-the-loop designs improve reliability and oversight.^{30,31}

Transparency also fosters trust and societal acceptance, as studies show it reduces opposition to AI and increases stakeholder confidence. However, the “right to explanation” recognized in frameworks like the GDPR has limited applicability to deep learning models, and excessive regulation may hinder innovation. A balanced strategy—combining independent audits, diverse training datasets, ethical standards, and robust transparency processes—is essential to ensure fairness and accountability (Figure 2).³²

In breast MRI, when a deep learning model predicting pCR was rendered explainable using SHAP, it was demonstrated that the decision primarily relied on early enhancement

patterns around the tumor; this explanation increased confidence in altering treatment intensity during the oncology board discussion.^{30,32}

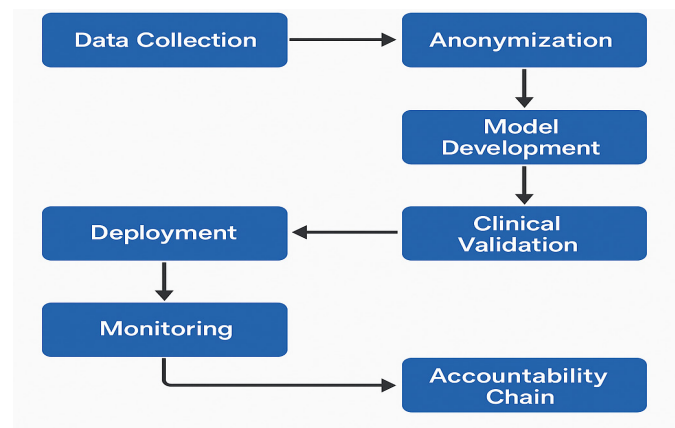


Figure 2. Ethical lifecycle of ai in radiology

ETHICAL DILEMMAS OF AI IN CLINICAL PRACTICE

In the practical implementation of AI-assisted clinical systems, ethical dilemmas often center on critical areas such as decision-support systems, algorithmic bias, patient privacy, and the allocation of responsibility. AI-CDSS tools have begun to be used in clinical processes such as resource allocation and care prioritization, which requires careful evaluation in terms of fairness, reliability, and patient trust.³³

Deep learning-based systems often operate as “black boxes,” with decision-making logic that may not be externally interpretable. This lack of transparency can undermine the confidence of both clinicians and patients, while also making it more difficult to identify the mechanisms underlying errors.

Another concern is that the datasets used to train AI systems may contain demographic imbalances or biased patterns. Such limitations can lead to misdiagnosis and inappropriate treatment planning, particularly for minority groups, thereby contributing to clinical inequities. Some dataset-based analyses have demonstrated higher error rates for non-white populations.³⁴

Informed consent is another important yet insufficiently addressed ethical issue in clinical scenarios involving AI. Patients may not fully understand how AI operates, what data it uses, or how it makes decisions, potentially undermining their right to autonomy. In addition, responsibility must be clearly defined. When AI contributes to an error, there must be explicit ethical and legal guidelines determining whether the developer, clinical personnel, or the institution is accountable.³³

A widely used U.S. healthcare algorithm disproportionately assigned lower risk scores to Black patients, limiting access to additional care; similar biases may threaten equity in radiology triage models.¹⁸

In a case where an AI-CDSS suggested chemotherapy modification, limited training data led to misclassification, underscoring the need for defined human-in-the-loop oversight and shared accountability.³³

To address these ethical challenges, three strategies stand out: continuous system monitoring and evaluation (silent evaluation), feedback loops incorporating clinician input, and the active involvement of multidisciplinary, multi-stakeholder ethics committees.³⁵

ETHICAL OVERSIGHT, REGULATIONS, AND THE SITUATION IN TURKIYE

The integration of AI into medicine necessitates a redefinition of ethical oversight and regulatory frameworks. The use of AI systems in decision-making processes introduces new dimensions—such as transparency, accountability, and data privacy—beyond the traditional ethical principles of autonomy, beneficence, non-maleficence, and justice.²

At the international level, several key documents, including the European Commission's Trustworthy AI ethics guidelines, the Asilomar Principles, and the Montreal Declaration, set standards for ensuring that AI is reliable, transparent, and human-centered.⁵ These frameworks emphasize that algorithms should not produce bias, decision-making processes should remain traceable, and human oversight must be maintained. The FDA has also established distinct regulatory models for "locked" and "adaptive" AI-based software, requiring both ethical and technical monitoring throughout the product life cycle.³⁶

In radiology, a multi-institutional international consortium's Ethics of AI in Radiology report addresses data ethics, algorithm auditing, and the ethical principles necessary for the clinical implementation of AI. This report underscores that humans should remain the ultimate decision-makers in clinical workflows, with ethical responsibility shared among developers, implementers, and decision-makers.⁹

In Turkiye, the KVKK provides the fundamental legal framework for the collection, processing, and sharing of health data. Under this law, anonymization of imaging data, establishment of dynamic consent procedures, and obtaining informed consent for data sharing are mandatory. However, AI-specific ethics committees or independent regulatory frameworks are still in early stages of development. Some university hospitals have nonetheless introduced dedicated review protocols for AI-based clinical research within their ethics committees.

In clinical use, an adaptive AI software exhibited performance drift following a protocol change, prompting notification to the manufacturer; in line with the FDA's lifecycle approach, recalibration with real-world data and post-market surveillance were initiated.³⁶

At a university hospital, the ethics committee began requiring an additional 'AI methodology appendix'—including data flow, anonymization, XAI outputs, and human-in-the-loop plans—for clinical studies involving AI.⁹

Turkiye's regulatory framework, particularly the KVKK and the 2024 Draft AI Law, shows substantial parallels with international instruments. Similar to the GDPR, the KVKK emphasizes consent, anonymization, and accountability

in health data use.¹⁶ Likewise, the draft law adopts a risk-based classification approach comparable to the EU AI Act.³⁷ However, while EU frameworks provide institutionalized oversight through dedicated AI governance bodies, Turkiye's mechanisms remain in early development, with ethics committees at university hospitals serving as the primary oversight structures. In contrast to global initiatives such as UNESCO's 2021 Recommendation on the Ethics of AI and the FUTURE-AI framework, which prioritize universal values and human rights, Turkiye's regulations currently place stronger emphasis on data protection and compliance.⁸

Looking ahead, establishing a common ethical language at both national and international levels will be crucial for the safe and equitable integration of AI. For Turkiye, aligning with this evolving ethical infrastructure will require the creation of independent oversight bodies encompassing both medical and engineering expertise.

FUTURE DIRECTIONS AND ETHICAL ROADMAP

For the safe integration of AI technologies into healthcare systems, a dynamic and continuously updated roadmap—rather than static principles—should be established. From the model development stage onward, ethical design principles must systematically incorporate core values such as transparency, fairness, accountability, and human oversight.

At the international level, the FUTURE-AI ethical framework is structured around six core principles—Fairness, Universality, Traceability, Usability, Robustness, and Explainability—and outlines 28 best practices covering the design, validation, regulation, and field implementation of medical AI systems.⁸ This model promotes reliable system development through risk-based assessments.

Recommendations from global organizations such as the United Nations (UN) and UNESCO emphasize the adoption of human rights-based, transparent, and accountable systems. The UN advisory group has proposed initiatives such as a "global AI standards exchange platform," an "international capacity-building network," and a "global AI data framework".² UNESCO's recommendations highlight algorithmic transparency, anti-discrimination system design, and accountable governance frameworks as key ethical action areas.¹⁶

In Turkiye, the Draft AI Law prepared in 2024 classifies systems by risk level, mandates the registration of high-risk AI applications, and requires the preservation of algorithmic transparency and human oversight.³⁷ Additionally, the National AI Strategy—which includes objectives such as developing domestic language models, promoting green AI solutions, and enabling the secure sharing of public data—aims to strengthen ethical governance.³⁸

The proposed ethical roadmap is summarized in [Table](#). This roadmap offers a sustainable framework to ensure that AI systems are developed not only with technical reliability but also with ethical soundness and a human-centered approach.

Table. Proposed ethical roadmap and implementation phases for AI systems

Phase	Recommended ethical measures
Design & development	Ethical-by-design approach, algorithmic impact assessment, bias testing
Regulation & compliance	AI system classification, conformity assessment for high-risk systems
Institutional oversight	AI-specific ethics committees, independent audit mechanisms
Social engagement	Stakeholder-based decision-making processes, collaboration with international ethics networks
Education & capacity	Integration of clinical and ethical training, performance and awareness education
Monitoring & updating	Continuous observation, adaptive feedback systems, local and global compliance monitoring

AI: Artificial Intelligence

CONCLUSION

The rapid rise of AI in radiology is shaped not only by technological advancements but also by the need for robust, multi-layered ethical approaches. Integrating AI into clinical decision-making requires reinterpreting principles such as transparency, explainability, fairness, and accountability so that they offer practical guidance in addressing issues like data privacy, algorithmic bias, responsibility allocation, and social justice. Ethical governance should place human-centered design at its core, ensuring that clinicians retain ultimate responsibility for patient care. A human-in-the-loop approach can preserve clinical autonomy, support professional accountability, and maintain system explainability.

In Türkiye, where AI-related regulations are still emerging, national strategies should be strengthened by interdisciplinary collaboration and the formation of independent ethical oversight structures. Governance models must be inclusive, engaging not only technical experts but also ethics committees, clinicians, patients, and the wider public.

Ultimately, the safe, fair, and effective use of AI in radiology will depend on more than algorithmic accuracy. Embedding ethical principles into the foundation of AI system design and deployment is essential to ensure that these technologies advance societal benefit, safeguard patient interests, and respect human dignity.

ETHICAL DECLARATIONS

Peer Review Process

This review was externally peer-reviewed.

Conflict of Interest

The author declare no conflicts of interest.

Financial Disclosure

No financial support was received for the preparation or publication of this article.

Author Contributions

The author is solely responsible for the conception, data collection, analysis, and writing of this manuscript.

REFERENCES

- Will AI replace doctors who read X-rays, or just make them better than ever? | AP News. <https://apnews.com/article/ai-algorithms-chatgpt-doctors-radiologists-3bc95db51a41469c390b0f1f48c7dd4e>
- Geis JR, Brady AP, Wu CC, et al. Ethics of Artificial Intelligence in radiology: summary of the joint European and North American multisociety Statement. *Can Assoc Radiol J.* 2019;70(4):329-334. doi:10.1016/j.carj.2019.08.010
- Brady AP, Neri E. Artificial Intelligence in radiology-ethical considerations. *Diagnostics (Basel).* 2020;10(4):231. doi:10.3390/diagnostics10040231
- D'antonoli TA. Ethical considerations for Artificial Intelligence: an overview of the current radiology landscape. *Diagn Interv Radiol.* 2020; 26(5):504-511. doi:10.5152/dir.2020.19279
- Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell.* 2019;1(9):389-399. doi:10.1038/s42256-019-0088-2
- Park SH. Ethics for artificial intelligence: focus on the use of radiology images. *J Korean Soc Radiol.* 2022;83(4):759-770. doi:10.3348/jksr.2022.0036
- Intelligence. HLEG on A. Ethics guidelines for trustworthy AI- Publications Office of the EU. 2019:39.
- Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable Artificial Intelligence in healthcare. *BMJ.* 2025;388:e081554. doi:10.1136/bmj-2024-081554
- RANZCR. Ethical Principles for AI in Medicine RANZCR 2023.
- Lawrence R, Dodsworth E, Massou E, et al. Artificial Intelligence for diagnostics in radiology practice: a rapid systematic scoping review. *EClinical Medicine.* 2025;83:103228. doi:10.1016/j.eclinm.2025.103228
- Nair A, Ong W, Lee A, et al. Enhancing radiologist productivity with Artificial Intelligence in magnetic resonance imaging (MRI): a narrative review. *Diagnostics (Basel).* 2025;15(9):1146. doi:10.3390/diagnostics15091146
- Lu D, Yan Y, Jiang M, et al. Predictive value of radiomics-based machine learning for the disease-free survival in breast cancer: a systematic review and meta-analysis. *Front Oncol.* 2023;13:1173090. doi:10.3389/fonc.2023.1173090
- Jorg T, Halfmann MC, Stoehr F, et al. A novel reporting workflow for automated integration of Artificial Intelligence results into structured radiology reports. *Insights Imaging.* 2024;15(1):80. doi:10.1186/s13244-024-01660-5
- Sun Z, Ong H, Kennedy P, et al. Evaluating GPT-4 on impressions generation in radiology reports. *Radiology.* 2023;307(5):e231259. doi:10.1148/radiol.231259
- Meşe İ, Altıntaş Taşlıçay C, Kuzan BN, Kuzan TY, Sivrioğlu AK. Educating the next generation of radiologists: a comparative report of ChatGPT and e-learning resources. *Diagn Interv Radiol.* 2024;30(3): 163-174. doi:10.4274/dir.2023.232496
- Ramos G. UNESCO's recommendation on the ethics of Artificial Intelligence (2021). In: *International Organization Initiatives.* 2025:203-214. doi:10.1093/9780197803325.003.0020
- The Council of Europe. The framework convention on Artificial Intelligence. Council of Europe Treaty Series-no. 225.
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. doi:10.1126/science.aax2342
- Floridi L, Cowls J, Beltrametti M, et al. An ethical framework for a good AI Society: opportunities, risks, principles, and recommendations. In: *Philosophical Studies Series.* Vol 144. Springer Netherlands. 2021:19-39. doi:10.1007/978-3-030-81907-1_3
- Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: an overview of interpretability of machine learning. In: *Proceedings-2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018.* Institute of Electrical and Electronics Engineers Inc.; 2018:80-89. doi:10.1109/DSAA.2018.00018
- Özer Deniz M. Sağlık verilerinin işlenmesinde aydınlatma yükümlülüğü. *Erciyes Akad.* 2022;36(3):1424-1445. doi:10.48070/erciyesakademi.1150085
- Aiello M, Esposito G, Pagliari G, Borrelli P, Brancato V, Salvatore M. How does DICOM support big data management? Investigating its use in medical imaging community. *Insights Imaging.* 2021;12(1):1-21. doi:10.1186/s13244-021-01081-8

23. Aryanto KYE, Oudkerk M, van Ooijen PMA. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy. *Eur Radiol.* 2015;25(12):3685. doi:10.1007/S00330-015-3794-0
24. Kondylakis H, Catalan R, Alabart SM, et al. Documenting the de-identification process of clinical and imaging data for AI for health imaging projects. *Insights Imaging.* 2024;15(1):1-8. doi:10.1186/s13244-024-01711-x
25. Van der Goten LA, Smith K. Privacy protection in MRI scans using 3D masked autoencoders. In: Lecture Notes in Computer Science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol 15007 LNCS. Springer, Cham; 2024:583-592. doi:10.1007/978-3-031-72104-5_56
26. Rempe M, Heine L, Seibold C, Hörst F, Kleesiek J. De-identification of medical imaging data: a comprehensive tool for ensuring patient privacy. *Eur Radiol.* 2025:1-10. doi:10.1007/s00330-025-11695-x
27. Hummel P, Braun M, Dabrock P. Own data? Ethical reflections on data ownership. *Philos Technol.* 2021;34(3):545-572. doi:10.1007/s13347-020-00404-9
28. Winkler EC, Jungkunz M, Thorogood A, Lotz V, Schickhardt C. Patient data for commercial companies? An ethical framework for sharing patients' data with for-profit companies for research. *J Med Ethics.* 2023; 51(5):337-344. doi:10.1136/jme-2022-108781
29. Søre SO, Mai JE. The ethics of sharing: privacy, data, and common goods. *Digit Soc.* 2023;2(2):1-20. doi:10.1007/s44206-023-00057-z
30. Cheong BC. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Front Hum Dyn.* 2024;6:1421273. doi:10.3389/fhumd.2024.1421273
31. Wagner B, Kuebler J, Zalnieriute M. Editorial: humans in the loop: exploring the challenges of human participation in automated decision-making systems. *Front Polit Sci.* 2025;7:1611563. doi:10.3389/fpos.2025.1611563
32. Park K, Young Yoon H. AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. *Humanit Soc Sci Commun.* 2025;12(1):1-13. doi:10.1057/s41599-025-05116-z
33. Elgin CY, Elgin C. Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: a qualitative study of healthcare professionals' perspectives. *BMC Med Ethics.* 2024;25(1):1-15. doi:10.1186/s12910-024-01151-8
34. Gallifant J, Bitterman DS, Celi LA, et al. Ethical debates amidst flawed healthcare Artificial Intelligence metrics. *NPJ Digit Med.* 2024;7(1):1-3. doi:10.1038/s41746-024-01242-1
35. Collins BX, Belisle-Pipon JC, Evans BJ, et al. Addressing ethical issues in healthcare artificial intelligence using a lifecycle-informed process. *JAMIA Open.* 2024;7(4):108. doi:10.1093/jamiaopen/ooae108
36. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device. FDA. Published January 6, 2025. Accessed March 7, 2026.
37. Kosterit SD. AI Policy Corner: The Turkish Artificial Intelligence Law Proposal. Montreal AI Ethics Institute. Published March 17, 2025. Accessed March 7, 2026.
38. Daily Sabah. Intelligence academy urges AI-backed roadmap for Türkiye. Daily Sabah. Published June 27, 2025. Accessed March 7, 2026.

The role of ADC value and ADC ratio in the diagnosis and prognostic evaluation of prostate cancer

 **Abdullah Enes Berksoy***¹,  **Mehmet Cihan Karacaoğlu**²

¹Department of Radiology, Göksun State Hospital, Kahramanmaraş, Türkiye
²Department of Radiology, Ankara Etlik City Hospital, Ankara, Türkiye

Received: 31.08.2025

Accepted: 22.10.2025

Published: 09.03.2026

Cite this article: Berksoy AE, Karacaoğlu MC. The role of ADC value and ADC ratio in the diagnosis and prognostic evaluation of prostate cancer. *J Radiol Med.* 2026;3(1):22-27.

*Corresponding Author: Abdullah Enes Berksoy, enesberksoy513@gmail.com

ABSTRACT

Prostate cancer is the second most common malignancy in men after lung cancer and remains a major cause of cancer-related mortality. Although prostate-specific antigen (PSA) testing has reduced mortality, its low specificity leads to overdiagnosis and unnecessary biopsies. Multiparametric magnetic resonance imaging (mpMRI) has advanced prostate cancer evaluation by integrating anatomical, functional, and vascular imaging. Diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) measurements provide objective markers of tumor cellularity, which correlate with Gleason score and International Society of Urological Pathology Grade Group. To summarize the diagnostic and prognostic role of absolute ADC values and the ADC ratio in prostate cancer. A narrative review of the literature from 2006 to 2025 was conducted through PubMed and Scopus using terms related to prostate cancer, DWI, ADC, ADC ratio, and PSA density. Thirty references were included, ranging from early technical reports to recent meta-analyses and guideline-based studies. Absolute ADC values decrease with increasing tumor aggressiveness but are limited by technical and patient-related variability. The ADC ratio, calculated as the lesion value relative to normal prostate tissue, reduces variability and improves diagnostic accuracy, particularly in Prostate Imaging Reporting and Data System (PI-RADS) category 3 lesions. When combined with PSA density, it increases accuracy for detecting clinically significant cancer and decreases unnecessary biopsies. ADC and the ADC ratio are promising non-invasive imaging biomarkers that improve lesion characterization, biopsy selection, and prognostic evaluation in prostate cancer. Future integration with artificial intelligence, radiomics, and radiogenomics may further enhance personalized patient management.

Keywords: Prostatic neoplasms, diffusion magnetic resonance imaging, apparent diffusion coefficient

INTRODUCTION

Prostate cancer is one of the most common malignancies in men worldwide and represents a major cause of cancer-related deaths.¹ Although prostate-specific antigen (PSA) screening, which is widely used in developed countries, has partially reduced mortality, its low specificity has led to overdiagnosis and unnecessary biopsies.² Therefore, more reliable, non-invasive, and prognostically valuable imaging methods have come to the forefront in diagnosis.

In the past decade, multiparametric magnetic resonance imaging (mpMRI) has emerged as a groundbreaking method in the diagnosis and management of prostate cancer.³ By combining T₂-weighted imaging, diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCE-MRI), mpMRI provides a comprehensive assessment of

the anatomical, functional, and vascular characteristics of the prostate. Furthermore, the Prostate Imaging Reporting and Data System (PI-RADS) guideline, developed for the standardization of mpMRI findings, has become an important reference point in clinical practice.⁴

In particular, DWI and its derivative apparent diffusion coefficient (ADC) values are quantitative parameters reflecting the microstructural properties of prostate tissue. ADC values are inversely correlated with tumor cellularity and show a significant reduction in malignant prostate lesions.^{5,6} Studies have demonstrated that ADC measurements are reliable biomarkers in differentiating benign from malignant lesions.⁷



However, absolute ADC measurements may vary due to scanner differences, magnetic field strength, selected b-values, and patient-related factors.⁸ For this reason, the ADC ratio parameter has been developed in recent years. Calculated as the ratio of the lesion ADC value to the ADC value of normal prostate tissue in the same patient, ADC ratio reduces the effects of individual differences and technical variability, thus providing more reliable results.⁹

Preliminary evidence suggests that the ADC ratio may improve diagnostic accuracy and help reduce unnecessary biopsies.¹⁰⁻¹² Furthermore, its potential integration with established clinical parameters like PSA density (PSAD) is an area of active investigation, with the goal of creating stronger, multi-parametric predictors for clinically significant cancer.¹⁰⁻¹² In this context, ADC and ADC ratio measurements not only provide diagnostic accuracy but also hold clinical importance in prognostic prediction and treatment response evaluation.^{13,14}

THE DIAGNOSTIC AND PROGNOSTIC ROLE OF ADC VALUE AND ADC RATIO IN PROSTATE CANCER

Basic Principles and Image Formation

DWI is a functional MRI technique that captures the Brownian motion of water molecules within tissues. Tissue microstructure-including cellular density, membrane integrity, and extracellular matrix architecture-modulates water mobility and thus the DWI signal.⁵ In malignant prostate tissue, increased cellularity and reduced extracellular space produce restricted diffusion, manifesting as a decrease in the ADC.⁶ In practice, DWI acquired with multiple b-values enables quantification of signal attenuation; pixel-wise monoexponential fitting yields ADC maps that allow quantitative assessment of intratissue water diffusion.^{7,8}

Practical Measurement Considerations

Accurate ADC quantification requires standardized acquisition and analysis:

- **Field strength and coils:** High-field MRI with optimized receiver coils, including endorectal coils when appropriate, improves signal-to-noise ratio and spatial resolution for lesion detection and characterization.⁷
- **Region-of-interest (ROI) strategy:** ADC can be measured as minimum, mean, or percentile values using whole-lesion segmentation. Whole-lesion histogram analysis provides a more reliable assessment of intratumoral heterogeneity than small, single ROIs.⁶
- **Quality control:** Geometric distortion, susceptibility to hemorrhage or post-biopsy changes, and motion artifacts must be recognized and minimized during interpretation.^{7,12}

Pathophysiology, Malignancy, and Quantitative Behavior

In the peripheral zone, malignant lesions typically have lower ADC values than benign tissue due to higher cell density and reduced extracellular space.^{6,9} Several studies confirm that prostate cancer shows significantly reduced ADC compared with benign conditions, supporting its role as a biomarker of malignancy.^{10,11} Moreover, ADC correlates with histopathology: minimum ADC values decrease with

increasing Gleason grade, demonstrating the biological link between diffusion restriction and tumor aggressiveness.^{5,6}

CLINICAL APPLICATIONS OF ABSOLUTE ADC

Lesion Characterization

Across benign mimickers-such as prostatitis, glandular hyperplasia, and stromal nodules-ADC values are generally higher than in cancer; by contrast, cancer foci demonstrate a marked ADC decrease.^{7,10} Nonetheless, atrophy and biopsy-related fibrosis can also restrict diffusion, so ADC must be integrated with T₂-weighted morphology and dynamic contrast-enhanced MRI (DCE-MRI) for robust characterization.¹²

Risk Stratification and Prognosis

ADC metrics track with histologic grade: increasing Gleason score and higher ISUP Grade Groups are associated with lower minimum ADC values.^{6,13} Pooled evidence indicates that ADC contributes prognostic information for grade upgrading and clinically significant disease.^{8,14} Quantitative ADC also relates to tumor volume and burden, supporting its role in disease assessment beyond binary detection.¹⁵

Treatment Response Assessment

Post-radiation, ADC tends to rise in treated cancers consistent with reduced cellularity, supporting its use as a response biomarker and potential surveillance parameter after therapy.¹⁶ After radiotherapy, DCE-MRI findings correlate with biopsy results, highlighting the value of multiparametric follow-up in which ADC serves as a key quantitative component.¹⁷

ADC Ratio: Rationale, Definition, and Advantages

Definition: The ADC ratio is the quotient of the lesion's ADC divided by the ADC of normal-appearing prostate tissue in the same patient.⁹ By normalizing to an internal reference, the ADC ratio mitigates interscan and interindividual variability introduced by scanner hardware, field strength, chosen b-values, and patient-related heterogeneity-limitations that challenge absolute ADC harmonization across centers.^{9,10,19} Through internal normalization, ADC ratio often yields more reliable, comparable measurements, especially in multi-institutional or longitudinal contexts.^{9,10}

How to Measure ADC Ratio in Practice

- **Internal reference placement:** Use normal-appearing tissue in the same zone (e.g., contralateral peripheral zone at a comparable slice level) to avoid zonal heterogeneity as a confounder.^{9,10}
- **Lesion sampling:** Prefer whole-lesion or carefully placed ROIs that capture the most restricted areas to avoid partial-volume effects and to reflect clinically relevant tumor biology.^{6,9,10}
- **Reporting:** Document the ROI strategy (size, location), the reference site, and summary statistics (minimum/mean ADC and the derived ADC ratio) alongside acquisition parameters, to facilitate reproducibility and longitudinal comparison.^{6,9,12}

ADC Ratio in PI-RADS 3 Lesions

PI-RADS 3 denotes indeterminate likelihood of clinically significant cancer and is a frequent clinical dilemma. Here, ADC ratio has shown particular value: multiple studies report that ADC ratio can outperform absolute ADC in differentiating benign from malignant PI-RADS 3 lesions, likely due to its robustness against interpatient and interscanner variability.^{10,11,12} Within a comprehensive mpMRI framework, integrating ADC ratio with T₂-weighted morphology and perfusion behavior supports more confident decisions about targeted biopsy versus surveillance.^{11,12,18,19}

Integration with the PI-RADS Framework

Quantitative measures can refine PI-RADS categorization when judiciously applied. Mean ADC quantification has been shown to improve PI-RADS v2 categorization and validation of PI-RADS v2.1 demonstrates solid reader agreement and diagnostic performance when standardized interpretation is maintained—an environment where internally normalized metrics like ADC ratio are naturally synergistic.^{18,19}

Pitfalls and Confounders

- **Benign entities with low ADC:** Prostatitis, fibrosis, and post-biopsy changes may mimic cancer by restricting diffusion; correlation with T₂ and DCE is essential.^{7,10,12}
- **Heterogeneity:** Focal necrosis, mucinous components, or hemorrhage can alter ADC distribution; whole-lesion analysis helps avoid sampling bias.⁶
- **Technical variability:** Differences in b-values, echo times, susceptibility artifacts, and field inhomogeneities all influence ADC; normalization by ADC ratio reduces, but does not eliminate, these effects.^{9,10,19}

Practical Reporting Recommendations

- **Acquisition summary:** Note field strength, coil configuration, and diffusion scheme (multiple b-values).⁷
- **Lesion quantification:** Provide minimum and/or mean ADC and (where available) ADC histogram descriptors for the index lesion.^{6,12,20}
- **Normalization:** Report an ADC ratio using same-zone normal tissue as reference and describe ROI placement.^{9,10}
- **Contextual interpretation:** Integrate T₂-weighted morphology and DCE kinetics, especially for PI-RADS 3 lesions, and state how quantitative metrics influenced management recommendations (e.g., biopsy targeting).^{11,12,18,19}
- **Longitudinal use:** In post-radiotherapy follow-up, track ADC trends as part of response assessment, recognizing that rising ADC generally aligns with reduced cellularity.^{16,17}

Limitations and Future Directions

Although absolute ADC and ADC ratio provide biologically meaningful, reproducible markers of malignancy and grade, heterogeneity persists across studies due to differences in acquisition protocols, ROI strategies, and reference-tissue

selection. Meta-analytic evidence supports diagnostic and prognostic utility, yet also highlights the need for standardized acquisition/analysis pipelines and prospective validation in PI-RADS 3 decision-making.^{8,14} Continued work on robust normalization schemes and automated whole-lesion analytics may further enhance reproducibility and clinical adoption.^{6,9,12,14}

The comparative features, advantages, and limitations of absolute ADC versus ADC ratio are summarized in **Table 1**.

Table 1. Comparison of definitions, advantages, and clinical values of absolute ADC vs. ADC ratio

Feature	Absolute ADC	ADC ratio
Definition	Direct measurement of the lesion's ADC value	Ratio of lesion ADC value to normal prostate tissue ADC value
Advantages	-Simple and rapid measurement -Can be directly calculated by most software-widely used in the literature	-Reduces inter-scanner variability -Minimizes individual differences within the same patient -More reliable for multi-center studies -Provides greater diagnostic contribution in PI-RADS 3 lesions
Disadvantages	-Influenced by scanner differences, magnetic field strength, and b-values -Patient-related factors (edema, hemorrhage, inflammation) may alter results	-Variability in reference region selection -Prostate zonal heterogeneity may cause challenges -Lack of standardized cut-off values in the literature
Clinical value	-Strong correlation with tumor aggressiveness -Decreases consistently with higher Gleason scores	-Improves diagnostic accuracy in PI-RADS 3 lesions -Reduces need for biopsy when combined with PSA density
Future perspective	-Limited role due to standardization issues	-Expected to become more prominent in multi-center studies and AI-assisted analyses

ADC: Apparent diffusion coefficient, PI-RADS: Prostate Imaging Reporting and Data System, PSA: Prostate-specific antigen, AI: Artificial Intelligence

Combination with PSA Density

PSAD, defined as the serum PSA level normalized to prostate volume, is a widely used clinical parameter for prostate cancer risk stratification and is particularly valuable in the assessment of indeterminate mpMRI findings such as PI-RADS 3 lesions. Although direct evidence evaluating the combined use of ADC ratio and PSAD is limited, the available literature suggests that integrating quantitative diffusion metrics with established clinical markers may enhance overall diagnostic confidence.

Studies have repeatedly demonstrated the prognostic and diagnostic utility of ADC-based measurements—including ADC ratio—as robust biomarkers of tumor cellularity and aggressiveness.⁹⁻¹⁴ Within this framework, PSAD offers complementary biological information that reflects glandular volume and PSA kinetics. Therefore, combining ADC-derived parameters with PSAD holds conceptual promise for improving risk stratification, guiding biopsy decisions more effectively, and potentially reducing unnecessary interventions; however, prospective validation is still required.

By decreasing the probability of unnecessary biopsies and focusing diagnostic work-up on patients at higher risk of clinically significant prostate cancer, such multiparametric approaches may ultimately contribute to more cost-effective care pathways. Formal cost-effectiveness studies and guideline-level recommendations regarding the integration of ADC ratio and PSAD have yet to be established, and further multicenter investigations will be essential to define their optimal role.

Limitations

Despite the growing interest in ADC ratio as a quantitative imaging biomarker, several limitations must be acknowledged.

- **Reference region variability:** There is no uniform consensus regarding the optimal site for measuring “normal” prostate tissue when calculating the ADC ratio. Different studies have used contralateral normal-appearing peripheral zone, transition zone, or whole-gland reference regions, which may introduce variability in ratio calculations.^{9,10}
- **Zonal heterogeneity:** ADC values differ between prostate zones due to intrinsic microstructural and cellular differences. As a result, ADC and ADC-ratio thresholds derived from peripheral-zone lesions may not be directly applicable to transition-zone tumors.²⁰
- **Heterogeneous cut-off values:** Thresholds proposed for both absolute ADC and ADC ratio vary considerably across studies, partly due to differences in scanners, field strengths, b-value schemes, and ROI strategies.^{12,14} This heterogeneity underscores the need for standardized acquisition protocols and large-scale prospective validation before ADC ratio can be implemented uniformly in clinical practice.

Clinical Applications

ADC and ADC ratio are among the most valuable quantitative parameters guiding clinical decision-making in the diagnosis and prognosis of prostate cancer today. Their clinical applications can be summarized as follows:

- **Differentiation of benign and malignant lesions:** ADC values play a critical role in distinguishing prostate cancer from benign conditions. Prostatitis and benign prostatic hyperplasia usually show higher ADC values, whereas malignant lesions present with marked decreases.^{7,10} However, fibrotic changes or atrophy may also cause restricted diffusion, so ADC measurements should always be interpreted together with T₂-weighted and DCE-MRI findings.¹²
- **Prediction of tumor aggressiveness:** There is a strong relationship between ADC measurements and histopathological aggressiveness. Meta-analysis and quantitative studies have confirmed that ADC is a reliable biomarker in predicting tumor aggressiveness.^{14,20} ADC ratio, compared with absolute ADC, stands out as a more stable parameter. Lower ADC (and, in some studies, ADC

ratio) values have been significantly associated with higher Gleason scores.^{13,21} Therefore, ADC ratio can be used as an additional prognostic tool for early identification of aggressive tumors.

- **Determining biopsy indications:** Adding ADC ratio increases diagnostic accuracy in detecting clinically significant prostate cancer.^{11,22} Its combination of mpMRI scores and PSAD reduces unnecessary biopsy rates and makes significant contributions to patient management.^{14,23}
- **Active surveillance and treatment response monitoring:** ADC measurements can be used to predict tumor progression during active surveillance of low-risk prostate cancer. Progressive decreases in ADC may indicate worsening tumor biology. In addition, increases in ADC values after radiotherapy and androgen deprivation therapy have been shown to correlate with decreased tumor cellularity.¹⁵⁻¹⁷
- **Potential impact on clinical guidelines:** Recent evidence suggests that ADC and ADC ratio, particularly when combined with PSA density, may be integrated into clinical guidelines.^{24,25} This approach may help base biopsy indications on more objective data, providing clinicians with a reliable roadmap.

A structured overview of the diagnostic, prognostic, and clinical roles of absolute ADC and ADC ratio is provided in **Table 2**.

Clinical Application	Role of absolute ADC	Role of ADC ratio
Differentiation between benign and malignant lesions	Markedly reduced ADC values in malignant lesions	More reliable than absolute ADC in distinguishing benign from malignant lesions
Prediction of tumor aggressiveness	-Minimum ADC values -Correlates with Gleason score and ISUP Grade Group	-Lower ADC ratio - Shows a stronger association with higher Gleason scores
Biopsy indication	Limited contribution in clinical practice	Supports biopsy decision in PI-RADS 3 lesions, reducing unnecessary biopsies
Active surveillance	Decrease in ADC may indicate progression	May represent a more stable and reliable parameter for follow-up
Assessment of treatment response	Increase in ADC after radiotherapy indicates treatment response	Currently limited data, but promising for the future due to normalization advantage
Integration into guidelines	Currently used indirectly	High potential for integration into clinical guidelines in combination with PSA density

ADC: Apparent diffusion coefficient, ISUP: International Society of Urological Pathology, PI-RADS: Prostate Imaging Reporting and Data System, PSA: Prostate-specific antigen

FUTURE PERSPECTIVES

Future applications of ADC values and ADC ratio in prostate cancer are expected to be strongly influenced by developments in quantitative imaging, artificial intelligence, and personalized medicine. Although these parameters already demonstrate diagnostic and prognostic value, several

domains require further exploration and validation before full clinical adoption.

Artificial Intelligence and Automated Measurements

Manual ROI placement is one of the main reasons for variability in ADC analysis. Differences in observer experience and lesion sampling strategies can lead to inconsistent results, especially in heterogeneous tumors. The integration of AI-based algorithms promises to automate ROI selection and standardize ADC quantification, thereby reducing interobserver variability. Initial studies integrating ADC maps into multiparametric machine learning pipelines have shown improved lesion classification and diagnostic accuracy.^{12,26,28} AI tools may also provide real-time biopsy guidance by highlighting regions with lowest ADC or abnormal ADC ratio, potentially redefining workflow in both diagnosis and active surveillance.

Radiomics and Multiparametric Integration

Conventional ADC reporting often relies on mean or minimum values, which may fail to capture intratumoral heterogeneity. Radiomics can extract a broad range of texture, shape, and histogram-based features from ADC maps, offering a more comprehensive representation of tumor biology. Several studies have demonstrated that radiomic models integrating ADC features with T₂-weighted or DCE-MRI improve the discrimination between clinically significant and insignificant prostate cancer.^{12,26,28} In the future, radiomics may not only refine PI-RADS classification but also contribute to individualized treatment planning, for example by predicting response to focal therapy or radiotherapy.

Radiogenomics and Molecular Correlations

Linking imaging features to genomic alterations is a promising field. Quantitative ADC metrics and ADC ratios may correlate with molecular subtypes, proliferation markers, and genetic signatures of aggressiveness. Early studies suggest that ADC-based measures can predict biochemical recurrence and treatment outcomes after radical prostatectomy.^{13,30} Radiogenomics could enable a “virtual biopsy,” allowing non-invasive assessment of tumor heterogeneity and guiding targeted treatment without relying exclusively on tissue sampling. In the long term, radiogenomic signatures combining ADC features with genomic risk scores may offer powerful prognostic tools in precision oncology.

Standardization and Multicenter Validation

One of the critical challenges is methodological heterogeneity. Scanner hardware, magnetic field strength, diffusion encoding (b-values), and ROI strategies lead to wide variability in ADC values across institutions. Current meta-analyses emphasize the lack of standardized cut-offs for distinguishing benign from malignant lesions or for predicting aggressiveness.^{8,14,29} Future research should focus on multicenter prospective trials with harmonized acquisition protocols, automated segmentation pipelines, and clearly defined reference tissues. Standardization will be essential for integrating ADC ratio into international guidelines, especially for PI-RADS 3 lesions where decision-making remains a clinical dilemma.

Clinical Workflow and Decision Support

In the next decade, ADC-derived metrics are likely to become part of multiparametric decision-making tools that combine imaging biomarkers with clinical and laboratory data. The prospective integration of quantitative MRI biomarkers like the ADC ratio with clinical data such as PSA density, Gleason score, and ISUP Grade Group holds the potential to refine integrated risk calculators.^{14,19,22} Such multi-parametric models could ultimately optimize biopsy decisions and reduce overtreatment, though this requires validation in future clinical trials. In active surveillance, monitoring ADC dynamics over time may allow early detection of tumor progression, enabling timely therapeutic intervention. Such integrative models could also be embedded into clinical information systems, supporting radiologists and urologists in daily practice.

Expanded Roles Beyond Diagnosis

ADC and ADC ratio are no longer limited to initial diagnosis. Evidence suggests they are valuable in treatment monitoring: post-radiotherapy and post-hormonal therapy, rising ADC values reflect reduced tumor cellularity and successful response.^{16,17,27} Furthermore, combining ADC with perfusion parameters from DCE-MRI may improve early prediction of therapeutic outcomes. Another promising area is the use of absolute ADC as a prognostic biomarker for biochemical recurrence after radical prostatectomy.³⁰ In the future, these parameters may guide selection of patients for focal therapies, immunotherapies, or novel targeted agents, broadening their role well beyond traditional diagnosis.

EXPANDED OUTLOOK

The future role of ADC and ADC ratio in prostate cancer management is likely to evolve from simple quantitative markers into multifunctional imaging biomarkers, integrated within AI-driven platforms, radiomics pipelines, and radiogenomic models. To realize this potential, rigorous standardization, multicenter validation, and incorporation into clinical guidelines will be essential. Ultimately, these developments may establish ADC-derived metrics as indispensable tools for personalized, non-invasive, and cost-effective prostate cancer care.

CONCLUSION

ADC values and ADC ratio represent promising quantitative biomarkers for the diagnosis and prognostic assessment of prostate cancer. By improving lesion characterization and supporting biopsy decisions—particularly in PI-RADS 3 cases—they have the potential to enhance clinical decision-making and reduce unnecessary interventions. Their association with tumor aggressiveness also highlights their role in prognostic evaluation and treatment monitoring.

Nevertheless, challenges such as variability in imaging protocols, lack of standardized thresholds, and zonal heterogeneity within the prostate remain barriers to routine clinical adoption. Addressing these limitations through prospective multicenter studies with harmonized methodology will be essential.

Looking ahead, integration with artificial intelligence, radiomics, and radiogenomics is expected to further strengthen the value of ADC-based metrics, paving the way for more objective and personalized prostate cancer management.

ETHICAL DECLARATIONS

Peer Review Process

This review was externally peer-reviewed.

Conflict of Interest

The authors declare no conflicts of interest.

Financial Disclosure

No financial support was received for the preparation or publication of this article.

Author Contributions

Conception/Planning: AEB, MCK; Analysis/Interpretation: AEB, MCK; Writing: AEB, MCK; Review and Editing AEB, MCK; Approval: AEB, MCK.

REFERENCES

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
- Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009; 360(13):1320-1328. doi:10.1056/NEJMoa0810084
- Turkbey B, Pinto PA, Choyke PL. Imaging techniques for prostate cancer: implications for focal therapy. *Nat Rev Urol*. 2009;6(4):191-203. doi:10.1038/nrurol.2009.27
- 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. doi: 10.1016/j.eururo.2019.02.033
- Hambrock T, Somford DM, Huisman HJ, et al. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology*. 2011;259(2):453-461. doi: 10.1148/radiol.11091409
- Donati OF, Mazaheri Y, Afaq A, et al. Prostate cancer aggressiveness: assessment with whole-lesion histogram analysis of the apparent diffusion coefficient. *Radiology*. 2014;271(1):143-152. doi:10.1148/radiol.13130973
- Vargas HA, Akin O, Franiel T, et al. Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness. *Radiology*. 2011;259(3):775-784. doi:10.1148/radiol.11102066
- Meyer HJ, Wienke A, Surov A. Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient—a systematic review and meta analysis. *BMC Cancer*. 2020; 20(1):482. doi: 10.1186/s12885-020-06942-x
- Barrett T, Priest AN, Lawrence EM, et al. Ratio of tumor to normal prostate tissue apparent diffusion coefficient as a method for quantifying DWI of the prostate. *AJR Am J Roentgenol*. 2015;205(6):W585-W593. doi:10.2214/AJR.15.14338
- Woo S, Kim SY, Cho JY, Kim SH. Preoperative evaluation of prostate cancer aggressiveness: using ADC and ADC ratio in determining Gleason score. *AJR Am J Roentgenol*. 2016;207(1):114-120. doi:10.2214/AJR.15.15894
- Gaur S, Harmon S, Rosenblum L, et al. Can apparent diffusion coefficient values assist PI-RADS version 2 DWI scoring? A correlation study using the PI-RADSV2 and International Society of Urological Pathology systems. *AJR Am J Roentgenol*. 2018;211(1):W33-W41. doi:10.2214/AJR.17.18702
- Bonekamp D, Kohl S, Wiesenfarth M, et al. Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values. *Radiology*. 2018;289(1):128-137. doi:10.1148/radiol.2018173064
- Lucarelli NM, Villanova I, Maggialelli N, et al. Quantitative ADC: an additional tool in the evaluation of prostate cancer? *J Pers Med*. 2023; 13(9):1378. doi:10.3390/jpm13091378
- Agrotis G, Pooch E, Abdelatty M, et al. Diagnostic performance of ADC and ADCratio in MRI-based prostate cancer assessment: a systematic review and meta-analysis. *Eur Radiol*. 2025;35(1):404-416. doi:10.1007/s00330-024-10890-6
- Donati OF, Afaq A, Vargas HA, et al. Prostate MRI: evaluating tumor volume and apparent diffusion coefficient as surrogate biomarkers for predicting tumor Gleason score. *Clin Cancer Res*. 2014;20(14):3705-3711. doi:10.1158/1078-0432.CCR-14-0044
- Scialpi M, Martorana E, Scialpi P, et al. MRI apparent diffusion coefficient (ADC): a biomarker for prostate cancer after radiation therapy. *Turk J Urol*. 2021;47(6):448-451. doi:10.5152/tud.2021.21274
- Haider MA, Chung P, Sweet J, et al. Dynamic contrast-enhanced magnetic resonance imaging for localization of recurrent prostate cancer after external beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;70(2):425-430. doi:10.1016/j.ijrobp.2007.06.029
- Wen J, Ji Y, Han J, Shen X, Qiu Y. Inter-reader agreement of the prostate imaging reporting and data system version v2.1 for detection of prostate cancer: a systematic review and meta-analysis. *Front Oncol*. 2022;12:1013941. doi:10.3389/fonc.2022.1013941
- Park KJ, Choi SH, Kim MH, Kim JK, Jeong IG. Performance of prostate imaging reporting and data system version 2.1 for diagnosis of prostate cancer: a systematic review and meta-analysis. *J Magn Reson Imaging*. 2021;54(1):103-112. doi:10.1002/jmri.27546
- Peng Y, Jiang Y, Yang C, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology*. 2013;267(3):787-796. doi:10.1148/radiol.13121454
- Woodfield CA, Tung GA, Grand DJ, Pezzullo JA, Machan JT, Renzulli JF 2nd. Diffusion-weighted MRI of peripheral zone prostate cancer: comparison of tumor apparent diffusion coefficient with Gleason score and percentage of tumor on core biopsy. *AJR Am J Roentgenol*. 2010; 194(4):W316-W322. doi:10.2214/AJR.09.2651
- Verma S, Rajesh A, Morales H, et al. Assessment of aggressiveness of prostate cancer: correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR Am J Roentgenol*. 2011; 196(2):374-381. doi:10.2214/AJR.10.4441
- Washino S, Okochi T, Saito K, et al. Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naïve patients. *BJU Int*. 2017;119(2):225-233. doi:10.1111/bju.13465
- Spadrotto N, Sauck A, Hainc N, Keller I, John H, Hohmann J. Quantitative evaluation of apparent diffusion coefficient values, ISUP grades and prostate-specific antigen density values of potentially malignant PI-RADS lesions. *Cancers (Basel)*. 2023;15(21):5183. doi:10.3390/cancers15215183
- Karaarslan E, Kus AA, Alis D et al. Performance of apparent diffusion coefficient values and ratios for the prediction of prostate cancer aggressiveness across different MRI acquisition settings. *Diagn Interv Radiol*. 2022;28(1):12-20. doi:10.5152/dir.2022.20732
- He D, Wang X, Fu C, et al. MRI-based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins. *Cancer Imaging*. 2021;21(1):46. doi:10.1186/s40644-021-00414-6
- Akin O, Gultekin DH, Vargas HA, et al. Incremental value of diffusion weighted and dynamic contrast enhanced MRI in the detection of locally recurrent prostate cancer after radiation treatment: preliminary results. *Eur Radiol*. 2011;21(9):1970-1978. doi:10.1007/s00330-011-2130-6
- Kwon D, Reis IM, Breto AL, et al. Classification of suspicious lesions on prostate multiparametric MRI using machine learning. *J Med Imaging (Bellingham)*. 2018;5(3):034502. doi:10.1117/1.JMI.5.3.034502
- Zhang L, Tang M, Chen S, Lei X, Zhang X, Huan Y. A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR imaging for the detection of prostate cancer. *Eur Radiol*. 2017;27(12):5204-5214. doi:10.1007/s00330-017-4843-7
- Park SY, Kim CK, Park BK, Lee HM, Lee KS. Prediction of biochemical recurrence following radical prostatectomy in men with prostate cancer by diffusion-weighted magnetic resonance imaging: initial results. *Eur Radiol*. 2011;21(5):1111-1118. doi:10.1007/s00330-010-1999-9