Original Article

Cukurova Anestezi ve Cerrahi Bilimler Dergisi

Diagnostic Sensitivity and Interobserver Variability of Pre-operative Knee MRI in Arthroscopically Confirmed Anterior Cruciate Ligament Tears

Onur Kaya[1], *Mert Erdem Yeşilli*[1], *Nevzat Gönder*[2], *Bilgin Bozgeyik*[1], *Tahir Yavuz*[1]

[1] Gaziantep City Hospital, Department of Orthopaedics and Traumatology, Gaziantep, Türkiye
[2] Gaziantep University Faculty of Medicine, Department of Orthopaedics and Traumatology, Gaziantep, Türkiye

***Correspondence:** Onur KAYA, MD; onurkaya27@hotmail.com

Abstract

Objective: Magnetic resonance imaging (MRI) is the primary modality for preoperative evaluation of anterior cruciate ligament (ACL) injuries. However, interobserver variability in MRI interpretation may influence diagnostic outcomes. This study aimed to compare the diagnostic sensitivity of preoperative MRI radiology reports with independent orthopaedic surgeon evaluations for detecting complete ACL tears, and to assess interobserver agreement in meniscal assessment. **Methods:** This single-centre retrospective observational agreement study included 96 patients who underwent ACL reconstruction with arthroscopically confirmed complete ACL tears. Preoperative MRI images were independently evaluated by three orthopaedic surgeons blinded to clinical data. Agreement with the arthroscopic reference standard was analysed using the McNemar test. Interobserver agreement was assessed using Cohen's kappa, Fleiss' kappa, and Spearman correlation coefficients. **Results:** Radiology reports demonstrated a sensitivity of 28.1% for detecting complete ACL tears, compared with 82.3–87.5% for orthopaedic surgeons ($p<0.001$). Agreement between radiology reports and orthopaedic surgeons was poor ($\kappa=0.052–0.120$). In meniscal assessment, interobserver agreement among orthopaedic surgeons ranged from poor to substantial for both the medial ($\kappa=0.099–0.585$) and lateral meniscus ($\kappa=0.207–0.770$), whereas agreement with radiology reports remained poor. **Conclusion:** In this selected patient cohort, preoperative MRI reports demonstrated substantially lower sensitivity than orthopaedic surgeon evaluations for detecting complete ACL tears. Interobserver agreement in meniscal assessment was variable. These findings emphasise the importance of integrating MRI findings with clinical evaluation and highlight the impact of observer-dependent variability in MRI interpretation.

Keywords: anterior cruciate ligament; magnetic resonance imaging; observer variation; sensitivity; menisci

Received: 14.04.2026
Accepted: 22.06.2026
Published: 30.06.2026

1. Introduction

Magnetic resonance imaging (MRI) is widely accepted as the primary non-invasive modality for the evaluation of anterior cruciate ligament (ACL) injuries and associated intra-articular pathologies, and plays a pivotal role in clinical decision-making [1]. Although the reported sensitivity and specificity of MRI for ACL tears are 87% and 93%, respectively [1], these values are known to vary considerably depending on factors such as tear type and location, the presence of intra-articular haemarthrosis, fibrous band formation in chronic injuries, mucoid degeneration, and the anatomical attachment site of the residual stump [2,3].

In clinical practice, MRI studies are predominantly interpreted by radiologists. However, it has been reported that orthopaedic surgeons' experience in image interpretation and clinical practice may influence how MRI findings are assessed [4].

Meniscal pathology commonly accompanies ACL injuries and is also of considerable diagnostic importance. The reported sensitivity of MRI for medial meniscal tears is 89%, and for lateral meniscal tears 78%; however, these values have been shown to vary markedly depending on tear location and type, the presence of a concomitant ACL injury, and the time interval between MRI acquisition and surgery [1,5,6]. Nevertheless, data on interobserver agreement between radiologists and orthopaedic surgeons in meniscal assessment remain limited [4].

This issue extends beyond academic debate and may have direct implications for clinical practice. Within the framework of healthcare system regulations, surgical decisions are expected to be consistent with radiological reports, which may complicate decision-making in certain clinical scenarios [7].

The aim of this study was to compare the sensitivity of preoperative MRI reports and independent MRI assessments by orthopaedic surgeons in detecting complete ACL tears in patients with arthroscopically confirmed complete ACL injuries, and to analyse interobserver agreement between radiology reports and orthopaedic surgeons regarding the assessment of the medial and lateral menisci.

2. Materials and Methods

2.1. Study design and patient selection

This study was conducted as a single-centre retrospective observational agreement study at a tertiary care training and research hospital. Ethical approval was obtained from the institutional review board (approval no: 492/2026), and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Patients who underwent anterior cruciate ligament (ACL) reconstruction between February 2024 and February 2026 were identified through the hospital information system. Inclusion criteria were defined as having a preoperative knee MRI performed at the same institution and the availability of imaging data through the Picture Archiving and Communication System (PACS).

Exclusion criteria were predefined at the beginning of the study to ensure imaging standardisation, to create a homogeneous and comparable patient cohort, and to minimise the impact of potential confounding factors. Patients were excluded if MRI examinations were performed at an external centre, if image quality did not allow reliable assessment of the ACL and meniscal structures or was reported as suboptimal in the radiology report, if there was a history of revision ACL surgery, or if concomitant major ligament injuries were present (Figure 1).

2.2. MRI protocol

All preoperative knee MRI examinations were performed using a 3 Tesla magnetic resonance system with a standard knee imaging protocol. The protocol included sagittal

proton density-weighted fat-suppressed (PD FS), sagittal T2-weighted, coronal proton density-weighted fat-suppressed (PD FS), and axial T2-weighted sequences. Slice thickness was 3 mm with an interslice gap of 0.5 mm. All images were evaluated through the PACS system.

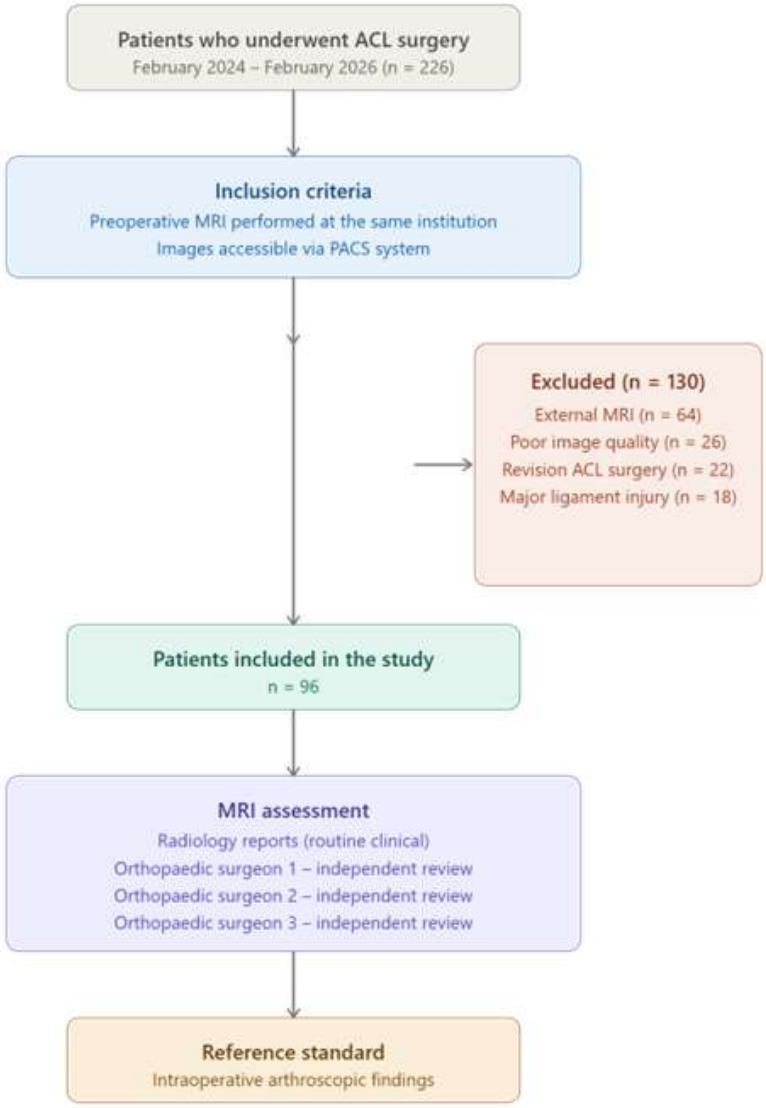

Figure 1: Flowchart of patient selection. Of 226 eligible patients, 96 were included in the final analysis after application of the exclusion criteria.

2.3. Radiology reports

Official preoperative MRI reports were obtained from the hospital information system. These reports had been prepared by radiologists as part of routine clinical practice. The reports represent standard written radiological assessments used during preoperative evaluation. No additional re-evaluation by a radiologist was performed within the scope of this study. For study purposes, radiology reports were categorised based on the explicit terminology used by the reporting radiologist, without any re-interpretation or

standardisation. Reports describing complete discontinuity, complete rupture, or non-visualisation of ACL fibres were classified as complete tears. Reports describing partial fibre disruption or partial tears were classified as partial tears. Reports indicating normal ACL signal and continuity, or the absence of ACL-related pathological findings, were classified as normal. Categorisation was based solely on the written conclusion of each radiology report as retrieved from the hospital information system.

2.4. Orthopaedic assessment

MRI images were independently re-evaluated by three orthopaedic surgeons with at least 5 years of clinical experience in sports medicine and orthopaedic surgery and active involvement in MRI-based decision-making in routine clinical practice. All evaluations were performed blinded to arthroscopic intraoperative findings, patient clinical information, radiology reports, and each other's assessments. In addition, the reviewers had no access to patients' surgical records, operative indications, or information regarding subsequent ACL reconstruction surgery at the time of image interpretation. All reviewers evaluated the images using the same PACS system and computer environment at different time points. The evaluation results were recorded using a standardised data collection form.

2.5. ACL assessment

ACL integrity was classified into three categories - normal, partial tear, and complete tear - based on ligament continuity, fibre integrity, signal characteristics, and anatomical course. Arthroscopic intraoperative findings were accepted as the reference standard. Since only patients with arthroscopically confirmed complete ACL tears were included in the study, a homogeneous reference standard was established for all cases.

2.6. Meniscal assessment

The medial and lateral menisci were evaluated separately. Meniscal signal characteristics were graded from 0 to 3 according to the classification system described by Stoller et al. [8], which is widely used in the literature. In this system, Grade 0 indicates a normal meniscus; Grade 1 indicates globular intrameniscal signal not reaching the articular surface; Grade 2 indicates linear intrameniscal signal not reaching the articular surface; and Grade 3 indicates signal extending to the articular surface, considered a tear.

In this study, Grade 0 and Grade 1 findings were classified as normal, whereas Grade 2 and Grade 3 findings were considered pathological. The inclusion of Grade 2 lesions in the pathological group was based on literature suggesting their association with meniscal degeneration or intrameniscal tears and their potential clinical relevance in symptomatic patients [9]. To ensure comparability between groups, orthopaedic evaluators were instructed to use terminology consistent with that used in radiology reports.

2.7. Statistical analysis

All statistical analyses were performed using SPSS software (IBM SPSS Statistics, version 26.0; IBM Corp., Armonk, NY, USA). A p-value <0.05 was considered statistically significant. Continuous variables were expressed as mean +/- standard deviation (SD), and categorical variables were presented as counts and percentages (%).

For ACL assessment, the McNemar test was used to compare the agreement of each evaluator with the arthroscopic reference standard. This test is used to evaluate differences between paired proportions within the same patient group. ACL status was coded as a binary variable (complete tear = 1, others = 0).

Interobserver agreement was assessed using Cohen's kappa coefficient for pairwise comparisons and Fleiss' kappa for multiple raters (three orthopaedic surgeons and four evaluators including the radiology report). Kappa values were interpreted according to

Landis and Koch criteria: <0.20 poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, >0.80 almost perfect.

Since an ordinal grading system (Grade 0-3) was used for meniscal assessment, Spearman rank correlation coefficients were also calculated, in addition to Cohen's and Fleiss' kappa, to evaluate the strength and direction of associations between observers.

3. Results

3.1. Patient demographics

A total of 96 patients were included in the study. Of these, 78 (81.3%) were male and 18 (18.7%) were female. The mean age was 29.4 +/- 9.9 years (range, 17-46 years). Complete ACL tears were confirmed intraoperatively by arthroscopy in all patients. Detailed demographic characteristics are presented in Table 1.

Table 1. Demographic and Clinical Characteristics of the Study Population

Variable	n (%) or Mean ± SD
Total patients	96
Sex	
Male	78 (81.3%)
Female	18 (18.7%)
Age (years), Mean ± SD	29.4 ± 9.9
Range	17-46
Arthroscopic ACL finding	
Complete tear	96 (100%)

ACL = anterior cruciate ligament.

3.2. ACL assessment

According to the arthroscopic reference standard, all 96 patients (100%) had complete ACL tears. In preoperative MRI reports, only 27 patients (28.1%) were classified as having a complete tear, whereas 48 patients (50.0%) were reported as partial tears and 21 patients (21.9%) as normal.

Independent MRI evaluations by orthopaedic surgeons demonstrated markedly higher detection rates for complete tears: Orthopaedic Surgeon 1 classified 84 patients (87.5%), Orthopaedic Surgeon 2 classified 79 patients (82.3%), and Orthopaedic Surgeon 3 classified 82 patients (85.4%) as having complete ACL tears (Table 2, Figure 2).

Table 2. Sensitivity for Complete ACL Tear Detection by Each Rater Compared to Arthroscopic Gold Standard (n=96)

Rater	Normal n (%)	Partial tear n (%)	Complete tear n (%)	Sensitivity* (%)	p value†
Radiology Report	21 (21.9%)	48 (50.0%)	27 (28.1%)	28.1%	—
Orthopaedic Surgeon 1	4 (4.2%)	8 (8.3%)	84 (87.5%)	87.5%	<0.001
Orthopaedic Surgeon 2	3 (3.1%)	14 (14.6%)	79 (82.3%)	82.3%	<0.001
Orthopaedic Surgeon 3	4 (4.2%)	10 (10.4%)	82 (85.4%)	85.4%	<0.001

** Sensitivity = proportion of complete ACL tear cases correctly identified as complete tear. Since all 96 patients had arthroscopically confirmed complete ACL tears, specificity and overall accuracy cannot be calculated in this single-class reference standard dataset. † McNemar test comparing each orthopaedic surgeon vs radiology report (binary classification: complete tear vs other). OS = Orthopaedic Surgeon.*

In paired comparisons using the McNemar test, all three orthopaedic surgeons showed significantly higher agreement with the arthroscopic reference standard compared to the radiology reports (p<0.001 for all comparisons). No statistically significant

differences were found among the orthopaedic surgeons (OS1 vs OS2: p=0.267; OS1 vs OS3: p=0.724; OS2 vs OS3: p=0.547).

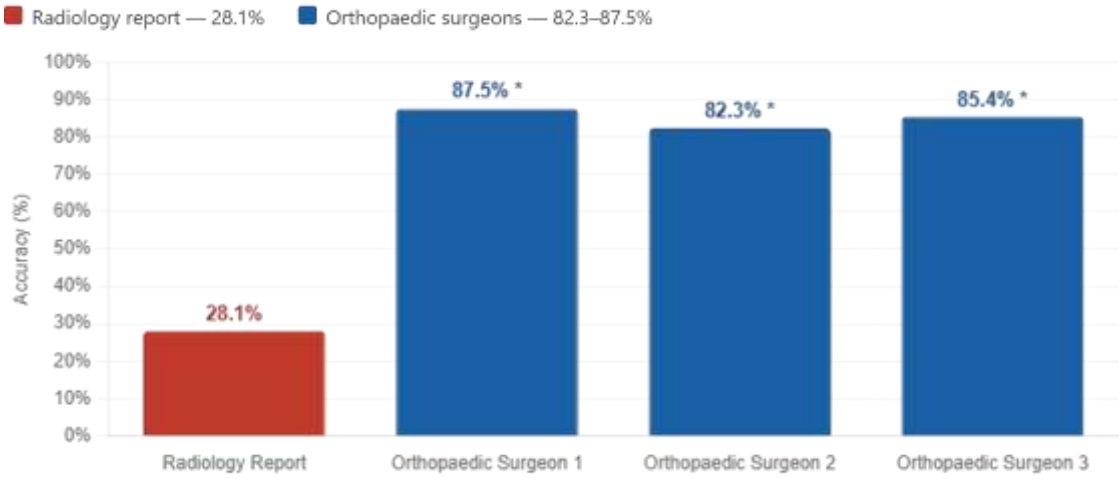

Figure 2. Sensitivity of ACL complete tear detection by each rater compared to arthroscopic gold standard (n=96). All patients had arthroscopically confirmed complete ACL tears. * p<0.001 vs radiology report (McNemar test)

Pairwise Cohen's kappa analysis revealed poor agreement between radiology reports and orthopaedic surgeons (κ=0.052-0.120). Agreement among orthopaedic surgeons ranged from fair to moderate (κ=0.296-0.527). The Fleiss' kappa value for the three orthopaedic surgeons indicated moderate agreement (κ=0.431). When the radiology report was included as a fourth evaluator, overall agreement markedly decreased (κ=0.128) (Table 3).

Table 3. Inter-rater Agreement for ACL Assessment: Cohen's Kappa and Fleiss Kappa

Structure	Comparison	Cohen's κ	Interpretation	Fleiss κ (3 OS)	Fleiss κ (4 raters)
ACL	OS 1 vs OS 2	0.296	Fair	0.431	0.128
ACL	OS 1 vs OS 3	0.527	Moderate	0.431	0.128
ACL	OS 2 vs OS 3	0.480	Moderate	0.431	0.128
ACL	OS 1 vs Radiology	0.052	Slight	—	—
ACL	OS 2 vs Radiology	0.108	Slight	—	—
ACL	OS 3 vs Radiology	0.120	Slight	—	—

OS = Orthopaedic Surgeon. Interpretation of κ values (Landis & Koch): <0.20 = slight; 0.21-0.40 = fair; 0.41-0.60 = moderate; 0.61-0.80 = substantial; >0.80 = almost perfect. Fleiss κ (3 OS) = calculated for the 3 orthopaedic surgeons combined. Fleiss κ (4 raters) = calculated for radiology report + 3 orthopaedic surgeons. Due to the highly skewed prevalence distribution (100% complete tear), kappa values for ACL assessment may be subject to prevalence bias and should be interpreted accordingly.

Table 4. Interobserver Agreement for Medial Meniscus Assessment: Cohen's Kappa, Fleiss' Kappa, and Spearman Correlation

Comparison	Cohen's κ	Spearman r	p value	Fleiss κ (3 OS)	Fleiss κ (4 raters)
OS 1 vs OS 2	0.099	0.586	<0.001	0.357	0.268
OS 1 vs OS 3	0.454	0.664	<0.001		
OS 2 vs OS 3	0.585	0.783	<0.001		
OS 1 vs Radiology	0.155	0.546	<0.001	---	---
OS 2 vs Radiology	0.207	0.494	<0.001	---	---
OS 3 vs Radiology	0.232	0.538	<0.001	---	---

OS = Orthopaedic Surgeon. No arthroscopic reference standard was available for meniscal assessment; all comparisons reflect interobserver agreement only. Cohen's κ and Fleiss' κ reflect exact categorical agreement; Spearman r reflects rank-order association. Fleiss' κ (3 OS) = calculated

for the three orthopaedic surgeons combined. Fleiss' κ (4 raters) = radiology report + three orthopaedic surgeons. Kappa interpretation (Landis & Koch): <0.20 poor; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; >0.80 almost perfect.

Among the 48 cases classified as partial ACL tears in the radiology reports, 39 (81.3%), 37 (77.1%), and 40 (83.3%) were reclassified as complete tears by Orthopaedic Surgeons 1, 2, and 3, respectively. Only a minority of these cases continued to be classified as partial tears by the orthopaedic surgeons (Table 6).

3.3. Medial meniscus assessment

In medial meniscus evaluation, pairwise Cohen's kappa values among orthopaedic surgeons ranged from 0.099 to 0.585. The lowest agreement was observed between OS1 and OS2 (κ =0.099, poor), whereas the highest agreement was found between OS2 and OS3 (κ=0.585, moderate). Agreement between radiology reports and orthopaedic surgeons remained poor across all comparisons (κ =0.155-0.232). The Fleiss' kappa value among the three orthopaedic surgeons indicated fair-to-moderate agreement (κ =0.357). When all four evaluators were included, overall agreement further decreased (κ=0.268).

Spearman correlation analysis demonstrated statistically significant positive correlations among orthopaedic surgeons (OS1-OS2: r=0.586; OS1-OS3: r=0.664; OS2-OS3: r=0.783; all p<0.001). Correlations between radiology reports and orthopaedic surgeons were also significant but weaker (r=0.494-0.546, p<0.001) (Table 4).

Table 5. Interobserver Agreement for Lateral Meniscus Assessment: Cohen's Kappa, Fleiss' Kappa, and Spearman Correlation

Comparison	Cohen's κ	Spearman r	p value	Fleiss κ (3 OS)	Fleiss κ (4 raters)
OS 1 vs OS 2	0.207	0.270	0.008	0.458	0.274
OS 1 vs OS 3	0.422	0.418	<0.001		
OS 2 vs OS 3	0.770	0.848	<0.001		
OS 1 vs Radiology	0.060	0.143	0.164	---	---
OS 2 vs Radiology	0.074	0.222	0.029	---	---
OS 3 vs Radiology	0.098	0.217	0.034	---	---

OS = Orthopaedic Surgeon. No arthroscopic reference standard was available for meniscal assessment; all comparisons reflect interobserver agreement only. Cohen's κ and Fleiss' κ reflect exact categorical agreement; Spearman r reflects rank-order association. Fleiss' κ (3 OS) = calculated for the three orthopaedic surgeons combined. Fleiss' κ (4 raters) = radiology report + three orthopaedic surgeons. Kappa interpretation (Landis & Koch): <0.20 poor; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; >0.80 almost perfect.

Table 6. Orthopaedic Surgeon Classification of Cases Reported as Partial ACL Tears in Radiology Reports (n=48)

Rater	Normal n (%)	Partial tear n (%)	Complete tear n (%)
Orthopaedic Surgeon 1	4 (8.3%)	5 (10.4%)	39 (81.3%)
Orthopaedic Surgeon 2	2 (4.2%)	9 (18.8%)	37 (77.1%)
Orthopaedic Surgeon 3	1 (2.1%)	7 (14.6%)	40 (83.3%)

All 96 patients had arthroscopically confirmed complete ACL tears. Radiology report categories: Normal = no ACL pathology reported; Partial tear = partial fibre disruption reported; Complete tear = complete discontinuity or non-visualisation of ACL fibres reported. Orthopaedic surgeon classifications were performed independently and blinded to clinical information.

3.4. Lateral meniscus assessment

A distinct pattern of agreement was observed in lateral meniscus evaluation. Substantial agreement was found between OS2 and OS3 (κ =0.770, r=0.848, p<0.001). Agreement between OS1 and OS3 was moderate (κ =0.422), whereas agreement between OS1 and OS2 was poor to fair (κ =0.207). The Fleiss' kappa value among the three orthopaedic surgeons indicated moderate agreement (κ =0.458).

Agreement between radiology reports and orthopaedic surgeons remained poor in all comparisons (κ =0.060-0.098). The Spearman correlation between OS1 and radiology

did not reach statistical significance (r=0.143, p=0.164), whereas correlations between OS2 and radiology (r=0.222, p=0.029) and OS3 and radiology (r=0.217, p=0.034) were statistically significant at a nominal level. When all four evaluators were considered together, overall agreement remained poor (κ =0.274) (Table 5).

4. Discussion

In this study, preoperative MRI reports demonstrated low sensitivity in detecting complete ACL tears in patients with arthroscopically confirmed injuries, whereas orthopaedic surgeons who independently evaluated the same images showed markedly higher sensitivity. These findings indicate that the same MRI images may yield significantly different interpretations depending on the observer. The results suggest that MRI interpretation is influenced not only by image quality but also by observer-dependent factors.

Previous studies have reported the sensitivity of MRI for diagnosing ACL tears to range between 87% and 95% under controlled research conditions [1,3]. In our study, the sensitivity of radiology reports was considerably lower than these values, suggesting that routine clinical reporting may differ from research-based assessments. In contrast, the sensitivity values obtained by orthopaedic surgeons were consistent with the literature, supporting the notion that observer-related differences may have a meaningful impact in clinical practice [4,10].

Despite high sensitivity rates, kappa values for ACL assessment remained low in our study. This finding can be explained by the well-known prevalence bias, whereby kappa statistics tend to be artificially low in datasets with a single dominant category [11,12]. Therefore, the observed kappa values may reflect not only interobserver disagreement but also the homogeneous structure of the study population.

In this study, Grade 2 meniscal signal changes were classified as pathological based on literature suggesting an association with meniscal degeneration or intrameniscal tears [8]. Although the clinical significance of Grade 2 lesions remains controversial, Low et al. reported that such findings may be clinically relevant in symptomatic patients [9]. Additionally, consistency with the classification system used in radiology reports was maintained to ensure comparability between groups.

The discrepancy observed between radiology reports and orthopaedic surgeons in preoperative MRI evaluations is likely multifactorial. Radiology reports are typically generated within a standardised reporting framework, whereas orthopaedic surgeons' experience in image interpretation and clinical practice may influence how MRI findings are assessed[4]. In this study, orthopaedic surgeons performed their evaluations blinded to clinical information, suggesting that the observed differences may be related to interpretative approaches rather than clinical integration. Another factor that may have contributed to the observed differences is the distinction between routine clinical reporting and study-specific image review. The radiology reports analysed in this study were generated as part of routine clinical practice, potentially under substantial workload and time constraints, whereas the orthopaedic surgeons performed dedicated study-specific image reviews under standardised conditions. Therefore, the observed differences should not be interpreted as evidence of inferior professional performance by radiologists, but rather as a reflection of differences in evaluation context, reporting objectives, and observer perspective. Particularly in borderline cases, where the distinction between partial and complete tears is subtle, MRI interpretation inherently involves a degree of subjectivity, which may amplify interobserver variability [1,3]. Nevertheless, the higher sensitivity observed among orthopaedic surgeons may, in part, reflect a tendency to overinterpret complete tears in certain cases. This finding underscores that diagnostic interpretation is influenced not only by imaging features but also by observer-related judgement[10]. This interpretation is supported by our additional analysis of cases classified as partial ACL tears in the

radiology reports. More than three-quarters of these cases were subsequently classified as complete tears by each orthopaedic surgeon. This finding suggests that much of the disagreement between radiology reports and orthopaedic assessments occurred in borderline cases, where different diagnostic thresholds may have been applied when distinguishing partial from complete ACL tears.

Importantly, only surgically treated patients were included in this study, resulting in a cohort with a high pretest probability of complete ACL tears. It is well established that including only confirmed cases in diagnostic accuracy studies may artificially inflate sensitivity estimates [18]. Although all MRI evaluations were performed under blinded conditions, orthopaedic surgeons may still have been inclined to favour a diagnosis of complete tear in borderline cases. Therefore, our findings should not be interpreted as evidence of superior diagnostic performance of orthopaedic surgeons, but rather as an indication of differing interpretative tendencies within a selected patient population.

Interobserver agreement in meniscal assessment showed a more heterogeneous pattern. While kappa values among orthopaedic surgeons varied widely for the medial meniscus, higher agreement was observed between certain observer pairs in the lateral meniscus. In contrast, agreement between radiology reports and orthopaedic surgeons remained low for both menisci. These findings are consistent with previous studies reporting discrepancies between radiologists and orthopaedic surgeons in MRI interpretation and highlighting the role of observer-dependent variability [13].

The combined use of Cohen's kappa and Spearman correlation coefficients provides complementary insights into interobserver agreement. While kappa reflects categorical agreement, Spearman correlation captures the direction and strength of associations between ordinal variables; it is well recognised that these measures may yield different results within the same dataset [14]. In our study, a similar pattern was observed, particularly in the medial meniscus, where low kappa values (κ =0.099) coexisted with moderate Spearman correlation (r=0.586). This suggests that, although exact categorical agreement was limited, observers demonstrated similar grading tendencies.

Our findings may also have implications for healthcare systems such as that in Turkey. Within reimbursement and auditing frameworks, surgical decisions are often expected to be consistent with radiological reports [7]. In this context, the inability of preoperative MRI reports to adequately reflect complete ACL tears in some cases may increase the importance of comprehensive clinical documentation when surgical decisions are based on clinical and arthroscopic findings. This is particularly relevant in cases where discrepancies exist between clinical assessment and radiological reporting.

Overall, our results indicate that preoperative MRI reports alone may not be sufficient for surgical decision-making and should be interpreted in conjunction with clinical findings. Previous studies have shown that MRI performance in community practice may differ from that observed under controlled research conditions [15]. Therefore, integrating MRI findings with clinical examination and patient history appears essential for accurate diagnosis and appropriate surgical planning [4,15,16,17].

5. Limitations

This study has several limitations. First, due to its retrospective design, intraoperative documentation of meniscal findings was not standardised. As a result, arthroscopy could not be used as a reference standard for meniscal assessment, and analyses were limited to interobserver agreement.

Second, information regarding the experience level and subspecialty expertise of the radiologists who prepared the reports was not available, which limited subgroup analyses within the radiology cohort.

Third, inclusion of only patients with arthroscopically confirmed complete ACL tears resulted in a single-category reference standard. This prevented calculation of specificity, predictive values, and overall diagnostic accuracy and may have led to kappa values being affected by prevalence bias. Therefore, the findings should be interpreted primarily in terms of sensitivity and interobserver variability rather than overall diagnostic accuracy. Additionally, the high pretest probability within this cohort may have increased the tendency of orthopaedic surgeons to classify borderline cases as complete tears, thereby influencing the observed sensitivity values.

Finally, the single-centre design may limit the generalisability of the findings.

6. Conclusion

This study demonstrated that, in a selected patient cohort with arthroscopically confirmed complete ACL tears, the sensitivity of preoperative MRI reports for detecting complete tears may be lower than that of independent evaluations performed by orthopaedic surgeons. In meniscal assessment, interobserver agreement was found to be variable. These findings suggest that MRI should not be interpreted in isolation but rather in conjunction with clinical evaluation, and that meaningful interpretative differences may exist between radiologists and orthopaedic surgeons.

Ethical Approval: The study protocol was approved by the Ethics Committee of Gaziantep City Hospital (IRB Number: 492/2026). All procedures were conducted in accordance with the Declaration of Helsinki. Informed consent was waived due to the retrospective nature of the study.

genAI: No artificial intelligence-based tools or generative AI technologies were used in the preparation of this manuscript. The entire content was originally prepared, reviewed, and approved by all authors.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest: The authors declare that they have no conflict of interest.

Data Availability: Due to institutional privacy policies, the datasets generated and/or analyzed during the current study are not publicly available, but are available from the corresponding author upon reasonable request.

Author Contributions: Conception and design: OK, MEY, NG. Analysis and interpretation of data: OK, MEY, BB, TY. Drafting the article: OK, MEY, NG. Revising the article: OK, NG, BB, TY. All authors read and approved the final manuscript.

References

1. Phelan N, Rowland P, Galvin R, O'Byrne JM. A systematic review and meta-analysis of the diagnostic accuracy of MRI for suspected ACL and meniscal tears of the knee. Knee Surg Sports Traumatol Arthrosc. 2016;24(5):1525-39. doi:10.1007/s00167-015-3861-8
2. Ng WHA, Griffith JF, Hung EHY, Paunipagar B, Law BKY, Yung PSH. Imaging of the anterior cruciate ligament. World J Orthop. 2011;2(8):75-84. doi:10.5312/wjo.v2.i8.75
3. Zhao M, Zhou Y, Chang J, Hu J, Liu H, Wang S, et al. The accuracy of MRI in the diagnosis of anterior cruciate ligament injury. Ann Transl Med. 2020;8(24):1657. doi:10.21037/atm-20-7391
4. Felli L, Garlaschi G, Muda A, Tagliafico A, Formica M, Zanirato A, et al. Comparison of clinical, MRI and arthroscopic assessments of chronic ACL injuries, meniscal tears and cartilage defects. Musculoskelet Surg. 2016;100(3):231-8. doi:10.1007/s12306-016-0427-y

5. Sharifah MIA, Lee CL, Suraya A, Johan A, Syed AFSK, Tan SP. Accuracy of MRI in the diagnosis of meniscal tears in patients with chronic ACL tears. Knee Surg Sports Traumatol Arthrosc. 2015;23(3):826-30. doi:10.1007/s00167-013-2766-7
6. Dufka FL, Lansdown DA, Zhang AL, Allen CR, Ma CB, Feeley BT. Accuracy of MRI evaluation of meniscus tears in the setting of ACL injuries. Knee. 2016;23(3):460-4. doi:10.1016/j.knee.2016.01.018
7. Sosyal Guvenlik Kurumu. Saglik Uygulama Tebligi (SUT) [Internet]. Ankara: SGK; 2008 [cited 2024 Jan 1]. Available from:
8. Stoller DW, Martin C, Crues JV 3rd, Kaplan L, Mink JH. Meniscal tears: pathologic correlation with MR imaging. Radiology. 1987;163(3):731-5. doi:10.1148/radiology.163.3.3575724
9. Low AK, Chia MR, Carmody DJ, Lucas P, Hale D. Clinical significance of intrasubstance meniscal lesions on MRI. J Med Imaging Radiat Oncol. 2008;52(3):227-30. doi:10.1111/j.1440-1673.2008.01951.x
10. van Grinsven S, Nijenhuis TA, Konings PC, van Kampen A, van Loon CJ. Are radiologists superior to orthopaedic surgeons in diagnosing instability-related shoulder lesions on magnetic resonance arthrography? A multicenter reproducibility and accuracy study. J Shoulder Elbow Surg. 2015;24(9):1405-12. doi:10.1016/j.jse.2015.02.029
11. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993;46(5):423-9. doi:10.1016/0895-4356(93)90018-V
12. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-9. doi:10.1016/0895-4356(90)90158-L
13. Figueiredo S, Sa Castelo L, Pereira AD, Machado L, Silva JA, Sa A. Use of MRI by radiologists and orthopaedic surgeons to detect intra-articular injuries of the knee. Rev Bras Ortop. 2018;53(1):28-32. doi:10.1016/j.rboe.2016.12.013
14. Warrens MJ, Bosker RJ, Kiers HAL. A comparison of reliability coefficients for ordinal rating scales. J Classif. 2021;38:519-43. doi:10.1007/s00357-021-09386-5
15. Hardy JC, Evangelista GT, Grana WA, Hunter RE. Accuracy of magnetic resonance imaging of the knee in the community setting. Sports Health. 2012;4(3):222-31. doi:10.1177/1941738111434396
16. Shu HT, Wegener NR, Connors KM, Yang DS, Lockey SD, Thomas JM, et al. Accuracy of magnetic resonance imaging in predicting anterior cruciate ligament tear location and tear degree. J Orthop. 2021;25:129-33. doi:10.1016/j.jor.2021.05.002
17. Nguyen JC, De Smet AA, Graf BK, Rosas HG. MR imaging-based diagnosis and classification of meniscal tears. RadioGraphics. 2014;34(4):981-99. doi:10.1148/rg.344125202
18. Whiting PF, Rutjes AW, Westwood ME, Mallett S; QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol. 2013;66(10):1093-104. doi:10.1016/j.jclinepi.2013.05.014