

COMPARATIVE EVALUATION OF ARTIFICIAL INTELLIGENCE AND RADIOLOGIST INTERPRETATION OF SHOULDER MRI

Piyush Yadav^{1*}, Dr. Monu Sarin², Rajdeep Thidwar³, Archi⁴, Vikas Kumar⁵

¹Research Scholar, Department of Radio-Imaging Technology, Faculty of Allied Health Sciences, SGT University, Gurugram, Haryana, India, Email: piyushi021995@gmail.com , Orcid id: 0000-0002-9210-7862

²Professor, Department of Radio-Diagnosis, Faculty of Medicine and Health Sciences, SGT University, Gurugram, Haryana, India, Email: monusarin11@gmail.com, Orcid id: 0000-0002-1321-1953

³Research Scholar, Department of Radio-Imaging Technology, Faculty of Allied Health Sciences, SGT University, Gurugram, Haryana, India, Email: rehanhts@gmail.com , Orcid id: 0009-0003-6135-4021

⁴Assistant Professor, Department of Radiology & Imaging Technology, Guru Nanak Paramedical College, Dhahan Kaleran, Punjab, India, Email: archibhateja15@gmail.com, Orcid Id: 0009-0005-6549-1052

⁵Assistant Professor, Department of Radiology & Imaging Technology, Guru Nanak Paramedical College, Dhahan Kaleran, Punjab, India, Email: vikasbhardwaj150@gmail.com, Orcid Id: 0000-0002-8958-8968

ABSTRACT

Background: Shoulder MRI provides detailed assessment of the rotator cuff, labrum, and other periarticular structures, but interpretation can be challenging in examinations with multiple or subtle abnormalities. Artificial intelligence (AI) has emerged as a potential adjunct to conventional radiological interpretation. This study evaluated the diagnostic performance of AI in shoulder MRI using the recorded final diagnosis as the reference standard.

Methods: A comparative observational study was conducted on 60 patients who underwent shoulder MRI at SGT Hospital, Gurugram. We recorded MRI acquisition parameters, including repetition time (TR), echo time (TE), flip angle, and sequence information. The same MRI examinations were analyzed using a designated AI system, and AI findings were compared with the documented final/reference diagnosis. Diagnostic performance was assessed using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and Cohen's kappa.

Results: The study included 38 (63.3%) men and 22 (36.7%) women, with a mean age of 44.18 ± 15.15 years. 54 examinations (90.0%) were abnormal and six (10.0%) were normal. AI correctly identified the main diagnostic category in 53/60 examinations (88.3%). For binary classification of normal versus abnormal examinations, AI demonstrated 100.0% sensitivity, 83.3% specificity, 98.2% PPV, 100.0% NPV, and 98.3% accuracy, with substantial agreement ($\kappa \approx 0.90$). Sensitivity was less for full-thickness supraspinatus tears (66.7%).

Conclusion: AI demonstrated high sensitivity and substantial agreement with the reference diagnosis for detecting abnormalities on shoulder MRI. However, limitations in lesion-specific characterization, particularly full-thickness supraspinatus tears, indicate that AI should currently be considered an adjunct to radiologist interpretation rather than a replacement.

KEYWORDS: Shoulder MRI, Artificial intelligence, Diagnostic accuracy, Radiologist, Rotator cuff, Cohen's kappa.

INTRODUCTION

Shoulder disorders are commonly encountered in clinical practice and can have a considerable impact on a patient's daily activities, mobility, and quality of life. The shoulder is a complex anatomical region that includes the glenohumeral and acromioclavicular joints, rotator cuff tendons, labrum, muscles, ligaments, and surrounding soft tissues. Abnormalities involving these structures may produce persistent pain, restricted movement, and functional impairment. Common examples include rotator cuff tendinopathy and tears, labral lesions, adhesive capsulitis, subacromial impingement, joint effusion, and traumatic or degenerative changes. Identifying these abnormalities accurately is important for establishing the diagnosis and selecting an appropriate treatment approach. [1–3]

Magnetic resonance imaging (MRI) has an important role in the evaluation of shoulder disorders because of its ability to provide excellent soft tissue contrast and multiplanar images. It allows detailed assessment of the rotator cuff, labrum, cartilage, muscles, tendons, joint capsule, and periarticular soft tissues. For this reason, MRI is particularly valuable when rotator cuff or labral pathology is suspected. [9,20] Earlier studies have also demonstrated the diagnostic value of MRI in the assessment of rotator cuff tears and other shoulder abnormalities, including its relationship with clinical symptoms and functional outcomes. [6,14]

The quality of a shoulder MRI examination depends on several factors, including the imaging protocol, sequence selection, and acquisition parameters. Appropriate parameter selection is necessary to obtain sufficient tissue contrast and adequate visualization of the anatomical structures of interest. Parameters such as repetition time (TR), echo time (TE), number of excitations (NEX), and other sequence characteristics influence image contrast, signal intensity, acquisition time, and overall image quality. [9,16] Quantitative measures such as the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) may also be used to assess image quality and the visibility of different tissues.

In recent years, considerable attention has been given to reducing MRI acquisition time while maintaining diagnostically useful images. Accelerated imaging techniques and deep learning-based reconstruction methods have been investigated for this purpose. Herrmann et al. demonstrated the feasibility of a rapid shoulder MRI protocol using accelerated turbo spin-echo sequences with deep learning reconstruction at 1.5 and 3 T. [13] Similarly, a prospective comparison of standard and deep-learning-reconstructed turbo spin-echo shoulder MRI reported the potential of this approach to improve acquisition efficiency while maintaining diagnostically acceptable image quality. [23] These developments have contributed to the growing use of artificial intelligence (AI) in MRI, initially in areas such as image reconstruction and acceleration, and more recently in image analysis and interpretation.

Despite these advances, the interpretation of shoulder MRI remains dependent on radiological expertise. Some findings, including subtle tendon abnormalities, partial-thickness tears, and small labral lesions, may be difficult to identify. The presence of multiple abnormalities in the same examination can further complicate interpretation. Technical limitations, such as motion artifacts or inadequate image quality, may also affect the assessment of these findings. Previous studies have evaluated the diagnostic performance of MRI by comparing imaging findings with surgical findings and clinical outcomes. [4,6] Such comparisons highlight the importance of accurate interpretation and appropriate correlation of MRI findings with the underlying clinical or pathological diagnosis.

AI-based methods, particularly machine learning and deep learning, have therefore attracted interest as potential tools for assisting image interpretation. These methods can be trained to recognize patterns within large imaging datasets and may support the identification of abnormalities. In MRI, AI has been investigated for image reconstruction, acquisition acceleration, image analysis, and computer-assisted interpretation. Deep-learning reconstruction in shoulder MRI is one example of how AI can improve the technical aspects of imaging, while also providing a basis for further research into its diagnostic applications. [13,23]

The assessment of AI-based interpretation requires comparison with an appropriate reference standard. Computational performance alone does not establish whether an AI system can identify clinically relevant abnormalities accurately. Diagnostic measures such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy can be used to evaluate its performance. Measures of agreement, including Cohen's kappa, may also be useful when AI-generated findings are compared with expert interpretation using appropriately structured reference data.

Image quality is an additional consideration when evaluating AI performance in MRI. Reduced signal, inadequate tissue contrast, motion artifacts, and suboptimal acquisition parameters may influence the visibility of subtle abnormalities and, consequently, the interpretation of the examination. On the other hand, improvements in acquisition and reconstruction techniques may allow diagnostically useful images to be obtained in a shorter examination time. [13,23] This raises an important question: does the diagnostic performance of AI remain consistent when shoulder MRI examinations are acquired using different technical parameters and when image quality varies?

Shoulder MRI is a useful setting in which to investigate this question because a single examination may demonstrate abnormalities involving several anatomical structures. Rotator cuff pathology, particularly supraspinatus involvement, is frequently encountered, while subscapularis abnormalities, labral lesions, adhesive capsulitis, joint effusion, and impingement-related changes may also be present. [4,19] Accurate interpretation therefore requires not only the detection of abnormalities but also their appropriate anatomical and pathological characterization.

Although AI applications in medical imaging are developing rapidly, evidence regarding their performance in routine shoulder MRI interpretation remains limited, particularly when diagnostic findings are considered alongside the technical characteristics of image acquisition. A comparison of AI-generated findings with established diagnostic findings, while also considering relevant MRI acquisition parameters and image quality, may provide a more clinically meaningful assessment of its performance.

Therefore, the present study aimed to compare the diagnostic interpretation of shoulder MRI by an artificial intelligence (AI) system and a radiologist using the same MRI examinations. The study also assessed their agreement with the recorded reference diagnosis and evaluated the performance of AI in detecting normal and abnormal examinations and characterizing selected shoulder abnormalities. Available MRI acquisition parameters were documented to explore their potential relevance to diagnostic interpretation, where supported by the available data.

Aim of the Study

The present study aimed to compare artificial intelligence (AI)-based interpretation with radiologist interpretation of shoulder magnetic resonance imaging (MRI), with particular emphasis on the identification and classification of shoulder abnormalities. The study also evaluated the diagnostic performance of AI in distinguishing normal from abnormal examinations and in detecting selected shoulder pathologies, particularly rotator cuff tears, using the recorded reference diagnosis as the comparator.

Objectives of the Study

- To compare AI-based interpretation with radiologist interpretation of shoulder MRI examinations, based on the available interpretation records.
- To evaluate the diagnostic performance of AI in differentiating normal from abnormal shoulder MRI examinations using the recorded reference diagnosis.
- To assess the ability of AI to identify selected shoulder abnormalities, particularly supraspinatus and other rotator cuff tears.

- To evaluate the agreement between AI-generated findings and the recorded reference diagnosis using appropriate diagnostic performance measures and agreement statistics.

MATERIALS AND METHODS

Study Design and Study Population: A comparative observational study was conducted on 60 patients who underwent shoulder magnetic resonance imaging (MRI) at SGT Hospital, Gurugram, during the study period. Patients referred for MRI because of clinically suspected shoulder pathology were included according to the predefined eligibility criteria. Both right- and left-shoulder examinations were considered. The same MRI examinations were evaluated by a radiologist and an artificial intelligence (AI) system. Their diagnostic interpretations were compared with a predefined reference diagnosis, where available, to assess diagnostic performance and agreement.

Inclusion Criteria

1. Shoulder MRI was performed at SGT Hospital during the study period.
2. A complete MRI examination was performed for diagnostic interpretation.
3. Adequate MRI images were available for review.
4. The relevant MRI acquisition parameters and image quality data were available.
5. Had a documented final or reference diagnosis against which the radiologist's and AI interpretations could be compared.

Exclusion Criteria

1. Incomplete or technically inadequate MRI examinations.
2. Severe motion or other artifacts affecting image interpretation.
3. Incomplete clinical or diagnostic information.
4. Missing MRI acquisition parameters or image quality measurements.
5. Images inadequate for reliable assessment of shoulder pathology.

MRI Examination and Image-Quality Assessment

All MRI examinations were performed using the routine shoulder MRI protocol followed at SGT Hospital for the assessment of the rotator cuff, tendons, muscles, labrum, articular structures, and the surrounding soft tissues.

The MRI acquisition parameters available from the examinations were recorded, including repetition time (TR), echo time (TE), and flip angle. These parameters were analyzed in relation to the available image-quality data.

Image quality was assessed from the MRI images based on their diagnostic adequacy and the visualization of the relevant anatomical structures. The assessment considered the clarity of the images, visualization of the structures of interest, presence of artifacts, and ability to identify pathological abnormalities.

Where region-of-interest (ROI) measurements were available from the MRI images, the recorded values were included in the analysis. The corresponding mean and standard deviation (SD) values were used as available in the study dataset. These image-derived measurements were analyzed along with the MRI acquisition parameters to examine their relationship with image quality.

Only measurements available from the MRI examinations were included in the analysis. No additional image quality parameters were derived when they were not available in the original dataset.

Radiologist Interpretation

A radiologist reviewed the shoulder MRI examinations and recorded the diagnostic findings, including the presence or absence of abnormalities and the relevant anatomical structures involved. The recorded interpretations were used for comparison with the AI-generated findings. The radiologist's level of experience, the method of interpretation, and whether the assessment was performed independently of the AI output should be reported. The radiologist's findings were compared with the predefined reference diagnosis where available.

Artificial Intelligence Interpretation

The same 60 shoulder MRI examinations were analyzed using the designated artificial intelligence (AI) system. The AI-generated findings and diagnostic classifications were recorded and compared with the corresponding radiologist interpretations and predefined reference diagnoses, where available. The AI system's name, version, intended function, input requirements. The independence of the AI and radiologist assessments.

Diagnostic Performance

Diagnostic performance was evaluated using the recorded reference diagnosis as the comparator. For binary classification, examinations were categorized as normal or abnormal. True-positive, false-positive, true-negative, and false-negative classifications were determined using the predefined classification rules. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy were calculated from the corresponding confusion matrix.

Where patient-level radiologist interpretations were available, AI and radiologist performance were assessed separately against the same reference standard. Agreement between categorical interpretations was evaluated using Cohen's kappa, with the categories and comparison pairs clearly specified. Lesion-specific performance was calculated only for diagnostic categories with sufficient and verifiable case-level data.

Statistical Analysis

Continuous variables were summarized as mean $\pm$ standard deviation (SD) or median and interquartile range, as appropriate. Categorical variables were summarized as frequencies and percentages. Diagnostic performance was calculated from verified patient-level classifications. When AI and radiologist interpretations were paired for the same examinations, an appropriate paired comparison, such as McNemar's test for binary outcomes, was considered where applicable. Cohen's kappa was used to assess agreement between categorical classifications. Correlation and group-comparison analyses were performed only where the required variables and assumptions were satisfied. All tests were two-sided, and a p-value below 0.05 was considered statistically significant.

Ethical Considerations

Patient confidentiality was maintained throughout the study period. Patient-identifying information was excluded from the study analysis, and MRI data were used only for the purposes of the study. This study was conducted in accordance with the applicable institutional ethical requirements governing the use of clinical imaging data.

RESULTS

The study included 60 shoulder MRI examinations. The mean age of the participants was 44.18 ± 15.15 years. Of the participants, 38 (63.3%) were male and 22 (36.7%) were female. According to the reference or final diagnosis, 54 examinations (90.0%) demonstrated abnormalities, whereas 6 (10.0%) were classified as normal.

The recorded MRI acquisition parameters showed a mean repetition time (TR) of 2680.46 ± 1493.53 ms and a mean echo time (TE) of 65.66 ± 38.90 ms. A flip angle of 90° was used in 59 examinations (98.3%), while a flip angle of 10° was used in one examination (1.7%).

The diagnostic performance of the AI system was evaluated by comparing its interpretations with the reference or final diagnoses. Following standardization of the diagnostic categories, the AI system correctly identified the main diagnostic category in 53 of the 60 examinations (88.3%), while discrepancies were observed in 7 examinations (11.7%). Cohen's kappa was reported as 0.868, indicating strong agreement beyond chance, subject to verification using the original case-level classifications.

The AI system demonstrated high diagnostic performance across several diagnostic categories, including adhesive capsulitis, massive rotator cuff tears, fractures, shoulder instability or dislocation, osteochondroma, osteomyelitis, and inferior glenohumeral ligament (IGHL) tears. The reported sensitivity and specificity for these categories were both 100%; however, these estimates should be interpreted in the context of the number of cases available in each category.

For partial-thickness rotator cuff tears, the reported sensitivity was 90.0%, specificity was 98.0%, positive predictive value (PPV) was 90.0%, and negative predictive value (NPV) was 98.0%. For tendinopathy, the sensitivity was 85.7%, specificity was 100%, PPV was 100%, and NPV was 95.8%.

For binary classification into normal and abnormal examinations, the AI system correctly identified all 54 reference-abnormal examinations. Of the six reference-normal examinations, five were correctly classified as normal, while one was incorrectly classified as abnormal. Accordingly, the sensitivity was 100.0%, specificity was 83.3%, PPV was 98.2%, NPV was 100.0%, and overall accuracy was 98.3%.

Overall, the findings indicated high agreement between the AI interpretations and the reference diagnoses, with 88.3% agreement at the standardized diagnostic-category level. The binary normal-versus-abnormal analysis also demonstrated high overall accuracy. Nevertheless, diagnostic discrepancies in specific categories, particularly full-thickness supraspinatus tears, warrant further investigation. Validation using a larger, independently assessed dataset is necessary to establish the generalizability of the system's diagnostic performance.

MRI Acquisition Parameters

The recorded repetition time (TR) ranged from 3.6 to 4908 ms, with a mean of 2680.46 ± 1493.53 ms. The echo time (TE) ranged from 1.7 to 100 ms, with a mean of 65.66 ± 38.90 ms.

A flip angle of 90° was used in 59 examinations (98.3%), whereas one examination (1.7%) used a flip angle of 10° . The most frequently recorded sequence was the PD fat-suppressed (PD FS) sequence, which was used in 33 examinations (55.0%). Other sequences included axial PD FS, PD, T2 FS/STIR, STIR, T1, T1+PD, PD/T1, and combined-sequence protocols.

The acquisition parameters demonstrated considerable variations in TR and TE across examinations, whereas the flip angle was relatively consistent.

Table 1: Demographic Characteristics of the Study Population

Characteristic	Value
Total number of patients	60
Mean age (years)	44.18 ± 15.15
Age range (years)	18–76
Male	38 (63.3%)
Female	22 (36.7%)
Male-to-female ratio	1.73:1

Distribution of MRI Diagnoses

A broad spectrum of shoulder abnormalities was identified. Supraspinatus tendinosis was the most frequently recorded specific final diagnosis, occurring in seven patients (11.7%). A normal shoulder MRI was documented in six patients (10.0%).

Massive rotator cuff tears were identified in four patients (6.7%), while partial subscapularis tears were also recorded in four patients (6.7%). Partial-width full-thickness supraspinatus tears were identified in three patients (5.0%).

Other diagnoses included adhesive capsulitis, partial rotator cuff tears, scapular fractures associated with multiple rotator cuff tears, supraspinatus tendinopathy, supraspinatus tendinosis with bursitis, shoulder instability/dislocation with Bankart and Hill-Sachs lesions, osteochondroma, Ewing's sarcoma, osteomyelitis, lipoma, IGHL tear, inflammatory synovitis, and other tendon and rotator cuff abnormalities.

Overall, the dataset demonstrated a predominance of rotator cuff and tendon-related abnormalities, particularly involving the supraspinatus tendon, as well as additional traumatic, inflammatory, degenerative, and neoplastic conditions.

Table 2. Distribution of Principal MRI Diagnoses in the Study Population

Final MRI Diagnosis	Number of cases, n (%)
Supraspinatus tendinosis	7 (11.7%)
Normal MRI	6 (10.0%)
Massive rotator cuff tear	4 (6.7%)
Partial subscapularis tear	4 (6.7%)
Partial-width full-thickness supraspinatus tear	3 (5.0%)
Other shoulder abnormalities	36 (60.0%)
	60 (100%)

Values are presented as the number of examinations (percentage of the total cohort). Each examination should be assigned to one principal diagnostic category for this table. Associated or coexisting abnormalities should be reported separately where applicable.

AI Interpretation and Agreement with Reference Diagnosis

The binary diagnostic performance of the AI system was evaluated by comparing its classification of shoulder MRI examinations as normal or abnormal with the recorded reference diagnosis. Of the 60 examinations, 54 were classified as abnormal and six as normal according to the reference diagnosis. Based on the recorded classifications, the AI system identified all 54 abnormal examinations, resulting in 54 true-positive and zero false-negative classifications. Among the six reference-normal examinations, five were classified as normal and one was classified as having minimal fluid or a possible minor finding, resulting in five true-negative and one false-positive classification.

The resulting sensitivity was 100.0%, specificity 83.3%, PPV 98.2%, NPV 100.0%, and overall binary accuracy 98.3%. Cohen's kappa was 0.90, indicating high agreement beyond chance for this binary classification in the study sample.

These results describe the ability of the AI system to distinguish broadly between normal and abnormal examinations. They should not be interpreted as demonstrating equivalent performance for individual shoulder disorders. In some cases, the AI system identified abnormalities but did not reproduce the principal diagnosis or the complete lesion characterization documented in the reference diagnosis. Accordingly, binary abnormality detection and lesion-specific diagnostic performance were evaluated as distinct outcomes.

Table 3. Binary Diagnostic Performance of the AI System for Normal-versus-Abnormal Shoulder MRI Classification

Reference diagnosis	AI abnormal	AI normal	Total
Abnormal	54	0	54
Normal	1	5	6
Total	55	5	60

The reference diagnosis was used as the comparator. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated from the confusion matrix. Cohen's kappa represents agreement for the binary classification.

DISCUSSION

Magnetic resonance imaging (MRI) is an important non-invasive imaging modality for evaluating shoulder disorders because it provides detailed visualization of the rotator cuff, labrum, cartilage, muscles, tendons, ligaments, bone marrow, and surrounding soft tissues in multiple planes. Shoulder MRI can demonstrate a wide range of abnormalities, including rotator cuff tendinopathy and tears, adhesive capsulitis, instability-related lesions, bursitis, fractures, inflammatory conditions, and neoplastic lesions. Previous studies have emphasized the importance of good-quality MRI acquisition and systematic interpretation for accurate assessment of these abnormalities [1,3,8,9]. In the present study, shoulder MRI findings were evaluated in 60 patients, and the agreement between AI-based interpretation and the recorded final diagnosis was assessed.

Demographic characteristics

The present study included 60 patients, with a mean age of 44.18 ± 15.15 years and an age range of 18–76 years. Males comprised 63.3% of the study population, while females accounted for 36.7%, giving a male-to-female ratio of approximately 1.73:1. The higher proportion of males may partly reflect the clinical population referred for shoulder MRI, particularly as several cases in the dataset involved trauma and rotator cuff-related abnormalities. Previous literature shows that shoulder MRI is used for a broad range of traumatic, degenerative, and pain-related conditions rather than for one specific clinical presentation [1,2,8].

The age distribution in the present study also reflects the varied nature of shoulder pathology. The patients ranged from young adults to individuals in the seventh decade and beyond. Younger patients included cases of recurrent anterior instability, osteochondroma, inflammatory synovitis, and traumatic injuries, whereas degenerative tendon abnormalities, rotator cuff tears, and adhesive capsulitis were more commonly seen among older patients. This wide age-related distribution is consistent with the established role of MRI in both traumatic and degenerative shoulder conditions [1,8,12].

Spectrum of shoulder abnormalities

Supraspinatus tendinosis was the most frequently recorded final diagnosis in the present study, accounting for 11.7% of cases. This was followed by normal studies (10.0%), massive rotator cuff tears (6.7%), partial subscapularis tears (6.7%), and partial-width full-thickness supraspinatus tears (5.0%). Overall, rotator cuff abnormalities formed a major part of the study population.

The predominance of supraspinatus-related pathology is consistent with previous studies describing the supraspinatus tendon as one of the structures most commonly affected in rotator cuff disease [6,8,14]. Hu et al. reported a higher frequency of supraspinatus involvement among patients with rotator cuff tears and also found that increasing MRI-based tear severity was associated with poorer shoulder outcome scores [14]. In the present study, supraspinatus abnormalities also covered a broad range, from tendinosis and tendon sprain to partial- and full-thickness tears and massive rotator cuff tears.

The master chart further showed that shoulder abnormalities often occurred together rather than as isolated lesions. Several patients had supraspinatus pathology associated with subscapularis tendinopathy, bursitis, adhesive capsulitis, acromioclavicular joint changes, or bone abnormalities. This highlights the need for a systematic approach when interpreting shoulder MRI. Amroudi emphasized that clinically useful shoulder MRI reporting should consider the location and extent of a tear, associated muscle atrophy and fatty degeneration, acromial morphology, acromioclavicular joint changes, and labral abnormalities [3]. Similar combinations of primary and secondary findings were observed in the present dataset.

Rotator cuff pathology

Rotator cuff abnormalities were an important part of the present study. Partial- and full-thickness supraspinatus tears, massive rotator cuff tears, subscapularis tears, and rotator cuff tendinopathy were identified. In several patients, complete supraspinatus and infraspinatus tears were accompanied by muscle atrophy, tendon retraction, and other associated abnormalities, showing the range of severity of rotator cuff disease within the study population.

These findings are in agreement with previous studies supporting MRI as an important modality for detecting rotator cuff tears and assessing their extent. Darai et al. compared routine MRI findings with surgical findings and emphasized the importance of assessing the site, size, shape, and muscle involvement in rotator cuff tears [6]. Arnold et al. similarly evaluated the diagnostic value of MRI in the preoperative assessment of rotator cuff and labral lesions [4]. The present findings therefore support the need to evaluate shoulder MRI beyond simply identifying whether a tear is present or absent.

An important finding from the AI analysis was that AI generally recognized major rotator cuff abnormalities but did not consistently reproduce every part of the recorded final diagnosis. For example, in one case of a partial-width full-thickness supraspinatus tear, AI identified the tear and associated tendinopathy, whereas in another case with the same major diagnosis, the AI mainly reported subscapularis bursal distension. Similarly, a case recorded as a partial subscapularis tear was interpreted by AI mainly as supraspinatus tendinopathy with minimal bursal fluid. These observations show that detecting an abnormality and providing its complete characterization are separate aspects of diagnostic performance.

Adhesive capsulitis and associated findings

Adhesive capsulitis was identified either as an isolated diagnosis or along with rotator cuff pathology in several patients. The master chart included cases of early adhesive capsulitis, adhesive capsulitis with supraspinatus tendinosis, and adhesive capsulitis associated with rotator cuff tendinosis or subscapularis pathology.

MRI can demonstrate structural changes related to adhesive capsulitis and provide additional information about the structures involved. Coşar et al. evaluated MRI findings in patients with adhesive capsulitis and examined their relationship with treatment outcomes [5]. The presence of adhesive capsulitis together with rotator cuff abnormalities in the present study also emphasizes the importance of recognizing coexisting findings instead of attributing the patient's symptoms to a single MRI abnormality.

The AI system recognized adhesive capsulitis in several cases where features such as ligamentous or capsular thickening and associated rotator cuff changes were present. However, the present master chart does not contain a standardized lesion-by-lesion AI scoring system. Therefore, the ability of AI to identify all individual MRI features of

adhesive capsulitis cannot be reliably quantified. This should be considered when interpreting the overall AI performance.

Traumatic and instability-related abnormalities

Several traumatic abnormalities were present in the study population, including scapular fractures, greater tuberosity avulsion fractures, bone contusions, rotator cuff tears, Bankart lesions, Hill-Sachs lesions, and anterior shoulder instability. These findings demonstrate the usefulness of MRI in patients presenting with both acute and chronic traumatic shoulder disorders.

Allam et al. emphasized the role of MRI in acute shoulder trauma and its ability to demonstrate a wide range of soft-tissue and osseous abnormalities [1]. In the present dataset, MRI demonstrated complex traumatic patterns, including scapular fractures associated with multiple rotator cuff tendon tears and greater tuberosity avulsion associated with partial supraspinatus tearing. Such findings show the importance of comprehensive assessment in traumatic shoulder MRI.

Instability-related abnormalities were represented by Bankart and Hill-Sachs lesions. Amroudi emphasized that evaluation of shoulder instability should include the type and extent of labral abnormality, glenoid rim changes, Hill-Sachs lesions, and associated rotator cuff pathology [3]. The present master chart included a case of recurrent anterior shoulder dislocation with both bony Bankart and Hill-Sachs lesions, as well as another case of chronic anterior shoulder instability with Bankart and Hill-Sachs lesions. AI identified these major instability-related abnormalities, indicating its potential usefulness in recognizing prominent structural lesions.

MRI pulse sequences and acquisition parameters

The MRI acquisition parameters showed considerable variation in repetition time (TR) and echo time (TE). The mean TR was 2680.46 ± 1493.53 ms, ranging from 3.6 to 4908 ms, while the mean TE was 65.66 ± 38.90 ms, with values ranging from 1.7 to 100 ms. A 90° flip angle was used in 59 of the 60 examinations, while one examination used a 10° flip angle. Proton density fat-suppressed imaging was the most frequently used sequence in the dataset.

The frequent use of PD fat-suppressed imaging is appropriate for assessing tendon, ligamentous, marrow, and soft-tissue abnormalities commonly encountered in shoulder MRI. Farjoodi et al. emphasized that understanding pulse-sequence characteristics and using a systematic approach are important for obtaining diagnostically useful musculoskeletal MRI examinations [9]. Farber et al. also discussed the importance of appropriate MRI techniques for evaluating different shoulder abnormalities [8].

The present dataset also indicates that individual sequences or parameters should not be considered separately from the clinical and anatomical context. Fat-suppressed sequences were commonly represented in examinations involving rotator cuff tears, tendinopathy, bursitis, marrow edema, and traumatic abnormalities. Axial, coronal, and combined sequences also helped in demonstrating different anatomical structures.

Guerhazi et al. showed that sequence selection can influence the assessment of specific shoulder structures, particularly the cartilage and labrum, and evaluated T1-weighted and T2-weighted fat-suppressed protocols for MR arthrography [11]. Although MR arthrography was not specifically evaluated in the present study, these findings support the general importance of protocol selection when a particular anatomical structure is the main diagnostic target.

Relationship between MRI quality and AI interpretation

The available MRI acquisition data included repetition time (TR), echo time (TE), flip angle, and sequence information. However, numerical measurements of the number of excitations (NEX), signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR) were not available in the study dataset. Consequently, a reliable quantitative assessment of the relationship between these image-quality parameters and AI diagnostic performance could not be performed. The recorded acquisition parameters are therefore presented descriptively, and no independent association between acquisition characteristics and diagnostic performance is inferred from the available data. Future studies should incorporate standardized quantitative image-quality measurements and predefined scoring criteria to evaluate these relationships. This limitation is relevant because MRI image quality depends on several acquisition and reconstruction factors. Herrmann et al. demonstrated that accelerated MRI using deep-learning reconstruction can reduce examination time while maintaining image quality and diagnostic performance comparable to standard imaging protocols [13]. Similarly, compressed-sensing techniques have increased the possibilities for accelerated MRI acquisition [10]. These developments suggest that AI and newer reconstruction methods may have an impact not only on interpretation but also on MRI acquisition and image quality.

Despite this, the present study should not assume an association between TR or TE and diagnostic performance based only on the available master chart. Such an assessment would require standardized quantitative image-quality measurements, particularly SNR and CNR, for each examination.

AI interpretation and diagnostic agreement

The present study demonstrated high binary abnormality-detection performance by the AI system when compared with the recorded reference diagnosis. The sensitivity was 100.0%, specificity 83.3%, PPV 98.2%, NPV 100.0%, and overall accuracy 98.3%, with a Cohen's kappa of 0.90 for normal-versus-abnormal classification. These findings indicate that the AI system identified all examinations classified as abnormal in this dataset, although one reference-normal examination was classified as abnormal.

The high binary accuracy should be interpreted in the context of the study population, in which 90.0% of examinations were classified as abnormal. Accuracy and predictive values depend on the distribution of normal and abnormal

examinations and may differ in populations with a lower prevalence of shoulder pathology. The specificity estimate is based on only six reference-normal examinations and is therefore imprecise.

Furthermore, correct detection of an abnormal examination does not necessarily indicate correct identification of the underlying lesion. The case-wise findings showed that the AI system did not consistently reproduce the principal diagnosis or the complete characterization of associated abnormalities. Therefore, binary abnormality detection and lesion-specific diagnostic performance should be considered separate outcomes. Larger studies with predefined diagnostic categories and independent reference assessments are required to establish the generalizability of these findings.

AI recognition of specific abnormalities

The qualitative comparison between AI interpretation and the final diagnosis provides information beyond the binary performance measures. In several cases, AI reproduced the principal diagnosis with considerable detail. Supraspinatus tendinosis, osteochondroma, full-thickness supraspinatus tear, massive rotator cuff tear, Bankart and Hill-Sachs lesions, scapular fracture, and adhesive capsulitis were recognized in multiple cases [present dataset].

However, some cases showed only partial agreement with the final diagnosis. In one patient with supraspinatus tendinosis and bursitis, AI emphasized bursitis and described the supraspinatus tendon as grossly intact. In another patient with a lipoma over the left upper back, AI reported mild supraspinatus tendinopathy and bursal fluid rather than the recorded lipoma. Similarly, one partial subscapularis tear was interpreted mainly as supraspinatus tendinopathy with minimal bursal fluid. These examples indicate that AI may identify secondary or associated findings while failing to reproduce the principal diagnosis.

This is particularly relevant in complex shoulder MRI, where useful interpretation requires not only detecting abnormalities but also prioritizing and characterizing them. Amroudi emphasized that clinically relevant reporting should communicate the type, location, and extent of tears and associated structural abnormalities [3]. Future AI assessment should therefore extend beyond binary abnormality detection and include lesion-specific sensitivity, specificity, and agreement for structures such as the supraspinatus, subscapularis, infraspinatus, labrum, bursa, and osseous structures.

Comparison with previous literature

The findings of the present study are broadly consistent with previous literature highlighting the importance of MRI in evaluating shoulder pain, trauma, rotator cuff disease, instability, and adhesive capsulitis [1–9]. The high frequency of supraspinatus-related abnormalities in the present dataset is also consistent with previous observations that the supraspinatus is commonly involved in rotator cuff pathology [6,14]. The presence of bursitis, impingement-related changes, and acromioclavicular abnormalities is similarly compatible with studies examining MRI findings in relation to shoulder symptoms [15].

Jäschke et al. reported that several MRI parameters, including subacromial and subdeltoid bursal dimensions and other structural features, may be associated with shoulder pain and functional impairment [15]. The current master chart also contains several cases in which supraspinatus tendinopathy was associated with bursitis, a reduced acromiohumeral interval, AC joint arthropathy, or impingement-related changes. These findings support the view that shoulder symptoms may be related to a combination of structural abnormalities rather than one isolated lesion.

The present study also emphasizes the importance of MRI protocol optimization. Farjoodi et al. highlighted systematic interpretation and knowledge of pulse sequences [9], while Guerhazi et al. showed that sequence selection can influence the assessment of specific shoulder structures [11]. More recent work by Herrmann et al. has also demonstrated the potential of accelerated acquisition and deep-learning reconstruction to maintain diagnostic performance while improving scanning efficiency [13]. Together, these findings provide useful context for the integration of AI into shoulder MRI workflows.

Clinical implications

The findings suggest that AI-based interpretation may have potential as an adjunct to shoulder MRI assessment, particularly for identifying examinations containing abnormalities. However, the results do not establish that the AI system can replace radiologist interpretation. The observed discrepancies in principal diagnosis and lesion characterization emphasize the importance of reviewing AI-generated findings in the anatomical and clinical context.

If independent radiologist interpretations are available, the clinical implications should additionally discuss the observed areas of agreement and disagreement between the radiologist and AI, supported by the relevant comparison statistics. Without such a direct comparison, the clinical conclusions should remain limited to AI performance against the recorded reference diagnosis. Further validation in larger, preferably multicentre cohorts is necessary before drawing conclusions about clinical implementation.

LIMITATIONS

Although the study included radiologist interpretation, the analysis was limited by the availability and standardization of the recorded diagnostic classifications. The methods used to compare the radiologist's findings with AI outputs and the reference diagnosis should be clearly documented. Any missing or non-standardized lesion-level data may limit the reliability of direct diagnostic comparisons. Future studies should use a predefined reporting template and independent, blinded interpretation procedures. **Future Directions**

Future research should use larger prospective cohorts with standardized MRI protocols and independent AI–radiologist comparison. Lesion-specific evaluation and assessment of sensitivity, specificity, accuracy, and agreement are needed.

Combining AI interpretation with accelerated MRI and deep-learning reconstruction may further improve diagnostic reliability and workflow efficiency.

CONCLUSION

This study evaluated AI-based interpretation of 60 shoulder MRI examinations using the recorded final diagnosis as the reference. For binary normal-versus-abnormal classification, the AI system demonstrated a sensitivity of 100.0%, specificity of 83.3%, PPV of 98.2%, NPV of 100.0%, and overall accuracy of 98.3%, with a Cohen's kappa of 0.90. These findings indicate high abnormality-detection performance within the study sample.

However, the case-wise assessment identified discrepancies in lesion-specific characterization, including the identification of full-thickness supraspinatus tears. Binary abnormality detection should therefore not be interpreted as equivalent to accurate characterization of individual shoulder disorders. The findings are limited by the small sample size, the distribution of diagnostic categories, and the availability of standardized comparison data.

AI may have a supportive role in shoulder MRI interpretation, but its comparative performance should be established through appropriately designed studies involving independent radiologist interpretation, predefined reference standards, and lesion-specific diagnostic assessment. Larger studies with standardized MRI acquisition and quantitative image-quality measurements are needed to determine the generalizability and clinical relevance of the observed results.

REFERENCES

1. Ibrahim AM, Allam ES, Kamar SA. Recent MRI study of acute shoulder trauma. *Med J Cairo Univ.* 2019;87(7):4247-4254. doi:10.21608/mjcu.2019.77413.
2. Shruthi GT, Amith R, Chowdhary AS, Sundee Sa C, Alemao NN, et al. Role of MRI in rotator cuff injuries. *J Evid Based Med Healthc.* 2018;5(32):2394-2401. doi:10.18410/jebmh/2018/494.
3. Nakhaei Amroudi M. What the orthopedic surgeon wants to know in shoulder MRI? *Iran J Radiol.* 2014;11(S1):e21299. doi:10.5812/iranjradiol.21299.
4. Arnold E, Zollner H, Schank T, Schaeffeler C, Greiner S, Gohlke F, et al. Aussagekraft der Magnetresonanztomographie bei der präoperativen Diagnostik von Rotatorenmanschetten- und Labrumläsionen der Schulter [Diagnostic value of magnetic resonance imaging in preoperative diagnosis of rotator cuff and labral lesions of the shoulder]. *Obere Extrem.* 2009;4(3):164-169. doi:10.1007/s11678-009-0034-8.
5. Coşar SN, Ozen S, Kurtcebe AN. Prognostic value of magnetic resonance imaging in the evaluation of physical therapy outcomes in patients with adhesive capsulitis. *Turk J Phys Med Rehabil.* 2021;67(2):146-154. doi:10.5606/tftrd.2021.5139.
6. Darai N, Pokhrel S, Shu R, Zhang X, Liu J, Teng G. Comparison with surgical findings for the accuracy of routine MRI in rotator cuff tears. *Open J Radiol.* 2016;6(2):73-83. doi:10.4236/ojrad.2016.62011.
7. Dirim B. Shoulder radiology. In: Bozkurt M, Açar HI, editors. *Clinical anatomy of the shoulder: an atlas.* Cham: Springer; 2017. p. 25-56. doi:10.1007/978-3-319-53917-1.
8. Farber JM, Fayad L, Johnson T, Cascio B, Shindle M, Neubauer PR, et al. Magnetic resonance imaging of the shoulder: current techniques and spectrum of disease. *J Bone Joint Surg Am.* 2006;88 Suppl 4:64-79. doi:10.2106/JBJS.F.00583.
9. Farjoodi P, Boone S, Listerneck Z, Lubowitz JH. Magnetic resonance imaging of the musculoskeletal system: basic science, pulse sequences, and a systematic approach to image interpretation. *J Bone Joint Surg Am.* 2010;92(5):1282-1293. doi:10.2106/JBJS.J.00783.
10. Fritz J, Gross JS, Matcuk GR. Compressed sensing MRI. *Adv Clin Radiol.* 2020;2:257-271. doi:10.1016/j.yacr.2020.06.008.
11. Guerhazi A, Hayashi D, Roemer FW, Niu J, Quinn EK, Crema MD, et al. MR arthrography of the shoulder: optimizing pulse sequence protocols for the evaluation of cartilage and labrum. *Eur J Radiol.* 2014;83(6):1051-1054. doi:10.1016/j.ejrad.2014.04.030.
12. Hegazi TM, Wu JS. Shoulder. In: *Musculoskeletal MRI: a case-based approach to orthopedic, sports, and emergency radiology.* Cham: Springer; 2019.
13. Herrmann J, Feng YS, Gassenmaier S, Grunz JP, Koerzdoerfer G, Lingg A, et al. Fast 5-minute shoulder MRI protocol with accelerated TSE-sequences and deep learning image reconstruction for the assessment of shoulder pain at 1.5 and 3 Tesla. *Eur J Radiol Open.* 2024;12:100557. doi:10.1016/j.ejro.2024.100557.
14. Hu X, Wang X, Mao W, Ying L, Huang Z. Magnetic resonance imaging classifications of rotator cuff tear are associated with different shoulder outcome scores. *J Nanomater.* 2021;2021:9918812. doi:10.1155/2021/9918812.
15. Jäschke M, Laux CJ, Seeli F, Scherr J, Zingg PO, Farshad M. Subacromial impingement syndrome: association of multiple magnetic resonance imaging parameters with shoulder function and pain. *Arch Orthop Trauma Surg.* 2021;141(12):2271-2279. doi:10.1007/s00402-021-04032-6.
16. Jo SY, Sammet S, Thomas S, Stacy GS. Musculoskeletal MRI pulse sequences: a review for residents and fellows. *Radiographics.* 2019;39(7):2038-2039. doi:10.1148/rg.2019190075.
17. Joo Y, Cho HR, Kim YU. Evaluation of the cross-sectional area of acromion process for shoulder impingement syndrome. *Korean J Pain.* 2020;33(1):60-65. doi:10.3344/kjp.2020.33.1.60.
18. Koff MF, Burge AJ, Potter HG. Clinical magnetic resonance imaging of arthroplasty at 1.5 T. *J Orthop Res.* 2020;38(7):1455-1464. doi:10.1002/jor.24606.
19. Kvalvaag E, Anvar M, Karlberg AC, Brox JJ, Engebretsen KB, Soberg HL, et al. Shoulder MRI features with clinical correlations in subacromial pain syndrome: a cross-sectional and prognostic study. *BMC Musculoskelet Disord.* 2017;18:469. doi:10.1186/s12891-017-1827-3.

20. Palmer WE. Imaging of the shoulder I. In: Imaging of the shoulder. Cham: Springer; 2009. p. 1-24. doi:10.1007/978-88-470-1378-0_1.
21. Park HJ, Lee SY, Choi YJ, Park JH, Kim E. Association between subacromial impingement and acromiohumeral distance on MRI. Iran J Radiol. 2018;15(2):e13811. doi:10.5812/iranjradiol.13811.
22. Saraya S, El Bakry R. Ultrasound: can it replace MRI in the evaluation of rotator cuff tears? Egypt J Radiol Nucl Med. 2016;47(1):193-201.
23. Xie H, Shan H, Wang G, Zhang Y, Cai W, Guo J, et al. Prospective comparison of standard and deep learning-reconstructed turbo spin-echo MRI of the shoulder. Radiology. 2024;310(1):e231405. doi:10.1148/radiol.231405.