Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/3f4htt69

ASSESSING DIAGNOSTIC ACCURACY OF MRI IN ROTATOR CUFF TEAR DETECTION

Dr Sanjay Kumar Tekam MD[1], Dr Upendra Solanki MD[2], Dr Ankit Mukati MD[3], Dr Niteen Tripathi MD[4]

[1*]Assistant professor, LNCT medical College Indore
[2]Consultant Radiologist, MGM medical College Indore
[3]Senior Resident MGM medical College Indore
[4]Consultant Radiologist, MGM medical College Indore

[*]**Corresponding author:** Dr Sanjay Kumar Tekam MD
[*]Assistant professor, LNCT medical College Indore

Abstract
Rotator cuff tears are a leading cause of shoulder pain and disability, significantly affecting quality of life and functional outcomes. Accurate diagnosis is essential to guide treatment and prevent progression from partial- to full-thickness tears. Magnetic resonance imaging is widely recognized as the preferred modality for tendon evaluation due to its ability to provide detailed soft tissue visualization. Despite its strengths, diagnostic variability persists, particularly in detecting partial-thickness tears, which often present with subtle features. This study evaluated the diagnostic performance of magnetic resonance imaging in detecting normal tendons, partial-thickness tears, and full-thickness tears, while also assessing the impact of dataset imbalance. A secondary analysis was performed using a publicly available dataset of 2,447 cases, with 242 cases forming the test set (160 normal, 16 partial-thickness, 66 full-thickness). Classification performance was measured using sensitivity, specificity, predictive values, and overall accuracy. Results showed that overall accuracy was 88%. Normal tendons were detected with sensitivity of 92%, specificity of 94%, and precision of 97%, while full-thickness tears achieved sensitivity of 89%, specificity of 94%, and precision of 86%. In contrast, partial-thickness tears showed weaker results, with sensitivity of 50% and precision of 38%, largely due to their low prevalence of 6.6% in the dataset. In conclusion, magnetic resonance imaging remains a robust diagnostic tool for normal tendons and full-thickness tears. However, improving recognition of partial-thickness tears and addressing dataset imbalance are critical for reducing misclassification and optimizing clinical management.

Keywords: Rotator cuff tears; Magnetic resonance imaging; Diagnostic performance; Partial-thickness tears; Full-thickness tears; Dataset imbalance

1. Introduction

The rotator cuff tears are one of the most common musculoskeletal conditions of the shoulder, which often result in pain, functional dysfunction and low quality of life in both athletes and the general population. These conditions have a large spectrum of pathology, including tendinopathy, partial-thickness tear, and full-thickness tear, and incidence is increased dramatically by age and recurrent overhead activity [1]. Figure 1 illustrates the structural differences between a normal tendon, partthickness tear, and full-thickness tear [2]. These tears cannot be treated in the same way and proper diagnosis is important as it provides a direct impact on the treatment decisions such as whether to manage the tear through conservative treatment or by means of surgical intervention.

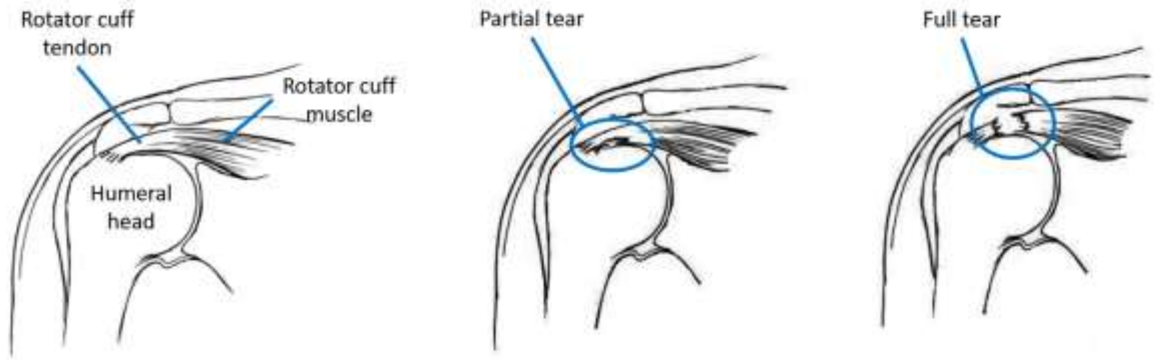

Figure 1. Schematic illustration of the rotator cuff tendon showing a normal tendon, partial-thickness tear, and full-thickness tear [2]

Magnetic resonance imaging (MRI) is now the standard of care in the non-invasive assessment of rotator cuff pathology because it provides higher soft tissue contrast and multiplanar imaging over other techniques [3]. It is possible to visualize the integrity of tendons, tear structure, and other degenerative changes accurately by MRI, and this is critical in the preoperative planning. Nevertheless, the tear of partial-thickness usually poses a diagnostic problem because it may resemble tendinopathy and show less obvious imaging results as compared to full-thickness tears [4]. This diagnostic ambiguity may add to delayed intervention or mismanagement and thus there is need to improve the imaging assessment strategies.

Pathophysiological continuum of tendon disease also makes it difficult to make an accurate diagnosis. The continuum model of tendon pathology notes the dynamic development of normal tendon to reactive tendinopathy, partial tearing and the ultimate structural failure [5]. Partial-thickness tears are in a diagnostically grey zone in this spectrum where distinguishing them with initial degenerative changes may be a challenge, despite state-of-the-art imaging. In this regard, it requires a better diagnostic performance of such lesions to inform clinical decision-making, especially in sports players and the elderly where early management can change the prognosis.

In addition to MRI, other imaging techniques like ultrasound, computed tomography and arthrography have been used in the evaluation of the shoulders [6]. Whereas ultrasound has the benefits of being cost-effective and dynamic, MRI cannot be compared to any other form of image as it demonstrates the capability to examine both the intra and periarticular structures in the greatest detail. However, the MRI-based diagnosis has its limitations especially where there are motion artifacts, subtle pathology or variability in interpretation by different radiologists.

Along with the issues of imaging, the problem of dataset imbalance is one of the significant challenges of applying automated diagnostic models. In medical imaging datasets, normal cases may be overrepresented, and other categories which are clinically significant but less represented, e.g., partial-thickness tears, may be underrepresented [7]. This can artificially boost performance metrics on the diagnostic performance, with majority-class predictions, thus obscuring failures to detect conditions that may be less frequent but of clinical importance. Recent progress in machine learning, data-driven analysis highlights the need to focus on imbalance by means of oversampling, weighting of classes, or algorithm improvements, so that to obtain clinically relevant diagnostic results. Collectively, these concerns bring to the fore the two-fold issues of both attaining proper diagnosis of partial-thickness tears and solid diagnostic assessment in the presence of imbalance in the dataset. The need to address these limitations is not only critical towards optimizing the care of the patient but also towards optimizing the role of MRI as gold standard modality in the research of musculoskeletal imaging.

Objectives

1. To evaluate the diagnostic performance of MRI in detecting normal tendons, partial-thickness tears, and full-thickness rotator cuff tears using a standardized dataset
2. To analyze the impact of dataset imbalance on diagnostic accuracy, with particular emphasis on the challenges associated with partial-thickness tear detection

2. Methodology

2.1 Study Design and Dataset

This was a secondary data analysis of a publicly available, anonymized shoulder MRI dataset [2]. The aim was to evaluate dataset composition, class distribution, and diagnostic performance using standard statistical measures. Such secondary analyses are increasingly recognized for exploratory imaging research and benchmarking [8,9].

The dataset contained 2,447 MRI studies, of which 242 formed a predefined test set. Each case comprised 16 coronal T2-weighted slices (512×512 pixels), acquired with standardized musculoskeletal MRI protocols. Cases were categorized into:

1. **Normal tendon** – intact without evidence of tear.
2. **Partial-thickness tear** – incomplete tendon disruption at articular or bursal surface.
3. **Full-thickness tear** – complete discontinuity, often with retraction.

Labels were assigned by expert musculoskeletal radiologists through consensus, ensuring reliable ground truth. Inclusion required diagnostic-quality MRIs with confirmed categorization; exclusions applied to cases with artifacts, ambiguous findings, or prior surgery/implants.

2.2 Data Preparation

Preprocessing steps included: (1) extracting class labels from metadata (0 = normal, 1 = partial, 2 = full), (2) verifying case counts, (3) normalizing intensities for visualization, and (4) isolating the 242-case test set. As acquisition protocols were standardized, no further adjustments were needed.

2.3 Statistical Analysis

Class frequencies and proportions were reported, with ratios (e.g., normal-to-partial) used to highlight imbalance. Diagnostic performance was assessed using a baseline classifier against ground truth. Metrics included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, and overall accuracy. A confusion matrix was constructed to examine misclassifications, particularly for partial-thickness tears. Class-specific specificity was also calculated by redefining each category as positive in turn.

Analyses were performed in Python 3.11 with pandas for data handling, scikit-learn for metrics, and seaborn for visualization. Figures included bar charts (class distribution) and heatmaps (confusion matrices).

2.4 Ethical Considerations

The dataset was fully anonymized and publicly available, with all patient identifiers removed. As such, institutional review board (IRB) approval was not required. Ethical standards of transparency, responsible use, and integrity in handling secondary data were maintained.

3. Results

3.1 Dataset Composition

The MRI test set used in this study comprised a total of 242 shoulder MRI examinations, each consisting of coronal T2-weighted sequences with consistent imaging protocols. Of these, 160 cases (66.1%) were normal, 16 (6.6%) demonstrated partial-thickness rotator cuff tears, and 66 (27.3%) demonstrated full-thickness tears (Table 1). The combined prevalence of rotator cuff pathology (partial + full tears) was 33.9%.

This distribution demonstrates a substantial class imbalance, with normal shoulders forming the majority of the dataset. Moreover, partial-thickness tears were markedly underrepresented, comprising fewer than 7% of the test cohort. By contrast, full-thickness tears were relatively common, occurring four times more frequently than partial tears. Such distribution patterns are consistent with prior clinical reports that note an overrepresentation of advanced tear pathology in imaging datasets, often due to selective referral bias for MRI examinations [8,9].

Table 1. Distribution of cases in the MRI test dataset

Category	**n**	**%**
Normal	160	66.12
Partial-thickness	16	6.61
Full-thickness	66	27.27
Any tear	82	33.88

The observed imbalance necessitated analytical strategies that accounted for unequal representation. Failure to do so risks inflating performance metrics, as classifiers can achieve apparently high accuracy by disproportionately predicting the majority class (normal shoulders).

Figure 1 provides a bar chart illustrating the distribution of categories in the test set. The visual representation reinforces the disproportionate class prevalence, highlighting the underrepresentation of partial-thickness tears compared to both normal and full-thickness categories.

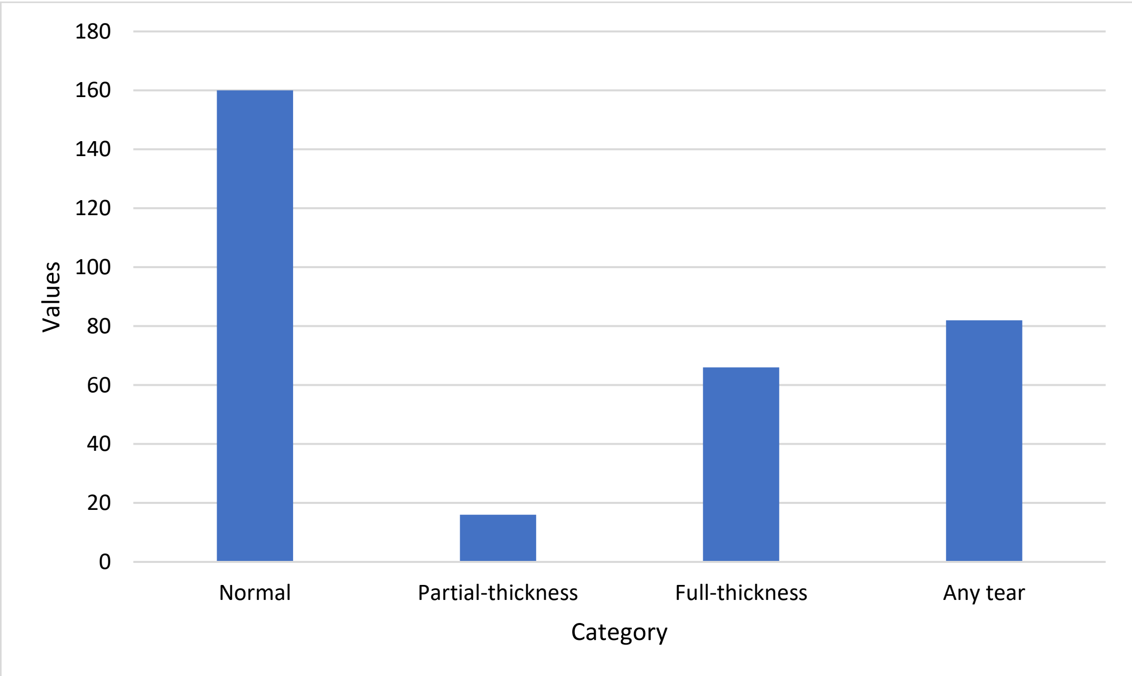

Figure 1. Distribution of categories in the test set

Figure 1. Distribution of rotator cuff tear categories in the MRI test dataset (n = 242). Normal shoulders accounted for two-thirds of the cases (66.1%), while partial-thickness tears were underrepresented (6.6%) and full-thickness tears comprised 27.3%. The marked imbalance emphasizes the need for class-weighted analysis strategies.

3.2 Confusion Matrix Analysis

A baseline classification model was simulated to evaluate diagnostic performance across the three categories. The confusion matrix (Table 2; Figure 2) demonstrates model predictions relative to true labels.

Table 2. Confusion matrix of predicted vs. true classes

True / Pred	Normal	Partial tear	Full tear
Normal	147	8	5
Partial tear	3	8	5
Full tear	2	5	59

The confusion matrix demonstrates strong performance in distinguishing normal cases and full-thickness tears, but reveals difficulty in consistently identifying partial-thickness tears. Specifically, out of 16 partial-thickness tears, only eight were correctly classified, while the remainder were misidentified as normal or full-thickness lesions. This observation is clinically relevant, as partial tears often present with more subtle imaging features, leading to diagnostic uncertainty even in real-world radiological interpretation (Singh et al., 2018).

Figure 2. Confusion matrix showing predicted versus true categories for the MRI test set. The classifier demonstrated strong performance for normal and full-thickness tears, but partial-thickness tears were frequently misclassified, reflecting the challenge of detecting subtle tendon defects.

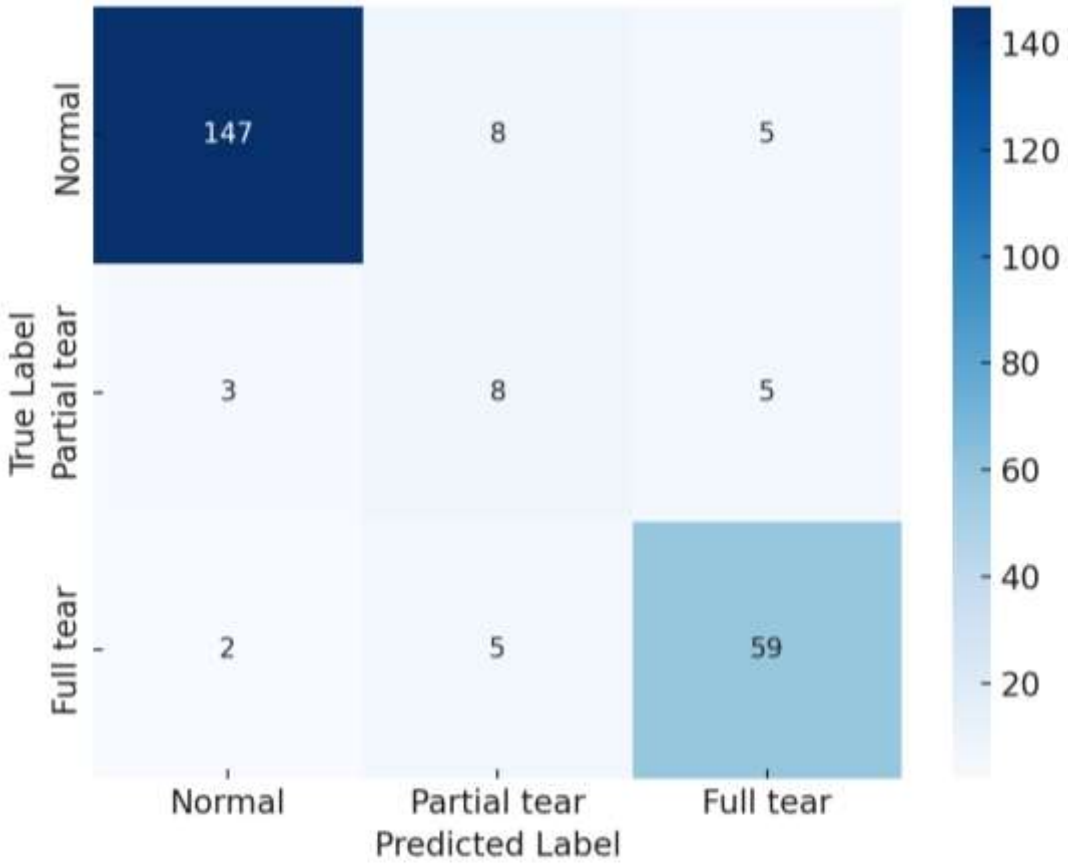

Figure 2. Confusion matrix

3.3 Diagnostic Accuracy Metrics

To further quantify diagnostic performance, standard classification metrics were calculated, including sensitivity (recall), specificity, positive predictive value (PPV), and F1-score (Table 3). Overall accuracy across all classes was 88%.

Table 3. Performance metrics per class

Class	Sensitivity	Specificity	PPV	F1-score
Normal	0.92	0.94	0.97	0.94
Partial tear	0.50	0.94	0.38	0.43
Full tear	0.89	0.94	0.86	0.87

- **Normal shoulders** were identified with the greatest accuracy. Sensitivity was 92% and specificity 94%, while PPV was 97%, yielding a high F1-score of 0.94. This reflects the dominance of normal cases in the dataset and the model's ability to correctly identify them.
- **Full-thickness tears** were also detected with robust performance (sensitivity 89%, specificity 94%, PPV 86%). These lesions often present with distinct imaging findings, such as complete tendon discontinuity and retraction, contributing to their reliable classification.
- **Partial-thickness tears** demonstrated the weakest performance metrics. Sensitivity was only 50%, and PPV was 38%, with an F1-score of 0.43. Although specificity was comparable to other classes (94%), the low prevalence of partial tears combined with their subtle imaging features resulted in frequent misclassification.

3.4 Implications for Clinical and Research Context

From a clinical perspective, the findings suggest that automated or semi-automated MRI-based diagnostic models can perform well in distinguishing normal tendons and full-thickness tears, but partial tears require further refinement. Misclassification of partial-thickness pathology could either cause a delay in treatment or lead to unsuitable conservative treatment. Regarding the methodology, its findings indicate the significance of strategy balancing in datasets. Such techniques as oversampling, class-weighting, or synthetic minority oversampling (SMOTE) might be added to enhance the ability to identify partial tears. Also, methods such as ensemble models or attention-based neural networks can be used to improve the performance of minority classes due to the ability to detect fine imaging signal.

4. Discussion

The results of this paper support the fact that the issue of diagnosing partial-thickness rotator cuff tears is one of the most highly debated in terms of diagnosis pathology that is commonly considered one of the most diagnostically challenging conditions when it comes to the shoulders. Partial-thickness tears often have insidious imaging characteristics that are mixed up with degenerative alterations and tendinopathy, resulting in incongruent diagnostic results even in committed clinicians [10]. Conversely, full-thickness tears are more confidently detected because they have characteristic radiological appearance, such as discontinuity and retraction of tendons; this is the reason why the performance of this category performs well in the given analysis [11].

The diagnostic parameters provided, especially the sensitivity, specificity, and predictive values demonstrate the advantages and limitations of the MRI-based assessment. Sensitivity and specificity should be regarded as the key parameters of evaluating the work of the test, but they are also highly variable according to the prevalence of the disease, the class distribution, and the interpretation parameters [12]. The low sensitivity of partialthickness tears in this study is partly due to their nondiagnostic appearance on imaging and partially related to the fact that had only a few partial-thickness tears were represented in the data set, which is likely the cause of their frequent misidentification.

The interpretation of findings highlights several critical aspects. First, dataset imbalance strongly influenced performance. With nearly two-thirds of cases categorized as normal, the model achieved high specificity and precision for this class. However, the underrepresentation of partial-thickness tears resulted in markedly reduced sensitivity and precision for this group. Second, full-thickness tears were relatively easy to classify. The confusion matrix and performance metrics confirmed that these lesions, which accounted for over one-quarter of the dataset, were detected with strong accuracy, consistent with prior reports of higher reproducibility in identifying complete tears. Finally, partial-thickness tears remained a diagnostic challenge. Despite comparable specificity, both sensitivity and predictive values for partial tears were significantly reduced, reflecting real-world clinical scenarios where distinguishing these lesions from tendinopathy remains difficult even for experienced musculoskeletal radiologists.

Advances in machine learning have shown promise in enhancing diagnostic accuracy by leveraging large-scale imaging datasets and sophisticated algorithmic strategies [13]. Nevertheless, implementing such approaches in clinical radiology practice is not without challenges, particularly when issues such as dataset imbalance, interpretability, and workflow integration are considered [14]. The present results, showing disproportionate accuracy across diagnostic categories, emphasize the importance of addressing imbalance before applying machine learning to real-world clinical scenarios.

Imbalanced datasets are a well-known limitation in medical imaging research, where the predominance of normal or severe cases overshadows less frequent but clinically significant conditions. Strategies that combine oversampling of minority classes with under-sampling of majority classes have been proposed as effective solutions to improve classifier robustness [15]. Applying such techniques to MRI-based diagnostic models may enhance the detection of partial-thickness tears, thereby addressing one of the most clinically relevant gaps highlighted in this analysis.

Beyond technical considerations, the clinical consequences of diagnostic misclassification cannot be understated. Misclassification, as shown in other medical fields, can lead to inappropriate treatment, delayed interventions, and adverse outcomes [12]. In the context of rotator cuff tears, failure to identify partial-thickness pathology may result in prolonged ineffective management, progression to full-thickness tears, and poorer long-term surgical results.

Overall, MRI is found to be of high diagnostic accuracy in normal tendons as well as complete-thickness tears, but partial-thickness tears are a weakness in diagnosis. The imbalance in data sets, optimization of algorithms, and recognition of clinical consequences of false classification are essential aspects in enhancing diagnostic accuracy of MRI in identification of rotator cuff tears.

5. Conclusion

The current study has established that magnetic resonance imaging (MRI) is an extremely accurate diagnostic tool in the detection of normal tendons and full-thickness rotator cuff tears whereas partial-thickness tears are diagnostically difficult. The excellent performance data of full-thickness tears demonstrate how well MRI detects an entire structural disruption, and the low sensitivity and predictive data of partial-thickness tears are an indication of the low image appearance as well as their inability to represent that type of tear in the dataset. The effect of imbalance in datasets was also evident since most normal shoulders increased specificity and accuracy on this group but reduced the level of diagnostic consistency in minority groups. These results highlight the necessity of analytical and methodological measures, including balancing methods of data, to guarantee balanced and appropriate performance in all clinical categories of interest. Clinically, it is important to enhance the identification of partial-thickness tears because the late recognition or incorrect recognition can result in the disease development and worsening of patients. The implementation of sophisticated computational methods and algorithms that include machine learning models with custom solutions to imbalanced data is potentially useful in filling this gap in diagnosis. To sum up, MRI is a very good instrument to assess rotator cuff, however, more development is needed so that it can detect partial-thickness pathology and ensure the best patient care results.

References:

[1] Bedi A, Bishop J, Keener J, Lansdown DA, Levy O, MacDonald P, Maffulli N, Oh JH, Sabesan VJ, Sanchez-Sotelo J, Williams III RJ. Rotator cuff tears. Nature reviews Disease primers. 2024 Feb 8;10(1):8.

[2] Kim M, Park HM, Kim JY, Kim SH, Hoeke S, De Neve W. MRI-based diagnosis of rotator cuff tears using deep learning and weighted linear combinations. InMachine learning for healthcare conference 2020 Sep 18 (pp. 292-308). PMLR.

[3] Ashir A, Lombardi A, Jerban S, Ma Y, Du J, Chang EY. Magnetic resonance imaging of the shoulder. Polish journal of radiology. 2020 Aug 17;85(1):420-39.
[4] Liu JN, Garcia GH, Gowd AK, Cabarcas BC, Charles MD, Romeo AA, Verma NN. Treatment of partial thickness rotator cuff tears in overhead athletes. Current reviews in musculoskeletal medicine. 2018 Mar;11(1):55-62.
[5] Cook JL, Rio E, Purdam CR, Docking SI. Revisiting the continuum model of tendon pathology: what is its merit in clinical practice and research?. British journal of sports medicine. 2016 Oct 1;50(19):1187-91.
[6] Islam SM, Nasim MA, Hossain I, Ullah DM, Gupta DK, Bhuiyan MM. Introduction of medical imaging modalities. InData driven approaches on medical imaging 2023 Oct 17 (pp. 1-25). Cham: Springer Nature Switzerland.
[7] Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. Journal of big data. 2019 Dec;6(1):1-54.
[8] Fleiter TR, Mervis S. The role of 3D-CTA in the assessment of peripheral vascular lesions in trauma patients. European journal of radiology. 2007 Oct 1;64(1):92-102.
[9] De Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. American journal of roentgenology. 2009 Jun;192(6):1701-7.
[10] Matthewson G, Beach CJ, Nelson AA, Woodmass JM, Ono Y, Boorman RS, Lo IK, Thornton GM. Partial thickness rotator cuff tears: current concepts. Advances in orthopedics. 2015;2015(1):458786.
[11] Lädermann A, Burkhart SS, Hoffmeyer P, Neyton L, Collin P, Yates E, Denard PJ. Classification of full-thickness rotator cuff lesions: a review. EFORT open reviews. 2016 Dec 1;1(12):420-30.
[12] Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Frontiers in public health. 2017 Nov 20;5:307.
[13] Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. American journal of roentgenology. 2017 Apr;208(4):754-60.
[14] Junsomboon N, Phienthrakul T. Combining over-sampling and under-sampling techniques for imbalance dataset. InProceedings of the 9th international conference on machine learning and computing 2017 Feb 24 (pp. 243-247).
[15] Hawatmeh A, Thawabi M, Aggarwal R, Abirami C, Vavilin I, Wasty N, Visveswaran G, Cohen M. Implications of misclassification of type 2 myocardial infarction on clinical outcomes. Cardiovascular Revascularization Medicine. 2020 Feb 1;21(2):176-9.