

Accuracy of imaging of BI-RADS 4 subcategorizations in breast lesion diagnosis

Radiologic-pathologic correlation

Hadeel A. Ghunaim, MBBS, Rana E. Alatawi, MBBS, Walaa M. Borhan, MBBS, Tareef S. Daqqaq, MBBS, Facharzt, Ayman S. Alhasan, MBBS, DES, Mervat M. Aboualkheir, MBBS, Reem M. Elkady, MD.

ABSTRACT

الأهداف: ربط تقارير تصوير الثدي ونظام البيانات (BI-RADS) لأمراض الثدي من الفئة 4 مع نتائج تشريح النسيج المرضي لتقييم دقة التصنيف الفرعي.

المنهجية: أجريت دراسة بأثر رجعي في الفترة من سبتمبر 2021 إلى يونيو 2022م. حيث قام اثنان من استشاريين الأشعة بمراجعة صور الموجات فوق الصوتية والماموغرام لـ 247 كتلة في الثدي وتصنيفها إلى فئات فرعية 4A، 4B، و4C ومقارنتها مع نتائج تحليل الأنسجة التي تم الحصول عليها من خلال الخزعة تحت الأشعة الصوتية. واعتبرت قيمة $p < 0.05$ ذات دلالة إحصائية.

النتائج: من بين 247 ورم، تم تصنيف 135 ضمن الفئة الفرعية 4A، و68 ضمن الفئة 4B، و44 ضمن الفئة 4C بشكل عام، كان 41 (16.6%) يعانون من أورام خبيثة، في حين أن 206 (83.4%) لديهم أورام حميدة. كان متوسط عمر المرضى الذين يعانون من الأورام الحميدة مقابل الأورام الخبيثة (43.18±14.02) مقابل (51.24±14.15) سنة، ($p < 0.001$)، على التوالي. حيث كان متوسط حجم الأورام الحميدة مقابل الأورام الخبيثة (1.93±1.65) مقابل (3.82±3.89) سم؛ ($p < 0.001$) على التوالي. وتمت مقارنة النتائج مع التشريح النسيجي، وتقع القيمة التنبؤية الإيجابية ضمن النطاق المرجعي للفئات الفرعية 4C أكثر من 70% (ولوحظت موثوقية عالية بين القارئتين، حيث بلغت قيمة متوسط كرونباخ ألفا=0.79، 0.73-0.85). وقد لوحظت خلافات كبيرة في تعيين الخصائص المتعلقة بتصنيف كتلة الثدي في وسائل الأشعة المستخدمة بين القارئتين فيما يتعلق بكثافة الكتلة، والشكل، والنمط، والأوعية الدموية، والحدود.

الخلاصة: تساهم نتائج هذه الدراسة على تأكيد الحاجة إلى خصائص توجيهية موحدة لتوصيف الفئات الفرعية 4 BI-RADS لأورام الثدي وتحسين دقة التشخيص.

Objectives: To correlate breast imaging-reporting and data system (BI-RADS) category 4 lesions with histopathology results to assess the accuracy of subcategorization.

Methods: A retrospective study was carried out from September 2021 to June 2022. A total of 247 breast lesions were reviewed categorized as BI-RADS 4 using ultrasound (US) and digital mammography. Feature analysis of the lesions were obtained using BI-RADS terminology and assigned to subcategories (4A, 4B, and 4C). Pathological analysis was carried out on tissue obtained through US-guided core biopsy. A p -value of < 0.05 was considered significant.

Results: Of the 247 lesions, 135 were categorized as subcategory 4A, 68 as 4B, and 44 as 4C. Overall, 41

(16.6%) had malignant lesions, while 206 (83.4%) had benign lesions. The mean age of the patients with benign versus malignant lesions was (43.18±14.02 vs. 51.24±14.15 years; $p < 0.001$). Mean size of benign versus malignant lesions was (1.93±1.65 vs. 3.82±3.89 cm; $p < 0.001$). Findings were compared with histopathology, and the positive predictive value fell within the reference range for subcategories 4C (>70%). High reliability was observed between the 2 readers, with a weighted Cohen's Kappa value of 0.79 (0.73-0.85). Significant disagreements in the assignment of features on radiological lesion characterization were observed between the 2 readers regarding lesion density, shape, echo pattern, vascularity, and borders.

Conclusion: The results of this study contribute to the existing body of knowledge, emphasizing the need for standardized guidelines for the characterization of BI-RADS 4 subcategories and improved diagnostic accuracy in the management of breast lesions.

Keywords: BI-RADS 4, malignancy, histopathology, positive predictive value

Saudi Med J 2024; Vol. 45 (11): 1228-1237
doi: 10.15537/smj.2024.45.11.20240001

From the Department of Internal Medicine (Ghunaim, Daqqaq, Alhasan); from the Department of Basic Medical Science (Borhan), College of Medicine; from the Department of Diagnostic Radiology (Elkady), College of Applied Medical Science, Taibah University, from the Department of Radiology (Alatawi), King Fahad Medical Hospital, from the Department of Radiology (Alhasan), King Faisal Specialist Hospital and Research Centre, Al-Madinah Al-Munawwarah, from the Department of Internal Medicine (Aboualkheir), College of Medicine, Princess Nourah bint Abdulrahman University, Riyadh, Kingdom of Saudi Arabia, and from the Department of Radiology (Elkady), Faculty of Medicine, Assiut University, Assiut, Egypt.

Received 23rd May 2024. Accepted 17th October 2024.

Address correspondence and reprint request to: Dr. Hadeel A. Ghunaim, Department of Internal Medicine, College of Medicine, Taibah University, Al-Madinah Al-Munawwarah, Kingdom of Saudi Arabia.
E-mail: dr.had.g@botmail.com
ORCID ID: <https://orcid.org/0000-0002-9036-4603>

Breast cancer is a worldwide health challenge and a leading cause of death among women. In 2020, 2.3 million women were diagnosed with breast cancer, resulting in 685,000 deaths globally. In Saudi Arabia, breast cancer is a major health problem with a cumulative incidence rate of 14.8% and mortality rate of 8.5% among both gender.¹ The incidence rate among women in Saudi Arabia was reported to be 29.7% in 2018.¹

The breast imaging-reporting and data system (BI-RADS) is a standardized risk estimation and reporting organization for breast pathology encountered in ultrasound (US). It includes mammography and magnetic resonance imaging (MRI). Its fifth edition was updated by the American College of Radiology (ACR) in 2013, incorporating a few changes.² The BI-RADS facilitates communication between radiologists and referring clinicians through easily understood terminology and international standardization.

The BI-RADS is categorized from 0-6, with each category linked to specific management approaches that play a significant role in the outcome of breast pathology. Suspicious breast lesions are classified as BI-RADS 4, which encompasses a wide range of likelihood of malignancy (>2 to <95%) and should be managed through tissue biopsy. It is further subcategorized into: 4A (mostly benign pathology [>2 to $\leq 10\%$] likelihood of malignancy), 4B (moderate suspicion [>10 to $\leq 50\%$] likelihood of malignancy), and 4C (high suspicion [>50 to $<95\%$] likelihood of malignancy).³

Ultrasound-guided core needle biopsy (CNB) is a primary interventional procedure used for breast pathology. It offers several advantages compared to open biopsy, such as lower cost, invasiveness, patient discomfort, and procedure time.⁴ It also demonstrates sensitivity and specificity close to open biopsy.⁵ However, CNB has a false negative rate of approximately 0.1-2.5%.⁶⁻⁸ Radiological-pathological concordance is a cornerstone that significantly affects patient outcomes and prevents delays in management.⁹ Hence, a thorough radiologist assessment of the lesion and effective communication with the pathologist helps identify false negative pathology and minimize radiologic-pathologic discordance. This is particularly important due to the lack of objective criteria for BI-RADS 4 subcategorization and poor inter-observer agreement.¹⁰

Disclosure. Authors have no conflict of interests, and the work was not supported or funded by any drug company.

Our study has 2 objectives. First, to ensure the accurate subcategorization of BI-RADS 4 (4A, 4B, and 4C) lesions by comparing them with histopathology results. Second, to determine the inter-observer variability between breast radiologists' judgment on BI-RADS 4 subcategorization.

Methods. This retrospective study was carried out from September 2021 to June 2022. A total of 3,193 women underwent mammography and US, resulting in the classification of 274 lesions as BI-RADS 4. Of these, 247 lesions were included in this study, and correlations were subsequently established with histopathology reports from the same hospital. Exclusion criteria encompassed lesions without available images (9 patients) or pathology results (17 lesions) as shown in **Appendix 1**.

The ethical committee of the institutional review board, General Directorate of Health Affairs approved this study (national registration number with NCBE-KACST: H-03-M-84. IRB log No:044-22). Suspicious lesions classified as BI-RADS 4 were independently reviewed and blindly scored as BI-RADS 4A, 4B, or 4C. Lesion features were evaluated based on ACR BI-RADS lexicon 5th edition. Mammography assessed density, mass shape (oval, lobular, and irregular), suspicious calcification, asymmetry, architectural distortion, and nipple retraction. While the US assessed margin, echo pattern, posterior characteristics, and associated features such as vascularity, and skin thickening.

In our institution, mammography and US are routine diagnostic tools for symptomatic women over 35 years old for provisional diagnosis while tissue biopsy with histology is used for confirmatory diagnosis. The US alone is carried out for younger symptomatic women. Also, screening mammography is carried out for women who are 40 years old or more; in addition to US if indicated. Breast radiologists carried out US using Philips EPIQ 7G and GE Healthcare machines equipped with high-frequency linear probes (5-12 and 6-15 MHz). Mammography was executed by skilled technologists (with 16 years of experience) using the Hologic Selenia Dimensions system, capturing mediolateral oblique (MLO) and craniocaudal (CC) views alongside tomosynthesis for all patients. Additional views, such as magnification and compression, were employed as needed. The BI-RADS 4 lesions were sampled under US guidance using a 14-gauge core needle through an automated gun, with at least 3 samples obtained from each lesion.

Hematoxylin and eosin-stained (H&E) sections of the core biopsies were examined. Results were categorized as benign or malignant. In accordance with the the BI-RADS ACR criteria, atypical ductal hyperplasia was grouped among benign lesions; while carcinoma in situ was considered a high-risk “cancerous” lesion and grouped with malignancies to calculate PPV. Patients with atypical ductal hyperplasia on histology were recommended for complete surgical excision.

Statistical analysis. Data analysis was carried out using the Statistical Package for the Social Sciences, version 27.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were calculated, including means \pm standard deviations (SD), medians (interquartile range [IQR]), and percentages. Weighted Cohen’s Kappa with quadratic weighting was used in estimating the inter-rater reliability between the 2 raters. Percentage positive predictive value (%PPV) for malignancy in the lesions was estimated with their 95% confidence intervals (CI) for the 2 raters. Continuous variables were compared using the t-test; while categorical groups were compared with the Chi-square test or Fisher’s exact test and Monte Carlo exact test as suitable. A *p*-value of <0.05 was considered significant.

Results. The mean age for the 247 lesions was 44.51 ± 14.33 years, with a mean size of 2.26 ± 2.29 cm. Mammography was carried out for 197 (79.8%) lesions. Among them, 84 (43.0%) patients had category B breast density, 75 (38.5%) had category C, 28 (14.4%) had category A, and 8 (4.1%) patients had category D. The left breast was more frequently affected compared to the right side 127 (51.6%) vs. 120 (48.4%). The upper outer quadrant was the most common location of the lesions 133 (53.8%), followed by the retro-areolar area 57 (23.1%), as shown in **Table 1**.

Of the 247 BI-RADS 4 lesions, the first reader categorized 135 (54.7%) as BI-RADS 4A, 68 (27.5%) as 4B, and 44 (17.8%) as 4C. Meanwhile, the second reader classified 151 (61.1%) as BI-RADS 4A, 48 (19.4%) as 4B, and 48 (19.4%) as 4C, as presented in **Table 2**.

The imaging characteristics of the lesions are summarized in **Table 1**. Mass constituted 164 (66.4%) of the cases, mass with calcification accounted for 14 (5.7%), calcification only for 5 (2.0%), asymmetry for 39 (15.8%), architectural distortion for 26 (10.5%), intraductal mass for 37 (15.0%), complex mass for 18 (7.3%), skin thickening for 6(2.4%), and nipple retraction for 2 (0.8%).

Histopathological diagnoses of the lesions are presented in **Figures 1 & 2**. Benign lesions comprised

Table 1 - Baseline characteristics of the studied sample (N=247).

Variables	n (%)
Age (years), mean \pm SD	44.47 \pm 14.35
Age (years), median (IQR)	44.5 (35.0-54.0)
Ultrasound	
Yes	246 (99.6)
No	1 (0.4)
Mammography	
Yes	197 (79.8)
No	50 (20.2)
Affected side	
Left	127 (51.4)
Right	120 (48.6)
Lesion size (cm), mean \pm SD	2.25 \pm 1.80
Lesion size (cm), median (IQR)	1.5 (1.0-2.8)
Pathology	
Benign	199 (80.6)
Malignant	41 (16.6)
Hyperplasia	7 (2.8)
Imaging findings	
Mass	164 (66.4)
Mass with calcification	14 (5.7)
Calcification only	5 (2.0)
Asymmetry	39 (15.8)
Architectural distortion	26 (10.5)
Intraductal mass	37 (15.0)
Complex mass	18 (7.3)
Skin thickening	6 (2.4)
Nipple retraction	2 (0.8)
Mammography breast density (n=197)	
A	28 (14.4)
B	84 (43.0)
C	75 (38.5)
D	8 (4.1)
Lesion location	
Retroareolar	57 (23.1)
Upper outer quadrant	133 (53.8)
Lower outer quadrant	22 (8.9)
Upper inner quadrant	20 (8.1)
Lower inner quadrant	15 (6.1)

Values are presented as numbers and percentages (%), or mean \pm standard deviation (SD), or median interquartile range (IQR).

Table 2 - Validity of Rater 1 and Rater 2 decision according to pathology of the lesion.

BIRADS	Pathology		<i>P</i> -value*
	Benign	Malignant	
Rater 1			
4A (n=135)	133 (98.5)	2 (1.5)	<0.001
4B (n=68)	63 (92.6)	5 (7.4)	
4C (n=44)	10 (22.7)	34 (77.3)	
Rater 2			
4A (n=151)	149 (98.7)	2 (1.3)	<0.001
4B (n=48)	44 (91.7)	4 (8.3)	
4C (n=48)	13 (27.1)	35 (72.9)	

Values are presented as numbers and percentages (%). *Chi-square test for proportions was used to estimate the *p*-values. BIRADS: breast-imaging reporting and data system

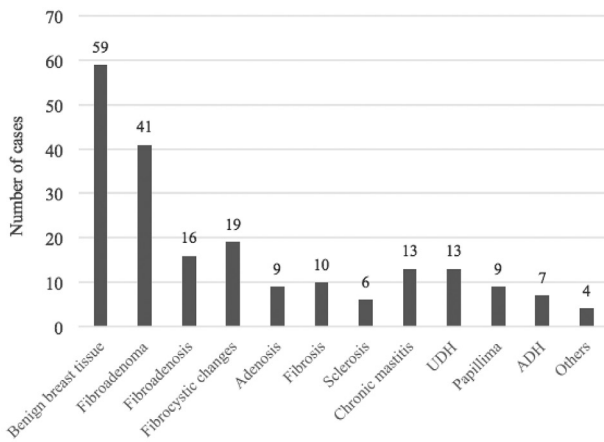


Figure 1 - Pathological distribution of benign lesions found (n=206). UDH: usual ductal hyperplasia, ADH: atypical ductal hyperplasia

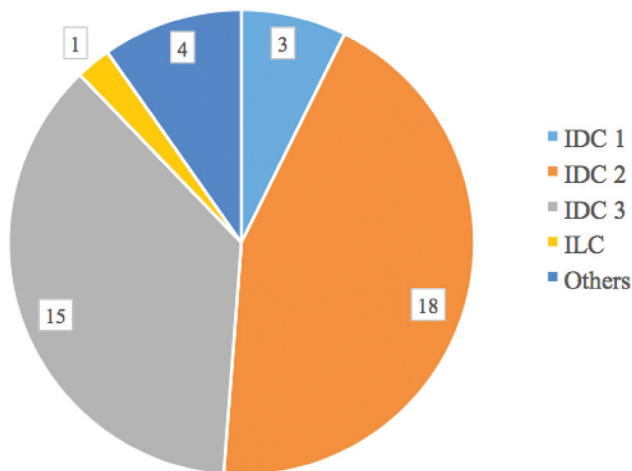


Figure 2 - Pathological distribution of malignant lesions found (n=41). IDC: invasive ductal carcinoma, ILC: invasive lobular carcinoma

206 (83.4%) and malignant lesions accounted for 41 (16.6%) of the cases. The mean age of the patients with benign lesions was 43.18 ± 14.02 years compared to 51.24 ± 14.15 years for patients having malignant lesions ($p < 0.001$). Mean size of benign versus malignant lesions was $(1.93 \pm 1.65$ vs. 3.82 ± 3.89 cm; $p < 0.001$). There were no significant differences in breast density ($p = 0.07$) and the affected side of the body ($p = 0.98$) in the prevalence of benign and malignant lesions. Among the benign cases (n=206), the most common benign lesions were benign breast tissue (n=59 [28.6%]), followed by fibroadenoma (n=41 [19.9%]), fibrocystic changes (n=19 [9.2%]), fibroadenosis (n=16 [7.8%]), usual ductal hyperplasia (n=13 [6.3%]), chronic mastitis (n=13 [6.3%]), fibrosis

(n=10 [4.9%]), papilloma (n=9 [4.4%]), adenosis (n=9 [4.4%]), atypical ductal hyperplasia (n=7 [3.4%]), and sclerosis (n=6 [2.9%]) as shown in **Figure 1**. Among the malignant lesions (n=41), the most common lesions were invasive ductal carcinoma type-2 (n=18 [43.9%]) and invasive ductal carcinoma type-3 (n=15 [36.6%]) as shown in **Figure 2**.

Our analysis revealed 1.5% of cancer-PPV for BI-RADS 4A, 7.4% for 4B, and 77.3% for 4C lesions, by rater 1; and cancer-PPV of 1.3% for BI-RADS 4A, 8.3% for 4B, and 72.9% for 4C lesions, by rater 2 as shown in **Table 3**.

Very good reliability was observed between the 2 readers in the subcategorization of the lesions, with a weighted Cohen's Kappa values of 0.79 (0.73-0.85) as shown in **Table 4**. Additionally, a significant agreement was observed between the 2 readers in subcategorization, particularly in subcategory 4A and 4C ($p < 0.001$).

Differences in the feature assignment between the 2 readers in the imaging characterization, with significant p -values, were observed in mammography including lesion density of ($p = 0.025$), and shape ($p = 0.04$), as well as US included; vascularity of the mass ($p = 0.01$), and lesion borders, including circumscribed ($p = 0.046$), angular ($p = 0.015$), and micro lobular borders ($p = 0.048$), as detailed in **Table 5**.

Discussion. Our study focused on the analysis and interpretation of various aspects of BI-RADS category 4 lesions. The results revealed valuable insights into lesion categorization, patient demographics, imaging characteristics, histopathological diagnoses, and inter-reader reliability for malignancy.

Patient demographics revealed an average age of 44.57 years, aligning with previous research that showed similar age distribution among women with BI-RADS 4 lesions.¹¹ Also, consistent with previous reports, we found a higher prevalence of benign lesions in the younger age group and a higher burden of malignant breast lesions among older women.^{12,13} Most mammography cases were associated with category B and C breast density. The prevalence of breast density categories indicated a potential association between higher density and risk of breast cancer, a trend consistent with previous studies.¹⁴

Our analysis demonstrated cancer PPV of BI-RADS 4 subcategories 4A as 1.5%, 4B as 7.4%, and 4C as 77.3% for rater 1; and 4A as 1.3%, 4B as 8.3%, and 4C as 72.9% for rater 2. Our findings were within the ACR 2013 reference range for subcategories 4C but not for 4A and 4B. The results obtained in our study were consistent with the existing literature that

Table 3 - Comparison of positive predictive value of breast imaging-reporting and data system 4 subcategories with reference value.

Rater BI-RADS subcategory	Total	Benign	Malignant	PPV for malignancy Actual (95% CI)	BIRADS PPV Reference
<i>Rater 1</i>					
4A	135	133 (98.5)	2 (1.5)	1.5 (0.2-5.2)	>2 - ≤10
4B	68	63 (92.6)	5 (7.4)	7.4 (2.4-16.3)	>10 - ≤50
4C	44	10 (22.7)	34 (77.3)	77.3 (62.2-88.5)	>50 - <95
Total	247	206 (83.4)	41 (16.6)		>2 - <95
<i>Rater 2</i>					
4A	151	149 (98.7)	2 (1.3)	1.3 (0.2-4.7)	>2 - ≤10
4B	48	44 (91.7)	4 (8.3)	8.3 (2.3-20.0)	>10 - ≤50
4C	48	13 (27.1)	35 (72.9)	72.9 (58.2-84.7)	>50 - <95
Total	247	206 (83.4)	41 (16.6)		>2 - <95

Values are presented as numbers and percentages (%). BI-RADS: breast-imaging reporting and data system, PPV: positive predictive value, CI: confidence interval

Table 4 - Inter-rater agreement between the 2 raters of the lesions.

Inter-rater Rater 2 Observer 2	Rater 1 BIRADS			K-values [†]
	4A	4B	4C	
BIRADS 4A	121 (89.6)	30 (44.1)	0 (0.0)	0.79 (0.73-0.85)
BIRADS 4B	12 (8.9)	30 (44.1)	6 (13.6)	
BIRADS 4C	2 (1.5)	8 (11.8)	38 (86.4)	<i>p</i> <0.001
Total	135	68	44	

Values are presented as numbers and percentages (%). [†]Weighted Cohen's Kappa (K) value and 95% confidence interval (CI). BIRADS: breast-imaging reporting and data system

Table 5 - Issues of differences in the feature assignment reported By Rater 1 and Rater 2.

Mammography	Rater (n=52)		P-values
	Rater 1	Rater 2	
<i>Density</i>			
Hypo-	1 (1.9)	0 (0.0)	0.025 [†]
Iso-	42 (77.8)	31 (58.5)	
Hyper-	11 (20.4)	22 (41.5)	
<i>Shape (n=24)</i>			
Oval	7 (29.2)	13 (54.2)	0.041 ^{**}
Round	11 (45.8)	4 (16.7)	0.044 ^{**}
Irregular	6 (25.0)	7 (29.1)	0.124 ^{**}
<i>Ultrasound (n=54)</i>			
<i>Echo</i>			
Hypo-	36 (66.7)	43 (76.8)	0.264 [†]
Iso-	1 (1.9)	3 (7.1)	
Mixed	17 (31.5)	11 (16.1)	
Vascularity (n=24)	20 (83.3)	3 (12.5)	0.017 ^{**}
<i>Borders (n=54)</i>			
Circumscribed	13 (24.1)	24 (44.4)	0.046 ^{**}
Indistinct	13 (24.1)	18 (33.3)	0.102 ^{**}
Angular	23 (42.6)	1 (1.9)	0.015 ^{***}
Micro-lobular	29 (53.7)	21 (38.9)	0.048 ^{**}
Spiculated	2 (3.7)	1 (1.9)	0.214 ^{***}
Anti-parallel	7 (13.9)	3 (5.6)	0.084 ^{***}

Values are presented as numbers and percentages (%). [†]Monte Carlo exact test was used to compare the frequency differences between groups. ^{**}The Chi-square test was used to compare the frequency differences between groups. ^{***}Fisher's exact test was used to compare the frequency differences between groups.

suggests a substantially higher PPV for BI-RADS 4C subcategory.¹⁵ However, compared to the same study, the PPV of subcategories 4A and 4B were significantly lower. The considerably higher PPV for subcategory 4C substantiates its association with a higher likelihood of malignancy, reflecting the importance of accurate subcategorization in guiding clinical decisions. It is important that both rater 1 and rater 2 in our study hit the ACR PPV benchmark for subcategory 4C. This indicates that they each have a PPV of >50% of correctly predicting the likelihood of malignancy in cancerous lesions. Therefore, such patients can go straight to biopsy without further adjunctive diagnostic tests. Masses that are classified as BI-RADS 4A and 4B often require consistent objective criteria for BI-RADS 4 subcategorization or in the case of 4A additional diagnostic tests before being downgraded to BI-RADS 3. A potential improvement that could lead to an increased PPV for BI-RADS categorization is using artificial intelligence through training observers in deep-learning computer-based diagnostic systems.¹⁶

The findings highlighted the distribution of BI-RADS 4 lesions among 2 readers, with variations observed in their subcategorizations. The study assessed inter-rater reliability rating of the lesions between rater 1 and rater 2, revealing substantial agreement between the 2 raters. The high weighted Cohen's Kappa value of 0.79 suggests high consistency in BI-RADS 4 subcategorization. In addition, there is an observed difference in the feature assignment by the raters in imaging characterization, especially regarding lesion density, shape, vascularity, and borders. The reliability between the 2 readers in the subcategorization of BI-RADS 4 is substantial. In the context of significant differences in lesion characterization, it is worth noting that the PPV fell within the reference range established by the ACR in 2013 for subcategory 4C. However, the PPV for subcategories 4A and 4B did not align with the reference range. Furthermore, both raters had mean PPVs for subcategory 4A that lie within the BI-RADS 3 benchmark PPV range (namely, <2%) and both raters had mean PPVs for subcategory 4B that lie within the BI-RADS 4A benchmark PPV range (namely, 2 to <10%). The reasons for these observations in the patient population may be because the overall prevalence of malignant lesions are small, the size of malignant lesions are large, and therefore, the difference in mean diameters between benign and malignant lesions are large. This indirectly suggests that it is likely that the majority of the patients studied presents clinically with palpable breast lesions with a relatively small percentage of the lesions of the patients being detected early through

mammography screening. Also, the large and significant discordances in feature assignment between the 2 raters may likely contribute significantly to inability of the raters to achieve PPV benchmarks for subcategory 4A and 4B. Therefore, the discrepancy between the 2 raters in the subcategorization of BI-RADS 4A and 4B lesions highlight the potential need for consistent criteria in characterizing BI-RADS 4 lesions. Examples of the disagreements are presented in [Appendices 2 and 3](#). In the first case, one of the readers classified the lesion as 4C, whereas the other reader classified it as 4B. For the lesion in [Appendix 3](#), the first reader classified it as 4B, and the second reader classified it as 4A.

These differences could be attributed to perception, experience, or interpretation variations. It could also be due to the fact that BI-RADS recommendations are non-strict or non-rigid, which could lead to improper categorization in BI-RADS 4 subcategories.¹⁷ Such discrepancies in subcategorization underscore the inherent subjectivity in mammographic interpretation. Comparable studies have documented similar inter-rater variations in BI-RADS assessments.^{15,18}

This study unveiled a diverse range of imaging characteristics in BI-RADS 4 lesions, with masses being the most common presentation (66.4%), followed by other features such as calcifications, architectural distortion, and intraductal masses. This aligns with the characteristic diversity expected within the BI-RADS 4 categories established by ACR guidelines.² The various imaging characteristics underscore the importance of a comprehensive assessment approach to evaluate lesions for potential malignancy accurately.

Histopathological examination of the lesions shows a predominantly benign nature (83.4%), with fibroadenomas and benign breast tissue being the most frequent findings. These results concur with previous research that reported a higher prevalence of benign lesions within the BI-RADS 4 category.¹¹ However, the notable presence of malignant lesions (16.6%) emphasizes the significance of accurate classification and management of these lesions. Also, there was a marked significant difference in the mean diameter of benign lesions (1.93 cm) compared to malignant lesions (3.82 cm). The sizes and the marked variation in the sizes of benign and malignant lesions suggest that majority of the patients studied might have presented with clinically palpable lesions and were not detected early through mammography screening or other imaging tests. A previous survey of 1,135 women aged 50 years or older in the study setting revealed that 92% of the women reported never having a mammogram despite the availability of free mammography screening

services; and a more recent survey of 3,245 women aged ≥ 40 years, revealed that only 40% of them reported ever having a mammogram.^{19,20} In addition, several studies in the setting have highlighted substantial perceived barriers to the free mammography screening services among women.²¹⁻²⁴

Study limitations. Although our study has provided valuable information regarding the correlation between the different BI-RADS 4 subcategories and histopathology results, some limitations should be addressed. Unlike prospective studies, our study has a retrospective design, and the absence of predefined protocols in our retrospective study could introduce various limitations, such as potential selection bias, limitations in reading US images, and the inability to control variables. Our study was carried out exclusively at a single hospital. The single-center approach could limit the generalizability of our findings to a broader population. Moreover, our study focused primarily on the correlation of BI-RADS 4 subcategorization with immediate histopathological findings. Having clinical examination data like palpable lump, thickening, and nipple discharge as well as post-surgical excision and long-term follow-up data, including patient outcomes and progression of identified lesions, could provide a more comprehensive assessment of the accuracy of subcategorization.

In conclusion, this study comprehensively analyzed BI-RADS 4 lesions, encompassing categorization, demographics, imaging characteristics, histopathological diagnoses, and inter-reader reliability. The results contribute to the existing body of knowledge, emphasizing the need for standardized guidelines and improved diagnostic accuracy in managing breast lesions.

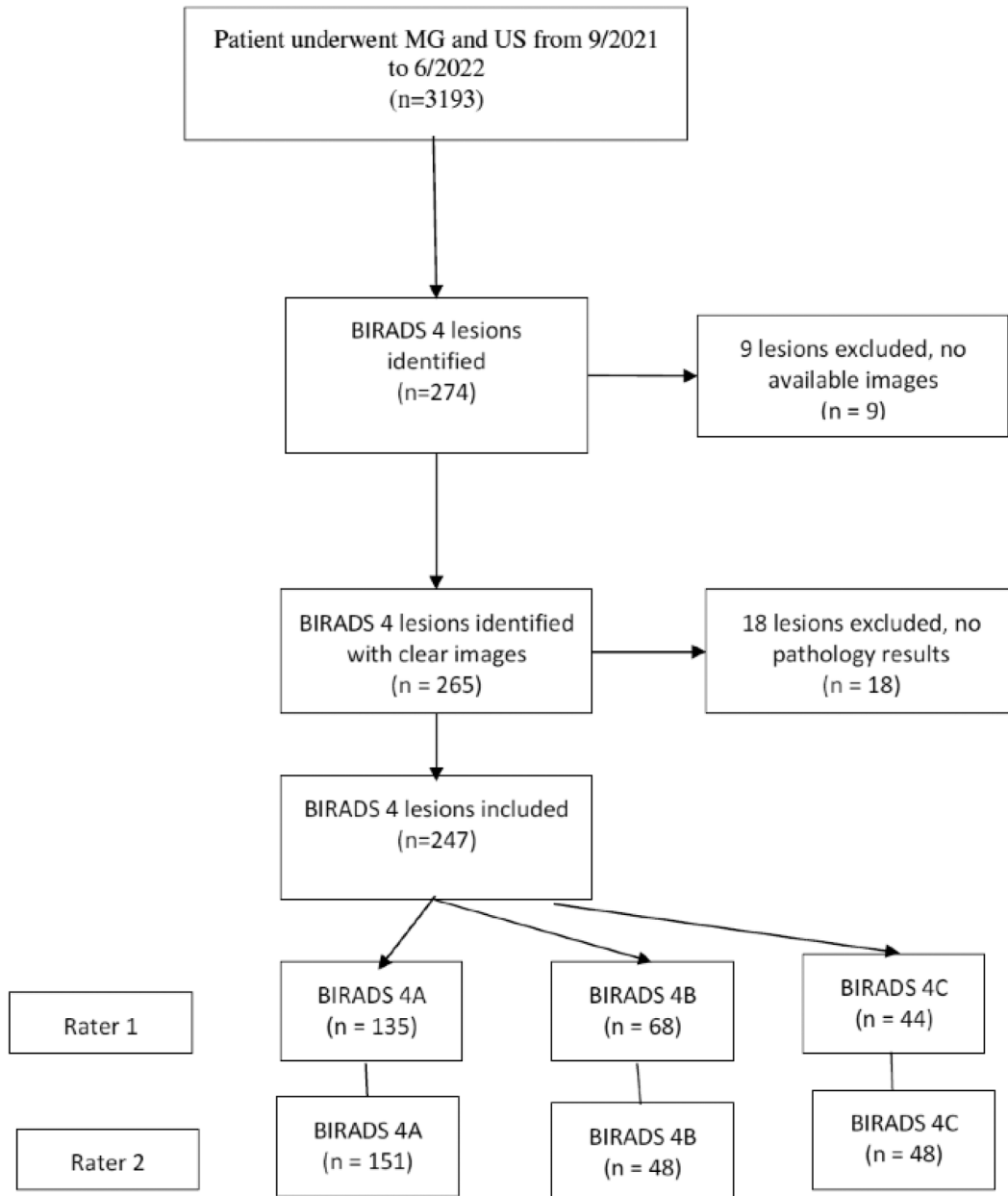
Acknowledgment. *The authors gratefully acknowledge Research Medics for the English language editing.*

References

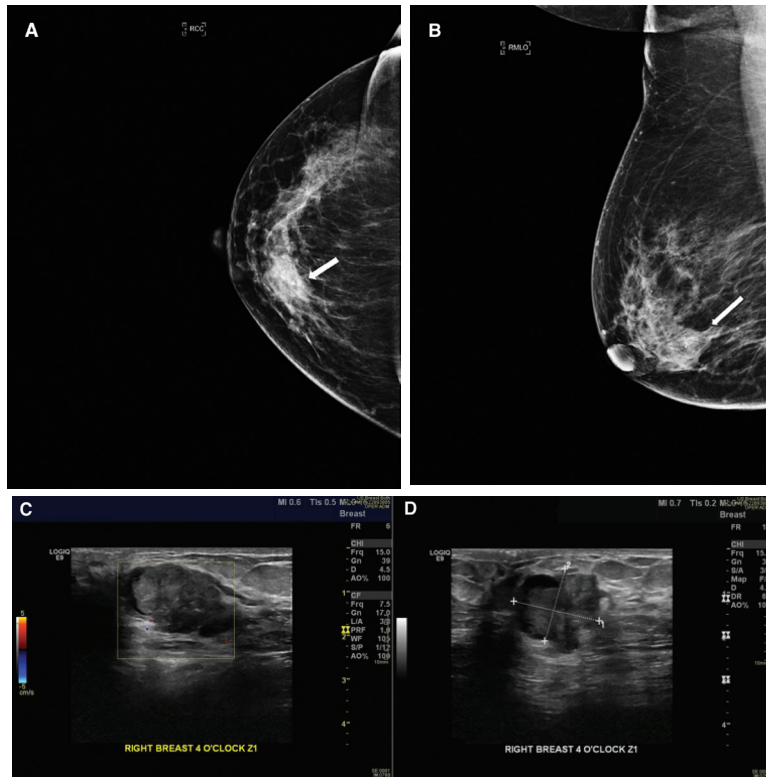
- Alqahtani WS, Almufareh NA, Domiaty DM, Albasher G, Alduwish MA, Alkhalaf H, et al. Epidemiology of cancer in Saudi Arabia thru 2010-2019: a systematic review with constrained meta-analysis. *AIMS Public Health* 2020; 7: 679-696.
- Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: a summary of changes. *Diagn Interv Imaging* 2017; 98: 179-190.
- D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. ACR BI-RADS® Atlas: breast imaging reporting and data system. [Updated 2013; accessed 2024 Jun 18]. Available from: <https://www.scirp.org/reference/ReferencesPapers?ReferenceID=2274695>
- Schueller G, Schueller-Weidekamm C, Helbich TH. Accuracy of ultrasound-guided, large-core needle breast biopsy. *Eur Radiol* 2008; 18: 1761-1773.
- Dahabreh IJ, Wieland LS, Adam GB, Halladay C, Lau J, Trikalinos TA. Core needle and open surgical biopsy for diagnosis of breast lesions: an update to the 2009 report. [Updated 2014; accessed 2024 Jun 18]. Available from: <https://pubmed.ncbi.nlm.nih.gov/25275206/>
- Park VY, Kim EK, Moon HJ, Yoon JH, Kim MJ. Evaluating imaging-pathology concordance and discordance after ultrasound-guided breast biopsy. *Ultrasonography* 2018; 37: 107-120.
- Zhang C, Lewis DR, Nasute P, Hayes M, Warren LJ, Gordon PB. The negative predictive value of ultrasound-guided 14-gauge core needle biopsy of breast masses: a validation study of 339 cases. *Cancer Imaging* 2012; 12: 488-496.
- Zhou JY, Tang J, Wang ZL, Lv FQ, Luo YK, Qin HZ, et al. Accuracy of 16/18G core needle biopsy for ultrasound-visible breast lesions. *World J Surg Oncol* 2014; 12: 7.
- Youk JH, Kim EK, Kim MJ, Ko KH, Kwak JY, Son EJ, et al. Concordant or discordant? Imaging-pathology correlation in a sonography-guided core needle biopsy of a breast lesion. *Korean J Radiol* 2011; 12: 232-240.
- Jörg I, Wieler J, Elfgen C, Bolten K, Hutzli C, Talimi J, et al. Discrepancies between radiological and histological findings in preoperative core needle (CNB) and vacuum-assisted (VAB) breast biopsies. *J Cancer Res Clin Oncol* 2021; 147: 749-754.
- Maltez de Almeida JR, Gomes AB, Barros TP, Fahel PE, de Seixas Rocha M. Subcategorization of suspicious breast lesions (BI-RADS category 4) according to MRI criteria: role of dynamic contrast-enhanced and diffusion-weighted imaging. *AJR Am J Roentgenol* 2015; 205: 222-231.
- Saadat R, Abdul-Ghafar J, Haidary AM, Rahmani S, Atta N. Age distribution and types of breast lesions among Afghan women diagnosed by fine needle aspiration cytology (FNAC) at a tertiary care centre in Afghanistan: a descriptive cross-sectional study. *BMJ Open* 2020; 10: e037513.
- Mukhtar R, Hussain M, Mukhtar MA, Haider SR. Prevalence of different breast lesions in women of southern Punjab, Pakistan, characterized on high-resolution ultrasound and mammography. *Egypt J Radiol Nucl Med* 2021; 52: 245. doi:10.1186/s43055-021-00633-8.
- Noonpradej S, Wangkulangkul P, Woodtichartprecha P, Laohawiriyakamol S. Prediction for breast cancer in BI-RADS category 4 lesion categorized by age and breast composition of women in Songklanagarind hospital. *Asian Pac J Cancer Prev* 2021; 22: 531-536.
- Zou X, Wang J, Lan X, Lin Q, Han F, Liu L, et al. Assessment of diagnostic accuracy and efficiency of categories 4 and 5 of the second edition of the BI-RADS ultrasound lexicon in diagnosing breast lesions. *Ultrasound Med Biol* 2016; 42: 2065-2071.
- Xiao M, Zhao C, Li J, Zhang J, Liu H, Wang M, et al. Diagnostic value of breast lesions between deep learning-based computer-aided diagnosis system and experienced radiologists: comparison the performance between symptomatic and asymptomatic patients. *Front Oncol* 2020; 10: 1070.
- Mohapatra SK, Mishra A, Sahoo TK, Nayak RB, Das PK, Nayak B. The positive predictive values of the breast imaging reporting and data system (BI-RADS) 4 lesions and its mammographic morphological features. *Indian J Surg Oncol* 2021; 12: 182-189.

18. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 2006; 239: 385-391.
19. El Bcheraoui C, Basulaiman M, Wilson S, Daoud F, Tuffaha M, AlMazroa MA, et al. Breast cancer screening in Saudi Arabia: free but almost no takers. *PLoS One* 2015; 10: e0119051.
20. Al-Wassia RK, Farsi NJ, Merdad LA, Hagi SK. Patterns, knowledge, and barriers of mammography use among women in Saudi Arabia. *Saudi Med J* 2017; 38: 913-921.
21. AlAbdulKader A, Gari D, Al Yousif G, Alghamdi A, AlKaltham S, AlDamigh F, et al. Perceived barriers and facilitators to breast cancer screening among women in Saudi Arabia. *Breast Cancer (Dove Med Press)* 2023; 15: 505-513.
22. Abdel-Salam DM, Mohamed RA, Alyousef HY, Almasoud WA, Alanzi MB, Mubarak AZ, et al. Perceived barriers and awareness of mammography screening among Saudi women attending primary health centers. *Risk Manag Healthc Policy* 2020; 13: 2553-2561.
23. Abdel-Aziz SB, Amin TT, Al-Gadeeb MB, Alhassar AI, Al-Ramadan A, Al-Helal M, et al. Perceived barriers to breast cancer screening among Saudi women at primary care setting. *J Prev Med Hyg* 2018; 59: E20-E29.
24. Aljehani AJ, Alomar MA, Albalawi AM, Alsultan AS, Alhussini FO, Alshehri RF, et al. Perceptions and behaviors regarding early screening and breast-related complaints among Saudi women. *Cureus* 2024; 16: e61891.

Appendix 1 - Data handling, imaging, and biopsy of the raters.



Appendix 2 - Mammography and sonography of a 65-year-old woman presented with clear right nipple discharge. A) Standard craniocaudal and B) mediolateral oblique views were obtained. Focal asymmetry in the right inner lower anterior to middle third (arrows) was detected. C-D) Sonography. It showed an avascular branching mass with posterior acoustic enhancement measuring 2x1.7x3 cm at 4 O'clock. Bilateral Abnormal axillary lymph nodes with cortical thickening approximately 0.4 cm. The histology of US-guided biopsy revealed intraductal papilloma.



Appendix 3 - Mammography and sonography of 43-year-old women. A) Standard craniocaudal and B) mediolateral oblique views were obtained. Focal asymmetry was detected in the upper outer quadrant left breast (arrows). C) Sonography. It showed a heterogenous mass with an indistinct border measuring 7mm at 2 O'clock (arrow). The histology of the US-guided biopsy revealed benign breast tissue.

