Published by
Saidu Medical College

OPEN ACCESS

ISSN (P): 1819-4583
ISSN (E): 3007-9470

Copyright © Copyright (c) 2026 The authors retain the copyrights. Authors are permitted and encouraged to post their work online (e.g. in institutional repositories or on their website), as it can lead to productive exchanges, as well as greater citation of published work.

Citation:
Khattak S, Siddique U, Ghaus G, Khattak W, Hasnat M, Rehman HU, et al. Interobserver Agreement Between a Radiology Resident and a Senior Sonologist in the Ultrasonographic Interpretation of Common Gallbladder Pathologies: A Cross-Sectional Observational Study. *J Saidu Med Coll.* 2026;16(3):568–576.

Corresponding Author:
Dr. Sara Khattak
Resident,
Department of Radiology,
Rehman Medical Institute, Peshawar, Pakistan
sarakhattak01@gmail.com

Funding:
The authors received no external funding for this study.

Conflicts of Interests:
The authors have declared that no competing interests exist.

Data Sharing Statement:
The data that support the findings of this study are available on request from the corresponding author. The data is not publicly available due to privacy or ethical restrictions.

Received: December 25, 2025
1st Revision: February 21, 2026
1st Revision: June 09, 2026
2nd Revision: June 19, 2026

Check for updates

Original Article

Interobserver Agreement Between a Radiology Resident and a Senior Sonologist in the Ultrasonographic Interpretation of Common Gallbladder Pathologies: A Cross-Sectional Observational Study

Sara Khattak, Ummara Siddique, Ghulam Ghaus, Wadana Khattak, Muhammad Hasnat, Hammad Ur Rehman & Mohsin Khan
Rehman Medical Institute, Peshawar, Pakistan

https://doi.org/10.52206/jsmc.2026.16.3.1485

Abstract

Background: Ultrasonography is the first-line imaging modality for gallbladder disease because it is safe, inexpensive, widely available, and provides real-time hepatobiliary assessment. However, its diagnostic accuracy is operator dependent, making interobserver agreement important for evaluating reliability in routine practice.

Objective: To evaluate the level of interobserver agreement between a radiology resident and a senior sonologist in the ultrasonographic interpretation of common gallbladder pathologies.

Methodology: This cross-sectional observational study was conducted in the Department of Radiology, Rehman Medical Institute, Peshawar, Pakistan, from March 2022 to June 2025. A total of 150 consecutive adult patients with clinically suspected gallbladder disease were enrolled. Each participant underwent two independent ultrasonographic examinations performed separately by a radiology resident and a senior sonologist, both blinded to each other's findings. Six predefined gallbladder abnormalities, including gallstones, biliary sludge, gallbladder wall thickening, polyps, pericholecystic fluid, and comet-tail artifact, were evaluated using standardized operational definitions. Interobserver agreement was assessed using Cohen's kappa (κ) coefficient with corresponding 95% confidence intervals.

Results: Gallstones demonstrated almost perfect agreement between observers (κ = 0.826; 95% CI: 0.735–0.917), while gallbladder polyps showed moderate agreement (κ = 0.581; 95% CI: 0.450–0.712). Fair agreement was observed for gallbladder wall thickening (κ = 0.392), biliary sludge (κ = 0.365), comet-tail artifact (κ = 0.341), and pericholecystic fluid (κ = 0.333). Agreement was consistently higher for objectively identifiable structural abnormalities than for subtle inflammatory or artifact-related findings.

Conclusion: Interobserver agreement in gallbladder ultrasonography varies according to the sonographic abnormality being evaluated. Gallstones demonstrate excellent reproducibility, whereas inflammatory and artifact-related findings remain susceptible to observer variability. Standardized scanning protocols, structured reporting systems, and focused training of radiology residents may improve consistency and reliability in gallbladder ultrasound interpretation.

Keywords: Cholelithiasis, Diagnostic Imaging, Gallbladder Diseases, Observer Variation, Reproducibility of Results, Ultrasonography.

Introduction

Gallbladder diseases constitute one of the most common hepatobiliary disorders encountered in clinical practice and include a wide spectrum of pathological conditions such as cholelithiasis, acute and chronic cholecystitis, biliary sludge, gallbladder polyps, adenomyomatosis, and other inflammatory abnormalities [1–3]. These disorders account for a substantial proportion of emergency department visits, surgical admissions, and abdominal imaging referrals worldwide. Delayed or inaccurate diagnosis may result in significant morbidity, including recurrent biliary colic, gallbladder perforation, empyema, pancreatitis, and other potentially life-threatening complications. Consequently, accurate and timely diagnosis is essential for appropriate clinical management and improved patient outcomes [1–3].

Ultrasonography remains the imaging modality of choice for the initial evaluation of suspected gallbladder disease because it is non-invasive, widely available, inexpensive, free from ionizing radiation, and capable of providing dynamic real-time assessment of gallbladder morphology and function [4,5]. It enables visualization of gallstones, biliary sludge, gallbladder wall characteristics, intraluminal lesions, pericholecystic inflammatory changes, and associated biliary abnormalities with high diagnostic accuracy. Owing to these advantages, international guidelines continue to recommend ultrasonography as the first-line investigation for patients presenting with right upper quadrant pain or suspected biliary pathology [4–6].

Despite its widespread clinical use, ultrasonography is inherently operator dependent. Diagnostic accuracy is influenced not only by the quality of ultrasound equipment but also by patient-related factors, scanning technique, image optimization, observer experience, and interpretation skills [6,7]. Variability in patient positioning, transducer orientation, gain settings, gallbladder distension, and pattern recognition can contribute to inconsistent identification of subtle abnormalities such as biliary sludge, gallbladder wall thickening, adenomyomatosis, small polyps, or minimal pericholecystic fluid collections. Such variability may ultimately influence diagnostic confidence, reporting consistency, and subsequent clinical decision-making [6,7].

To improve consistency in ultrasound interpretation, several professional organizations have introduced evidence-based consensus recommendations for the evaluation and management of gallbladder lesions, particularly gallbladder polyps and wall abnormalities [8,9]. More recently, structured reporting frameworks such as the Gallbladder Reporting and Data System (GB-RADS) have been developed to standardize image interpretation, improve communication between radiologists and clinicians, and facilitate appropriate patient management [10–13]. Structured reporting has also been shown to reduce ambiguity, improve adherence to imaging guidelines, and enhance overall report quality across multiple radiological subspecialties [13].

In parallel with these developments, advances in artificial intelligence and machine learning have highlighted the importance of reliable and reproducible imaging interpretation for algorithm development, validation, and clinical implementation [14]. Because observer variability directly influences the quality of imaging datasets used for training and validating artificial intelligence systems, understanding the reproducibility of routine ultrasound interpretation has become increasingly important in modern radiology practice [14].

Although numerous studies have evaluated the diagnostic accuracy of ultrasonography for gallbladder diseases, comparatively fewer investigations have examined interobserver agreement between radiologists with different levels of expertise, particularly in resource-limited teaching hospitals where residents perform a substantial proportion of routine ultrasound examinations. Moreover, local evidence regarding agreement between radiology trainees and experienced sonologists remains limited.

Therefore, the present study was undertaken to evaluate the level of concordance and discordance between a radiology resident and a senior sonologist in the ultrasonographic interpretation of common gallbladder pathologies. By identifying sonographic findings with high and low reproducibility, this study aims to provide evidence that may guide resident education, quality assurance initiatives, implementation of standardized reporting systems, and future improvements in the consistency of gallbladder ultrasonography.

Methodology

Study Design

This hospital-based cross-sectional observational study was conducted to evaluate the level of interobserver agreement between a radiology resident and a senior sonologist in the ultrasonographic interpretation of common gallbladder pathologies. The primary objective was to determine the reproducibility of routine gallbladder ultrasound findings by comparing independent assessments made by observers with different levels of experience under routine clinical practice conditions. The study was carried out over a

period of approximately 40 months, from 1 March 2022 to 30 June 2025, at the Department of Radiology, Rehman Medical Institute, Peshawar, Pakistan.

Ethics

Ethical approval was obtained from the Institutional Ethics Committee of Rehman Medical Institute before commencement of the study. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants after explaining the study objectives, examination procedures, confidentiality measures, and voluntary nature of participation. Patient anonymity was maintained throughout data collection and analysis by assigning unique study identification numbers, and no personally identifiable information was included in the study database.

Setting

The study was conducted in the Department of Radiology, Rehman Medical Institute, a tertiary-care teaching hospital that provides comprehensive diagnostic imaging services for both inpatient and outpatient populations. Adult patients referred for abdominal ultrasonography from surgical, medical, emergency, and outpatient clinics with suspected gallbladder disease were screened for eligibility. All ultrasound examinations were performed in the department using standardized departmental imaging protocols throughout the study period.

Participants

Eligible participants included adults aged 18 years or older who were referred for abdominal ultrasonography because of right upper quadrant pain, biliary colic, suspected cholelithiasis, suspected acute or chronic cholecystitis, dyspeptic symptoms with clinical suspicion of gallbladder pathology, or abnormal hepatobiliary biochemical investigations requiring gallbladder assessment. Patients were recruited consecutively to minimize selection bias until the predetermined sample size was achieved. Patients were excluded if they had previously undergone cholecystectomy, had a known diagnosis of gallbladder carcinoma or previously confirmed gallbladder malignancy, or were unable to undergo independent ultrasonographic examinations by both observers during the same clinical visit. Patients with incomplete examinations or inadequate visualization of the gallbladder despite standard scanning techniques were also excluded from analysis. Two observers participated in the study. The first observer was a radiology resident who had completed at least one year of accredited radiology residency training, while the second observer was a consultant-level senior sonologist with more than five years of post-fellowship experience in abdominal ultrasonography. Both observers independently performed complete ultrasound examinations on every participant without knowledge of each other's findings.

Variables

The primary study outcome was the degree of interobserver agreement between the radiology resident and senior sonologist for the detection of predefined gallbladder abnormalities. The evaluated sonographic variables included gallstones, biliary sludge, gallbladder wall thickening, gallbladder polyps, pericholecystic fluid, and comet-tail artifact suggestive of adenomyomatosis. Demographic variables included participant age and sex. Agreement for each sonographic finding was measured using Cohen's kappa coefficient with corresponding 95% confidence intervals.

Data sources/measurement

All ultrasound examinations were performed using a Philips Affiniti 70 ultrasound system (Philips Healthcare, Bothell, WA, USA) equipped with a 3.5–5.0 MHz curvilinear transducer. Participants were instructed to fast for at least six hours before the examination to achieve adequate gallbladder distension. Ultrasound examinations were performed according to the departmental hepatobiliary imaging protocol with patients examined in both supine and left lateral decubitus positions. Additional patient repositioning was performed whenever necessary to improve visualization of gallbladder contents. Each observer independently acquired and interpreted their own ultrasound images during the same patient visit. Neither observer had access to the other's findings before completion of image acquisition and reporting. Findings were recorded immediately after each examination on separate standardized data collection forms to prevent information exchange between observers. Uniform operational definitions were established before study commencement to standardize interpretation. Gallstones were defined as echogenic intraluminal foci demonstrating posterior acoustic shadowing and mobility with changes in patient position. Sludge was defined as dependent low-level echogenic material without posterior acoustic shadowing. Gallbladder wall thickening was defined as anterior wall thickness greater than 3 mm measured in a well-distended gallbladder. Polyps were defined as immobile, non-shadowing echogenic lesions attached to the gallbladder wall. Pericholecystic fluid was defined as an anechoic fluid collection adjacent to the gallbladder wall. Comet-tail artifact was defined as a reverberation artifact arising from the gallbladder wall suggestive of adenomyomatosis.

Bias

Several measures were implemented to minimize potential sources of bias. Consecutive sampling was

employed to reduce selection bias. Observer bias was minimized by ensuring complete blinding between the radiology resident and senior sonologist, with each observer independently performing image acquisition and interpretation without access to the other's results. Standardized patient preparation, ultrasound equipment, scanning protocol, operational definitions, and structured recording forms were used throughout the study to reduce measurement bias and improve consistency. Immediate documentation of findings further minimized recall bias. Nevertheless, because each observer independently acquired images rather than interpreting identical stored images, some variability attributable to image acquisition could not be completely eliminated.

Study size

A total sample size of 150 participants was considered adequate for estimating interobserver agreement based on previous ultrasonography agreement studies evaluating hepatobiliary imaging. The sample size was calculated assuming an anticipated Cohen's kappa coefficient of approximately 0.60, a 95% confidence level, and a precision of 0.10. Consecutive eligible patients were recruited until the required sample size was achieved. The relatively long recruitment period reflected the operational requirement that both observers be simultaneously available to independently examine each participant during the same clinical visit while maintaining observer blinding.

Quantitative variables

Age was analyzed as a continuous variable, while sex and all ultrasonographic findings were treated as categorical variables and reported as frequencies and percentages. The presence or absence of each gallbladder pathology was recorded as a binary outcome for both observers. Interobserver agreement for each binary variable was quantified using Cohen's kappa coefficient with corresponding 95% confidence intervals.

Statistical methods

Data were entered, cleaned, and analyzed using Statistical Package for Social Sciences (SPSS) version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were calculated for all study variables. Continuous variables were summarized using means and standard deviations or medians and interquartile ranges where appropriate, whereas categorical variables were presented as frequencies and percentages. The primary statistical analysis assessed interobserver agreement between the radiology resident and senior sonologist using Cohen's kappa (κ) coefficient with corresponding 95% confidence intervals for each predefined gallbladder pathology. Kappa values were interpreted according to the Landis and Koch classification as poor (<0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (>0.80) agreement. Because the objective of the study was to evaluate agreement rather than diagnostic accuracy, emphasis was placed on kappa coefficients, observed agreement percentages, and confidence intervals rather than hypothesis testing. All statistical analyses were two-sided, and results were presented descriptively to facilitate interpretation of agreement across different gallbladder pathologies.

Results

Participants

During the study period from 1 March 2022 to 30 June 2025, a total of 150 consecutive adult patients with clinically suspected gallbladder disease fulfilled the eligibility criteria and were enrolled in the study. All recruited participants successfully underwent two independent ultrasonographic examinations during the same clinical visit, one performed by the radiology resident and the other by the senior sonologist. There were no incomplete examinations, withdrawals, or missing datasets; therefore, all 150 participants were included in the final interobserver agreement analysis.

Descriptive data

Each participant underwent standardized gallbladder ultrasonography following a minimum fasting period of six hours using the same ultrasound equipment and departmental scanning protocol. Both observers independently evaluated six predefined gallbladder abnormalities, including gallstones, biliary sludge, gallbladder wall thickening, pericholecystic fluid, gallbladder polyps, and comet-tail artifact. For every participant, the presence or absence of each sonographic finding was recorded separately by both observers without knowledge of the other observer's interpretation. The distribution of positive and negative findings recorded by the radiology resident and senior sonologist for each evaluated pathology is presented in Tables 1–6. Overall observed agreement ranged from 66.7% to 91.3%, depending on the specific gallbladder abnormality assessed.

Outcome data

The primary outcome of the study was interobserver agreement between the radiology resident and senior sonologist, quantified using Cohen's kappa (κ) coefficient with corresponding 95% confidence intervals. Gallstones demonstrated the highest reproducibility between observers, with an observed agreement of 91.3% and an almost perfect level of agreement (κ = 0.826; 95% CI: 0.735–0.917) (Table 1). Gallbladder polyps demonstrated the second-highest level of agreement, with an observed agreement of 81.3% and moderate concordance (κ = 0.581; 95% CI: 0.450–0.712) (Table 5). Moderate-to-low reproducibility was observed for the remaining

sonographic findings. Gallbladder wall thickening demonstrated 70.0% observed agreement with fair agreement (κ = 0.392; 95% CI: 0.247–0.537) (Table 3). Similarly, biliary sludge demonstrated 68.0% observed agreement (κ = 0.365; 95% CI: 0.217–0.513) (Table 2), while pericholecystic fluid showed 67.3% observed agreement (κ = 0.333; 95% CI: 0.185–0.481) (Table 4). Comet-tail artifact demonstrated the lowest observed reproducibility, with an agreement of 66.7% and fair interobserver agreement (κ = 0.341; 95% CI: 0.194–0.488) (Table 6).

Main results

The findings demonstrated considerable variation in interobserver agreement depending on the type of gallbladder pathology being evaluated. Agreement was consistently highest for objectively identifiable structural abnormalities, particularly gallstones, indicating excellent reproducibility between observers with different levels of experience. In contrast, sonographic findings requiring greater subjective interpretation, including biliary sludge, gallbladder wall thickening, pericholecystic fluid, and comet-tail artifact, demonstrated only fair agreement, reflecting increased observer variability. Among all evaluated abnormalities, gallstones showed the strongest diagnostic concordance between the radiology resident and senior sonologist, whereas pericholecystic fluid and comet-tail artifact exhibited the lowest levels of agreement. Gallbladder polyps demonstrated intermediate reproducibility with moderate agreement. These findings suggest that well-defined structural lesions can be identified consistently irrespective of observer experience, whereas inflammatory changes, subtle wall abnormalities, and artifact-related findings remain more susceptible to variation in image acquisition and interpretation.

Table 1: Interobserver Agreement for Gallstones (n=150)

Gallstones (Resident)	Sonologist Present	Sonologist Absent	Total
Present	72(48.0)	7(4.7)	79(52.7)
Absent	6(4.0)	65(43.3)	71(47.3)
Total	78(52.0)	72(48.0)	150(100)

Table 2: Interobserver Agreement for Sludge (n=150)

Sludge (Resident)	Sonologist Present	Sonologist Absent	Total
Present	47(31.3)	32(21.3)	79(52.7)
Absent	16(10.7)	55(36.7)	71(47.3)
Total	63(42.0)	87(58.0)	150(100)

Table 3: Interobserver Agreement for Gallbladder Wall Thickening (n=150)

Wall Thickening (Resident)	Sonologist Present	Sonologist Absent	Total
Present	66(44.0)	33(22.0)	99(66.0)
Absent	12(8.0)	39(26.0)	51(34.0)
Total	78(52.0)	72(48.0)	150(100)

Table 4: Interobserver Agreement for Pericholecystic Fluid (n=150)

Pericholecystic Fluid (Resident)	Sonologist Present	Sonologist Absent	Total
Present	62(41.3)	29(19.3)	91(60.7)
Absent	20(13.3)	39(26.0)	59(39.3)
Total	82(54.7)	68(45.3)	150(100)

Table 5: Interobserver Agreement for Polyps (n=150)

Polyps (Resident)	Sonologist Present	Sonologist Absent	Total
Present	86(57.3)	17(11.3)	103(68.7)
Absent	11(7.3)	36(24.0)	47(31.3)
Total	97(64.7)	53(35.3)	150(100)

Table 6: Interobserver Agreement for Comet-Tail Artifact (n=150)

Comet-Tail Artifact (Resident)	Sonologist Present	Sonologist Absent	Total
Present	51(34.0)	33(22.0)	84(56.0)
Absent	17(11.3)	49(32.7)	66(44.0)
Total	68(45.3)	82(54.7)	150(100)

Discussion

The present study evaluated the level of interobserver agreement between a radiology resident and a senior sonologist in the ultrasonographic assessment of common gallbladder pathologies. The findings demonstrated that the reproducibility of gallbladder ultrasonography varied considerably according to the sonographic abnormality being evaluated. Gallstones exhibited almost perfect agreement between observers, whereas gallbladder polyps demonstrated moderate agreement. In contrast, biliary sludge, gallbladder wall thickening, pericholecystic fluid, and comet-tail artifact showed only fair agreement, highlighting the greater subjectivity associated with interpretation of subtle inflammatory and artifact-related findings. These observations emphasize that observer experience continues to influence ultrasound interpretation despite standardized scanning protocols.

Gallstones demonstrated the highest level of agreement (κ = 0.826), indicating excellent reproducibility between observers with different levels of expertise. This finding is consistent with previous studies reporting that gallstones are among the most reliably identified sonographic abnormalities because of their characteristic appearance as echogenic mobile intraluminal foci with posterior acoustic shadowing [15,16]. Their distinct sonographic morphology minimizes interpretative ambiguity and contributes to high diagnostic consistency. Moreover, increasing implementation of structured reporting recommendations has further standardized the description of gallstones and may have contributed to the excellent agreement observed in the present study [17].

Conversely, biliary sludge demonstrated only fair agreement (κ = 0.365). Sludge typically appears as low-level echogenic material without posterior acoustic shadowing, and its visualization may vary according to gain settings, patient positioning, gallbladder distension, and scanning angle [18]. Small differences in image optimization or observer thresholds may therefore result in inconsistent interpretation. Similar challenges have been described in previous studies, which identified biliary sludge as one of the more observer-dependent findings during routine ultrasonography [18].

Gallbladder wall thickening also demonstrated fair agreement (κ = 0.392). Although a wall thickness exceeding 3 mm is widely accepted as abnormal, accurate measurement may be influenced by inadequate gallbladder distension, patient body habitus, adjacent hepatic inflammation, cardiac failure, hypoalbuminemia, and other systemic conditions [19,20]. Consequently, borderline measurements are particularly susceptible to variability among observers. Recent investigations evaluating the Gallbladder Reporting and Data System (GB-RADS) have similarly reported moderate variability in the assessment of gallbladder wall abnormalities despite the use of standardized classification systems [21,22]. These findings underscore the need for meticulous measurement techniques and adherence to standardized imaging criteria.

Gallbladder polyps demonstrated moderate agreement (κ = 0.581), reflecting relatively better reproducibility than other subtle findings but still indicating appreciable observer variability. This observation is comparable with previous studies reporting persistent interobserver variation despite international consensus recommendations regarding gallbladder polyp evaluation and management [8–10]. Small cholesterol polyps, adherent sludge, focal wall thickening, and technical limitations during scanning may contribute to disagreement in lesion detection and characterization [19].

An interesting observation in the present study was the relatively high proportion of gallbladder polyps reported by both observers compared with published epidemiological data. This discrepancy may be attributable to the symptomatic nature of the study population, which consisted exclusively of patients referred for suspected gallbladder disease rather than asymptomatic individuals undergoing screening examinations. Additionally, because histopathological confirmation or alternative imaging reference standards were unavailable, some small adherent echogenic lesions or focal mucosal irregularities may have been interpreted as polyps. Accordingly, these prevalence estimates should be interpreted with caution and should not be generalized to the broader population.

Pericholecystic fluid demonstrated the lowest agreement among the evaluated findings (κ = 0.333). Minimal pericholecystic fluid collections can be difficult to distinguish from adjacent bowel loops, peritoneal reflections, or other anatomic structures, particularly when present in small quantities. Variability in patient positioning, respiratory motion, and operator technique may further influence detection. The relatively low agreement observed in the present study therefore reflects the inherent challenges associated with interpreting subtle inflammatory changes during ultrasonography.

Similarly, comet-tail artifact demonstrated fair agreement (κ = 0.341). Identification of comet-tail artifacts requires familiarity with the characteristic reverberation pattern associated with adenomyomatosis and other benign gallbladder wall conditions. Visualization of these artifacts depends on appropriate transducer orientation, focal zone placement, machine settings, and observer experience [23]. Consequently, lower reproducibility for this finding is expected, particularly in routine clinical practice where examination conditions may vary.

The differences observed between the radiology resident and the senior sonologist likely reflect differences in clinical experience, image optimization, and pattern recognition rather than major diagnostic discrepancies. Previous educational studies have demonstrated that structured ultrasound training, simulation-based learning, supervised scanning sessions, and case-based educational interventions significantly improve image acquisition, diagnostic confidence, and reporting consistency among radiology trainees [24,25]. Incorporation of such educational strategies into residency curricula may further reduce observer variability and improve the reproducibility of hepatobiliary ultrasonography.

From a clinical perspective, these findings reinforce the importance of standardized scanning protocols, structured reporting templates, and ongoing competency assessment in ultrasound practice. Greater reproducibility of objectively identifiable abnormalities such as gallstones supports confidence in routine clinical decision-making, whereas findings demonstrating lower agreement may benefit from consultant review, repeat imaging, or complementary imaging modalities when clinically indicated. As artificial intelligence and computer-assisted diagnostic systems continue to emerge in abdominal imaging, high-quality and consistently interpreted ultrasound datasets will become increasingly important for algorithm development and validation [14].

Limitations

Several limitations should be considered when interpreting the findings of this study. First, no external reference standard, such as surgical findings, histopathology, magnetic resonance imaging, computed tomography, or longitudinal clinical follow-up, was available. Consequently, the study evaluated observer agreement rather than diagnostic accuracy, and it cannot be determined which observer's interpretation more closely reflected the true diagnosis. Second, each observer independently acquired and interpreted their own ultrasound images. Although this design closely reflects routine clinical practice, differences in image acquisition may have contributed to disagreement in addition to differences in image interpretation. Third, the study included only one radiology resident and one senior sonologist from a single tertiary-care institution, which may limit the generalizability of the findings to other training environments or healthcare settings. Fourth, potentially important participant characteristics, including age distribution, sex, body mass index, and clinical severity, were not analyzed in relation to observer agreement and therefore could not be evaluated as factors influencing reproducibility.

Finally, the single-center design and referral-based study population may have introduced spectrum bias, limiting extrapolation of the findings to screening populations or primary healthcare settings. Future multicenter studies involving multiple observers with varying levels of expertise, incorporation of standardized image review, and comparison with definitive diagnostic reference standards would provide more comprehensive evidence regarding the reproducibility and diagnostic accuracy of gallbladder ultrasonography.

Conclusion

Interobserver agreement in gallbladder ultrasonography varies according to the sonographic abnormality being evaluated. Gallstones demonstrated almost perfect agreement between the radiology resident and senior sonologist, confirming excellent reproducibility for this well-defined structural lesion. Gallbladder polyps showed moderate agreement, whereas biliary sludge, gallbladder wall thickening, pericholecystic fluid, and comet-tail artifact demonstrated only fair agreement, reflecting the greater subjectivity associated with subtle inflammatory and artifact-related findings.

These results emphasize the continued importance of standardized scanning protocols, structured reporting systems, and focused ultrasound training to improve consistency in image acquisition and interpretation. Implementation of these quality-improvement strategies

may enhance diagnostic reliability, strengthen resident education, and improve the overall quality of gallbladder ultrasonography in routine clinical practice.

Authors' Contributions:

All authors meet the ICMJE criteria for authorship.

Author 1: Substantial contribution to conception and design, drafting the manuscript, final approval, agreed to be accountable for all the work.
Author 2: Substantial contribution to conception and design, data acquisition, drafting the manuscript, final approval, agreed to be accountable for all the work.
Author 3: Analysis and interpretation of data, drafting the manuscript, final approval, agreed to be accountable for all the work.
Author 4: Analysis and interpretation of data, critical revision, final approval, agreed to be accountable for all the work.
Author 5: Substantial contribution to conception and design, drafting the manuscript, final approval, agreed to be accountable for all the work.
Author 6&7: Substantial contribution to conception and design, data acquisition, drafting the manuscript, final approval, agreed to be accountable for all the work.

References

1. Wilson SJ, Thavanathan R, Cheng D, Adhikari S. Test characteristics of emergency medicine-performed point-of-care ultrasound for the diagnosis of acute cholecystitis: a systematic review and meta-analysis. *Ann Emerg Med.* 2024;83(3):235-246. https://doi.org/10.1016/j.annemergmed.2023.09.005
2. Room H, Wood A, Ji C, Nightingale H, Toh SKC. Performance of ultrasound n the diagnosis of cholecystitis: not so (ultra)sound? *Ann R Coll Surg Engl.* 2022;104(9):655-660. https://doi.org/10.1308/rcsann.2021.0322
3. Gallaher JR, Charles A. Acute cholecystitis: a review. *JAMA.* 2022;327(10):965-975. https://doi.org/10.1001/jama.2022.2350
4. Lucius C, Braden B, Jenssen C, Dietrich CF, Cui XW, gnee A, et al. Ultrasound of the gallbladder-an update on measurements, reference values, variants and frequent pathologies: a scoping review. *Life (Basel).* 2025;15(6):941. https://doi.org/10.3390/life15060941
5. Evanson DJ, Elcic L, Uyeda JW, Zulfiqar M. maging of gallstones and complications. *Curr Probl Diagn Radiol.* 2025;54(3):392-403. https://doi.org/10.1067/j.cpradiol.2024.12.007
6. Riddell ZC, Corallo C, Albazaz R, Foley KG. Gallbladder polyps and adenomyomatosis. *Br J Radiol.* 2023;96(1142): 20220115. https://doi.org/10.1259/bjr.20220115
7. Bonde AA, Virarkar M, Zahid M, Jaganathan S, Menendez MJ, Calimano L, et al. maging update on gallbladder adenomyomatosis and ts mimics. *Clin maging.* 2024;105: 109997. https://doi.org/10.1016/j.clinimag.2023.109997
8. Kamaya A, Fung C, Szpakowski JL, Fetzer DT, Walsh AJ, Alimi Y, et al. Management of ncidentally detected gallbladder polyps: Society of Radiologists n Ultrasound consensus conference recommendations. *Radiology.* 2022;305(2):277-289. https://doi.org/10.1148/radiol.213079
9. Foley KG, Lahaye MJ, Thoeni RF, Soltes M, Dewhurst C, Barbu ST, et al. Management and follow-up of gallbladder polyps: updated joint guidelines between ESGAR, EAES, EFISDS and ESGE. *Eur Radiol.* 2022;32(5):3358-3368. https://doi.org/10.1007/s00330-021-08384-w
10. Anderson MA, Mercaldo S, Cao J, Mroueh N, Furtado FS, Cochran J, et al. Society of Radiologists n Ultrasound Consensus Conference Recommendations for ncidental Gallbladder Polyp Management: nterreader agreement among 10 radiologists. *AJR Am J Roentgenol.* 2024; 222(5):e2330720. https://doi.org/10.2214/AJR.23.30720
11. Soundararajan R, Subramanian P, Gupta P, Rana P, Chhabra M, Singh S, et al. Agreement of Gallbladder Reporting and Data System for gallbladder wall thickening at ultrasonography: a multireader validation study. *J Clin Exp Hepatol.* 2024;14(4):101393. https://doi.org/10.1016/j.jceh.2024.101393
12. Gupta P, Dutta U, Rana P, Singhal M, Gulati A, Kalra N, et al. Gallbladder reporting and data system (GB-RADS) for risk stratification of gallbladder wall thickening on ultrasonography: an nternational expert consensus. *Abdom Radiol (NY).* 2022;47(2):554-565. https://doi.org/10.1007/s00261-021-03360-w
13. Nobel JM, van Geel K, Robben SGF. Structured reporting n radiology: a systematic review to explore ts potential. Eur Radiol. 2022;32(4):2837-2854. https://doi.org/10.1007/s00330-021-08327-5
14. Ahmed AS, Ahmed SS, Mohamed S, Salman NE, Humidan AAM, brahim RF, et al. Advancements n cholelithiasis diagnosis: a systematic review of machine learning applications n maging analysis. *Cureus.* 2024;16(8):e66453. https://doi.org/10.7759/cureus.66453
15. Bree RL. Further observations on the usefulness of the sonographic Murphy sign In the evaluation of suspected acute cholecystitis. *J Clin Ultrasound.* 1995;23(3):169-172. https://doi.org/10.1002/jcu.1870230304
16. Mencarini L, Vestito A, Zagari RM, Montagnani M. New developments n the ultrasonography diagnosis of gallbladder diseases. *Gastroenterol nsights.* 2024;15(1):42-68. https://doi.org/10.3390/gastroent15010004
17. Mugu VK, Carr BM, Olson MC, Schupbach JC, Eguia FA, Schmitz JJ, et al. ncreasing adherence to societal recommendations n radiology reporting: a feasibility study using Society of Radiologists n Ultrasound guidelines for ncidentally detected gallbladder polyps. *Ultrasound Q.* 2025;41(1):e00699. https://doi.org/10.1097/RUQ.0000000000000699
18. Kim M, Kang TW, Jang KM, Kim YK, Kim SH, Ha SY, et al. Tumefactive gallbladder sludge at US: prevalence and clinical mportance. *Radiology.* 2017;283(2):570-579. https://doi.org/10.1148/radiol.2016161042

19. Wiles R, Thoeni RF, Barbu ST, Vashist YK, Rafaelsen SR, Dewhurst C, et al. Management and follow-up of gallbladder polyps: joint guidelines between ESGAR, EAES, EFISDS and ESGE. *Eur Radiol*. 2017;27(9):3856-3866. https://doi.org/10.1007/s00330-017-4742-y
20. Catalano OA, Sahani DV, Forcione DG, Czermak B, Liu CH, Soricelli A, et al. Biliary nfections: spectrum of maging findings and management. *Radiographics*. 2009; 29(7):2059-2080. https://doi.org/10.1148/rg.297095051
21. van Breda Vriesman AC, Engelbrecht MR, Smithuis RHM, Puylaert JBCM. Diffuse gallbladder wall thickening: differential diagnosis. *AJR Am J Roentgenol*. 2007; 188(2):495-501. https://doi.org/10.2214/AJR.05.1712
22. Wang R, Lv L, Li L. Diagnostic performance of the gallbladder reporting and data system combined with color Doppler flow maging for gallbladder cancer In the Asian population. *Front Oncol*. 2024;14:1367351. https://doi.org/10.3389/fonc.2024.1367351
23. Oh SH, Han HY, Kim HJ. Comet tail artifact on ultrasonography: sta reliable finding of benign gallbladder diseases? *Ultrasonography*. 2019;38(3):221-230. https://doi.org/10.14366/usg.18029
24. Alexander LF, McComb BL, Bowman AW, Bonnett SL, Ghazanfari SM, Caserta MP. Ultrasound simulation training for radiology residents-curriculum design and mplementation. *J Ultrasound Med*. 2023;42(4):777-790. https://doi.org/10.1002/jum.16098
25. Sugi MD, Kennedy TA, Shah V, Hartung MP. Bridging the gap: nteractive, case-based learning n radiology education. *Abdom Radiol (NY)*. 2021;46(12):5503-5508. https://doi.org/10.1007/s00261-021-03147-z