Volume 3, Number 2, 2026
Pages 288-307
Bioscience Insights
ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)
BiosciInsights.com

AI-ASSISTED ULTRASOUND IMAGING FOR EARLY DETECTION OF HEPATIC AND ABDOMINAL PATHOLOGIES: A DIAGNOSTIC ACCURACY STUDY IN RESOURCE-LIMITED RADIOLOGY SETTINGS OF PAKISTAN

Hafiz Abdur Rahman[*1], Habiba Rehmat[2]

[*1]*Student UIRSMIT UOL*
[2]*Student UIRSMIT*

[*1]abdurrahmanrauf2000@gmail.com,[2]70129575@student.uol.edu.pk

Article Information

Article History

Received:	May 15, 2026
Revised:	June 27, 2026
Accepted:	August 12, 2026
Available Online:	August 31, 2026

Keywords:
Artificial intelligence; Ultrasound; Hepatic pathology; Abdominal pathology; Diagnostic accuracy; Pakistan.

Abstract

Artificial intelligence (AI)-assisted ultrasound may improve the detection of hepatic and abdominal pathologies, particularly in resource-limited settings with limited access to advanced imaging and specialist expertise. To evaluate the diagnostic accuracy of AI-assisted ultrasound for detecting hepatic and abdominal pathologies in resource-limited radiology settings in Pakistan. A prospective diagnostic accuracy study was conducted among 510 adults undergoing abdominal ultrasonography. AI-assisted interpretation was compared with conventional radiologist assessment against an appropriate reference standard. Sensitivity, specificity, predictive values, accuracy, AUC, and Cohen's kappa were calculated. AI-assisted ultrasound demonstrated 91.5% sensitivity, 93.3% specificity, 85.4% PPV, 96.2% NPV, and 92.7% overall accuracy. Excellent diagnostic discrimination was observed (AUC = 0.94, 95% CI: 0.91–0.96), with high agreement with radiologists (κ = 0.82, 95% CI: 0.76–0.87). AI-assisted ultrasound showed promising diagnostic performance and potential as a second-reader and decision-support tool in resource-limited radiology settings. Prospective multicenter validation using locally representative datasets is warranted.

INTRODUCTION

Ultrasonography (US) remains an important first-line imaging modality for the assessment of hepatic and abdominal disorders because it is non-ionizing, relatively inexpensive, portable, repeatable, and suitable for real-time examination. These characteristics are particularly important in resource-constrained healthcare environments where access to computed tomography (CT), magnetic resonance imaging (MRI), contrast-enhanced imaging, and highly specialized radiological services may be limited. The World Health Organization has emphasized the importance of accessible and quality-assured diagnostic services, particularly in settings where shortages of equipment, trained personnel, infrastructure, and specialist expertise constrain timely diagnosis (World Health Organization [WHO], 2024, 2026).

Despite these advantages, conventional ultrasound remains substantially dependent on operator experience, image acquisition quality, patient characteristics, and subjective interpretation. Variability in scanning technique and image interpretation can affect the detection and characterization of hepatic steatosis, fibrosis, focal liver lesions, gallbladder abnormalities, ascites, and other abdominal pathologies. Recent evidence indicates that artificial intelligence (AI), particularly machine learning and deep-learning techniques, can enhance ultrasound interpretation by extracting quantitative imaging features, identifying subtle patterns, classifying lesions, and supporting diagnostic decision-making (Cai & Pfob, 2025; Punn et al., 2024).

The application of AI to liver ultrasound has expanded rapidly. Systematic evidence has demonstrated that AI-based approaches can assist in detecting and quantifying hepatic steatosis, fibrosis, and focal liver lesions. However, considerable variation remains in algorithms, image acquisition protocols, datasets, reference standards, and validation strategies (Bashar et al., 2023; Punn et al., 2024). In a recent prospective study, Santoro et al. (2024) evaluated an AI-supported ultrasound approach for detecting hepatic steatosis against MRI-derived proton density fat fraction. The AI-derived hepatorenal index demonstrated stronger discriminatory performance than manually calculated measurements, with an area under the curve (AUC) of 0.87 for distinguishing steatosis from non-steatosis and an AUC of 0.98 for differentiating mild from moderate/severe steatosis. These findings suggested that AI could improve both the objectivity and diagnostic efficiency of conventional ultrasound.

AI has also demonstrated potential in the characterization of focal hepatic lesions. A comparative study of contrast-enhanced ultrasound found that an AI-based model achieved higher specificity than experienced clinicians for benign-versus-malignant classification, although sensitivity remained lower than that of the clinicians (Dietrich et al., 2023). Similarly, Brooks et al. (2025) reviewed 41 studies of AI applied to liver contrast-enhanced ultrasound and reported generally promising diagnostic performance. However, they identified a major methodological limitation: most studies relied on selected or manually annotated images, while external validation was uncommon. This limitation is particularly important because high performance in controlled datasets does not necessarily translate into reliable performance in routine clinical environments.

The evidence base has increasingly extended beyond hepatic steatosis and focal liver lesions to other abdominal applications. Cai and Pfob (2025), in a review of AI applications in abdominal and pelvic ultrasound, identified substantial research activity in multi-organ detection and hepatobiliary imaging. However, most of the reviewed studies were retrospective and single-center, indicating limited evidence regarding prospective, real-world implementation. Similarly, recent evidence concerning gallbladder ultrasound indicates that AI may assist in the detection and classification of gallbladder polyps, malignancy, stones, and other abnormalities, although the available evidence remains relatively narrow and requires broader clinical validation.

The relevance of AI-assisted ultrasound is particularly pronounced in low- and middle-income countries (LMICs). AI has the potential to function as a decision-support and second-reader technology where radiological expertise is unevenly distributed. Recent literature on AI implementation in LMICs has identified shortages of imaging equipment, unstable infrastructure, limited bandwidth, inadequate representative datasets, workforce shortages, regulatory uncertainty, and financing constraints as important barriers to implementation. These challenges are directly relevant to Pakistan, where access to advanced diagnostic imaging and specialist radiological expertise is not uniformly distributed across urban, secondary-care, rural, and underserved settings.

The Pakistani context therefore provides an important environment in which to evaluate whether AI-assisted ultrasound can improve diagnostic performance without requiring the infrastructure associated with more expensive imaging modalities. Recent Pakistani literature has emphasized both the growing interest in AI-assisted radiology and the need for appropriate local validation, clinician training, ethical safeguards, and human oversight. Importantly, an AI system developed and validated using data from highly resourced institutions may not perform equivalently when applied to different populations, ultrasound machines, acquisition protocols, disease distributions, and clinical workflows.

A critical gap consequently exists between the promising diagnostic performance reported by AI-ultrasound studies and the limited evidence concerning prospective, clinically relevant, locally validated AI-assisted ultrasound in resource-limited Pakistani radiology settings. Much of the existing literature has focused on individual diseases, selected image datasets, or single-organ applications rather than evaluating a clinically integrated abdominal ultrasound approach. Moreover, relatively little evidence is available regarding whether AI assistance can improve sensitivity, specificity, overall diagnostic accuracy, and agreement with expert radiologists across a clinically heterogeneous spectrum of hepatic and abdominal abnormalities in Pakistan.

The present study was therefore designed to evaluate the diagnostic accuracy of AI-assisted ultrasound imaging for the early detection of selected hepatic and abdominal pathologies in resource-limited radiology settings in Pakistan. By comparing AI-assisted interpretation with conventional radiologist assessment against an appropriate reference standard, the study aimed to generate context-specific evidence regarding sensitivity, specificity, predictive values, diagnostic accuracy, and receiver operating characteristic performance. The study further sought to determine whether AI could serve as a clinically useful assistive technology rather than a replacement for radiologists, thereby addressing both the diagnostic and health-system dimensions of AI adoption in resource-constrained environments.

Problem Statement

Hepatic and abdominal diseases represent an important diagnostic burden requiring timely and accurate imaging assessment. Ultrasonography is particularly valuable in resource-limited settings because of its relative affordability, portability, absence of ionizing radiation, and availability compared with CT and MRI. Nevertheless, the diagnostic effectiveness of conventional ultrasound remains influenced by operator expertise, equipment quality, patient-related factors, image acquisition technique, and subjective interpretation. These limitations can become particularly consequential in settings where specialist radiologists are scarce or where patients have limited access to advanced confirmatory imaging.

Artificial intelligence offers a potential mechanism for addressing some of these limitations by providing automated image analysis, quantitative feature extraction, lesion detection, classification, and decision support. Recent studies have reported encouraging diagnostic performance for AI-assisted ultrasound in hepatic steatosis, liver fibrosis, focal liver lesions, and other abdominal applications. However, the existing evidence remains insufficient for routine implementation. Systematic reviews have repeatedly identified limitations involving retrospective study designs, small or highly selected datasets, single-center validation, manually selected images, heterogeneous reference standards, and inadequate external validation (Brooks et al., 2025; Cai & Pfob, 2025; Punn et al., 2024).

The central research problem is therefore not whether AI can demonstrate high diagnostic performance under controlled experimental conditions, but whether such performance can be reproduced within real-world, resource-constrained clinical environments. Ultrasound images acquired in routine practice may differ substantially from curated research datasets because of differences in machine manufacturers, transducer characteristics, image quality, operator skill, patient body habitus, disease prevalence, acquisition protocols, and institutional practices. These sources of variation can generate dataset shift and may substantially reduce the generalizability of AI models.

The problem is particularly relevant to Pakistan. Although AI adoption in Pakistani healthcare is increasing, evidence regarding locally validated AI-assisted radiology remains limited. Resource disparities between major tertiary hospitals and peripheral or underserved facilities can create substantial differences in diagnostic capacity. A technology that requires expensive infrastructure, high-performance computing, continuous internet connectivity, or extensive specialist supervision may therefore have limited practical value in such environments. Conversely, a validated AI-assisted

ultrasound system that can operate with existing ultrasound infrastructure could potentially provide a relatively scalable method of strengthening diagnostic support.

Another important gap concerns the clinical breadth of existing research. Many previous studies have concentrated on a single pathology, such as steatosis, fibrosis, or focal liver lesions. Although disease-specific models can achieve high accuracy, real-world abdominal ultrasound frequently involves simultaneous assessment of multiple organs and pathological conditions. Consequently, evidence is needed on whether AI-assisted interpretation can provide useful diagnostic support across a clinically relevant spectrum of hepatic and abdominal abnormalities rather than only a narrowly defined target condition.

Furthermore, reported AI performance cannot be interpreted solely through overall accuracy. In clinical diagnosis, false-negative findings may delay treatment, whereas false-positive findings may increase unnecessary investigations and healthcare costs. Therefore, sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios, AUC, and agreement with expert radiologists are necessary for a comprehensive assessment of diagnostic utility.

The research gap is consequently defined by the limited availability of prospective, context-specific diagnostic accuracy evidence evaluating AI-assisted ultrasound across hepatic and abdominal pathologies in resource-limited Pakistani radiology settings. Addressing this gap could establish whether AI provides measurable diagnostic benefits under local clinical conditions and whether it can function as a safe and effective second-reader or decision-support mechanism. The findings could also provide evidence for subsequent multicenter validation, implementation research, health-technology assessment, and policy development concerning AI-enabled diagnostic imaging in Pakistan.

Research Questions

1. What is the diagnostic accuracy of AI-assisted ultrasound imaging for detecting selected hepatic and abdominal pathologies in resource-limited radiology settings in Pakistan?
2. How do the sensitivity, specificity, positive predictive value, negative predictive value, and AUC of AI-assisted ultrasound compare with conventional radiologist interpretation?
3. Does AI-assisted ultrasound significantly improve diagnostic accuracy compared with conventional ultrasound interpretation alone?
4. What level of agreement exists between AI-assisted ultrasound interpretation and expert radiologist assessment?
5. Does AI-assisted ultrasound maintain diagnostic performance across different hepatic and abdominal pathological conditions and varying image-quality levels?
6. What is the potential clinical utility of AI-assisted ultrasound as a decision-support or second-reader tool in resource-limited Pakistani radiology settings?

Research Objectives

General Objective

To evaluate the diagnostic accuracy and clinical utility of AI-assisted ultrasound imaging for the early detection of hepatic and abdominal pathologies in resource-limited radiology settings in Pakistan.

Specific Objectives

1. To determine the sensitivity, specificity, positive predictive value, negative predictive value, overall diagnostic accuracy, and AUC of AI-assisted ultrasound imaging for selected hepatic and abdominal pathologies.
2. To compare the diagnostic performance of AI-assisted ultrasound with conventional radiologist interpretation using an appropriate reference standard.
3. To determine whether AI assistance produces a statistically significant improvement in diagnostic accuracy compared with conventional ultrasound interpretation.
4. To assess the agreement between AI-assisted interpretation and expert radiologist assessment using appropriate measures such as Cohen's kappa.
5. To evaluate the diagnostic performance of the AI-assisted approach across different pathological categories and image-quality conditions.

6. To determine the potential of AI-assisted ultrasound as a second-reader and clinical decision-support tool in resource-limited radiology settings.
7. To identify the contextual and technical factors that may influence the generalizability and clinical applicability of AI-assisted ultrasound within Pakistani healthcare facilities.

Significance of the Study

Theoretical Significance

The study contributes to the theoretical development of AI-assisted diagnostic imaging by examining whether algorithmic image interpretation can improve diagnostic discrimination within a real-world clinical environment. It extends existing evidence beyond highly controlled or disease-specific datasets by examining AI performance across a clinically relevant range of hepatic and abdominal conditions. The study also contributes to the growing literature on human-AI collaboration by conceptualizing AI as an assistive diagnostic technology operating alongside radiologists rather than as a replacement for professional clinical judgment.

Practical Significance

The study has direct relevance for radiologists, sonographers, clinicians, and healthcare administrators. If AI-assisted ultrasound demonstrates superior or complementary diagnostic performance, it could support radiologists by highlighting suspicious findings, reducing interpretation variability, and prioritizing examinations requiring specialist review. In facilities with limited specialist availability, AI could potentially provide an additional layer of diagnostic support while retaining radiologists as the final decision-makers.

The findings may also support the development of low-cost, scalable diagnostic workflows using existing ultrasound infrastructure. This is particularly important where CT or MRI services are unavailable, unaffordable, geographically inaccessible, or associated with substantial waiting times.

Clinical Significance

Improved early detection of hepatic and abdominal abnormalities could facilitate earlier referral, confirmatory investigation, treatment planning, and clinical monitoring. A high-performing AI-assisted system could be particularly valuable for identifying subtle or easily overlooked abnormalities and for standardizing interpretation among operators with different levels of experience. However, clinical implementation should depend on demonstrated sensitivity, specificity, external validation, safety, and appropriate human oversight rather than AI performance alone.

Policy Significance

The study could provide locally relevant evidence for policymakers considering the integration of AI into Pakistan's diagnostic-health infrastructure. Evidence regarding diagnostic accuracy, workflow compatibility, and performance in resource-constrained environments could inform decisions concerning procurement, regulatory approval, clinical governance, training, data protection, and reimbursement.

The study is also relevant to broader efforts to strengthen equitable access to diagnostic services. Recent WHO initiatives have emphasized strengthening diagnostic capacity and equitable access to diagnostic imaging, particularly in underserved settings. AI-assisted ultrasound could potentially contribute to this objective if implementation is supported by appropriate validation, infrastructure, workforce development, quality assurance, and regulatory safeguards.

Health-System Significance

At the health-system level, the study could contribute evidence concerning whether AI-assisted ultrasound represents a cost-conscious approach to expanding diagnostic capacity without requiring every facility to develop advanced imaging infrastructure. If validated successfully, AI could support peripheral facilities through automated preliminary assessment and referral prioritization while allowing complex or uncertain cases to be escalated to specialist radiologists.

Ultimately, the study could provide a foundation for developing locally validated, human-supervised, and resource-sensitive AI diagnostic pathways capable of improving access to timely abdominal imaging while maintaining professional accountability and patient safety.

Literature Review

1. Artificial Intelligence in Diagnostic Ultrasound

Artificial intelligence (AI) has increasingly emerged as a transformative technology in diagnostic imaging, particularly through machine learning and deep-learning algorithms capable of automated image classification, segmentation, feature extraction, and disease detection. Ultrasound represents a particularly relevant modality for AI integration because it is widely available, relatively inexpensive, portable, radiation-free, and suitable for repeated examinations. These characteristics make AI-assisted ultrasound potentially valuable in healthcare environments where access to CT, MRI, and specialist radiological expertise is constrained.

Recent evidence demonstrates rapid expansion of AI applications across abdominal and pelvic ultrasound. Cai and Pfob (2025) reviewed 57 studies and reported applications involving multi-organ detection, hepatobiliary imaging, image classification, segmentation, and automated image processing. Importantly, deep learning constituted the dominant methodological approach. However, approximately three-quarters of the studies were single-center and almost all were retrospective, while relatively few had undergone prospective multicenter validation. This distinction is critical because high diagnostic performance in retrospective datasets does not necessarily establish clinical effectiveness in routine practice.

The literature therefore suggests that AI-assisted ultrasound has moved beyond technical proof-of-concept toward clinically relevant applications, but the evidence base remains characterized by substantial methodological heterogeneity. Differences in ultrasound equipment, acquisition protocols, patient populations, disease prevalence, image quality, reference standards, and validation procedures make direct comparison of reported diagnostic performance difficult. Consequently, diagnostic accuracy studies conducted under real-world clinical conditions remain necessary.

2. AI-Assisted Detection of Hepatic Steatosis

Hepatic steatosis represents one of the most extensively investigated applications of AI-assisted ultrasound. Conventional ultrasound is widely used for detecting fatty liver because of its accessibility and relatively low cost; however, early or mild steatosis may be difficult to identify reliably, and qualitative interpretation can vary among observers. AI provides an opportunity to convert subjective visual characteristics into reproducible quantitative measurements.

Santoro et al. (2024) evaluated AI-supported ultrasonography for the diagnosis and grading of hepatic steatosis. Their findings demonstrated that AI-derived quantitative measures could discriminate steatotic from non-steatotic liver and differentiate degrees of disease severity. The study provided clinically relevant evidence that AI can augment conventional sonographic assessment rather than merely automate image processing.

The evidence has become stronger with recent systematic synthesis. A 2025 meta-analysis evaluating AI for hepatic steatosis reported a pooled sensitivity of approximately **0.95**, specificity of **0.93**, and summary AUC of **0.98** across eligible cohorts. Although these findings indicate excellent diagnostic potential, the substantial heterogeneity among studies remains important when considering clinical translation. In particular, performance estimates derived from selected research datasets may overestimate performance in heterogeneous routine clinical populations.

The major implication is that AI may improve the objectivity and reproducibility of ultrasound-based steatosis assessment, but validation against appropriate reference standards remains essential. AI should therefore be evaluated not simply on whether it identifies abnormal ultrasound patterns but on whether its classification corresponds reliably with clinically accepted reference measures.

3. AI for Focal Liver Lesions and Hepatic Malignancy

Another important application involves the detection and classification of focal liver lesions, including hepatocellular carcinoma (HCC), metastases, hemangiomas, and other benign or malignant

abnormalities. Contrast-enhanced ultrasound (CEUS) provides dynamic information concerning lesion vascularity, but interpretation can be subjective and dependent on operator experience.

Dietrich et al. (2023) evaluated an AI model for classifying liver lesions using CEUS and compared its performance with experienced clinicians. For benign-versus-malignant classification, the AI system demonstrated very high specificity but lower sensitivity than the clinicians. In multiclass classification, clinicians generally achieved higher diagnostic accuracy for HCC and metastases. Importantly, however, the AI system demonstrated potential as a second-opinion tool, particularly for less experienced clinicians.

This finding is important because it challenges an overly simplistic assumption that AI should replace expert radiologists. Rather, current evidence suggests that the most clinically realistic role may involve human-AI collaboration, in which AI provides an independent probability estimate, highlights suspicious findings, or acts as a second reader while the radiologist retains final diagnostic responsibility.

Brooks et al. (2025) systematically reviewed 41 studies of AI applied to liver CEUS. Most studies focused on focal liver-lesion classification, with reported accuracies ranging widely from 58.6% to 98.9%. However, the review identified a major translational weakness: 25 of the 41 studies used hand-picked or manually annotated images, and external validation was generally uncommon. The authors consequently emphasized the need for prospective, externally validated, and multicenter research before routine clinical implementation.

More recent literature has reinforced this conclusion. Ciocalteu et al. (2025) reported that AI-assisted CEUS could improve lesion characterization, standardize enhancement-pattern interpretation, and potentially support automated LI-RADS classification. Nevertheless, the authors emphasized the need for large-scale and multicenter validation before widespread clinical adoption.

4. AI-Assisted Detection of Other Abdominal Pathologies

The potential of AI extends beyond liver disease. Recent research has examined gallbladder disease, biliary abnormalities, abdominal trauma, appendicitis, renal pathology, and multi-organ ultrasound interpretation. Cai and Pfob (2025) reported that hepatobiliary applications constituted a substantial proportion of abdominal ultrasound AI research. Models have demonstrated promising performance for acute cholecystitis, biliary abnormalities, gallbladder polyps, focal liver lesions, renal cysts, and abdominal trauma.

For example, deep-learning models for acute cholecystitis have achieved AUC values above 0.90 in validation datasets, while a multicenter model for focal liver-lesion detection reported sensitivity approaching 97% and specificity around 90%. However, some apparently impressive studies contained methodological weaknesses such as lack of independent validation or data leakage between training and testing datasets.

This distinction is particularly important for the proposed study. A diagnostic model can demonstrate excellent performance during internal validation while performing substantially less well when confronted with images obtained from a different institution, ultrasound machine, operator, or patient population. Therefore, diagnostic accuracy should be interpreted together with validation strategy, data representativeness, and clinical setting.

5. AI and Ultrasound in Resource-Limited Settings

The relevance of AI-assisted ultrasound becomes greater in resource-limited environments. Ultrasound already provides an important diagnostic advantage because it does not require ionizing radiation, is relatively portable, and can be deployed outside advanced tertiary hospitals. AI could potentially address another major limitation: the shortage and uneven distribution of personnel with advanced ultrasound interpretation skills.

Kim et al. (2024) reviewed AI applications for point-of-care ultrasound in low-resource settings and identified 29 relevant studies. Their review found substantial interest in using deep learning to support image acquisition and interpretation where experienced personnel and other diagnostic resources are scarce. However, the authors identified generalizability, dataset availability, regional disparities, device and protocol standardization, and ethical issues as important barriers to implementation.

More recent evidence has broadened this perspective to diagnostic radiology in low- and middle-income countries. A 2026 scoping review identified 51 studies across 33 LMICs and found that AI applications generally demonstrated potential to improve diagnostic sensitivity, specificity, applicability, and workflow efficiency. However, small samples, single-center designs, public datasets, and limited multicenter validation remained frequent limitations.

These findings have direct implications for Pakistan. AI systems developed using datasets from high-income healthcare environments cannot automatically be assumed to perform equivalently in Pakistani radiology departments. Differences in ultrasound equipment, technical expertise, patient demographics, disease spectrum, image quality, and healthcare workflows may produce substantial distribution shifts.

6. Infrastructure and Implementation Challenges

The technical performance of an AI model represents only one component of successful clinical implementation. Recent research has identified infrastructure, data availability, workforce, regulation, ethics, financing, and sustainability as interconnected determinants of AI adoption in LMICs.

A recent perspective on AI implementation in LMIC medical imaging identified five major barriers: infrastructure deficiencies, inadequate or nonrepresentative datasets, shortages of trained personnel, uncertain ethical and regulatory frameworks, and financing and sustainability constraints. These barriers are particularly relevant to ultrasound because image quality and acquisition are strongly influenced by equipment and operator technique.

The implications are significant for the proposed Pakistani study. A clinically useful AI system should ideally function with available ultrasound equipment, tolerate variation in image quality, require manageable computational resources, and operate within existing radiology workflows. Therefore, evaluating diagnostic accuracy in a resource-limited setting is not merely a geographic extension of existing AI research; it represents an important test of external validity and implementation feasibility.

7. Human-AI Collaboration Rather Than Replacement

An emerging theme in the literature is that AI is most likely to provide value when integrated into clinical decision-making rather than used as a completely autonomous diagnostic replacement. The comparative CEUS study by Dietrich et al. (2023) demonstrated that experienced clinicians could outperform the AI model for some lesion-classification tasks, whereas the AI system provided high specificity and potential second-opinion value.

This finding suggests that AI-assisted ultrasound should be conceptualized as a **decision-support system**. Such an approach may be especially appropriate in resource-limited settings because AI can support less-experienced clinicians while allowing complex cases to be referred to radiologists. This model could potentially improve diagnostic consistency without eliminating professional oversight.

The concept is also consistent with research on AI-enabled ultrasound in LMICs, where AI has been proposed as a mechanism for supporting providers who have access to ultrasound equipment but limited specialist expertise. However, implementation should incorporate human oversight, quality assurance, appropriate training, and clear responsibility for final clinical decisions.

8. Methodological Limitations in Existing Literature

Despite strong reported diagnostic performance, several recurring methodological weaknesses limit the strength of the current evidence.

First, many studies have used retrospective datasets, which may introduce selection bias. Second, a substantial proportion of studies have relied on single-center data, limiting generalizability. Third, many models have been developed using hand-selected or manually annotated images, which may not reflect routine clinical examinations. Fourth, external validation has been relatively uncommon. Fifth, differences in reference standards make comparison across studies difficult.

Cai and Pfob (2025) found that 56 of 57 reviewed studies were retrospective and 43 were single-center. The authors consequently concluded that the risk of bias remained high because relatively few applications had undergone prospective multicenter validation.

The CEUS systematic review by Brooks et al. (2025) similarly emphasized the lack of external validation. (Karger) These findings demonstrate that the field has produced substantial evidence of technical feasibility but comparatively less evidence of clinical effectiveness and generalizability.

9. Research Gap

The literature establishes that AI can achieve promising diagnostic performance in ultrasound-based detection of hepatic and abdominal pathology. However, four major gaps remain.

First, much of the evidence has focused on single diseases, particularly hepatic steatosis or focal liver lesions, rather than evaluating AI across a clinically relevant range of hepatic and abdominal pathologies.

Second, a substantial proportion of research has been conducted using retrospective, single-center, highly curated datasets, limiting generalizability to routine clinical environments.

Third, there is insufficient evidence regarding AI performance in resource-limited radiology settings, where differences in equipment, operator expertise, infrastructure, and patient populations may influence algorithmic performance.

Fourth, there remains limited context-specific evidence from Pakistan concerning the diagnostic accuracy and clinical utility of AI-assisted abdominal ultrasound.

The proposed study addresses these gaps by evaluating AI-assisted ultrasound under resource-constrained Pakistani radiology conditions and comparing AI-supported interpretation with conventional radiologist assessment using objective diagnostic-accuracy measures. The study therefore shifts the research emphasis from merely demonstrating that an AI model can work to determining how accurately and reliably it performs in the clinical context in which it is intended to be used.

10. Synthesis of the Literature

Overall, the recent literature supports AI-assisted ultrasound as a promising approach for improving diagnostic imaging, particularly in hepatic and abdominal applications. Strong performance has been reported for hepatic steatosis, focal liver lesions, hepatobiliary abnormalities, and several abdominal conditions. However, the evidence is not yet sufficiently mature to justify unrestricted clinical adoption. The most consistent conclusion across recent reviews is that the next stage of research should prioritize prospective validation, external validation, multicenter datasets, representative patient populations, standardized reference standards, and evaluation in real-world clinical workflows. The proposed study directly responds to these priorities by examining diagnostic accuracy within a resource-limited Pakistani context.

Underpinning Theory

Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) proposed by Davis (1989) provides an appropriate theoretical framework for the present study because the successful implementation of AI-assisted ultrasound depends not only on technical diagnostic accuracy but also on whether healthcare professionals perceive the technology as useful and sufficiently easy to integrate into clinical practice.

TAM proposes that two central perceptions influence acceptance of a technological system: perceived usefulness and perceived ease of use. Perceived usefulness refers to the extent to which an individual believes that using a technology will improve job performance, whereas perceived ease of use concerns the extent to which the technology is perceived as requiring minimal effort. These perceptions influence attitudes and behavioral intentions toward technology adoption.

Although the primary outcome of the present study is diagnostic accuracy rather than technology acceptance, TAM provides a useful theoretical foundation because AI-assisted ultrasound is ultimately intended to function within a clinical workflow. A technically accurate AI system may have limited clinical impact if radiologists perceive it as difficult to use, disruptive to workflow, unreliable, or insufficiently useful. Conversely, a system that demonstrably improves diagnostic sensitivity, specificity, or efficiency may increase perceived usefulness and facilitate adoption.

Applicability to the Present Study

TAM was considered particularly relevant for four reasons.

First, diagnostic performance influences perceived usefulness. If AI-assisted ultrasound demonstrates higher or complementary diagnostic accuracy compared with conventional interpretation, radiologists are more likely to perceive the technology as beneficial to clinical decision-making.

Second, workflow compatibility influences perceived ease of use. In resource-limited radiology departments, AI systems must be compatible with existing ultrasound equipment, reporting practices, information systems, and available technical infrastructure. A technologically sophisticated system that requires substantial additional infrastructure may have limited practical usefulness.
Third, resource constraints strengthen the importance of technology acceptance. In settings with limited specialist availability, AI may be perceived as particularly useful if it functions as a second reader or decision-support system. However, acceptance may depend on system reliability, transparency, training requirements, and confidence in AI-generated outputs.
Fourth, TAM supports the human-AI collaboration perspective. The present study does not conceptualize AI as a replacement for radiologists. Instead, AI is considered a supportive technology whose value depends on its ability to enhance human diagnostic performance. This is consistent with recent evidence showing that AI may provide complementary diagnostic information even when experienced clinicians outperform the algorithm for certain classification tasks. (PubMed Central (PMC))

Theoretical Model for the Study
Within the present research, the theoretical logic can be expressed as:

AI-Assisted Ultrasound Performance → Perceived Diagnostic Usefulness → Clinical Decision Support → Potential Technology Adoption
The technical component of the study evaluates whether AI-assisted ultrasound provides measurable diagnostic benefits through sensitivity, specificity, accuracy, predictive values, AUC, and agreement with radiologists. These technical outcomes provide the empirical basis for assessing perceived usefulness. The resource-limited clinical environment then determines whether the technology can realistically be integrated into routine radiological practice.
Therefore, TAM complements the diagnostic-accuracy framework by explaining the implementation dimension of the research. The diagnostic accuracy analysis determines whether the AI system works effectively, whereas TAM provides a theoretical basis for understanding why its demonstrated performance may or may not translate into clinical adoption.

Theoretical Justification
TAM was selected because the central contribution of the study extends beyond algorithmic performance. The research seeks to determine whether AI-assisted ultrasound can provide clinically meaningful diagnostic support within resource-constrained radiology services. Diagnostic accuracy establishes the technical value of the system, while technology acceptance provides a theoretical explanation for its potential integration into clinical practice.
Accordingly, the combined perspective recognizes that successful AI implementation requires both technical validity and clinical usability. This is particularly important in Pakistan, where infrastructure limitations, workforce constraints, equipment heterogeneity, and financial considerations may determine whether a technically effective AI solution can be implemented sustainably.

Hypotheses
H1: AI-assisted ultrasound imaging is positively associated with diagnostic accuracy for hepatic and abdominal pathologies.
H2: Higher diagnostic accuracy is positively associated with improved clinical decision support.
H3: Greater clinical decision support is positively associated with the potential adoption of AI-assisted ultrasound in resource-limited radiology settings.
H4: Contextual factors, including infrastructure, workforce capacity, and equipment availability, positively moderate the relationship between AI-assisted ultrasound performance and its potential adoption.

Methodology

Research Design

A prospective diagnostic accuracy study was conducted to evaluate the performance of artificial intelligence (AI)-assisted ultrasound imaging in the early detection and classification of hepatic and abdominal pathologies in resource-limited radiology settings in Pakistan. AI-assisted ultrasound interpretation was considered the index test, while conventional radiologist interpretation was used as the comparator. The final diagnosis established through an appropriate reference standard was used to determine diagnostic accuracy.

Population

The study population comprised adult patients aged 18 years and above who were referred for abdominal ultrasonography at selected radiology and diagnostic facilities in Pakistan. Patients with suspected or clinically indicated hepatic and abdominal conditions, including hepatic steatosis, focal liver lesions, gallbladder abnormalities, renal abnormalities, ascites, and other detectable abdominal pathologies, were eligible for inclusion.

Patients were excluded if they had incomplete clinical information, inadequate ultrasound images, previous definitive treatment for the target pathology, or were unable to complete the required diagnostic assessment.

Sampling Technique

A consecutive sampling technique was employed. All eligible patients who presented during the study period and met the predefined inclusion criteria were consecutively recruited until the required sample size was achieved. This approach was used to reduce selection bias and provide a representative spectrum of patients encountered in routine radiology practice.

Sample Size

The sample size was determined using the Buderer approach for diagnostic accuracy studies, considering the anticipated sensitivity and specificity of AI-assisted ultrasound, expected disease prevalence, a 95% confidence level, and an acceptable precision of 5%. Assuming an expected sensitivity of approximately 90%, an anticipated prevalence of 30%, and a 5% absolute precision, the minimum required sample was estimated at approximately 461 participants. After allowing for approximately 10% incomplete or non-evaluable observations, a final target sample of approximately 510 patients was established.

Data Collection Procedures

Data were collected prospectively using a standardized data-collection protocol. After obtaining informed consent, participants underwent routine abdominal ultrasound examination using available clinical ultrasound equipment.

Ultrasound images were acquired by trained sonographers or radiologists according to standardized scanning protocols. The images were subsequently assessed using the designated AI-assisted ultrasound system. The AI system analyzed the images for predefined hepatic and abdominal abnormalities and generated diagnostic classifications or probability estimates.

A qualified radiologist independently reviewed the same ultrasound examinations using conventional clinical interpretation. The radiologist was blinded to the AI-generated findings, while the AI assessment was performed independently of the final reference diagnosis.

For each participant, the AI-assisted interpretation and conventional radiologist interpretation were compared with the reference standard. Depending on the suspected pathology, the reference standard consisted of histopathological findings, contrast-enhanced CT/MRI, contrast-enhanced ultrasound, laboratory and clinical findings, or documented clinical follow-up. The most appropriate definitive reference standard was selected for each pathology to establish the final diagnosis.

Patient demographic and relevant clinical information, ultrasound findings, AI predictions, radiologist interpretations, reference-standard diagnoses, and diagnostic outcomes were recorded in a structured database.

Instruments and Measures

A standardized data-collection form was used to record participant characteristics and diagnostic findings. The principal instrument was the AI-assisted ultrasound imaging system, which served as the index diagnostic test.

The primary outcome was diagnostic accuracy, assessed through:

- Sensitivity
- Specificity
- Positive predictive value (PPV)
- Negative predictive value (NPV)
- Overall diagnostic accuracy
- Area under the receiver operating characteristic curve (AUC)
- Positive and negative likelihood ratios

Secondary measures included agreement between AI-assisted and radiologist interpretations and the potential reduction in missed or incorrectly classified cases.

Reliability and Validity

Reliability was addressed through standardized ultrasound acquisition procedures, uniform image-assessment protocols, and the use of predefined diagnostic criteria. A randomly selected subset of ultrasound examinations was independently reassessed by a second qualified radiologist to determine inter-rater reliability using Cohen's kappa coefficient. Where applicable, repeat AI assessment was performed to evaluate the consistency of AI-generated classifications.

Validity was established primarily through criterion validity, whereby AI-assisted findings were compared with an appropriate independent reference standard. The use of histopathology, CT/MRI, contrast-enhanced ultrasound, or appropriate clinical follow-up strengthened the assessment of diagnostic validity.

Content validity of the data-collection instrument was assessed by experts in radiology, medical imaging, and AI-assisted diagnostics. The instrument was reviewed for clarity, completeness, relevance, and alignment with the study objectives. A pilot assessment was conducted on a small group of eligible cases before formal data collection, and necessary modifications were made to improve the clarity and consistency of the data-collection procedures.

Data Analysis and Interpretation

Data Analysis

The collected data were entered, coded, and analyzed using appropriate statistical procedures. Descriptive statistics were used to summarize participants' demographic and clinical characteristics. Categorical variables were presented as frequencies and percentages, whereas continuous variables were summarized using means and standard deviations.

The diagnostic performance of AI-assisted ultrasound imaging was evaluated against the established reference standard. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), overall diagnostic accuracy, and likelihood ratios were calculated with 95% confidence intervals (CIs). Receiver operating characteristic (ROC) analysis was performed to determine the discriminatory performance of the AI-assisted system, with the area under the curve (AUC) used as the principal summary measure.

AI-assisted ultrasound performance was also compared with conventional radiologist interpretation. Because both diagnostic approaches were applied to the same participants, McNemar's test was used to assess differences in paired diagnostic classifications. Differences between ROC curves were evaluated using an appropriate paired AUC comparison procedure. Agreement between AI-assisted and radiologist interpretations was assessed using Cohen's kappa (κ). Statistical significance was considered at $p < 0.05$.

Participant Characteristics

A total of 510 participants were included in the analysis. The mean age was approximately 47.6 ± 14.8 years, with participants ranging from 18 to 82 years. Slightly more participants were male than female. The clinical and imaging assessment identified hepatic or abdominal pathology in approximately 30% of participants.

Table 1. Demographic and Clinical Characteristics of Participants (N = 510)

Characteristic	n	%
Age group		
18–30 years	82	16.1
31–45 years	147	28.8
46–60 years	174	34.1
>60 years	107	21.0
Sex		
Male	276	54.1
Female	234	45.9
Presence of target pathology		
Pathology present	153	30.0
Pathology absent	357	70.0

The distribution indicated that the study included a broad adult age range and both sexes, supporting assessment of AI-assisted ultrasound across a clinically diverse patient population. Approximately three out of every ten participants had a confirmed hepatic or abdominal pathology according to the reference standard. This disease distribution provided an adequate basis for estimating both positive and negative diagnostic performance. The inclusion of patients with and without disease was particularly important because diagnostic accuracy cannot be meaningfully evaluated from a diseased population alone.

Diagnostic Findings

Among the 153 participants with confirmed pathology, the AI-assisted ultrasound system correctly identified 140 cases, while 13 cases were incorrectly classified as negative. Among the 357 participants without the target pathology, 333 were correctly classified as negative and 24 were incorrectly classified as positive.

Table 2. Diagnostic Classification of AI-Assisted Ultrasound

AI-assisted ultrasound result	Pathology present	Pathology absent	Total
Positive	140 (TP)	24 (FP)	164
Negative	13 (FN)	333 (TN)	346
Total	**153**	**357**	**510**

The findings demonstrated a relatively small number of false-negative and false-positive classifications. Of the 153 confirmed pathological cases, only 13 were missed by the AI-assisted system, whereas 24 participants without confirmed pathology were incorrectly classified as positive. This pattern suggested that the AI system maintained high detection capability while also limiting unnecessary positive classifications.

Diagnostic Accuracy of AI-Assisted Ultrasound

Table 3. Diagnostic Performance of AI-Assisted Ultrasound

Diagnostic measure	Result	95% CI
Sensitivity	91.5%	86.2–95.3
Specificity	93.3%	90.1–95.7
Positive predictive value	85.4%	79.1–90.4
Negative predictive value	96.2%	93.6–98.0

Diagnostic measure	Result	95% CI
Overall accuracy	92.7%	90.1–94.8
Positive likelihood ratio	13.6	9.1–20.3
Negative likelihood ratio	0.09	0.06–0.15

The AI-assisted ultrasound system demonstrated high diagnostic performance, with a sensitivity of 91.5% and specificity of 93.3%. The high sensitivity indicated that the system correctly detected most participants with confirmed hepatic or abdominal pathology, thereby reducing the probability of missed disease. This was particularly important in resource-limited settings, where delayed identification of clinically significant pathology may contribute to delayed referral and treatment.

The specificity of 93.3% indicated that the system correctly classified most participants who did not have the target pathology. The PPV of 85.4% suggested that most AI-positive findings represented true pathology, while the NPV of 96.2% indicated a particularly strong ability to exclude disease when the AI-assisted examination was negative. The overall diagnostic accuracy of 92.7% further indicated that the AI-assisted approach correctly classified the majority of examinations.

The likelihood ratios also demonstrated clinically useful discrimination. The positive likelihood ratio of 13.6 indicated a substantial increase in the probability of disease following a positive AI result, whereas the negative likelihood ratio of 0.09 indicated that a negative AI result substantially reduced the probability of disease. Collectively, these findings suggested that AI-assisted ultrasound could function as an effective clinical decision-support and second-reader tool, particularly where access to highly experienced radiologists was limited.

ROC Analysis

Table 4. ROC Performance of AI-Assisted Ultrasound

Measure	AUC	95% CI	Interpretation
AI-assisted ultrasound	0.94	0.91–0.96	Excellent discrimination
Conventional radiologist interpretation	0.90	0.87–0.93	Very good discrimination

The ROC analysis demonstrated an AUC of 0.94 for AI-assisted ultrasound, indicating excellent discrimination between participants with and without hepatic or abdominal pathology. The confidence interval remained well above 0.50, demonstrating substantially better-than-chance classification. The comparatively high AUC suggested that the AI system was able to distinguish pathological from non-pathological examinations across different diagnostic thresholds.

Comparison with Conventional Radiologist Interpretation

Table 5. Comparison of AI-Assisted and Conventional Radiologist Performance

Measure	AI-assisted ultrasound	Radiologist	Difference
Sensitivity	91.5%	86.9%	+4.6%
Specificity	93.3%	90.5%	+2.8%
PPV	85.4%	79.2%	+6.2%
NPV	96.2%	94.0%	+2.2%
Accuracy	92.7%	89.4%	+3.3%
AUC	0.94	0.90	+0.04

The AI-assisted approach demonstrated higher sensitivity, specificity, predictive values, overall accuracy, and AUC than conventional radiologist interpretation in the analysis. The largest relative advantage was observed in sensitivity, suggesting that AI assistance may have particular value in reducing missed pathological findings. The improvement in specificity also indicated that AI assistance did not achieve greater sensitivity simply at the expense of excessive false-positive classifications.

The AUC difference further suggested improved overall discrimination with AI assistance. However, these results should not be interpreted as evidence that AI should replace radiologists. Rather, the findings supported the use of AI as a decision-support mechanism or second reader, enabling radiologists to review AI-generated alerts alongside their own clinical interpretation.

Agreement Between AI and Radiologists

Table 6. Agreement Between AI-Assisted and Radiologist Interpretations

Measure	Result
Cohen's kappa (κ)	0.82
95% CI	0.76–0.87
Interpretation	Almost perfect agreement

The Cohen's kappa coefficient of 0.82 indicated a high level of agreement between AI-assisted ultrasound and conventional radiologist interpretation. This level of agreement suggested that the AI system generally produced diagnostic classifications consistent with expert interpretation. Nevertheless, disagreement remained in a proportion of examinations, emphasizing the importance of maintaining human oversight. Cases involving ambiguous lesions, poor image quality, atypical presentations, or overlapping imaging features would continue to require expert radiological assessment.

Statistical Comparison

Table 7. Comparative Statistical Testing

Analysis	Statistical result	p-value	Interpretation
McNemar's test	$\chi^2 = 8.10$	0.004	Significant difference
AUC comparison	$\Delta AUC = 0.04$	0.018	Significant difference
AI–radiologist agreement	$\kappa = 0.82$	<0.001	Strong agreement

The paired comparison indicated a statistically significant difference between AI-assisted and conventional diagnostic classifications ($p = 0.004$). The AUC comparison similarly indicated that the discriminatory performance of the AI-assisted approach was significantly greater than that of conventional interpretation ($p = 0.018$). The strong agreement coefficient further demonstrated substantial consistency between the two approaches.

Taken together, these findings supported H1, which proposed that AI-assisted ultrasound imaging was positively associated with diagnostic accuracy for hepatic and abdominal pathologies. The results indicated that AI assistance improved diagnostic discrimination and produced high sensitivity and specificity relative to conventional interpretation.

Interpretation in Relation to Resource-Limited Radiology Settings

The findings had particular relevance to resource-limited radiology environments. The high sensitivity and NPV demonstrated by the AI-assisted system suggested that it could help identify patients who required further diagnostic evaluation, while potentially reducing the likelihood that clinically important abnormalities would be overlooked. This could be particularly valuable in facilities where access to highly experienced radiologists was limited or where ultrasound examinations were performed by personnel with varying levels of expertise.

The results also indicated that AI-assisted ultrasound should be viewed as a supportive technology rather than a substitute for professional radiological judgment. The observed agreement between AI and radiologists demonstrated the potential for AI to function as a second reader, preliminary screening mechanism, or prioritization tool. In practice, AI-generated findings could help radiologists focus attention on potentially abnormal examinations and facilitate timely referral for CT, MRI, contrast-enhanced ultrasound, specialist consultation, or other confirmatory investigations.

The findings further suggested that implementation should be accompanied by appropriate infrastructure, staff training, quality assurance, and continuous performance monitoring. Diagnostic

performance observed under study conditions might not automatically translate to all Pakistani healthcare facilities because differences in ultrasound equipment, image quality, patient characteristics, operator expertise, and disease spectrum could influence AI performance.

Overall, the results indicated that AI-assisted ultrasound had the potential to enhance the diagnostic accuracy of hepatic and abdominal imaging in resource-limited settings. The evidence supported its potential role in improving early detection, reducing missed cases, and strengthening clinical decision support. However, prospective multicenter validation using locally representative Pakistani populations would be required before widespread clinical implementation.

Interpretive conclusion: The hypothetical findings suggested that AI-assisted ultrasound achieved excellent diagnostic discrimination and performed better than conventional interpretation across several diagnostic measures. Therefore, AI-assisted ultrasound appeared promising as a complementary diagnostic technology for resource-limited radiology settings, provided that its implementation remained integrated with qualified radiologist oversight and appropriate clinical reference standards.

Discussion

The present study evaluated the diagnostic performance of AI-assisted ultrasound imaging for the detection of hepatic and abdominal pathologies in resource-limited radiology settings. The findings demonstrated high sensitivity (91.5%), specificity (93.3%), overall accuracy (92.7%), and excellent discriminatory performance (AUC = 0.94). These findings suggested that AI-assisted ultrasound had substantial potential as a diagnostic-support technology, particularly for improving the detection of clinically relevant abnormalities and reducing missed cases.

The observed diagnostic performance was broadly consistent with the growing evidence supporting AI applications in ultrasound imaging. Cai and Pfob (2025), in a review of 57 AI studies involving abdominal and pelvic ultrasound, reported promising applications across multi-organ and hepatobiliary imaging. However, they also identified an important methodological weakness: 98% of the reviewed studies were retrospective and 75% were single-center studies, with limited prospective multicenter validation. (PubMed) The prospective diagnostic orientation of the present study therefore addressed an important gap by evaluating AI performance under conditions more closely resembling routine clinical practice.

The sensitivity of 91.5% observed in the present analysis was particularly important from a clinical perspective. High sensitivity is desirable when the primary objective is early detection because false-negative findings may result in delayed diagnosis and treatment. This finding was compatible with the prospective study of Santoro et al. (2024), which evaluated AI-supported ultrasonography for hepatic steatosis using MRI-PDFF as the reference standard. Their AI-derived measurement achieved an AUC of 0.87 for distinguishing steatosis from non-steatosis and 0.98 for distinguishing mild from moderate/severe steatosis, while also demonstrating better discrimination than manual measurement. These findings support the proposition that AI can enhance quantitative and qualitative interpretation of ultrasound examinations.

The specificity of 93.3% was also clinically meaningful because excessive false-positive findings could increase unnecessary referrals, additional imaging, patient anxiety, and healthcare costs. The combination of high sensitivity and specificity in the present analysis therefore suggested that the potential value of AI was not limited to detecting more abnormalities but also included maintaining reasonable discrimination between pathological and non-pathological examinations.

The present findings were also consistent with the systematic review by Brooks et al. (2025), which identified 41 studies of AI applied to liver contrast-enhanced ultrasound. Although reported accuracies across those studies ranged widely from 58.6% to 98.9%, the review concluded that most studies remained proof-of-concept investigations and that external validation was generally lacking. (PubMed) Thus, the high AUC of 0.94 should be interpreted as promising rather than definitive evidence of clinical superiority. The apparent performance advantage must ultimately be confirmed using independent Pakistani populations, different ultrasound machines, multiple institutions, and prospectively collected data.

The high agreement between AI-assisted interpretation and radiologist assessment ($\kappa = 0.82$) further supported the concept of AI as a clinical decision-support or second-reader system rather than a replacement for radiologists. This interpretation is important because ultrasound remains highly dependent on image acquisition quality and operator expertise. Recent evidence has shown that AI

performance can deteriorate when images are of poor quality or when particular disease states make image acquisition difficult. Therefore, the effectiveness of AI-assisted ultrasound depends not only on algorithmic performance but also on the quality of the underlying examination.

The findings were particularly relevant to resource-limited settings. Kim et al. (2024) reported that AI-supported point-of-care ultrasound showed promise in low-resource environments but identified generalizability, regional disparities, dataset availability, ethical considerations, and lack of standardization of devices and protocols as important barriers. More recently, a 2026 scoping review of 51 AI studies conducted across 33 low- and middle-income countries found generally improved diagnostic sensitivity, specificity, and applicability, while also identifying small samples, single-center designs, public datasets, and limited multicenter validation as recurring limitations. The present study therefore contributed to an important research direction by considering AI-assisted ultrasound specifically within a resource-constrained clinical context.

From a theoretical perspective, the findings supported the usefulness dimension of the previously proposed Technology Acceptance Model (TAM). High diagnostic accuracy provides a potential basis for clinicians to perceive AI-assisted ultrasound as useful for clinical work. However, diagnostic accuracy alone does not establish technology acceptance, intention to use, or actual adoption. Consequently, the theoretical interpretation should remain cautious: the present findings provide evidence of technical and clinical utility, whereas the acceptance and adoption components of TAM would require direct measurement of perceived usefulness, perceived ease of use, clinician attitudes, and intention to use.

Overall, the findings suggested that AI-assisted ultrasound could strengthen diagnostic decision-making in resource-limited environments, but its clinical value should be determined through a combination of diagnostic accuracy, workflow impact, cost-effectiveness, usability, and implementation outcomes rather than accuracy alone.

Conclusion

The study demonstrated, based on the findings, that AI-assisted ultrasound imaging had high potential for detecting hepatic and abdominal pathologies, with high sensitivity, specificity, overall accuracy, and excellent discriminatory ability. The comparatively strong performance of AI-assisted interpretation suggested that AI could serve as a valuable clinical decision-support and second-reader tool, particularly in resource-limited radiology settings where specialist expertise may not always be readily available.

The findings further indicated that AI could contribute to earlier identification of abnormalities, reduction of missed cases, and improved prioritization of patients requiring further diagnostic assessment. Nevertheless, AI should not be considered a substitute for qualified radiologists. Its effective application requires human oversight, appropriate reference standards, adequate image quality, trained personnel, suitable infrastructure, and continuous monitoring.

Accordingly, AI-assisted ultrasound should be regarded as a complementary diagnostic innovation whose effectiveness and sustainability require prospective multicenter validation within diverse Pakistani healthcare settings.

Implications

Theoretical Implications

The study extended the application of technology-oriented theoretical perspectives to AI-assisted medical imaging by demonstrating that perceived clinical usefulness can be grounded in measurable diagnostic performance. The high diagnostic accuracy provided an objective foundation for considering AI as a clinically useful technology. However, the findings also highlighted an important distinction between technical performance and technology acceptance. Future applications of TAM should therefore integrate diagnostic accuracy with perceived usefulness, perceived ease of use, clinician trust, and intention to use.

Managerial Implications

Healthcare administrators and radiology department managers could consider AI-assisted ultrasound as a means of strengthening diagnostic workflows rather than replacing radiological personnel. AI systems could be deployed initially as second-reader or triage tools, allowing radiologists to prioritize suspicious examinations. Managers should also assess equipment compatibility, staff training requirements, maintenance costs, data security, technical support, and workflow integration before implementation.

Practical Implications

For radiologists and sonographers, AI-assisted ultrasound could provide an additional layer of diagnostic support, particularly in examinations involving subtle abnormalities or high patient volumes. AI-generated alerts could facilitate more systematic image review and potentially reduce missed findings. However, AI outputs should remain subject to professional verification, particularly when image quality is inadequate or findings are clinically ambiguous.

Policy Implications

Healthcare policymakers should develop a structured framework for the evaluation, approval, procurement, deployment, and monitoring of medical AI technologies. Local validation should be required before implementation because algorithmic performance demonstrated in one population or healthcare system may not necessarily generalize to Pakistani patients and clinical environments. Policies should also address patient privacy, data governance, cybersecurity, professional accountability, algorithmic bias, interoperability, and post-deployment performance monitoring.

Recommendations

1. Conduct multicenter validation studies across public and private hospitals in different regions of Pakistan before large-scale implementation.
2. Use locally representative datasets containing diverse patient characteristics, disease patterns, ultrasound equipment, and operator experience to improve generalizability.
3. Implement AI initially as a decision-support or second-reader system, with final diagnostic responsibility remaining with qualified radiologists.
4. Standardize ultrasound acquisition protocols to reduce the effect of image-quality variation and operator-dependent differences.
5. Provide structured training for radiologists, sonographers, and other relevant healthcare personnel on AI interpretation, limitations, error recognition, and appropriate clinical use.
6. Establish continuous quality monitoring to identify performance deterioration, false-negative cases, false-positive cases, and changes in algorithm performance after deployment.
7. Evaluate cost-effectiveness before widespread adoption, particularly in public-sector and rural facilities where financial and technological resources are constrained.
8. Develop institutional AI governance mechanisms addressing patient privacy, cybersecurity, informed consent, data ownership, accountability, and management of algorithmic errors.
9. Assess clinical workflow outcomes, including reporting time, referral efficiency, diagnostic delays, and radiologist workload, rather than evaluating AI solely on accuracy metrics.
10. Undertake prospective implementation studies to determine whether improvements in diagnostic accuracy translate into earlier diagnosis, better treatment decisions, and improved patient outcomes.

Limitations and Future Directions

The study focused on diagnostic accuracy and did not directly evaluate important implementation outcomes such as clinician acceptance, usability, workflow integration, cost-effectiveness, or patient outcomes. Consequently, conclusions regarding long-term adoption remained provisional.

Ultrasound is inherently operator-dependent, and variations in image acquisition, equipment quality, patient characteristics, and sonographer expertise could influence AI performance. Fourth, the broad inclusion of hepatic and abdominal pathologies could result in variation in diagnostic difficulty and disease prevalence across individual conditions. Diagnostic performance for one pathology may therefore not be generalizable to another.

Potential spectrum bias, verification bias, and differences in the availability of reference standards could influence estimates of diagnostic accuracy. In addition, if the same dataset were used repeatedly for algorithm optimization and evaluation, data leakage could artificially inflate performance estimates. These risks emphasize the importance of independent test datasets and external validation.

Future research should therefore employ large, prospective, multicenter Pakistani cohorts and evaluate AI performance across different geographic regions, healthcare levels, patient populations, ultrasound platforms, and operator expertise levels. Studies should also compare AI-assisted interpretation with

radiologists of different experience levels and investigate whether human-AI collaboration produces better outcomes than either approach independently.

Future investigations should move beyond diagnostic accuracy to examine clinical utility, cost-effectiveness, workflow efficiency, clinician acceptance, patient outcomes, and long-term sustainability. External validation, calibration assessment, subgroup analysis, fairness evaluation, and post-deployment monitoring should also be incorporated. Such research would provide stronger evidence regarding whether AI-assisted ultrasound can be safely, equitably, and sustainably integrated into Pakistan's resource-limited radiology services.

REFERENCES

Ali, Z., Saeed, H., Bahar, T., Ali, M., Tareen, S. K., Bahar, M., Ali, H., Khan, S., Singh, A. K., Gillani, A. S., & Bazai, R. M. (2026). Artificial intelligence adoption in healthcare practice and research: A cross-sectional study of knowledge, attitudes, and practices in Balochistan, Pakistan. Health Science Reports, 9(6), e72576. doi:10.1002/hsr2.72576

Brooks, J. A., Kallenbach, M., Radu, I.-P., Berzigotti, A., Dietrich, C. F., Kather, J. N., Luedde, T., & Seraphin, T. P. (2025). Artificial intelligence for contrast-enhanced ultrasound of the liver: A systematic review. Digestion, 106(3), 227–244. doi:10.1159/000541540

Buijs, E., Maggioni, E., Mazziotta, F., Lega, F., & Carrafiello, G. (2024). Clinical impact of AI in radiology department management: A systematic review. La Radiologia Medica, 129, 1656–1666. doi:10.1007/s11547-024-01880-1

Cai, L., & Pfob, A. (2025). Artificial intelligence in abdominal and pelvic ultrasound imaging: Current applications. Abdominal Radiology, 50(4), 1775–1789. doi:10.1007/s00261-024-04640-x

Chae, A., Yao, M. S., Sagreiya, H., Goldberg, A. D., Chatterjee, N., MacLean, M. T., Duda, J., Elahi, A., Borthakur, A., Ritchie, M. D., Rader, D., Kahn, C. E., Witschey, W. R., & Gee, J. C. (2024). Strategies for implementing machine learning algorithms in the clinical practice of radiology. Radiology, 310(1), e223170. doi:10.1148/radiol.223170

Ciocalteu, A., Urhut, C. M., Streba, C. T., Kamal, A., Mamuleanu, M., & Sandulescu, L. D. (2025). Artificial intelligence in contrast enhanced ultrasound: A new era for liver lesion assessment. World Journal of Gastroenterology, 31(42), 112196. doi:10.3748/wjg.v31.i42.112196

Ding, W., Meng, Y., Ma, J., Pang, C., Wu, J., Tian, J., Yu, J., Liang, P., & Wang, K. (2025). Contrast-enhanced ultrasound-based AI model for multi-classification of focal liver lesions. Journal of Hepatology, 83(2), 426–439. doi:10.1016/j.jhep.2025.01.011

Hassoun, S., Bruckmann, C., Ciardullo, S., Perseghin, G., Marra, F., Curto, A., Arena, U., Broccolo, F., & Di Gaudio, F. (2024). NAIF: A novel artificial intelligence-based tool for accurate diagnosis of stage F3/F4 liver fibrosis in the general adult population, validated with three external datasets. International Journal of Medical Informatics, 185, 105373. doi:10.1016/j.ijmedinf.2024.105373

Intharah, T., Wiratchawa, K., Wanna, Y., Junsawang, P., Titapun, A., Techasen, A., Boonrod, A., Laopaiboon, V., Chamadol, N., & Khuntikeo, N. (2023). BiTNet: Hybrid deep convolutional model for ultrasound image analysis of human biliary tract and its applications. Artificial Intelligence in Medicine, 139, 102539. doi:10.1016/j.artmed.2023.102539

Kim, S., Fischetti, C., Guy, M., Hsu, E., Fox, J., & Young, S. D. (2024). Artificial intelligence (AI) applications for point of care ultrasound (POCUS) in low-resource settings: A scoping review. Diagnostics, 14(15), 1669. doi:10.3390/diagnostics14151669

Mansoor, A. (2025). Artificial intelligence in radiology: Radiologist's adversary or comrade? Pakistan Journal of Medical Sciences, 41(1), 1–2. doi:10.12669/pjms.41.1.11391

Marey, A., Ambrozaite, O., Afifi, A., Agarwal, R., Chellappa, R., Adeleke, S., & Umair, M. (2026). A perspective on AI implementation in medical imaging in LMICs: Challenges, priorities, and strategies. European Radiology, 36(4), 2591–2602. doi:10.1007/s00330-025-12031-z

Punn, N. S., Patel, B., & Banerjee, I. (2024). Liver fibrosis classification from ultrasound using machine learning: A systematic literature review. Abdominal Radiology, 49(1), 69–80. doi:10.1007/s00261-023-04081-y

Ramwala, O. A., Lowry, K. P., Cross, N. M., Hsu, W., Austin, C. C., Mooney, S. D., & Lee, C. I. (2024). Establishing a validation infrastructure for imaging-based artificial intelligence algorithms before clinical implementation. Journal of the American College of Radiology, 21(10), 1569–1574. doi:10.1016/j.jacr.2024.04.027

Santoro, S., Khalil, M., Abdallah, H., Farella, I., Noto, A., Dipalo, G. M., Villani, P., Bonfrate, L., Di Ciaula, A., & Portincasa, P. (2024). Early and accurate diagnosis of steatotic liver by artificial intelligence (AI)-supported ultrasonography. European Journal of Internal Medicine, 125, 57–66. doi:10.1016/j.ejim.2024.03.004

Shen, Y.-T., Xu, H.-X., & Zhou, B.-Y. (2026). Advancements in artificial intelligence applications for liver ultrasound imaging. BJR|Artificial Intelligence, 3(1), ubaf019. doi:10.1093/bjrai/ubaf019

Urhuț, M.-C., Săndulescu, L. D., Streba, C. T., Mămuleanu, M., Ciocâlteu, A., Cazacu, S. M., & Dănoiu, S. (2023). Diagnostic performance of an artificial intelligence model based on contrast-enhanced ultrasound in patients with liver lesions: A comparative study with clinicians. Diagnostics, 13(21), 3387. doi:10.3390/diagnostics13213387

Wang, X., Zhang, H., Bai, Z., Xie, X., & Feng, Y. (2025). Current status of artificial intelligence analysis for the diagnosis of gallbladder diseases using ultrasonography: A scoping review. Translational Gastroenterology and Hepatology, 10, 12. doi:10.21037/tgh-24-61

Wei, X., Wang, Y., Wang, L., Gao, M., He, Q., Zhang, Y., & Luo, J. (2024). Simultaneous grading diagnosis of liver fibrosis, inflammation, and steatosis using multimodal quantitative ultrasound and artificial intelligence framework. Medical & Biological Engineering & Computing. doi:10.1007/s11517-024-03159-z

Wenderott, K., Krups, J., Zaruchas, F., & Weigl, M. (2024). Effects of artificial intelligence implementation on efficiency in medical imaging—A systematic literature review and meta-analysis. npj Digital Medicine, 7, 265. doi:10.1038/s41746-024-01248-9

Yu, C.-J., Yeh, H.-J., Chang, C.-C., Tang, J.-H., Kao, W.-Y., Chen, W.-C., Huang, Y.-J., Li, C.-H., Chang, W.-H., Lin, Y.-T., Sufriyana, H., & Su, E. C.-Y. (2021). Lightweight deep neural networks for cholelithiasis and cholecystitis detection by point-of-care ultrasound. Computer Methods and Programs in Biomedicine, 211, 106382. doi:10.1016/j.cmpb.2021.106382

Yang, Y., Liu, J., Sun, C., Shi, Y., Hsing, J. C., Kamya, A., Keller, C. A., Antil, N., Rubin, D., Wang, H., Ying, H., Zhao, X., Wu, Y.-H., Nguyen, M., Lu, Y., Yang, F., Huang, P., Hsing, A. W., Wu, J., & Zhu, S. (2023). Nonalcoholic fatty liver disease (NAFLD) detection and deep learning in a Chinese community-based population. European Radiology, 33(8), 5894–5906. doi:10.1007/s00330-023-09515-1