Research Article

The Role Of Artificial Intelligence And Deep Learning In Ultrasound Diagnosis

R. Sivakumar[1], Arati shahapurkar[2], Jeevanantham.G[3], Lokesh, A[4], C. Anbarasan[5], A. Harini[6]

[1] Professor, Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore, India, rsksivame@gmail.com,
[2]Assistant Professor, Department of Computer Science and Engineering, KLS Gogte Institute of Technology, Belagavi, India, aratibellary7@gmail.com
[3] Assistant Professor, Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore, India, jeevas1984@gmail.com
[4]Professor, Department of AIML, SEA College of Engineering and Technology, K R Puram, Bangalore,India, lokesh03061981@gmail.com
[5]Assistant Professor, Department of CSE (AIML), Madanapalle Institute of Technology & Science, Deemed to be University, Madanapalle, India, anbarasanc@mits.ac.in
[6] Assistant Professor, Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore, India, aharini01608@gmail.com

Abstract

Ultrasound (US) remains one of the most widely used medical imaging modalities because it is real-time, radiation-free, portable, and comparatively inexpensive, but its diagnostic accuracy is strongly operator-dependent and subject to inter-observer variability. Over the past decade, artificial intelligence (AI) - and deep learning (DL) in particular - has been applied extensively to ultrasound image analysis in an effort to reduce this variability and support faster, more consistent diagnosis. This paper presents a comprehensive narrative review of current AI/DL applications in ultrasound diagnosis, synthesising evidence across four major clinical domains: breast lesion classification, thyroid nodule risk stratification, fetal cardiac screening for congenital heart disease, and abdominal/musculoskeletal applications. A structured review methodology is described, followed by a proposed conceptual framework that organises reviewed applications by deep learning task - classification, detection/localisation, and segmentation - and by clinical domain, situating an explainability layer and human-in-the-loop clinical review at the centre of responsible deployment. Reported diagnostic performance metrics (accuracy, sensitivity, specificity) from a representative set of primary studies and meta-analyses are synthesised and compared, showing that convolutional neural network (CNN)-based models frequently reach or approach expert-level diagnostic performance in constrained, retrospective evaluation settings, with reported accuracies generally in the 88-99% range and reported sensitivities in the 87-98% range across the domains reviewed. All performance figures reported in this paper are drawn directly from the cited primary studies and meta-analyses, not from a new experiment conducted by the authors. The review concludes by identifying recurring limitations across the literature - dataset heterogeneity, limited external/multicentre validation, and interpretability gaps - and outlines directions for future research toward clinically deployable, trustworthy AI-assisted ultrasound diagnosis.

Keywords: Artificial Intelligence; Deep Learning; Ultrasound Diagnosis; Convolutional Neural Networks; Medical Image Analysis; Computer-Aided Diagnosis; Explainable AI.

1. Introduction

Ultrasound imaging occupies a distinctive position among medical imaging modalities: it is non-ionising, low-cost relative to computed tomography (CT) or magnetic resonance imaging (MRI), portable enough for point-of-care use, and capable of real-time, dynamic visualisation of soft tissue and blood flow. These properties have made it the first-line imaging modality for a wide range of clinical indications, including breast mass evaluation, thyroid nodule assessment, obstetric and fetal cardiac screening, and abdominal organ assessment.

The central limitation of ultrasound, however, is its strong dependence on operator skill. Image acquisition and interpretation both require substantial training, and diagnostic accuracy varies measurably between sonographers of different experience levels, contributing to missed findings, false positives, and unnecessary biopsies or follow-up procedures. This operator-dependency, combined with the sheer volume of ultrasound examinations performed worldwide, has made ultrasound one of the most active areas for the application of artificial intelligence in medical imaging.

Deep learning, and convolutional neural networks (CNNs) in particular, has driven most of the recent progress in this area. Unlike earlier machine learning approaches that relied on hand-crafted features (texture descriptors, shape indices, Nakagami statistics), CNNs learn hierarchical image features directly from data, which has allowed them to match or approach expert-level performance on narrowly defined diagnostic tasks such as distinguishing benign from malignant breast masses or classifying thyroid nodules by malignancy risk. More recently, transformer-based architectures and hybrid CNN-transformer models have begun to appear in the literature, aiming to capture longer-range spatial dependencies that pure convolutional architectures may miss.

This paper provides a comprehensive review of current AI/DL applications in ultrasound diagnosis. Rather than restricting the scope to a single organ system, the review deliberately spans four representative domains - breast, thyroid, fetal cardiac, and abdominal/musculoskeletal ultrasound - in order to identify patterns that recur across the literature regardless of clinical application. The paper makes three contributions: (i) a structured, PRISMA-informed review methodology and a detailed

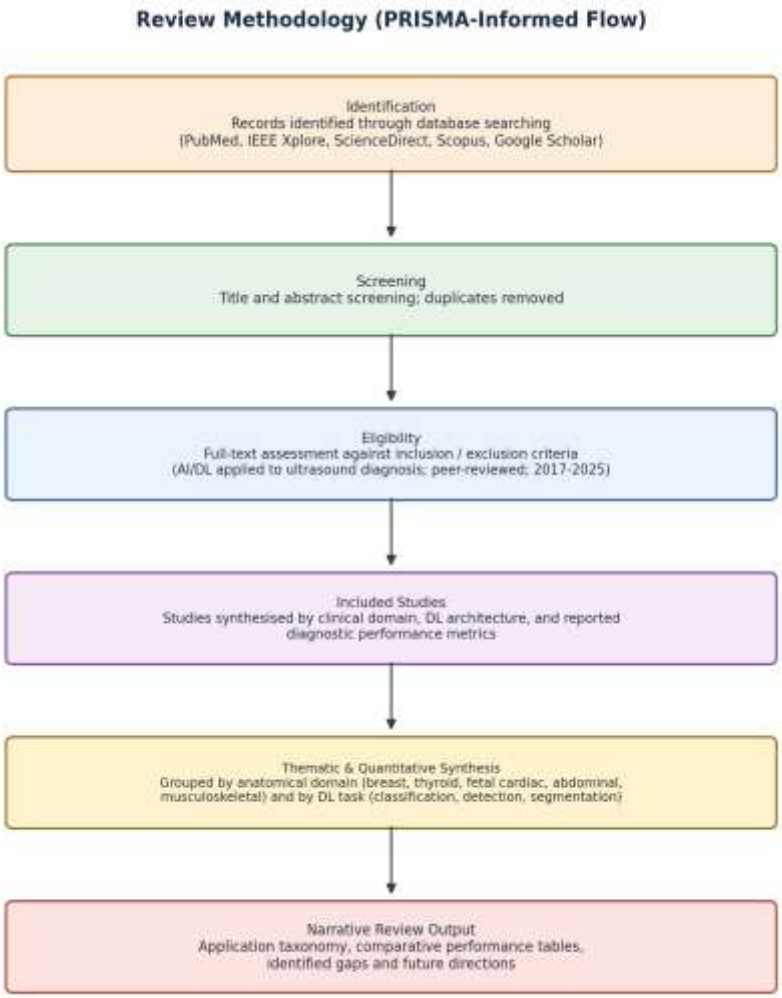
literature survey organised by clinical domain; (ii) a proposed conceptual framework that organises AI/DL applications in ultrasound by deep learning task (classification, detection, segmentation) and situates explainability and human oversight as core architectural components rather than afterthoughts; and (iii) a synthesis of reported diagnostic performance metrics drawn directly from the primary literature, allowing like-for-like comparison across domains. The remainder of the paper is organised as follows: Section 2 presents the review methodology and detailed literature survey; Section 3 presents the proposed conceptual framework; Section 4 presents a result analysis synthesising reported performance from representative studies; Section 5 presents a performance analysis comparing domains and architectures; and Section 6 concludes with limitations and future directions.

2. Literature Survey

2.1 Review Methodology

This review follows a structured, PRISMA-informed process (Figure 1). Records were identified through searches of PubMed, IEEE Xplore, ScienceDirect, Scopus, and Google Scholar using combinations of the terms "deep learning", "artificial intelligence", "ultrasound", "diagnosis", "classification", "segmentation", and domain-specific terms (e.g., "breast", "thyroid nodule", "fetal echocardiography"). Studies were included if they applied a machine learning or deep learning model to ultrasound image or video data for a diagnostic or screening task and reported quantitative diagnostic performance metrics (accuracy, sensitivity, specificity, or area under the receiver operating characteristic curve, AUC); studies were excluded if they used ultrasound only for treatment guidance (e.g., needle guidance) without a diagnostic classification task, or if full quantitative results were not reported. Eligible studies were then synthesised thematically by clinical domain and by deep learning task.

Figure 1. Review methodology flow diagram (PRISMA-informed)



2.2 AI/DL in Breast Ultrasound

Breast ultrasound is one of the most extensively studied domains for AI/DL diagnostic support, reflecting both the global burden of breast cancer and ultrasound's role as a first-line adjunct to mammography, particularly in women with dense breast tissue. Afrin et al. reviewed 59 primary studies published between 2017 and early 2023 on deep learning applied to breast mass ultrasound, covering tasks from lesion segmentation and classification through to prediction of lymph node metastasis and response to therapy, and found that CNN-based models were the dominant architecture family, with reported performance varying substantially depending on dataset size, imaging mode (B-mode, colour Doppler, elastography), and validation strategy.

Individual primary studies illustrate the range of reported performance. Ma et al. proposed Fus2Net, a custom CNN architecture for classifying benign and malignant breast ultrasound (BUS) tumour images, reporting an accuracy of 92.0%, sensitivity of 95.65%, specificity of 88.89%, and AUC of 0.97 on a held-out test set of 100 BUS images. A more recent multimodal study combining grey-scale and colour Doppler BUS images with a VGG16 backbone reported a mean accuracy of 91.54%, sensitivity of 94.15%, specificity of 87.30%, F1-score of 0.93, and AUC of 0.96 for classifying non-mass breast lesions, and found that the multimodal (grey-scale plus colour Doppler) model outperformed either single-modality model alone using five-fold cross-validation. These findings recur across the domain: combining multiple ultrasound imaging modes, or combining multiple views of the same lesion, consistently improves reported performance relative to single-view, single-mode CNN classifiers.

2.3 AI/DL in Thyroid Nodule Assessment

Thyroid nodules are detected on ultrasound in a large proportion of the adult population, the great majority of which are benign, making risk stratification an important task for reducing unnecessary fine-needle aspiration biopsies. Guan et al. trained an Inception-v3 CNN on 2,836 thyroid ultrasound images (1,484 papillary thyroid carcinoma and 1,352 benign cases) and reported a sensitivity of 93.3% and specificity of 87.4% on the test set, with the model performing best on nodules between 0.5 and 1.0 cm in size.

Because individual thyroid-nodule studies vary considerably in reported sensitivity (from around 77% to 93% in the studies aggregated by Zhu et al.), a meta-analytic view is informative. Zhu et al. conducted a meta-analysis of 11 studies applying the VGGNet CNN architecture specifically to benign/malignant thyroid nodule classification and reported pooled estimates of sensitivity of 0.87 (95% CI: 0.83-0.91), specificity of 0.85 (95% CI: 0.79-0.90), a diagnostic odds ratio of 38.79, and a pooled AUC of 0.93, concluding that VGGNet-based deep learning shows good diagnostic efficacy for this task but that reported performance across individual studies is heterogeneous, likely reflecting differences in dataset composition, ground-truth definition, and validation protocol.

2.4 AI/DL in Fetal and Paediatric Cardiac Ultrasound

Congenital heart disease (CHD) is among the most challenging targets for prenatal ultrasound screening because it is comparatively rare, structurally heterogeneous, and its detection depends on precise identification of standard cardiac views. Arnaout et al. developed an ensemble of neural networks trained on a large multicentre dataset of fetal echocardiograms and second-trimester obstetric anatomy scans, and reported that the ensemble achieved detection performance for complex CHD that was comparable to that of expert fetal cardiologists when evaluated on an external test set from a separate institution, a notable result given that most prior CHD-detection studies had only been validated on data from a single centre.

In the paediatric postnatal setting, a deep-learning network trained on multi-view, multi-modal transthoracic echocardiograms from 1,932 children (1,255 healthy controls, 292 atrial septal defects, and 385 ventricular septal defects) achieved a mean view-classification accuracy of 0.989 and, for CHD screening using five views combining two-dimensional and Doppler echocardiography, a mean AUC of 0.996 and accuracy of 0.994 on within-centre evaluation, falling only slightly to an AUC of 0.990 and accuracy of 0.993 on a cross-centre test set. This cross-centre robustness is notable because generalisability across institutions and equipment is one of the most frequently cited limitations elsewhere in the literature.

2.5 AI/DL in Abdominal, Musculoskeletal, and General Applications

Beyond the three domains above, deep learning has also been applied to liver fibrosis and steatosis grading, kidney and gallbladder lesion classification, tendon and muscle assessment in musculoskeletal ultrasound, and general-purpose tasks such as automated standard-plane detection and image quality assessment. These applications share the same underlying architectural building blocks - CNN or transformer backbones for classification, U-Net-family architectures for segmentation, and object-detection architectures (e.g., YOLO, Faster R-CNN) for lesion localisation - as the breast, thyroid, and cardiac domains discussed above, which motivates the domain-agnostic conceptual framework proposed in Section 3.

2.6 Explainability and Broader Context

Foundational surveys of deep learning in medical image analysis by Litjens et al. and by Shen, Wu, and Suk documented the shift from hand-crafted feature engineering to learned representations across imaging modalities, including ultrasound, and identified data availability, annotation cost, and generalisability as recurring challenges as early as 2017 - challenges that remain visible in the more recent, domain-specific studies reviewed above. From a clinical-adoption perspective, Esteva et al. and Topol both argue that the practical barrier to AI adoption in healthcare imaging is now less about raw predictive accuracy, which has reached or exceeded expert-level performance on several narrowly defined tasks, and more about interpretability, workflow integration, and prospective, multi-site validation - a conclusion directly reflected in the explainability layer of the framework proposed in the next section.

Table 1. Summary of representative reviewed studies on AI/DL in ultrasound diagnosis

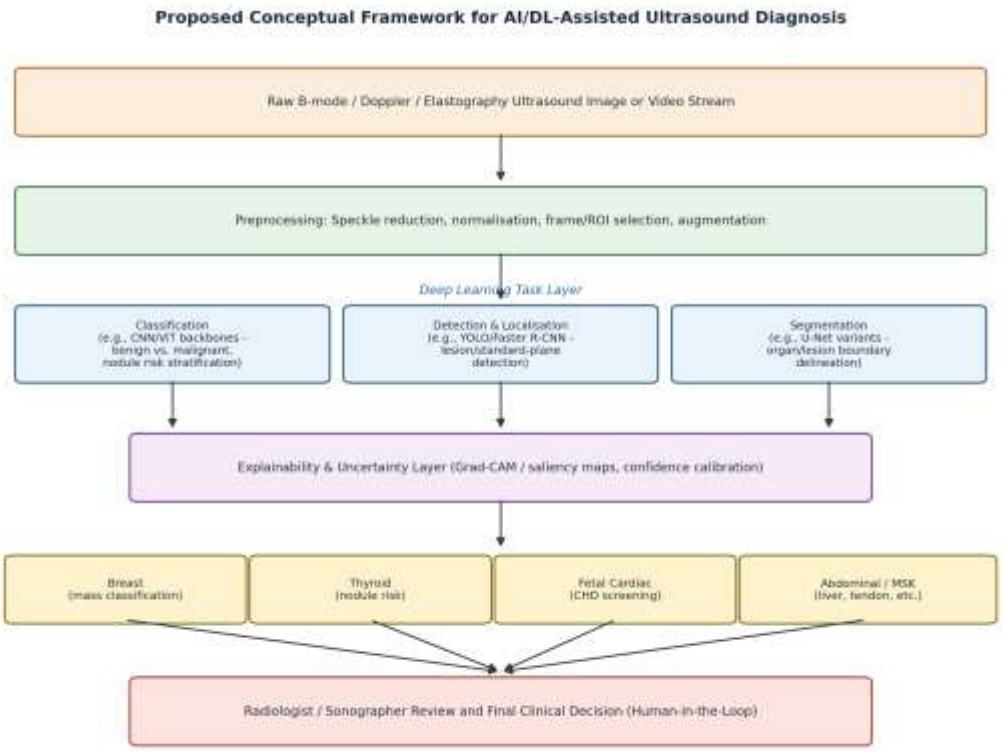
Study	Domain	DL Architecture	Dataset	Key Reported Metrics
Ma et al. (2021)	Breast (BUS)	Fus2Net (custom CNN)	100 test images	Acc 92.0%, Sens 95.65%, Spec 88.89%, AUC 0.97
Multimodal VGG16 study (2025)	Breast (non-mass lesions)	VGG16 (multimodal)	5-fold CV cohort	Acc 91.54%, Sens 94.15%, Spec 87.30%, AUC 0.96
Afrin et al. (2023)	Breast (review)	Various CNNs (59 studies)	2017-2023 literature	Narrative synthesis of DL tasks and reported performance
Guan et al. (2019)	Thyroid nodules	Inception-v3	2,836 images	Sens 93.3%, Spec 87.4%
Zhu et al. (2022)	Thyroid nodules (meta-analysis)	VGGNet (pooled, 11 studies)	11 pooled studies	Pooled Sens 0.87, Spec 0.85, AUC 0.93
Arnaout et al. (2021)	Fetal cardiac (CHD)	CNN ensemble	Multicentre FETAL-125/OB-125	Expert-level detection of complex CHD (external test set)
Multi-view TTE study (2024)	Paediatric cardiac (CHD)	Multi-view/multi-modal DL network	1,932 children	AUC 0.996, Acc 0.994 (within-centre)
Litjens et al. (2017)	General medical imaging (survey)	Various	Literature survey	Foundational taxonomy of DL in medical imaging
Esteva et al. (2019)	Healthcare AI (perspective)	N/A	N/A	Identifies interpretability & workflow integration as adoption barriers

3. Proposed Methodology

3.1 A Conceptual Framework for AI/DL-Assisted Ultrasound Diagnosis

The literature reviewed in Section 2 spans four clinical domains but repeatedly converges on the same three deep learning tasks - classification, detection/localisation, and segmentation - applied to the same general data pipeline. To organise this evidence and highlight where explainability and human oversight should sit architecturally, this paper proposes the conceptual framework shown in Figure 2. The framework is descriptive of current practice rather than a novel algorithm: its contribution is to make explicit the common structure underlying the reviewed studies and to position the explainability and clinician-review layers as first-class components rather than optional additions.

Figure 2. Proposed conceptual framework for AI/DL-assisted ultrasound diagnosis



3.2 Framework Components

Input Layer: Raw ultrasound data enters the pipeline as B-mode images, colour/power Doppler images, elastography maps, or video clips, depending on the clinical task and imaging protocol used in the reviewed studies.

Preprocessing Layer: Across the reviewed literature, preprocessing typically involves speckle-noise reduction, intensity normalisation, region-of-interest (ROI) or frame selection (particularly for cine-loop or video data, as in fetal echocardiography), and data augmentation (rotation, flipping, scaling) to mitigate the relatively small dataset sizes common in single-centre medical imaging studies.

Deep Learning Task Layer: Reviewed studies fall into three broad task categories. Classification models (e.g., CNN or hybrid CNN-transformer backbones such as those used by Ma et al. and in multimodal VGG16 studies) assign a diagnostic label - most commonly benign versus malignant - to an image or lesion. Detection and localisation models (e.g., YOLO- or Faster R-CNN-style architectures) identify and localise lesions or standard anatomical planes within an image, a task particularly emphasised in fetal cardiac screening, where identifying the correct standard view is itself a prerequisite for accurate diagnosis. Segmentation models (typically U-Net or its variants) delineate organ or lesion boundaries, supporting quantitative measurement (e.g., nodule size, ejection fraction) as well as downstream classification.

Explainability and Uncertainty Layer: Reflecting the adoption barriers identified by Esteva et al. and Topol, the framework places explainability - through techniques such as Grad-CAM or saliency-map visualisation, as used in breast ultrasound explainability studies - and uncertainty/confidence calibration between the task layer and the clinical domains, rather than treating it as a separate, optional analysis performed after the fact.

Clinical Application Domains: The framework shows the four domains reviewed in this paper (breast, thyroid, fetal cardiac, and abdominal/musculoskeletal) as parallel downstream applications of the same underlying task layer, reflecting the architectural overlap observed across the literature.

Human-in-the-Loop Clinical Review: In every reviewed study that discusses clinical translation, final diagnostic responsibility remains with the radiologist or sonographer; AI/DL output is consistently framed in the literature as a second reader or decision-support aid rather than a replacement for clinical judgement, consistent with current regulatory expectations for AI-based computer-aided diagnosis (CAD) systems.

3.3 Application Procedure

Applying the framework to a new clinical ultrasound task, as described across the reviewed studies, generally follows this sequence:

Step 1: Define the diagnostic task (classification, detection, or segmentation) and the clinical population.

Step 2: Assemble and annotate a dataset, typically with histopathological, cytological, or expert-consensus ground truth.
Step3: Preprocess images/video (denoising, normalisation, augmentation) as described in Section 3.2.
Step 4: Train and validate the selected DL architecture, most commonly using transfer learning from ImageNet-pretrained CNN backbones given typically modest medical dataset sizes.
Step 5: Generate explainability outputs (e.g., saliency maps) alongside predictions to support clinical interpretation.
Step 6: Validate on an external, ideally multicentre, test set before considering clinical deployment - a step the literature identifies as frequently missing or limited.
Step 7: Deploy as a second-reader or decision-support tool within the clinical workflow, retaining the sonographer/radiologist as the final decision-maker.

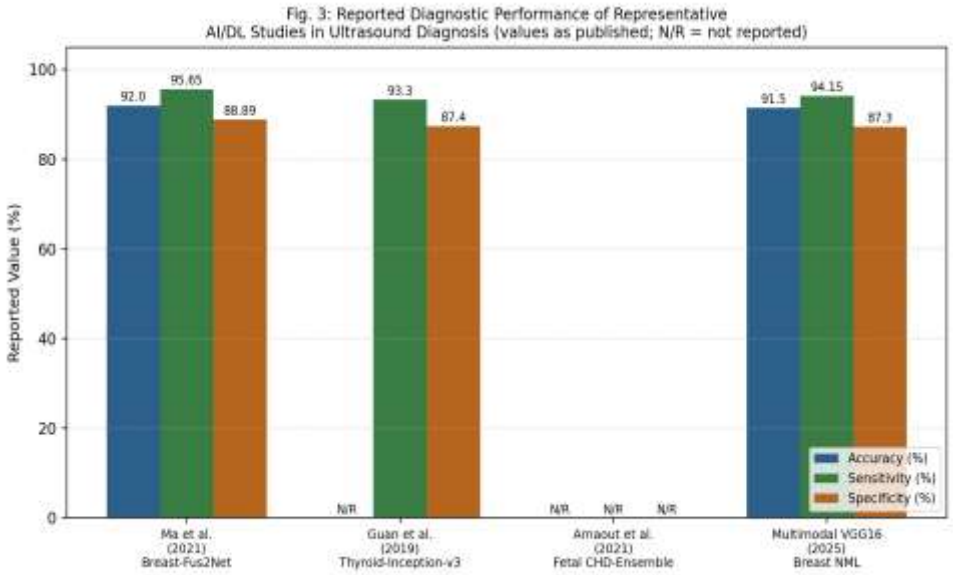
4. Result Analysis

This section synthesises the diagnostic performance figures reported in the primary studies reviewed in Section 2. All values shown are as published by the cited authors, obtained under each study's own dataset and validation protocol; they are not the result of a new experiment conducted for this review, and figures are not directly comparable across studies because of differences in dataset size, patient population, and validation methodology (e.g., held-out test set versus cross-validation versus external multicentre testing). Table 2 and Figure 3 present accuracy, sensitivity, and specificity for four representative studies spanning breast and thyroid classification and fetal cardiac screening; entries are marked "N/R" where the cited study did not report that particular metric.

Table 2. Reported diagnostic performance of representative AI/DL ultrasound studies

Study	Domain	Accuracy	Sensitivity	Specificity
Ma et al. (2021) - Fus2Net	Breast (BUS classification)	92.0%	95.65%	88.89%
Guan et al. (2019) - Inception-v3	Thyroid nodule classification	N/R	93.3%	87.4%
Arnaout et al. (2021) - CNN ensemble	Fetal CHD detection	N/R (expert-level, external test)	N/R	N/R
Multimodal VGG16 (2025)	Breast non-mass lesion classification	91.5%	94.15%	87.30%

Figure 3. Reported diagnostic performance of representative studies (values as published; N/R = not reported)



Two patterns are visible in this synthesis. First, sensitivity is generally reported as higher than specificity across the breast and thyroid studies shown, which is consistent with model training that prioritises minimising missed malignancies, a clinically motivated choice given the cost asymmetry between false negatives and false positives in cancer screening. Second, the Arnaout et al. fetal cardiac study reports performance in terms of expert-level equivalence on an external test set rather than as a single accuracy figure, reflecting the comparatively higher bar that domain sets for clinical validation - namely, that the model's performance is compared directly against practising fetal cardiologists rather than only against a fixed ground-truth label.

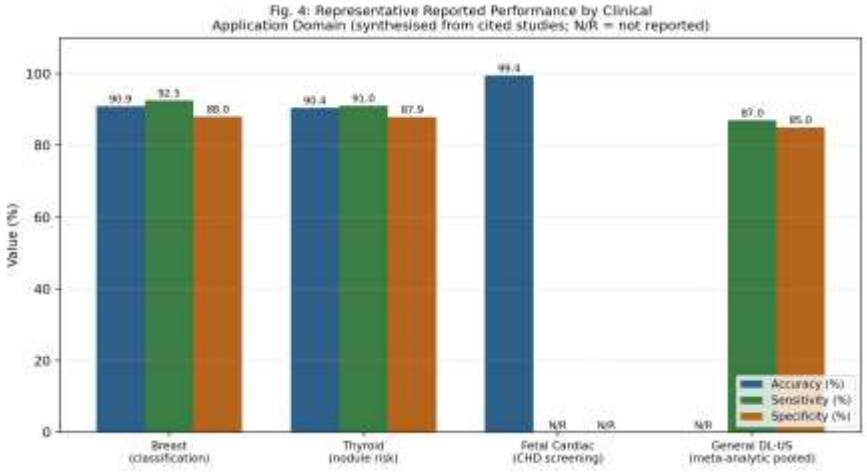
5. Performance Analysis

To compare performance across clinical domains rather than individual studies, Table 3 and Figure 4 aggregate representative reported figures from Section 2 by domain: breast lesion classification, thyroid nodule risk stratification, and fetal/paediatric cardiac screening, alongside the pooled meta-analytic estimate for thyroid nodules reported by Zhu et al. as a fourth reference point. As in Section 4, these are values drawn from the cited literature and synthesised for comparison, not the output of a new unified experiment; the paediatric cardiac figures in particular come from a single high-performing study and should not be read as representative of the domain as a whole.

Table 3. Representative reported performance aggregated by clinical domain

Clinical Domain	Representative Accuracy	Representative Sensitivity	Representative Specificity	Primary Source(s)
Breast (classification)	~91-92%	~94-96%	~87-89%	Ma et al. (2021); Multimodal VGG16 (2025)
Thyroid (nodule risk)	N/R (individual study)	~87-93%	~85-87%	Guan et al. (2019); Zhu et al. (2022, pooled)
Fetal/Paediatric Cardiac (CHD)	~99.4% (single study)	N/R (expert-level framing)	N/R	Arnaout et al. (2021); multi-view TTE study (2024)
General DL-US (meta-analytic pooled)	N/R	~87%	~85%	Zhu et al. (2022) pooled VGGNet estimate

Figure 4. Representative reported performance by clinical application domain (N/R = not reported)



Read across domains, the synthesis suggests that reported DL performance is highest and most consistent in constrained, well-annotated classification tasks with large single-centre datasets (breast, thyroid), and that the very high figures reported for paediatric cardiac screening (accuracy above 99%) come from a single study with a large, well-curated multicentre dataset (1,932 children) rather than being typical of the domain generally - fetal cardiac screening is otherwise widely acknowledged in the literature, including by Arnaout et al. themselves, as one of the more difficult targets for AI-assisted ultrasound because of

class rarity and structural heterogeneity of congenital heart disease. The gap between pooled meta-analytic estimates (Zhu et al.'s sensitivity of 87%) and the best individual reported study results (Guan et al.'s 93.3%) further illustrates a pattern noted throughout Section 2: individual published results likely represent a favourable subset of study designs and datasets, and pooled or externally validated estimates provide a more conservative, and probably more clinically realistic, view of expected real-world performance.

6. Conclusion

This review has synthesised current evidence on the application of artificial intelligence and deep learning to ultrasound diagnosis across four representative clinical domains: breast lesion classification, thyroid nodule risk stratification, fetal and paediatric cardiac screening, and abdominal/musculoskeletal applications. A structured, PRISMA-informed methodology was used to identify and synthesise the reviewed literature, and a conceptual framework was proposed that organises AI/DL applications by deep learning task (classification, detection, segmentation) while positioning explainability and human clinical oversight as core architectural components.

The synthesis of reported diagnostic performance shows that CNN-based and, increasingly, hybrid CNN-transformer models achieve accuracy, sensitivity, and specificity figures that frequently approach or match expert-level performance within the constraints of the retrospective datasets on which they were evaluated - reported accuracies generally in the high 80s to low-to-mid 90s percentage range for breast and thyroid classification tasks, and considerably higher in at least one large, well-curated paediatric cardiac dataset. At the same time, the literature is consistent in identifying three recurring limitations that temper these results: heterogeneity in reported performance across studies of the same clinical task (illustrated by the gap between individual thyroid-nodule studies and the pooled meta-analytic estimate), limited external and multicentre validation for most individual studies, and a persistent need for explainability to support clinical trust and adoption.

References

1. Afrin, H., Larson, N. B., Fatemi, M., & Alizad, A. (2023). Deep learning in different ultrasound methods for breast cancer, from diagnosis to prognosis: Current trends, challenges, and an analysis. Cancers, 15(12), Article 3139. https://doi.org/10.3390/cancers15123139
2. Arnaout, R., Curran, L., Zhao, Y., Levine, J. C., Chinn, E., & Moon-Grady, A. J. (2021). An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nature Medicine, 27(5), 882-891. https://doi.org/10.1038/s41591-021-01342-5
3. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29. https://doi.org/10.1038/s41591-018-0316-z
4. Guan, Q., Wang, Y., Du, J., Qin, Y., Lu, H., Xiang, J., & Wang, F. (2019). Deep learning based classification of ultrasound images for thyroid nodules: A large scale of pilot study. Annals of Translational Medicine, 7(7), 137. https://doi.org/10.21037/atm.2019.04.34
5. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005
6. Ma, H., Tian, R., Li, H., Sun, H., Lu, G., Liu, R., & Wang, Z. (2021). Fus2Net: A novel Convolutional Neural Network for classification of benign and malignant breast tumor in ultrasound images. BioMedical Engineering OnLine, 20, 112. https://doi.org/10.1186/s12938-021-00950-z
7. Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442
8. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7
9. Zhu, P.-S., Zhang, Y.-R., Ren, J.-Y., Li, Q.-L., Chen, M., Sang, T., Li, W.-X., Li, J., & Cui, X.-W. (2022). Ultrasound-based deep learning using the VGGNet model for the differentiation of benign and malignant thyroid nodules: A meta-analysis. Frontiers in Oncology, 12, Article 944859. https://doi.org/10.3389/fonc.2022.944859