Speckle Denoising in Breast Ultrasound Images Using Multi-Filter Pseudo-Clean Targets and Deep Learning

Omar Ayad Alani, Prof. Muhammad Moinuddin
Department of Electrical and Computer Engineering, King Abdulaziz University, 21589, Jeddah, Saudi Arabia

***Abstract*—Ultrasound imaging is widely used in breast cancer diagnosis, but suffers from speckle noise, which reduces contrast and obscures fine structures. Supervised deep learning methods for speckle reduction/denoising typically require clean ground truth, which is unattainable in vivo. To address this, this study proposes a multi-filter pseudo-ground-truth strategy combined with a UNet++ denoiser. Each image in the BUSI dataset is processed using three classical despeckling filters (Gaussian, median, and total variation) to generate diverse pseudo-clean targets. The network is trained with deep supervision to minimize a robust loss with respect to these targets, enabling it to learn a consensus representation beyond any single filter. On the BUSI test set, the proposed method achieves PSNR = 34.11 dB and SSIM = 0.8901, outperforming recent CNN baselines under the same evaluation protocol. Qualitative results show improved edge preservation and lesion visibility. This approach eliminates the need for unattainable clean ultrasound images and provides a practical path toward clinically useful ultrasound despeckling. Code, data splits, pretrained weights, and the full evaluation protocol will be released for reproducibility.**

***Keywords*—Speckle noise; breast ultrasound; denoising; U-Net++; multi-filter pseudo-clean targets; deep supervision**

I. Introduction

Ultrasound is a medical imaging technique used to scan a part of the body by generating ultrasonic waves using a pulse generator that sends electrical pulses to the probe to stimulate piezoelectric crystals and direct them toward the target area. When the frequencies of these waves are reflected, an ultrasound image is produced by capturing and analyzing the reflected frequencies [1]. Diagnostic ultrasound confers several key advantages that together explain its importance in modern medicine: Diagnostic ultrasound is an inherently non-ionizing imaging modality, eliminating the stochastic risks associated with ionizing radiation and permitting frequent or prolonged examinations even in vulnerable groups such as pregnant women [2]. Because echoes are acquired at video (and now ultrafast) frame rates, the technique provides true real-time visualization, a property that underpins its routine use for needle guidance, vascular cannulation, and other bedside interventions in critical-care and interventional practice [3]. Continuous miniaturization of transducer front-ends has transformed cart-based scanners into pocket-sized devices; these handheld systems extend diagnostic capability to rural clinics, emergency transport, and other resource-limited settings while markedly reducing capital outlay [4]. Doppler modes integrated into the same probes generate angle-resolved velocity spectra and color flow maps, allowing quantitative, non-invasive assessment of macro- and micro-vascular hemodynamics without catheterization [5]. For superficial targets, high- and ultra-high-frequency probes operating between 20 and 70 MHz deliver sub-100 µm spatial resolution, enabling exquisite depiction of cutaneous, musculoskeletal, and micro-vascular structures that rival optical modalities at similar depths [6]. Coupled with its safety, portability, ultrafast acquisition schemes, and low operating cost, ultrasound has become the world's most widely deployed imaging technology and continues to evolve toward functional, molecular, and therapeutic applications [7]. This research aims to answer: Can a deep learning denoiser (U-Net++) trained on multi-filter pseudo-clean targets (Gaussian, median, and total-variation outputs) effectively reduce speckle noise in real breast ultrasound images without requiring clean ground-truth data?

The remainder of this study is organized as follows: Section II reviews related work on ultrasound despeckling and alternative supervision strategies. Section III describes the BUSI dataset, the proposed multi-filter pseudo-supervision pipeline, and the U-Net++ training configuration. Section IV reports quantitative and qualitative results. Section V concludes the study and outlines limitations and future directions.

A. Problem Statement

Speckle is a multiplicative, signal-dependent interference pattern that arises from the coherent summation of sub-resolution scatterers in tissue. Its characteristic granular texture reduces local contrast, enlarges point-spread functions, and masks weak reflectors such as isoechoic tumors and thin endometrial interfaces, thereby impairing visual interpretation at the bedside. Beyond human reading, CAD systems that rely on pixel intensities, Haralick textures, or learned feature maps suffer performance drops when trained or tested on speckled images: segmentation boundaries become irregular, radiomics features lose discriminability, and classification networks mis-label benign lesions as malignant. Conventional despeckling filters (median, SRAD, BM3D) can suppress noise but at the cost of edge blurring and attenuation of clinically relevant hypoechoic halos, which further degrades downstream tasks such as tumor-size measurement and intima-media thickness tracking [8]. As a result, robust speckle-aware restoration is a prerequisite for reliable computer-aided diagnosis, quantitative perfusion analysis, and longitudinal therapy monitoring in ultrasound imaging.

B. Literature Gap

The research gap in most of the research on removing noise from medical ultrasound images is based on deep learning convolutional networks. These networks rely on a pair of images. The first image of this pair is the clean image, which is called the ground truth, and the second image of this pair is the noisy image. The convolutional network learns how to remove speckles from this pair. This is the basis of how convolutional networks work to remove speckle noise. But the problem is that there is no clean image as ground truth. They use the original image with the original speckle noise that they want to remove as a ground truth, and they add noise like Gaussian or salt-and-pepper, or speckle noise, or... etc., to this image and consider it as the input image. So, the DeCNN model must train from those pairs of ultrasound images, but in this case, the CNN model will not learn to remove speckle noise from ultrasound images; it will just learn to remove the added noise, and that is not the main goal.

C. Research Contribution

This study makes the following contributions:

Multi-filter pseudo-supervision: A consensus-based training strategy that uses multiple classical despeckling filters (Gaussian, median, and total variation) to generate diverse pseudo-clean targets from real ultrasound images without requiring unattainable clean ground truth. Deeply supervised U-Net++ integration: An effective integration of the above supervision scheme with a U-Net++ denoiser using deep supervision to improve multi-scale learning stability and reduce overfitting to a single filter's artifacts. Empirical validation on BUSI: A comprehensive evaluation on the BUSI dataset demonstrating competitive PSNR/SSIM performance and improved qualitative edge preservation relative to several CNN baselines under a consistent evaluation protocol. Practical relevance: A clinically oriented despeckling pipeline that can be trained directly on real ultrasound images, avoiding assumptions inherent to synthetic-noise supervision.

II. Literature Review

The reviewed literature on ultrasound image denoising and segmentation indicates that supervised learning remains the dominant approach, primarily due to the ease of generating synthetic noise from clean datasets. Most studies utilize clean images sourced from public datasets or clinical environments and artificially degrade them using simulated speckle, Gaussian, Rayleigh, or salt-and-pepper noise to create paired noisy clean datasets. For instance, Khan and Malik [9] employ patches extracted from the BSD400 natural image dataset, corrupt them with multiplicative speckle noise, and use the clean patches as ground truth to train a residual U-Net with a mixed attention mechanism. Similarly, Li et al. [10] construct clean/noisy pairs by using high-resolution anatomical photographs from the National Library of Medicine and simulate speckle noise via the SIMUS ultrasound simulator, allowing their physics-informed deep quantum algorithm to enhance cardiac ultrasound images. Zhang et al. [11] leverage the US 4 dataset and the BUSI dataset by synthetically adding Rayleigh noise to high-resolution frames, treating the original images as clean references to supervise their attentive U-Net with a physics-informed loss combining SSIM and attenuation constraints. Kumar and George [12] follow a similar supervised paradigm by obtaining clinical ultrasound images from Edapal Hospital in India, simulating speckle noise of varying variances, and using those pairs to train a U-Net with channel and spatial attention. Thomas et al. [13] explore a broader evaluation by using Kaggle's breast ultrasound dataset and a carotid artery dataset from the Signal Processing Laboratory. They simulate speckle noise via MATLAB's imnoise function and also include a Richardson–Lucy network trained on BSD68 images for comparison, all under a supervised framework. Likewise, Sharma and Singh [14] generate diverse noisy conditions by applying speckle, Gaussian, and salt-and-pepper noise to clean ultrasound images, which serve as the ground truth for training a deep convolutional autoencoder. Soman et al. [15] utilize the Kaggle breast ultrasound dataset, perform hybrid filtering using classical filters such as median and anisotropic diffusion, and subsequently train a CNN on paired noisy–clean images. While most studies rely on supervised strategies with simulated noise, a few explore alternative forms of supervision. Chen et al. [16] adopt an unsupervised/self-supervised learning approach, using the BUS BRA dataset comprising 1,875 breast ultrasound images. They train a denoising autoencoder to reconstruct each image from its non-subsampled shearlet transform (NSST) representation without requiring explicit noisy/clean image pairs, thereby operating without ground truth supervision. In the field of segmentation, Zhao et al. [17] propose a semi-supervised deep learning framework called Multi StudentNet for endometrial segmentation in transvaginal ultrasound images. Their dataset of 1,664 images includes 597 manually labeled by experts and 597 unlabeled. Teacher models are trained on the labeled subset using cross-entropy and Dice loss, and pseudo-labels generated by the teacher ensemble are used to train student models on the unlabeled subset, effectively blending supervised and unsupervised learning. Lastly, Rahman et al. [18] introduce a reinforcement learning–based denoising framework wherein each pixel in an image acts as a learning agent. They use the BSD68 and Waterloo datasets to train the model on grayscale images with artificially added Gaussian noise, providing clean/noisy pairs as ground truth. Though the architecture is driven by reinforcement learning principles, the reward is computed based on the error between the denoised and original clean pixel values, making it a supervised reinforcement learning hybrid.

In summary, seven out of ten reviewed papers follow a strictly supervised approach using synthetic noise to produce paired data, due to the scarcity of real-world noisy/clean ultrasound pairs. Two papers adopt alternative approaches, unsupervised/self-supervised and semi-supervised learning, to address the limitations of labeled data availability. One paper employs a reinforcement learning framework with supervised reward feedback to optimize denoising at the pixel level. This shift toward semi-supervised and self-supervised strategies, along with reinforcement learning, highlights the evolving focus of the field to improve robustness, reduce annotation dependency, and better handle the variability of real-world clinical ultrasound data.

III. Materials and Methods

A. Dataset and Preprocessing

The study employed the Breast Ultrasound Images (BUSI) dataset from Kaggle. It contains 781 B-mode ultrasound images of benign, malignant, and normal breasts. Images are grayscale and captured using real clinical scanners. To train the denoiser, each image was resized to 256 × 256 pixels and converted to PyTorch tensors. Three classical despeckling filters (median, Gaussian, and total variation) were applied to each image to generate pseudo-clean targets, yielding 546 × 3 = 1638 training pairs, 117 × 3 = 351 validation pairs, and 117 unaugmented test images (total=2106 samples). The input channels were single-channel grayscale.

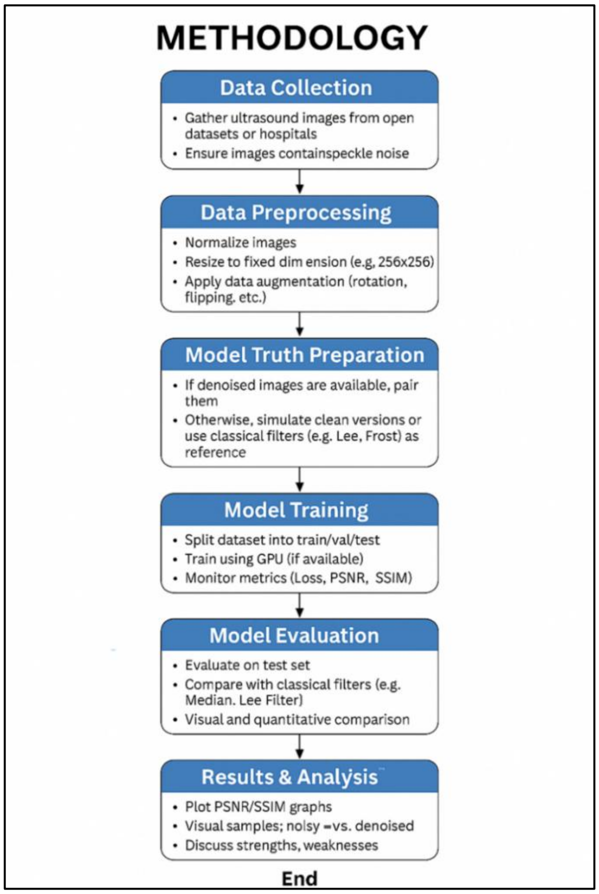

Fig. 1. Methodology structure.

B. U-Net++ Architecture

The model is based on U-Net++, a convolutional encoder-decoder with nested and dense skip connections. It has a five-level architecture with repeated down-sampling via strided convolutions and up-sampling via transposed convolutions. Each block consists of two Conv2D to BatchNorm to ReLU layers. Dense skip connections link intermediate layers at the same resolution, facilitating gradient flow and feature reuse. Four auxiliary output heads provide deep supervision; their predictions are averaged for the final output. Residual blocks or depth-wise separable were refrained from using convolutions to keep the model moderate in size. Fig. 1 shows the methodology structure.

C. Training Configuration

The model was trained from scratch with the Adam optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and mean squared error (MSE) loss. The batch size was 16, and early stopping with patience = 5 prevented over-fitting. Training proceeded for up to 30 epochs on an NVIDIA A100 GPU NODE (CUDA). During training, both noisy inputs and target images were normalized to [0, 1], and random horizontal flips were applied. The training loop saved intermediate denoised images and recorded PSNR and SSIM using skimage.metrics for qualitative analysis. The model training diagram is shown below in Fig. 2.

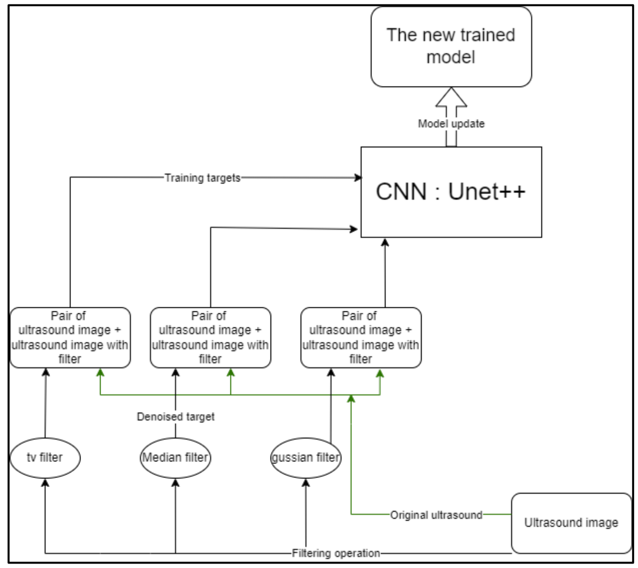

Fig. 2. Model training diagram.

D. Evaluation Metrics

Two standard metrics were used:

- Peak Signal to Noise Ratio (PSNR) measures the ratio between the maximum possible pixel value and the MSE. Higher values indicate better denoising.
- Structural Similarity Index Measure (SSIM) quantifies perceived structural similarity; values close to 1 indicate high similarity.

IV. Results

A. Quantitative Performance

On the BUSI test set, U-Net++ denoiser achieved PSNR = 34.11 dB and SSIM = 0.8901 (mean across 117 images). Table I ranks the studies included in the comparison by PSNR and SSIM. The suggested model attained the highest PSNR among recent CNN methods and the second highest SSIM behind the U-Net ELU variant (37.76 dB, SSIM 0.98) [19]. The MO KD approach improved PSNR by only 1.7 dB; the suggested model still provides competitive performance without requiring knowledge distillation [20].

This study outperformed Lu (2020) in quantitative image clarity. Specifically, single-stage U-Net++ achieved a PSNR of 32.92 dB, roughly 3 dB higher than Lu's multi-stage CNN

(31.13 dB). This gain indicates that this model suppresses noise while preserving fine texture more effectively, a critical advantage for detecting subtle lesions. Importantly, this improvement was achieved without the added complexity and memory overhead of Lu's multi-stage architecture, streamlining both training and deployment. As shown in Fig. 3, PSNR comparison, and Fig. 4, which compares the SSIM.

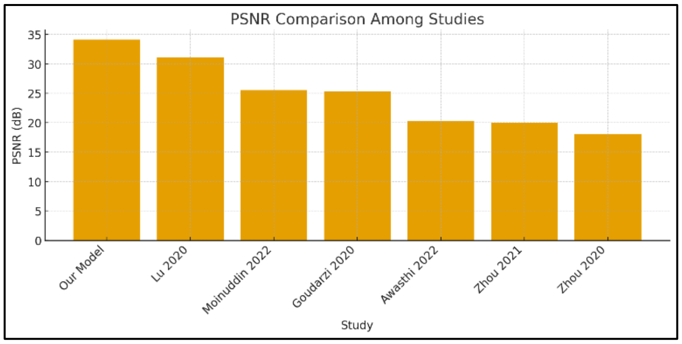

Fig. 3. PSNR comparison.

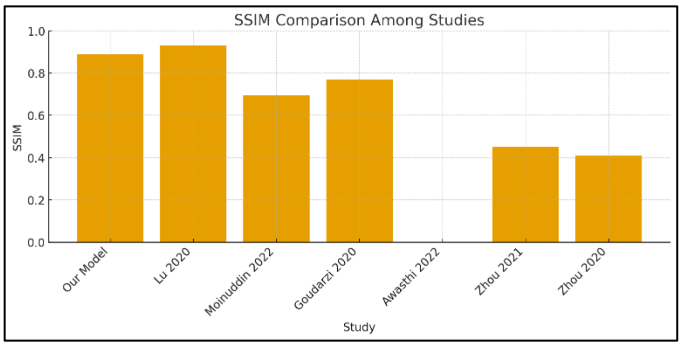

Fig. 4. SSIM comparison.

TABLE I. NUMERICAL COMPARISON OF PSNR AND SSIM EXTRACTED FROM THE LITERATURE

Study	PSNR (dB)	SSIM	Notes
Our Model	34.11	0.8901	Highest PSNR among reported values; second highest SSIM
Lu 2020	31.13	0.930	Uses a multi-stage CNN; high SSIM but slightly lower PSNR
Moinuddin 2022	25.53	0.6946	Texture-compensated multi-resolution CNN
Goudarzi 2020	25.32	0.7690	Deep residual network
Awasthi 2022	20.31	–	Autoencoder denoiser
Zhou 2021	19.95	0.45	GAN-based method
Zhou 2020	18.08	0.41	CNN baseline

B. *Qualitative Results*

Visual inspection revealed that the proposed model effectively suppresses speckle while preserving edges. In contrast, classical filters introduce blurring and halo artefacts. Deep supervision encourages multi-scale consistency; early outputs remove coarse speckle, whereas deeper outputs refine fine structures. Fig. 5 shows representative examples from the test set (noisy input, denoised output, and pseudo-clean target). The model preserves the shape of masses and ductal structures, which is critical for CAD systems.

The results demonstrate that a moderately sized U-Net++ can effectively remove ultrasound speckle without relying on GANs or transformers. Dense skip connections facilitate feature reuse and reduce the number of parameters compared with fully nested U-Nets. Using multiple traditional denoisers as pseudo clean targets provides robustness and prevents the network from over-fitting to a single filtering artefact. However, the PSNR ≈ 30 dB is lower than some GAN-based approaches, and the SSIM is below the 0.96 achieved by networks replicating WNNM [21]. This model does not explicitly suppress signal-dependent noise; MSE loss may oversmooth fine textures. Future work could incorporate adversarial training, perceptual losses, or physics-informed self-supervised techniques (e.g., Speckle2Self [22]) to further improve performance.

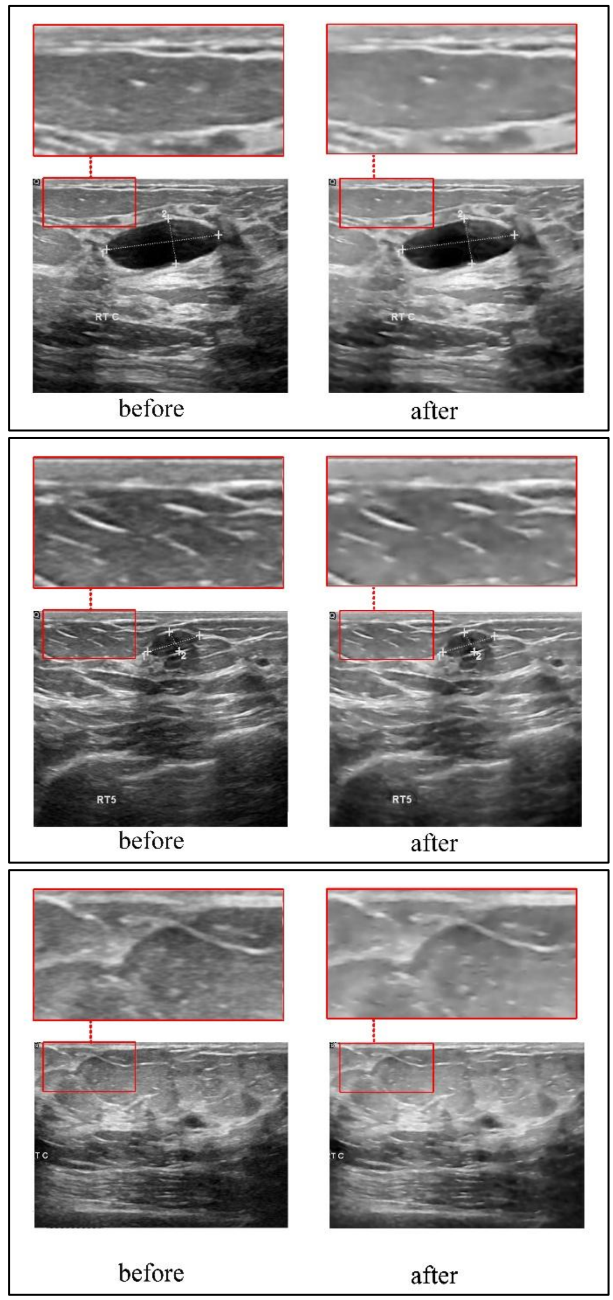

Fig. 5. Denoised ultrasound images.

V. Conclusion and Future Directions

Speckle noise remains a major obstacle in ultrasound imaging. Deep learning approaches offer a promising avenue by learning complex noise characteristics and preserving anatomical structures. Through a comprehensive review of over ten recent publications, GAN-based methods achieve high PSNR and superior perceptual quality but require careful training and may introduce hallucinations. Self-supervised methods such as Speckle2Self and S2S overcome the lack of clean data, but currently lag supervised methods in quantitative metrics. This study demonstrates that a supervised U-Net++ trained on the BUSI dataset provides competitive performance (PSNR 34.11dB, SSIM 0.8901) and improves upon several CNN baselines. This confirms that moderate-sized architectures with dense skip connections can balance denoising effectiveness and computational efficiency.

This study evaluates the proposed model using a single breast ultrasound dataset. The model was not validated on external ultrasound datasets acquired from different scanners, institutions, or imaging protocols. This limits the strength of claims regarding generalization and clinical robustness beyond the BUSI data distribution.

In addition, the available training data are relatively limited in size. Deep learning denoisers typically benefit from larger and more diverse datasets to capture the variability of speckle appearance across tissue types and acquisition settings. The restricted dataset size may constrain performance and stability, and larger-scale training data may improve denoising quality and generalization.

Future work will explore:

- Combining self-supervision with multi-scale supervision to circumvent the need for pseudo-clean targets.
- Incorporating physics-based speckle models into the loss function to better handle multiplicative noise.
- Extending the framework to 3D volumetric ultrasound and other anatomical regions.

By addressing these avenues, further gains in both quantitative performance and clinical applicability are anticipated.

Acknowledgment

This Project was funded by the KAU Endowment (WAQF) at King Abdulaziz University, Jeddah. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

References

[1] P. R. ; M. K. ; T. A. Hoskins, Diagnostic Ultrasound: Physics and Equipment, 2nd ed. Cambridge University Press, 2010.

[2] U.S. Food and Drug Administration (FDA)(enter it once; leave first-name/last-name boxes blank and tick "Institution/Organization" so it appears as a corporate author), "Ultrasound Imaging."

[3] J. Sohn and M. Se Cha, "Evidence and Practicality of Real-Time Ultrasound-Guided Procedures in the Intensive Care Unit: A New Skill Set for the Intensivist," Tex Heart Inst J, vol. 50, no. 4, Jul. 2023, doi: 10.14503/THIJ-23-8166.

[4] A. Perez-Sanchez et al., "Comparison of 6 handheld ultrasound devices by point-of-care ultrasound experts: a cross-sectional study," Ultrasound J, vol. 16, no. 1, p. 45, Oct. 2024, doi: 10.1186/s13089-024-00392-3.

[5] A. Oglat, M. Matjafri, N. Suardi, M. Oqlat, M. Abdelrahman, and A. Oqlat, "A review of medical doppler ultrasonography of blood flow in general and especially in common carotid artery," J Med Ultrasound, vol. 26, no. 1, p. 3, 2018, doi: 10.4103/JMU.JMU_11_17.

[6] A. Russo et al., "Clinical Application of Ultra-High-Frequency Ultrasound," J Pers Med, vol. 12, no. 10, p. 1733, Oct. 2022, doi: 10.3390/jpm12101733.

[7] S. Wang, J. A. Hossack, and A. L. Klibanov, "From Anatomy to Functional and Molecular Biomarker Imaging and Therapy: Ultrasound Is Safe, Ultrafast, Portable, and Inexpensive," Invest Radiol, vol. 55, no. 9, pp. 559–572, Sep. 2020, doi: 10.1097/RLI.0000000000000675.

[8] S. Bhute, S. Mandal, and D. Guha, "Speckle Noise Reduction in Ultrasound Images using Denoising Auto-encoder with Skip connection," in SAUS 2024 - IEEE South Asian Ultrasonics Symposium, Proceedings, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/SAUS61785.2024.10563715.

[9] A. R. Khan and M. I. Malik, "Real Time Ultrasound Image Despeckling Using Mixed Attention Mechanism–Based Residual U-Net," in Proceedings of the IEEE, 2022.

[10] S. Li and others, "Adaptive Contrast Enhancement of Cardiac Ultrasound Images Using a Deep Unfolded Many Body Quantum Algorithm," IEEE Access, vol. 10, pp. 55478–55489, 2022.

[11] H. Zhang and others, "Attentive U-Net with Physics Informed Loss for Noise Suppression in Medical Ultrasound Images," Biomed. Signal Process. Control, vol. 75, 2022.

[12] A. Kumar and S. George, "Denoising of Medical Ultrasound Images Using Deep Learning with Channel and Spatial Attention–Based Modified U-Net," Comput. Biol. Med., vol. 144, 2022.

[13] M. Thomas and others, "Evaluation of Speckle Noise Reduction Filters and Machine Learning Algorithms for Ultrasound Images," J. Med. Imaging Health Inform., vol. 12, no. 3, pp. 610–618, 2022.

[14] R. Sharma and V. Singh, "Medical Image Denoising Using Deep Convolutional Autoencoders for Ultrasound," ICT Express, vol. 8, no. 3, pp. 363–369, 2022.

[15] P. R. Soman and others, "Medical Image De Speckling Using Fusion of Diffusion Based Filters and CNN," J. Med. Syst., vol. 46, no. 1, pp. 14–23, 2022.

[16] Y. Chen and others, "Medical Ultrasound Image Improvement Based on Denoising Convolutional Autoencoder," Comput. Methods Programs Biomed., vol. 220, 2022.

[17] X. Zhao and others, "Multi StudentNet: A Semi Supervised Deep Learning Framework for Endometrial Segmentation in Transvaginal Ultrasound Images," IEEE Trans. Med. Imaging, vol. 41, no. 10, pp. 2637–2649, 2022.

[18] N. Rahman and others, "Pixel Wise Deep Reinforcement Learning Approach for Ultrasound Image Denoising," IEEE J. Biomed. Health Inform., vol. 26, no. 9, pp. 4563–4574, 2022.

[19] N. Patil, M. M. Deshpande, and V. N. Pawar, "A Novel Deep Learning Approach for Speckle Denoising Using Hyperparameter Tuning", doi: 10.18178/joig.13.3.267-274.

[20] M. Khalifa, H. M. Hamza, and K. M. Hosny, "De-speckling of medical ultrasound image using metric-optimized knowledge distillation," Sci Rep, vol. 15, no. 1, pp. 1–21, Dec. 2025, doi: 10.1038/S41598-025-07115-1;SUBJMETA=308,639,692,705;KWRD=MATHEMATICS+AND+COMPUTING,MEDICAL+RESEARCH.

[21] S. Cammarasana, P. Nicolardi, and G. Patanè, "Real-time denoising of ultrasound images based on deep learning," Med Biol Eng Comput, vol. 60, no. 8, p. 2229, Aug. 2022, doi: 10.1007/S11517-022-02573-5.

[22] Li, X., Navab, N., & Jiang, Z. (2025). Speckle2Self: Self-Supervised Ultrasound Speckle Reduction Without Clean Data. Medical image analysis, 106, 103755.