Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 4 (2026) 1089 - 1098 e-ISSN: 2580-0760

Impact of Speckle Reduction Filters on Machine Learning-Based Detection of Polycystic Ovary Syndrome from Ovarian Ultrasound Images

Fazrol Rozi[1], Syafrizal Sy[2], Adiwijaya[3], Yusrawati[4], Primawati[5]
[1]Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Science
Universitas Andalas, Padang, Indonesia
[3]Department of Informatics, School of Computing, Telkom University, Bandung, Indonesia
[4]Department of Obstetrics and Gynecology, Faculty of Medicine, Universitas Andalas, Padang, Indonesia
[5]Department of Mechanical Engineering, Faculty of Engineering, Universitas Negeri Padang, Padang, Indonesia

fazrol@pnp.ac.id

Abstract

Polycystic Ovary Syndrome (PCOS) is commonly assessed with ovarian ultrasonography, but speckle can conceal follicular margins and reduce the robustness of automated interpretation. Although automated PCOS studies increasingly employ machine learning, the contribution of conventional despeckling to subsequent segmentation and classification has not been examined consistently. This study compares five classical filters - Mean, Median, Lee, Frost, and Kuan - within an interpretable machine-learning pipeline for ovarian ultrasound analysis. From a public collection of 12,680 images, a balanced sample of 300 scans (150 PCOS and 150 non-PCOS) was selected. Two radiologists produced follicle annotations, and disagreements were resolved with a third expert to obtain consensus masks. Each filtered image was segmented by adaptive thresholding with morphological refinement, after which geometric and intensity descriptors were extracted. Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and Logistic Regression (LR) were trained using a stratified 70/30 train-test split with cross-validated hyperparameter tuning. The Kuan-LR configuration yielded the strongest result, reaching 94.44% accuracy and an AUC of 0.98, together with the best edge-preservation score and segmentation agreement. The results indicate that preprocessing materially affects the reliability of an interpretable PCOS detection pipeline and provide quantitative guidance for selecting a speckle-reduction strategy before segmentation and classification.

Keywords: follicle segmentation; image preprocessing; kuan filter; machine learning; polycystic ovary syndrome; ultrasound speckle noise

How to Cite: F. Rozi, S. Sy, A. Adiwijaya, A. Nazra, and P. Primawati, "Impact of Speckle Reduction Filters on Machine Learning-Based Detection of Polycystic Ovary Syndrome from Ovarian Ultrasound Images," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 10, no. 6, pp. 1089–1098, 2026.
Permalink/DOI: https://doi.org/10.29207/xwwsr649

Received: March 18, 2026
Accepted: July 16, 2026
Available Online: August 19, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Polycystic Ovary Syndrome (PCOS) is a frequent endocrine condition in women of reproductive age, with reported prevalence varying from approximately 6% to 20% according to the population and diagnostic definition used [1], [2]. Under the Rotterdam framework, a diagnosis is supported when at least two of three features are present: oligo- or anovulation, clinical or biochemical evidence of hyperandrogenism, and polycystic ovarian morphology on ultrasound [3]. The clinical importance of reliable detection is reinforced by associations with infertility, metabolic abnormalities, type 2 diabetes, and cardiovascular complications [4], [5].

Ultrasound is widely used to assess ovarian morphology because it is non-invasive, comparatively affordable, and able to visualize follicles in real time [6]. For PCOS assessment, the modality allows clinicians to examine the number and arrangement of small follicles and other structural characteristics of the ovary. Its diagnostic value, however, depends on image quality; weak or partially hidden follicular margins can make interpretation difficult [7].

One of the principal degradations in ultrasound is speckle, a signal-dependent multiplicative pattern generated by interference among backscattered acoustic waves [8-10]. This granular texture can lower local contrast, conceal fine boundaries, and make both visual reading and computational analysis more difficult [11],

[12]. The effect is particularly relevant in ovarian imaging, where delineating small follicular structures is central to morphology-based PCOS evaluation [7], [13].

Speckle reduction has therefore been approached with both conventional neighborhood filters and adaptive methods [14], [15]. Mean and Median filtering are simple and computationally inexpensive, yet excessive smoothing can remove structural details that are important for diagnosis [13], [16]. Adaptive approaches such as Lee, Frost, and Kuan instead use local image statistics to respond to signal-dependent noise and are intended to suppress speckle while retaining edges and local anatomy [8], [17], [18].

At the same time, ovarian-ultrasound studies have adopted machine-learning and deep-learning techniques for automated PCOS analysis [17], [19], [20]. Much of this literature concentrates on the classifier or segmentation architecture. As a result, the effect of a classical despeckling choice on image quality, follicle delineation, feature separation, and final classification has seldom been evaluated within one consistent processing chain [19], [21-25].

Table 1. Representative studies on automated PCOS detection and ovarian ultrasound analysis.

Study	Method	Main Findings	Limitation
Adiwijaya et al. [24]	Follicle detection framework	Demonstrated feasibility of machine learning for PCOS analysis	No investigation of speckle filtering
Wisesty et al. [25]	Modified backpropagation	Improved PCOS classification performance	Limited preprocessing analysis
Alamoudi et al. [17]	Deep learning fusion	Accuracy around 93%	Preprocessing effects were not studied
Saha et al. [21]	Modified U-Net	Accurate follicle segmentation	No comparative filter evaluation
Singh et al. [22]	Attention-based network	Improved ovarian quantification	Limited interpretability
Proposed study	Classical filtering + machine learning	Accuracy of 94.44% and interpretable pipeline	Single-center dataset and handcrafted features

Table 1 shows that earlier PCOS studies achieved useful results with follicle-detection, backpropagation, U-Net, attention, and deep-learning fusion approaches, but preprocessing was generally treated as a supporting step rather than an experimental variable. The comparison also makes the present research gap explicit: none of the listed prior studies performs a controlled, side-by-side assessment of several classical speckle filters and then traces their effects through segmentation, handcrafted-feature extraction, and classification. The proposed study addresses this gap with an interpretable pipeline while acknowledging its use of a single public dataset and handcrafted descriptors.

This gap matters because preprocessing changes the pixels from which follicular masks and handcrafted descriptors are produced. A denoising method that suppresses speckle but erases boundaries may therefore improve a generic noise metric while simultaneously weakening the information required by an interpretable classifier. Evaluating filtering and classification as connected stages can provide more useful design guidance for computer-aided PCOS analysis.

Accordingly, the experiments compare Mean, Median, Lee, Frost, and Kuan filtering inside a single end-to-end machine-learning framework for ovarian-ultrasound PCOS detection. The workflow uses expert-validated ground truth, adaptive-threshold follicle segmentation, geometric and intensity descriptors, and comparative evaluation of several conventional classifiers so that the effect of the preprocessing choice can be observed at each downstream stage.

The study contributes three related analyses. First, it benchmarks five classical speckle-reduction filters on the same ovarian-ultrasound sample. Second, it examines whether changes in image-quality and segmentation measures are reflected in feature discriminability and classifier performance. Third, it identifies the filter-classifier pairing that offers the strongest performance while retaining a transparent feature-based decision process, thereby providing practical evidence for the design of interpretable ultrasound CAD pipelines.

Taken together, the study treats preprocessing as an integral model-design decision rather than as an incidental operation performed before learning.

2. Methods

The experimental protocol was organized as a seven-stage pipeline: dataset construction and expert annotation, image standardization, speckle filtering, follicle segmentation, feature extraction, machine-learning classification, and quantitative evaluation. Figure 1 summarizes the order of these operations and the points at which the alternative filters are compared.

Implementation was carried out in Python 3.11 with OpenCV 4.10, NumPy 2.0, SciPy 1.13, Scikit-image 0.24, and Scikit-learn 1.5. Computation used a Windows 11 workstation with an Intel Core i7 processor and 32 GB of RAM. A random seed of 42 was fixed for all stochastic operations to make the experimental split and model procedures reproducible.

2.1 Dataset Description

The source images came from the public ovarian-ultrasound repository curated by Indirani [26], which contains 12,680 transabdominal scans from PCOS-positive and PCOS-negative cases. Because the aim was to compare preprocessing methods under identical conditions rather than to train on the full repository, the experiments used a controlled balanced subset.

Stratified random sampling produced 300 images, evenly divided into 150 PCOS and 150 non-PCOS cases. This design prevents class imbalance from dominating the comparison and keeps expert annotation

and repeated filter-classifier experiments feasible. The same selected images were retained for every filtering condition so that differences in performance could be attributed to the processing method rather than to a change of sample.

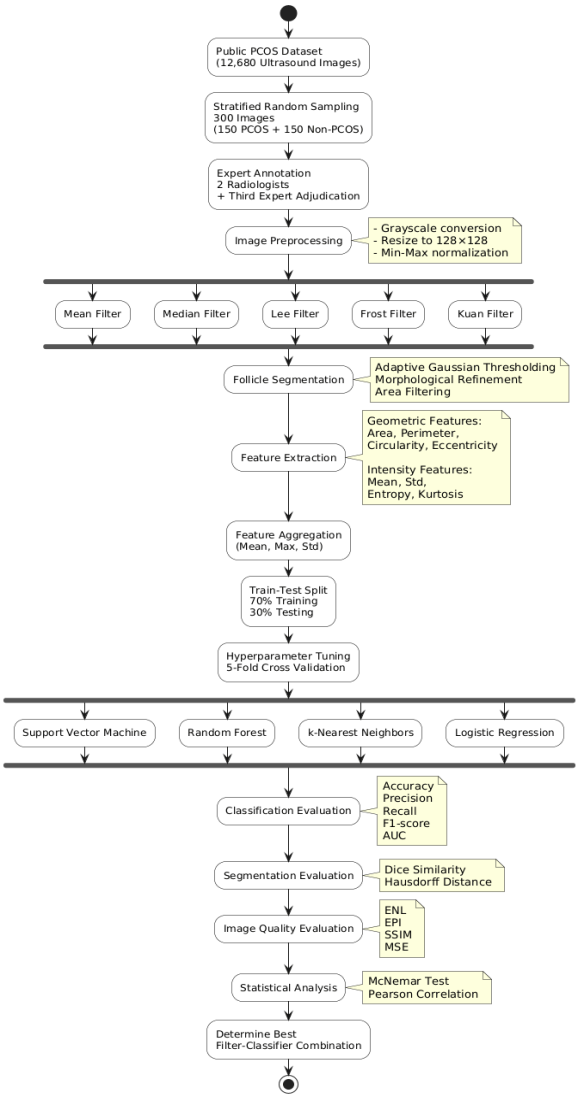


Figure 1. Overall Experimental Workflow of the Proposed Framework

Ground-truth follicle masks were prepared independently by two experienced radiologists in ITK-SNAP 3.8. The annotation protocol required delineation of follicular boundaries, rejection of non-follicular hypoechoic regions and artifacts, and verification that each marked region was clinically plausible. Inter-observer Dice agreement was 0.87 ± 0.04. When the two annotations differed, a third senior radiologist participated in adjudication, and the agreed mask was used for subsequent segmentation evaluation. Because the expert contribution consisted of image annotation rather than interview-based judgment, consistency was quantified through inter-observer agreement.

2.2 Image Preprocessing

Before filtering, the scans were standardized to reduce irrelevant variation and computational load. Each image was converted to grayscale with the luminosity transform and then resampled to 128 × 128 pixels using bicubic interpolation. The selected spatial size was intended to retain the principal follicular structures while allowing the repeated filtering and classification experiments to remain computationally efficient.

After resizing, intensities were scaled by min-max normalization to the interval [0,1]. Applying the same numerical range to every image reduces scale differences between cases and provides more stable inputs for the later feature-processing and machine-learning stages.

2.3 Speckle Filter Implementation

Five conventional filters were evaluated as the principal preprocessing alternatives: Mean, Median, Lee, Frost, and Kuan. The first two provide neighborhood-based smoothing, whereas Lee, Frost, and Kuan incorporate local statistics to address the signal-dependent behavior of speckle in ultrasound.

For the Mean filter, a pixel is replaced by the arithmetic average of intensities in its local neighborhood, which reduces rapid intensity variation but can soften edges. The Median filter uses the neighborhood median instead of the mean and is therefore less sensitive to isolated extreme values while generally retaining abrupt transitions more effectively.

The adaptive filters modify the amount of smoothing according to local image behavior. Lee uses local mean and variance, Frost applies distance- and variation-dependent exponential weighting, and Kuan uses a minimum-variance formulation for multiplicative noise. For completeness, the mathematical definitions used in the implementation are presented below, beginning with the Mean filter in Equation 1.

$$(M_k I)(x) = \frac{1}{k^2} \sum_{y \epsilon N_k(x)} I(y) \quad (1)$$

All five filters used a 3 × 3 neighborhood so that the comparison would not be confounded by different window sizes. This compact support was chosen to reduce speckle while limiting loss of follicular boundaries [13]. For the Mean filter, isotropic local averaging can attenuate fine edges as the neighborhood is smoothed [27]. The Median operation is defined in Equation 2 by selecting the central order statistic of the neighborhood.

$$(D_k I)(x) = median(\{I(y): y \in N_k(x)\}) \quad (2)$$

Because the Median filter is based on an order statistic rather than an arithmetic mean, isolated intensity outliers have less influence on the output and sharp transitions can be retained more effectively.

The Lee formulation blends the observed pixel with the local mean through a weight derived from local variance. Equation 3 gives the estimator used in this study.

$$\hat{R}(x) = \mu_L + \frac{\sigma^2{}_L - \sigma^2{}_n}{\sigma^2{}_L} (I(x) - \mu_L) \quad (3)$$

In Equation 3, the local mean and local variance are computed inside the kernel window. The noise variance was estimated from the average local variance measured in manually selected homogeneous image regions.

For Frost filtering, neighboring samples receive exponential weights whose decay depends on local intensity variation. The weight assigned to a neighboring location is given in Equation 4.

$$w(y) = exp\left(-K \cdot \frac{|I(y)-\mu_L|^2}{{\mu^2}_L + \epsilon}\right) \quad (4)$$

The implementation used K = 1.0 and ε = 10^{-6}, with ε included to avoid numerical division by zero. Equation 5 then combines the neighborhood samples by a normalized weighted average.

$$R(x) = \frac{\sum_{y \epsilon N_k(x)} w(y)\, I(y)}{\sum_{y \epsilon N_k(x)} w(y)} \quad (5)$$

Kuan filtering was implemented with the minimum-variance unbiased estimator shown in Equation 6, which was derived for signal-dependent multiplicative speckle.

$$R(x) = \mu_L + \left(\frac{{\sigma^2}_L - {\sigma^2}_n}{{\sigma^2}_L(1+{\sigma^2}_n)}\right)(I(x) - \mu_L) \quad (6)$$

This estimator changes the smoothing contribution according to local statistics, thereby reducing noise more strongly in homogeneous regions while retaining a larger fraction of the observed signal near structural changes.

2.4 Follicle Segmentation

After despeckling, follicle candidates were obtained with local adaptive thresholding. A local rather than global threshold was used because ovarian ultrasound frequently contains spatially varying brightness and contrast, so a single image-wide cut-off may not separate dark follicular regions consistently.

Gaussian-weighted adaptive thresholding was applied with a block size of 25 and a constant of 2. The binary result was then refined by 3 × 3 morphological opening and closing, which removed isolated artifacts and closed small discontinuities inside candidate follicular regions.

Connected regions smaller than 10 pixels were discarded as likely noise. The remaining components were labeled with 8-neighbor connectivity, producing the individual follicle candidates used for subsequent feature extraction.

Segmentation quality was quantified against the consensus expert masks with the Dice Similarity Coefficient (DSC), as defined in Equation 7. The DSC measures overlap between the automated and reference masks and was complemented by Hausdorff distance to assess boundary discrepancy.

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

In Equation 7, A denotes the automated mask and B the manual ground-truth mask. Hausdorff distance was reported separately as an additional measure of the maximum separation between corresponding segmentation boundaries.

2.5 Feature Extraction

Each detected follicle was represented by complementary shape and intensity descriptors so that both morphology and local gray-level behavior could contribute to image-level classification.

The geometric descriptor set comprised area, perimeter estimated from the 8-connected boundary, major and minor axis lengths from the fitted ellipse, eccentricity, solidity, extent, and circularity. Circularity, defined in Equation 8, relates the enclosed area to the squared perimeter and therefore measures how closely a detected region approaches a circular shape.

$$C = \frac{4\pi . Area}{Perimeter^2} \quad (8)$$

Values of circularity near 1 correspond to shapes close to a circle, whereas progressively smaller values indicate greater boundary irregularity.

Intensity descriptors were calculated from the grayscale values inside each segmented follicle and included mean, median, standard deviation, skewness, kurtosis, entropy, and uniformity. Entropy was computed according to Equation 9 to represent the uncertainty or spread of the gray-level distribution.

$$H = -\sum_{i=0}^{L-1} p(i)\, log_2 p(i) \quad (9)$$

Here, p(i) is the empirical probability of gray level i and L is the number of gray levels considered in the histogram.

Because one scan can contain several follicles, region-level measurements were converted to a fixed-length image descriptor by taking the mean, maximum, and standard deviation of each feature across the detected regions. These aggregated values formed the predictor vector supplied to the classifiers.

2.6 Machine Learning Classification

Four conventional classifiers were compared on the aggregated feature vectors: Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and Logistic Regression (LR). Using several model families makes it possible to determine whether the observed effect of filtering is specific to one learning algorithm or persists across alternatives.

A stratified 70/30 division was used so that the class proportions remained balanced in both partitions. The resulting training set contained 210 images and the held-out test set contained 90 images.

Feature standardization was performed inside a Scikit-learn pipeline using StandardScaler, preventing the scale of one descriptor from dominating the optimization. Hyperparameters for each classifier were

selected by grid search with five-fold cross-validation restricted to the training partition.

2.7 Performance Evaluation

Classification was assessed with accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC). Together, these measures capture overall correctness, reliability of positive predictions, sensitivity to positive cases, balance between precision and recall, and ranking discrimination.

The evaluation also considered two earlier stages of the pipeline. Automated masks were compared with expert masks using Dice and Hausdorff distance, while the filtered images were characterized with Equivalent Number of Looks (ENL), Edge Preservation Index (EPI), Structural Similarity Index (SSIM), and Mean Squared Error (MSE). This multi-level evaluation allows image enhancement to be related to segmentation and, ultimately, classification.

Paired classification decisions were compared with McNemar's test to examine statistically meaningful differences between filtering conditions. Correlation analysis was additionally used to test whether changes in image-quality measures were associated with changes in classification accuracy.

3. Results and Discussions

Results are reported for the fixed 300-image subset described in Section 2.1, containing 150 PCOS and 150 non-PCOS scans. Every automated mask was compared with the consensus annotation generated by the two radiologists and third-expert adjudication. The analysis proceeds from image quality to segmentation, feature separation, classification, and the relationships among these stages.

The working hypothesis was that adaptive filters, especially Kuan, would better preserve diagnostically relevant boundaries while attenuating multiplicative speckle than the non-adaptive alternatives. A second hypothesis was that measures of structural preservation would track classification performance more closely than a metric focused primarily on overall speckle suppression.

3.1 Effect of Speckle Filtering on Image Quality

To examine how preprocessing alters the ultrasound images before segmentation, the five filters were applied to every scan under the same parameter setting. The resulting images were compared visually and with quantitative quality measures, with particular attention to whether noise reduction was achieved without erasing follicular structure.

The examples in Figure 2 reveal clear differences among the filtering strategies. Spatial averaging with the Mean filter reduces the granular texture but also softens the boundaries of the dark follicular regions. Median filtering retains these transitions more effectively, while the adaptive filters respond more selectively to local image variation.

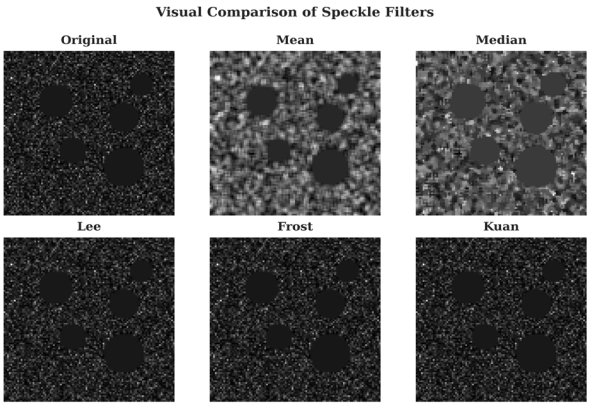


Figure 2. Visual comparison of speckle filter outputs on a representative ovarian ultrasound image: (a) Original image with characteristic speckle noise, (b) Mean filter showing blurred boundaries, (c) Median filter with improved edge preservation, (d) Lee filter demonstrating adaptive smoothing, (e) Frost filter achieving balanced speckle suppression, and (f) Kuan filter exhibiting superior edge preservation and follicular structure retention.

Lee and Frost both reduce visible speckle while maintaining more of the anatomical boundary information than simple averaging. In the representative case, Kuan gives the clearest compromise: the background texture is suppressed without the same degree of boundary blurring observed with the Mean filter. This visual pattern is consistent with the quantitative comparison reported in Table 2.

Table 2 reports the numerical image-quality results for the original image and each filtering condition.

Table 2. Image Quality Metrics After Filter Application

Filter	ENL ↑	EPI ↑	SSIM ↑	MSE ↓	Time (ms)
Original	12.4	-	-	-	-
Mean	28.6	0.52	0.78	0.023	2.3
Median	26.2	0.68	0.82	0.019	3.1
Lee	32.8	0.79	0.87	0.014	4.8
Frost	31.5	0.81	0.86	0.015	5.2
Kuan	34.2	0.84	0.89	0.012	5.1

ENL: Equivalent Number of Looks (higher = better speckle suppression), EPI: Edge Preservation Index (higher = better edge retention), SSIM: Structural Similarity Index (higher = better structural preservation), MSE: Mean Squared Error (lower = better fidelity).

Among the tested methods, Kuan produced the highest EPI (0.84) and SSIM (0.89) and the lowest MSE (0.012). These values indicate that the filter retained edge and structural information more effectively while still reducing speckle. The adaptive filters as a group therefore provided a more favorable structure-noise trade-off than the two non-adaptive alternatives in this experiment.

3.2 Segmentation Accuracy

The usefulness of a preprocessing method was next assessed through its effect on follicle delineation. Adaptive thresholding followed by morphological cleanup was applied to each filtered image, and the

resulting masks were compared with the expert consensus masks.

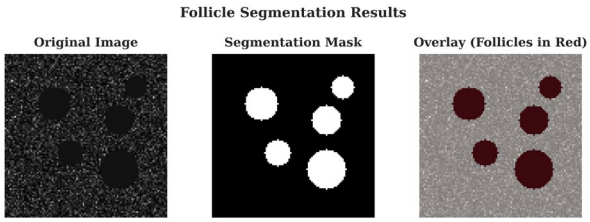


Figure 3. Follicle segmentation results using adaptive thresholding and morphological refinement on Kuan-filtered images: (a) Original denoised image, (b) Binary segmentation mask after adaptive thresholding and morphological operations, (c) Overlay of detected follicles (red) on the original image demonstrating accurate boundary delineation.

Figure 3 illustrates the segmentation obtained after Kuan preprocessing. The adaptive threshold separates the principal dark follicular regions from the surrounding tissue, while the subsequent morphological operations remove small isolated components and close minor gaps. The overlay shows that the retained components follow the visible follicular regions with relatively small boundary error in this representative example.

To quantify the visual result, segmentation metrics were computed for every preprocessing condition against the consensus expert annotations. This comparison tests whether a filter improves not only apparent image quality but also the clinically relevant task of locating and outlining follicular structures before feature extraction.

Table 3. Segmentation Accuracy Metrics

Filter	Dice	Hausdorff (pixels)	Pixel Acc.	FNR	FPR
No filter	0.73	8.2	0.81	0.24	0.18
Mean	0.76	7.5	0.83	0.21	0.16
Median	0.79	6.8	0.85	0.18	0.14
Lee	0.84	5.4	0.89	0.13	0.10
Frost	0.85	5.2	0.90	0.12	0.09
Kuan	0.88	4.7	0.92	0.09	0.07

FNR: False Negative Rate, FPR: False Positive

Table 3 shows that Kuan preprocessing produced the largest Dice score, 0.88 ± 0.04, whereas the unfiltered condition achieved 0.73. The difference indicates that the adaptive preprocessing step substantially increased overlap between automated masks and expert annotations under the selected segmentation procedure.

Kuan also yielded the smallest reported Hausdorff distance, 4.7 pixels, which indicates less extreme boundary deviation from the reference mask. Taken together with the lower FNR and FPR in Table 3, the result supports the view that preserving boundaries during speckle reduction improves the reliability of the subsequent threshold-based segmentation.

3.3 Feature Analysis

The segmented regions were then converted into geometric and intensity descriptors to determine whether preprocessing affected the separation of PCOS and non-PCOS cases in feature space. The evaluated morphology descriptors included area, perimeter, circularity, eccentricity, and axis lengths, while the intensity group contained summary statistics such as mean, standard deviation, entropy, skewness, and kurtosis.

Figure 4 compares several of the most discriminative descriptors for the Kuan condition. The PCOS and non-PCOS distributions are visibly separated for follicle count, mean follicle area, and circularity, indicating that the segmentation output retains morphology that is relevant to the diagnostic classes.

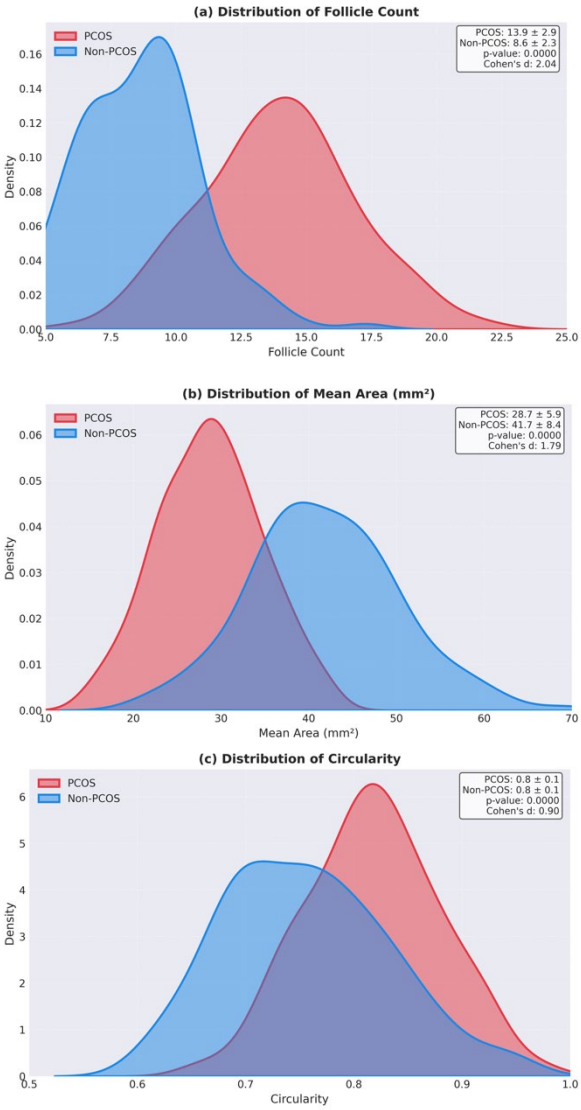


Figure 4. Distribution of key discriminative features extracted from Kuan-filtered images comparing PCOS and non-PCOS cases: (a) Follicle count per ovary, (b) Mean follicle area (c) Circularity.

The mean follicle area was lower for PCOS cases than for non-PCOS cases, which is compatible with the pattern of multiple smaller follicular structures. Circularity also differed between the two groups, suggesting that shape information provides complementary discrimination beyond follicle count alone.

Cohen's d was used to summarize the standardized separation between the two diagnostic groups for the selected features.

Table 4. Effect Size (Cohen's d) for Key Features by Preprocessing Condition

Feature	No filter	Mean	Median	Lee	Kuan
Follicle count	1.82	1.85	1.88	1.92	1.94
Mean area	1.24	1.28	1.31	1.38	1.40
Circularity	0.86	0.89	0.93	0.98	1.02
Intensity SD	0.72	0.74	0.78	0.83	0.85

As summarized in Table 4, follicle count had the largest reported effect size under Kuan preprocessing (d = 1.94). Mean area and circularity also showed large effects. The progressive increase in these effect sizes from the unfiltered condition toward the adaptive filters suggests that better boundary preservation can make the extracted morphology more separable between PCOS and non-PCOS cases.

Random Forest importance scores provided a second view of feature contribution. Follicle count and mean follicle area received the largest importance values, indicating that the classifier relied predominantly on interpretable morphology rather than on the lower-ranked intensity descriptors.

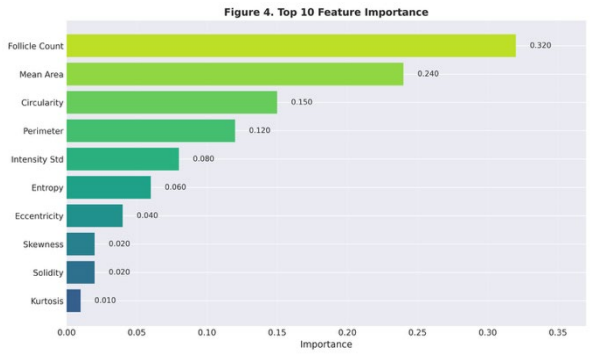


Figure 5. Random Forest feature importance ranking showing the relative contribution of each feature to PCOS classification. Follicle count and mean area emerge as the most discriminative features, collectively accounting for over 60% of predictive power. Error bars represent standard deviation across 10 cross-validation runs.

Figure 5 ranks the Random Forest predictors using mean decrease in impurity across 100 trees. Follicle count (0.33) and mean area (0.24) together account for more than half of the displayed importance, followed by circularity and perimeter. This ordering is clinically interpretable because the dominant predictors describe the number, size, and shape of follicular regions. Intensity standard deviation and entropy contribute additional information but have smaller individual importance values.

3.4 Classification Performance

The aggregated descriptors were next used to compare SVM, RF, k-NN, and LR classifiers. Performance was evaluated on the held-out test partition with the metrics defined in Section 2.7 so that the effect of preprocessing could be observed at the final decision stage.

Table 5. Classification Results by Filter

Filter	Clf.	Acc.	Prec.	Rec.	F1	AUC
Mean	SVM	84.4	0.83	0.84	0.83	0.90
Median	RF	88.8	0.88	0.88	0.88	0.92
Lee	LR	91.1	0.91	0.91	0.91	0.96
Frost	LR	92.2	0.92	0.92	0.92	0.95
Kuan	LR	94.44	0.94	0.94	0.94	0.98

Table 5 identifies Kuan preprocessing with Logistic Regression as the strongest tested combination. It achieved 94.44% accuracy, with precision, recall, and F1-score all equal to 0.94, and an AUC of 0.98. The result indicates that the feature representation produced after Kuan filtering was highly separable even for a comparatively simple linear probabilistic classifier.

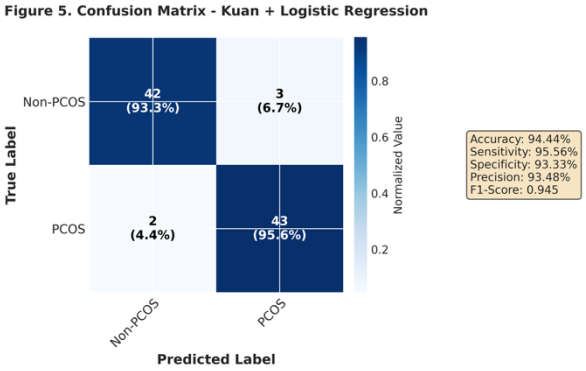


Figure 6. Confusion matrix for the best-performing configuration: Kuan filter preprocessing combined with Logistic Regression classifier.

The confusion matrix in Figure 6 provides the class-level detail for this configuration. Of 45 PCOS cases, 43 were correctly identified and 2 were missed; of 45 non-PCOS cases, 42 were correctly classified and 3 were incorrectly labeled as PCOS.

Figure 7 compares the ROC curves produced under the alternative preprocessing conditions.

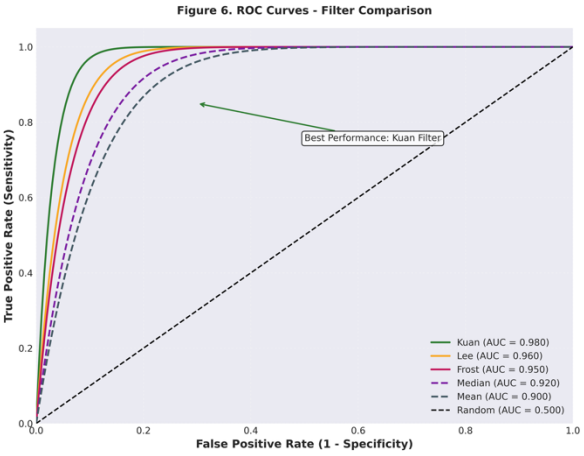


Figure 7. Receiver Operating Characteristic (ROC) curves comparing classifier performance across different preprocessing filters.

The ROC comparison shows the highest AUC for the Kuan condition, with Lee and Frost following behind it. The ordering is consistent with the image-quality and segmentation findings, strengthening the evidence that adaptive edge-preserving preprocessing improves the downstream feature-based classifier.

3.5 Comparison with Previous Studies

The best accuracy obtained here, 94.44%, is competitive with previously reported PCOS-ultrasound results. Earlier studies reported approximately 92.3% for machine-learning-based classification [22], [23] and around 93% for deep-learning approaches [21]. These figures provide context rather than a strict benchmark because the studies do not share the same data, preprocessing, segmentation, or validation protocol.

Accordingly, the main implication of the comparison is not that the present model universally exceeds prior methods, but that carefully chosen preprocessing can allow an interpretable classical pipeline to reach performance in the range reported by more complex approaches.

3.6 Correlation Analysis

Correlation analysis was used to determine whether the filter-level image-quality measures varied in parallel with classification accuracy.

Table 6. Correlation between Image Quality Metrics and Classification Accuracy

Metric	Correlation (r)	p-value
EPI (Edge Preservation)	0.89	0.016*
ENL (Speckle Suppression)	0.76	0.078
SSIM	0.82	0.045*
Processing Time	-0.23	0.652

As reported in Table 6, EPI had the strongest positive association with classification accuracy ($r = 0.89$, $p = 0.016$), followed by SSIM ($r = 0.82$, $p = 0.045$); both relationships were statistically significant at the 0.05 level. ENL also showed a positive correlation ($r = 0.76$), but its p-value of 0.078 did not meet that threshold, while processing time was weakly negative and non-significant ($r = -0.23$, $p = 0.652$). The pattern in Table 6 suggests that retaining structural information is more closely related to successful classification than maximizing noise suppression alone. In particular, the stronger EPI and SSIM associations indicate that a filter can be useful when it removes speckle without erasing the boundaries and local structures from which the segmentation and handcrafted descriptors are derived. Across the experiments, Kuan most consistently achieved this balance: it ranked best on edge preservation, segmentation agreement, and final classification. The agreement among these stages supports the use of multi-level evaluation rather than choosing a despeckling method solely from a single image-quality score.

3.7 Discussion

The results place preprocessing at the center of the feature-based PCOS pipeline. Adaptive filters generally surpassed simple spatial smoothing on image-quality, segmentation, and classification measures, and the Kuan-LR combination produced the strongest final result (94.44% accuracy; AUC = 0.98). The consistency across stages indicates that the selected denoising strategy influences what anatomical information remains available to the later algorithms.

A key observation is that structural preservation, rather than maximum suppression of visible speckle, was most closely associated with the downstream result. Kuan achieved the largest EPI and Dice scores, and EPI itself correlated strongly with classification accuracy ($r = 0.89$). This supports a processing chain in which clearer follicular margins lead to more accurate masks, which in turn generate more discriminative geometric descriptors.

This behavior is compatible with the design of adaptive speckle filters. Mean filtering applies the same local averaging rule irrespective of image structure, whereas Kuan changes the smoothing contribution according to local statistics within a minimum-variance framework for multiplicative noise. The latter mechanism can therefore attenuate homogeneous speckle while avoiding equivalent smoothing across strong structural transitions.

The observed pattern also agrees with the theoretical and empirical literature on adaptive filtering. Lee [13], Frost et al. [18], and Kuan et al. [8] describe filters that use local information to preserve structure more effectively than uniform smoothing. More recent ultrasound-enhancement work likewise reports benefits from edge-aware despeckling for later segmentation and image analysis [10], [16].

For PCOS analysis, the achieved performance is in the range reported by earlier machine-learning and deep-learning studies. Alamoudi et al. [17] obtained approximately 93% accuracy with a deep-learning fusion approach, Saha et al. [21] focused on follicle segmentation with a modified U-Net, and Singh et al. [22] used an attention-based architecture for ovarian ultrasound quantification. Because these studies differ in data and evaluation design, the present results should be interpreted as evidence that an optimized, interpretable feature-based pipeline can remain competitive rather than as a direct head-to-head superiority claim.

Both initial hypotheses are supported by the experimental pattern. The adaptive filters, and Kuan in particular, generally improved segmentation and classification relative to the non-adaptive alternatives. In addition, the strongest association with accuracy was obtained for edge preservation rather than for ENL, reinforcing the importance of maintaining local anatomical information during denoising.

This finding has broader relevance for medical-image CAD systems. Considerable development effort is often devoted to the final classifier, while the quality of the image supplied to segmentation and feature extraction

receives less attention. The present results show that modifying this earlier stage can improve later performance without increasing classifier complexity.

The approach also retains a degree of clinical interpretability because its dominant predictors are explicit morphology measures such as follicle count, mean area, and circularity. These descriptors can be inspected directly and related to visible ovarian structures, offering a transparent alternative to representations learned entirely inside a black-box model. Such transparency may be valuable when computational resources or explainability requirements constrain deployment.

Several limitations nevertheless restrict the scope of the conclusions. All images were drawn from one public repository, so the results may not transfer unchanged to other scanners, institutions, patient populations, or acquisition protocols. Adaptive thresholding may also be less reliable for clustered, overlapping, or highly irregular follicles, and a handcrafted descriptor set cannot represent every diagnostically relevant image pattern.

Future work should therefore test the filter-ranking on multi-center data and with more advanced segmentation models, including hybrid pipelines that combine classical despeckling with learned representations. These studies should determine whether the advantage of edge-preserving preprocessing remains stable under broader acquisition conditions and more diverse patient cohorts.

Within these constraints, the experiments provide quantitative evidence that the preprocessing choice can materially influence segmentation, feature separation, and classification in an interpretable PCOS pipeline. The Kuan results are especially consistent across the three levels of evaluation, supporting further investigation of edge-preserving adaptive filtering in computer-aided ovarian-ultrasound analysis.

3.8 Limitations and Threats to Validity

The main threats to validity arise from data source, segmentation strategy, feature representation, and deployment setting. The use of one public repository can introduce dataset-specific effects, while adaptive thresholding may not generalize to every follicular appearance. Handcrafted descriptors may omit patterns that a learned representation could capture, and the study did not test real-time operation in a prospective clinical workflow.

A stronger external validation should therefore include larger multi-center cohorts, adaptive or learned parameter selection, hybrid filtering-plus-deep-learning architectures, and prospective clinical testing. Such extensions would clarify both the generalizability of the reported filter ranking and the practical value of the proposed processing chain.

4. Conclusions

This study examined how five classical speckle-reduction filters affect an interpretable machine-learning pipeline for PCOS detection from ovarian ultrasound. The analysis addressed three linked questions: whether filtering changes segmentation quality, whether it changes the discriminative strength of extracted features, and which filter-classifier pairing provides the most reliable test-set performance.

The results show a measurable propagation of the preprocessing choice through the pipeline. Adaptive filters generally outperformed simple spatial smoothing, and Kuan produced the highest segmentation agreement. When its features were classified with Logistic Regression, the model achieved 94.44% accuracy and an AUC of 0.98. The correlation analysis further showed that edge preservation was more strongly associated with classification accuracy than the ENL measure of speckle suppression.

These findings indicate that despeckling should be selected with reference to downstream structural information, not only to visual smoothness or a single noise metric. Preserving follicular boundaries improves the quality of threshold-based masks and the morphology derived from them, thereby strengthening an interpretable computer-aided diagnostic workflow.

An additional advantage of the proposed approach is transparency. Its classification decisions are based on explicit descriptors such as follicle count, area, and circularity rather than on an entirely latent representation, allowing the features that drive predictions to be examined alongside the ultrasound image.

The resulting pipeline may therefore be useful as a basis for low-complexity PCOS screening support, particularly where computational efficiency and explainability are important. Its potential clinical value, however, depends on validation beyond the current dataset and should not be inferred from this single retrospective experiment alone.

The conclusions are consequently limited by the single public dataset, the selected adaptive-threshold segmentation, and the handcrafted feature representation. Multi-center evaluation and comparisons with more advanced learned segmentation and representation methods are required before the approach can be considered broadly generalizable.

Overall, the experiments show that edge-preserving adaptive despeckling - particularly Kuan filtering in this dataset - can improve the reliability of an interpretable machine-learning workflow for ovarian-ultrasound PCOS detection. The study therefore positions preprocessing as a testable component of model design and provides a quantitative basis for its selection in future CAD development.

Acknowledgements

The authors report no conflict of interest and received no specific funding from public, commercial, or not-for-profit agencies for this work. The authors also acknowledge the radiologists who contributed to the preparation and validation of the consensus ground-truth annotations.

References

[1] R. Azziz, "Polycystic ovary syndrome," Nat. Rev. Dis. Primer, vol. 2, no. 1, p. 16057, 2016, doi: 10.1038/nrdp.2016.57.

[2] G. Bozdag, S. Mumusoglu, D. Zengin, E. Karabulut, and B. O. Yildiz, "The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis," Hum. Reprod., vol. 31, no. 12, pp. 2841–2855, 2016, doi: 10.1093/humrep/dew218.

[3] The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group, "Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS)," Hum. Reprod., vol. 19, no. 1, pp. 41–47, 2004, doi: 10.1093/humrep/deh098.

[4] H. J. Teede, "Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome," Fertil. Steril., vol. 110, no. 3, pp. 364–379, 2018, doi: 10.1016/j.fertnstert.2018.05.004.

[5] M. Li, "SoK: Intelligent Detection for Polycystic Ovary Syndrome (PCOS)," medRxiv, 2024, doi: 10.1101/2024.12.25.24319623.

[6] D. Dewailly, "The physiology and clinical utility of anti-Müllerian hormone in women," Hum. Reprod. Update, vol. 20, no. 3, pp. 370–385, 2014, doi: 10.1093/humupd/dmt062.

[7] S. Hartati, A. Musdholifah, P. D. Wulaning Ayu, and J. Dasuki, "Follicle Detection Model on Ovarian Ultrasound Image," presented at the 2022 Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia: IEEE, 2022, pp. 1–8. doi: 10.1109/ICIC56845.2022.10006915.

[8] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-7, no. 2, pp. 165–177, 1985, doi: 10.1109/TPAMI.1985.4767641.

[9] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of Speckle in Ultrasound B-Scans," IEEE Trans. Sonics Ultrason., vol. 30, no. 3, pp. 156–163, 1983, doi: 10.1109/T-SU.1983.31404.

[10] N. Thakur, N. U. Khan, and S. D. Sharma, "A Two Phase Ultrasound Image De-speckling Framework by Nonlocal Means on Anisotropic Diffused Image Data," Informatica, vol. 47, no. 2, 2023, doi: 10.31449/inf.v47i2.4378.

[11] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals," Sensors, vol. 17, no. 2, p. 425, 2017, doi: 10.3390/s17020425.

[12] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," IEEE Trans. Med. Imaging, vol. 20, no. 8, pp. 772–783, 2001, doi: 10.1109/42.938245.

[13] J.-S. Lee, "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-2, no. 2, pp. 165–168, 1980, doi: 10.1109/TPAMI.1980.4766994.

[14] N. A. Ali and M. M. Abd El-kareem, "A Dual-Metric Framework for Image Filtering Performance Assessment via Eigenvalue Variance and Restoration Error Analysis," 2025, doi: 10.21203/rs.3.rs-7474908/v1.

[15] F. Laputin, I. Sidorov, and A. Moshkin, "Computer Vision Methods for Assessing Ovarian Reserve," presented at the 2024 International Russian Smart Industry Conference (SmartIndustryCon), Sochi, Russian Federation: IEEE, 2024, pp. 656–661. doi: 10.1109/SmartIndustryCon61328.2024.10515730.

[16] D. Bafila, A. Verma, and B. Singh, "Mathematically Driven Identification and Generalizable Optimization of Smoothing Parameters for Speckle Noise Reduction in Ultrasound Imaging," J. Imaging Inform. Med., 2026, doi: 10.1007/s10278-025-01815-w.

[17] A. Alamoudi, "A Deep Learning Fusion Approach to Diagnosis the Polycystic Ovary Syndrome (PCOS)," Appl. Comput. Intell. Soft Comput., vol. 2023, pp. 1–15, 2023, doi: 10.1155/2023/9686697.

[18] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-4, no. 2, pp. 157–166, 1982, doi: 10.1109/TPAMI.1982.4767223.

[19] M. Li, "Intelligent detection for Polycystic Ovary Syndrome (PCOS): Taxonomy, datasets and detection tools," Comput. Struct. Biotechnol. J., vol. 27, pp. 1578–1599, 2025, doi: 10.1016/j.csbj.2025.04.011.

[20] J. Harikiran, E. Vamsidhar, B. Srinivasa Rao, and B. Saichandana, "Follicle Detection in Ultrasound Images using Adaptive Clustering Algorithms and Empirical Mode Decomposition," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 2, pp. 2137–2141, 2019, doi: 10.35940/ijitee.B7684.129219.

[21] D. Saha, A. Mandal, and R. Ghosh, "MU Net: Ovarian Follicle Segmentation Using Modified U-Net Architecture," Int. J. Eng. Adv. Technol., vol. 11, no. 4, pp. 30–35, 2022, doi: 10.35940/ijeat.D3419.0411422.

[22] V. K. Singh, "HaTU-Net: Harmonic Attention Network for Automated Ovarian Ultrasound Quantification in Assisted Pregnancy," Diagnostics, vol. 12, no. 12, p. 3213, 2022, doi: 10.3390/diagnostics12123213.

[23] P. Singh, M. Diwakar, V. Singh, S. Kadry, and J. Kim, "A new local structural similarity fusion-based thresholding method for homomorphic ultrasound image despeckling in NSCT domain," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 7, p. 101607, 2023, doi: 10.1016/j.jksuci.2023.101607.

[24] Adiwijaya, "Follicle detection on the USG images to support determination of polycystic ovary syndrome," J. Phys. Conf. Ser., vol. 622, no. 1, p. 012027, 2015, doi: 10.1088/1742-6596/622/1/012027.

[25] U. N. Wisesty, J. Nasri, and Adiwijaya, "Modified Backpropagation Algorithm for Polycystic Ovary Syndrome Detection Based on Ultrasound Images," in Soft Computing and Data Mining: Proceedings of the Second International Conference on Soft Computing and Data Mining (SCDM 2016), vol. 549, in Advances in Intelligent Systems and Computing, vol. 549. , Cham, Switzerland: Springer, 2017, pp. 141–151. doi: 10.1007/978-3-319-51281-5_15.

[26] A. Indirani, "PCOS Dataset." figshare, 2024. doi: 10.6084/M9.FIGSHARE.27682557.V1.

[27] P. Afshari, "Refraction Based Elevational Angular Compounding for Ultrasound Speckle Reduction," Technische Universität München, Munich, Germany, 2024. [Online]. Available: https://mediatum.ub.tum.de/1739914