Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 843-860 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

HAREN: A Hybrid Attention Residual Ensemble Network for PCOS classification and Prediction

Pragati Patil[1], and Nandini Chaudhari[2]

[1] Faculty of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, INDIA.
[2] Department of Computer Science and Engineering, KSET, Drs. Kiran and Pallavi Patel Global University, Gujarat, INDIA

Corresponding author: Pragati Patil (email: phdscholar21010@kpgu.ac.in, **Author(s) Email**: Pragati Patil (email: phdscholar21010@kpgu.ac.in, Nandini Chaudhari (email: dir.kset@kpgu.ac.in)

Abstract **Polycystic Ovary Syndrome (PCOS) is one of the most prevalent endocrine disorders affecting women of reproductive age and is a leading cause of infertility. Ultrasound imaging is widely used for PCOS diagnosis; however, visual assessment of ovarian morphology is highly subjective, time-consuming, and dependent on clinical expertise. Quality differences in ultrasound images, very near to similar visual patterns among PCOS and NOT PCOS images, and noise in the images increase the threat of improper diagnosis. These problems suggest a need for an accurate, automatic, and computer-assisted PCOS diagnostic system. This research aims to create a deep learning-assisted automatic PCOS diagnostic system which can detect and classify the Polycystic Ovary Syndrome from the gray-scale ultrasound ovarian images. In addition to high classification accuracy, the proposed framework incorporates an explicit explainability pipeline that highlights diagnostically relevant ovarian regions, such as follicular distributions and stromal patterns, thereby supporting clinically interpretable decision making. The proposed HAREN framework addresses the limitations of single backbone models, and attention augmented variants, such as vanilla ResNet50 and ResNet50 with hybrid attention by leveraging ensemble learning and residual feature fusion. HAREN combines three architecturally diverse and complementary pretrained CNN backbones (ResNet50, DenseNet121, and EfficientNetB0) to enhance feature diversity. In addition, a novel hybrid attention mechanism combining channel, spatial, and cross-scale attention is introduced to emphasize diagnostically relevant ovarian regions. A residual fusion strategy is employed to preserve discriminative features and stabilize training, and an explicit explainability pipeline is incorporated to support Grad CAM-based visual interpretation. This network first converts the ultrasound grayscale ovarian images to RGB , followed by the extraction of important features applying backbones, which are augmented with attention mechanisms. The network, trained with categorical crossentropy loss, was evaluated using comprehensive performance metrics on 11,784 ultrasound images (6,784 PCOS and 5,000 NOT PCOS). HAREN achieved 99.33% accuracy, 98.96% precision, 98.97% recall, 98.96% F1 score, and an AUC of 99.93%, outperforming conventional models. Overall, it delivers an accurate, reliable, and interpretable solution for automated PCOS detection, demonstrating strong potential for clinical decision support systems.**

Keywords **Polycystic Ovary Syndrome (PCOS); Hybrid Attention; Ensemble Learning; Explainable AI; Deep Learning; Ultrasound Imaging.**

I. Introduction

The Polycystic Ovary Syndrome (PCOS) is one of the most common endocrine disorders affecting women of reproductive age. It is characterized by ovarian dysfunction, hyperandrogenism, and polycystic ovarian morphology, often leading to infertility as well as metabolic and cardiovascular complications. Epidemiological studies report that approximately 6%–20% of premenopausal women are affected by PCOS [1][2]. Pelvic and transvaginal ultrasound imaging is the primary clinical modality used to assess ovarian morphology. However, the interpretation of grayscale ultrasound images is highly operator-dependent and affected by speckle noise, low contrast, and heterogeneous image quality, which may result in inconsistent and subjective diagnoses [3]. These limitations emphasize the need for automated and reliable PCOS detection systems.

Recent advances in machine learning (ML) and deep learning (DL) models have significantly enhanced automated disease detection in medical imaging. Specifically, convolutional neural networks have shown very robust performance in prediction, classification, and feature extraction from grayscale ultrasound images. Kumar et al. [4] and Kang et al. [5] reported that CNN-based classifiers outperform traditional texture-based

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

approaches. Similarly, Zhang et al. [6] and Lingamaiah et al. [7] showed that deep learning models handle variations in grayscale ultrasound images more effectively than conventional CNN pipelines.

In spite of these enhancements, regular CNNs still encounter problems like imprecise localization of ovarian features, sensitivity to noise present in the images, artifacts, and limited interpretability. Some research [8][9] have outlined that convolution neural network (CNN) models frequently work as black box systems, which restricts their acceptance in medical practice. Attention mechanisms are introduced to tackle the above challenges produced by regular CNNs. This mechanism enhances the capacity of the CNNs to emphasize very relevant regions. Different attention mechanisms like cross-scale attention, spatial attention, and channel attention are also used and applied in medical imaging to emphasize salient anatomical structures by suppressing non-useful information. Fan et al. [10] demonstrated that feature hybrid attention improves lesion localization in ultrasound images. Zhao et al. [11] and Akindele et al. [12] enhanced the performance of the CNN model in low-quality grayscale imaging by reducing noise through the application of multiscale attention. Authors in [13][14] have integrated attention mechanisms with CNNs and UNET architectures to enhance follicle detection and ovarian segmentation, which leads to very accurate diagnosis. These works show that attention mechanisms can play a big role in enhancing the ability of the analysis and diagnosis from gray scale ultrasound images for PCOS and other ailments.

Another significant way is creating ensemble architectures by integrating some complementary deep learning architectures. This is known as ensemble learning. Ensemble learning improves robustness and generalization by using their complimentary characteristics. Authors in [15][16][17] have introduced ensemble variants of EfficientNet, VGG, DenseNet, and ResNet for the classification of the gray scale ultrasound medical images. They also achieved significant improvement in the performance. A multimodal ensemble variant is introduced by Pawar et al. [18]. They have detected PCOS by working on ultrasound images, and they have also achieved good, strong results. Approaches such as residual learning and hybrid attention have also been used to improve generalization by reducing overfitting. It has been mainly used for medium-sized datasets [19][20].

Silambarasan et al. [21] introduced an ensemble hybrid CNN–SVM model for classification and detection of PCOS from the ultrasound images. They also achieved a good improvement in classification performance. A lightweight ensemble model is introduced by Kaur et al. [22] for PCOS classification in ovarian ultrasound images and achieved encouraging results.

Another important aspect, such as interpretability, is also very important for any model to adopt it in clinical procedures. It improves trustworthiness in the patients. In this context, explainable AI (XAI) is incorporated to provide medically meaningful insights rather than solely predictive outputs. Specifically, the explainability module highlights diagnostically relevant ovarian regions such as follicular clusters, stromal echogenicity, and morphological patterns, that radiologists commonly assess PCOS evaluation. These visual explanations are intended to support clinicians in confirming automated predictions and improving diagnostic confidence [23][24][25].

However, only a small number of studies have incorporated hybrid attention-based ensemble networks and XAI techniques in PCOS ultrasound classification, indicating a clear gap between model performance and clinical usability [26]. The above challenges are further aggravated by methodological limitations in existing learning frameworks. Limited feature diversity in single backbone models restricts the representation of subtle morphological variations between PCOS and NOT PCOS ovaries, particularly when visual patterns appear highly similar under noisy imaging conditions. Similarly, the isolated or partial use of attention mechanisms often enhances only a single aspect of the feature space (e.g., spatial or channel-wise), which limits robustness against speckle noise and heterogeneous follicular distributions. These constraints collectively reduce discriminative capability in real-world ultrasound settings and motivate the need for a hybrid, ensemble-based attention framework.

To tackle these challenges, this research work introduces a Hybrid Attention Residual Ensemble Network (HAREN) for automatic PCOS classification and detection in grayscale ultrasound ovarian images. The introduced framework integrates three complementary pretrained backbones, EfficientNetB0, DenseNet121, and ResNet50, boosted with hybrid attention modules that integrate cross-scale, spatial, and channel. Residual fusion connections are also employed to preserve discriminative features and stabilize learning, while XAI techniques are applied to deliver clinically significant visual explanations. This work mainly aims to create a deep learning framework for robust, accurate, interpretable, and reliable PCOS classification. The key contributions of this paper are summarized as follows:

1. A hybrid attention-enhanced CNN framework tailored for grayscale ovarian ultrasound images.
2. A residual ensemble fusion strategy combining ResNet50, DenseNet121, and EfficientNetB0 to improve feature diversity and classification performance.

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

3. Integration of multiple XAI techniques to explore clinically relevant ovarian regions for improving interpretability and trust in the patients.

The rest of the paper is structured as follows: Section 2 discusses the introduced HAREN framework and research methodology. Section 3 discusses the dataset, data preprocessing, experimental setup, evaluation metrics, and results. Section 4 discusses the research outcomes in detail. Finally, Section 5 concludes the paper and outlines future research directions.

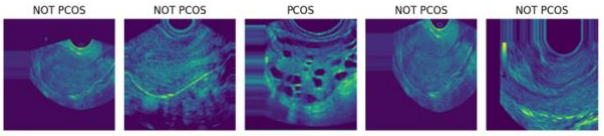


Fig. 1. Sample Dataset Images

II. Method

A. Dataset

The proposed HAREN model is evaluated using a publicly available ovarian ultrasound image dataset obtained from Kaggle. The dataset consists entirely of grayscale ultrasound images, which reflect real clinical imaging conditions [27]. The dataset consists of a total of 11,784 grayscale ovarian ultrasound images divided in two classes: images infected with PCOS (6784) and images not infected with PCOS (5000). Although the dataset contains 11,784 ultrasound images, which is relatively large compared to many existing PCOS studies, it does not provide patient-level identifiers. Consequently, strict patient wise data separation cannot be guaranteed, raising the possibility of correlated samples across training and testing splits. Additionally, the dataset may not fully capture population-level diversity in terms of ultrasound acquisition protocols and demographic variability. These factors are acknowledged as potential limitations affecting generalization. Fig. 1 shows some sample images.

To ensure consistency in the input needs of the deep learning models used, all images in the dataset are resized to 224 × 224 pixels. The intensity values of pixels in all images are normalized to the [0, 1] range. This normalization helps stabilize training and improves the network's convergence.

B. Data preprocessing

Since the selected pretrained CNN backbones are originally designed to process three-channel RGB images, grayscale ultrasound images are transformed into a three-channel representation using a learnable 1×1 convolutional layer. Unlike naive channel duplication, this approach allows the network to learn an optimal mapping from single-channel intensity information to a multichannel feature space, thereby enhancing feature expressiveness and improving adaptation to pretrained ImageNet weights.

The conversion is defined as Eq. (1) [34]:

$$I_{rgb} = Conv_{1\times1}(I_{gray}), \quad I_{gray} \in \mathbb{R}^{224\times224\times1} \tag{1}$$

In Eq. (1) [34], $I_{gray} \in \mathbb{R}^{224\times224\times1}$ denotes the grayscale image, where 224×224 represents the spatial resolution, and 1 indicates a single intensity channel. The operation $\mathrm{Conv}_{1\times1}(\cdot)$ refers to a learnable pointwise convolution. The term $I_{rgb} \in \mathbb{R}^{224\times224\times3}$ denotes the resulting pseudo-RGB image. Data augmentation is applied during the training process of the model to reduce overfitting and increase the robustness of the model. We performed random rotations of the images within ±15°, vertical and horizontal flipping, and random zooming up to 10%. These processes help the introduced model learn invariant features across different imaging conditions and improve generalization performance.

C. Proposed Methodology

This section presents the detailed methodology of the introduced Hybrid Attention Residual Ensemble Network (HAREN).

1. Overall Framework Description

The introduced HAREN network follows an ensemble-based classification strategy. The framework contains three key stages: conversion from grayscale images to RGB for preprocessing, application of attention-enhanced CNN backbones and residual fusing for parallel feature extraction, which is followed by classification. We also integrate explainable AI methods to improve clinical interpretability and trustworthiness.

Initially, each grayscale ultrasound image is passed through a preprocessing pipeline to standardize image size and intensity distribution. The processed images are then simultaneously fed into three different pretrained CNN backbones, such as ResNet50, DenseNet121, and EfficientNetB0, which are augmented with a hybrid attention module. The extracted features from all backbones are fused using a residual ensemble strategy and passed to a fully connected classification layer to predict the PCOS class.

2. Selection of Pretrained Backbone Networks

Three pretrained convolutional neural networks, EfficientNetB0, DenseNet121, and ResNet50, are selected as backbone feature extractors due to their established efficacy in different types of medical image analysis and their balanced learning characteristics. The term “balancing” refers to the complementary integration of pretrained CNN backbones with diverse architectural properties: ResNet50 for deep residual learning, DenseNet121 for feature reuse and gradient flow, and EfficientNetB0 for parameter-efficient representation,

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

thereby achieving a balance between feature richness, computational efficiency, and architectural diversity.

All backbone networks are initialized with ImageNet pretrained weights and subsequently fine-tuned on the PCOS ultrasound dataset. During fine-tuning, all layers are unfrozen to enable domain-specific adaptation, with a reduced learning rate to preserve pretrained representations while allowing effective feature refinement. This strategy ensures a balanced tradeoff between transfer learning stability and task-specific optimization.

3. Hybrid Attention Mechanism

Ultrasound medical images frequently have complex background patterns and speckle noise. This can generate ambiguous ovarian structures. To tackle this challenge, we have integrated a Hybrid Attention Block after the convolutional layers of each backbone network. This hybrid attention block sequentially integrates cross-scale attention, spatial attention, and channel attention to fine-tune feature representations.

The sequential ordering of the hybrid attention mechanism was intentionally designed to progressively refine feature representations. Cross-scale attention is applied first to integrate multiresolution contextual information, which is crucial for capturing follicles of varying sizes. Spatial attention is subsequently employed to localize anatomically relevant regions within the ovary. Finally, channel attention recalibrates feature responses by emphasizing diagnostically informative channels. This hierarchical refinement strategy was found to be more stable than alternative permutations during preliminary experimentation and aligns with the clinical process of coarse-to-fine visual assessment.

4. Residual Learning Strategy

Residual connections are employed at two distinct stages within HAREN. First, residual learning is applied within each attention-augmented backbone to preserve the original feature representations while enhancing discriminative information. Second, a residual fusion strategy is introduced during ensemble feature aggregation, where fused features are combined with backbone-specific representations via skip connections. This dual-level residual design stabilizes gradient propagation and mitigates information loss during ensemble integration.

5. Ensemble Feature Fusion and Classification

Each backbone network generates a high-level feature vector derived from the global average pooling of the last convolutional feature maps. These feature vectors encapsulate a variety of complementary information acquired by distinct architectures. The vectors are combined to create a cohesive ensemble representation that encompasses more comprehensive and resilient feature information than an individual backbone model.

6. Dropout regularization

Dropout regularization is applied to the fused feature vector before classification, and the final classification layer is a fully connected dense layer with a softmax activation function to output probabilities of the two classes, PCOS infected and PCOS not infected, which is trained using categorical cross-entropy loss.

7. Training Strategy and Optimization

The HAREN model is trained using the Adam optimizer with a learning rate of 0.001, and L2 regularization is applied to both the convolutional and dense layers to combat overfitting with a dropout rate of 0.3 in the fusion layer. The input image is 224 × 224 pixels with a single grayscale channel. The model is evaluated using multiple metrics, such as accuracy, precision, recall, F1 score, and AUC.

8. Explainable AI for Model Interpretation

For clinical reliability, explainability is an essential component of the proposed framework. Several XAI techniques, including Grad CAM, Grad CAM++, Layer CAM, Saliency Maps, and LIME, are used to visualize the regions that influence model predictions. Backbone-specific convolutional feature maps are explicitly extracted to avoid ambiguity in ensemble explainability. These visualizations consistently highlight clinically relevant ovarian regions, supporting the trustworthiness of the proposed system. Table. 1 depicts the summary of the HAREN model.

D. Mathematical Formulation of the HAREN

This section outlines the mathematical representation of the essential elements comprising the proposed Hybrid Attention Residual Ensemble Network (HAREN).

1. Input Representation and Grayscale to RGB Conversion using:

Let

$$I_{gray} \in \mathbb{R}^{224\times224\times1} \tag{2}$$

denote a normalized grayscale ovarian ultrasound image. Since the pretrained backbone networks are designed to accept three-channel RGB inputs, a 1×1 convolution is used to map the grayscale image to a three-channel representation:

$$I_{rgb} = Conv_{1\times1}(I_{gray}), \quad I_{rgb} \in \mathbb{R}^{224\times224\times3} \tag{3}$$

In Eq. (2) [34] and Eq. (3) [34], $I_{gray} \in \mathbb{R}^{224\times224\times1}$ denotes the grayscale image, where 224×224 represents the spatial resolution, and 1 indicates a single intensity channel. The operation $\mathrm{Conv}_{1\times1}(\cdot)$ refers to a learnable pointwise convolution. The term $I_{rgb} \in \mathbb{R}^{224\times224\times3}$ denotes the resulting pseudo-RGB image.

This operation preserves spatial information while enabling compatibility with ImageNet-pretrained CNNs.

2. Feature Extraction Using Pretrained Backbones

Let

$$\mathcal{B}_{\mathcal{K}}(\cdot), \qquad \mathcal{K} \in (1,2,3) \tag{4}$$

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

represents the pretrained backbone networks ResNet50, DenseNet121, and EfficientNetB0, respectively. For each backbone, deep feature maps are extracted as:

$$F_k = \mathcal{B}_{\mathcal{K}}(I_{rgb}), \quad F_k \in \mathbb{R}^{H \times W \times C} \tag{5}$$

In Eq. (4) [35] and Eq. (5) [35], $F_k \in \mathbb{R}^{H \times W \times C}$ represents the deep convolutional feature map extracted from the network. In this notation, H denotes the spatial height of the feature map, W represents the spatial width, and C corresponds to the total number of channels capturing the learned feature representations. These feature maps are then refined using the proposed hybrid attention mechanism.

Table. 1. The Proposed Model Summary

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 224, 224, 1)	0	–
resnet50_head (Functional)	[(None, 256), (None, 7, 7, 2048)]	121,843, 0xx*	input_layer[0][0]
densenet121_head (Functional)	[(None, 256), (None, 7, 7, 1024)]	31,785,927	input_layer[0][0]
efficientnetb0_head (Functional)	[(None, 256), (None, 7, 7, 1280)]	42,603,594	input_layer[0][0]
fused features (Concatenate)	(None, 768)	0	resnet50_head, densenet121_head, efficientnetb0_head
dropout (Dropout)	(None, 768)	0	fused features
classifier (Dense)	(None, 2)	1,538	Dropout
Total params: 196,234,138 (748.57 MB) Trainable params: 196,055,347 (747.89 MB) Non-trainable params: 178,791			

3. Channel Attention Mechanism

Channel attention emphasizes informative feature channels by modeling inter-channel dependencies. Given a feature map F, global average pooling is first applied:

$$Z = GAP(F) \tag{6}$$

The pooled vector is passed through two fully connected layers with ReLU and sigmoid activations:

$$S = \sigma(W_2 \cdot \delta(W_1 \cdot Z)) \tag{7}$$

The refined feature map is obtained by channel-wise multiplication:

$$F_{ch} = F \otimes s \tag{8}$$

In Eq. (6) [36], Eq. (7) [36], and Eq. (8) [36], $GAP(\cdot)$ denotes the global average pooling operation, and Z represents the aggregated channel description vector obtained from it. The symbols W_1 and W_2 denote learnable weight matrices, while δ refers to the ReLU activation function and σ represents the sigmoid function. The term s indicates the channel-wise attention vector, and $\otimes$ denotes element-wise multiplication.

4. Spatial Attention Mechanism

Spatial attention focuses on important spatial locations within the feature map. Average pooling and max pooling are applied along the channel dimension:

$$M_{avg} = AvgPool(F_{ch}), M_{max} = MaxPool(F_{ch}) \tag{9}$$

These maps are concatenated and convolved using a (7 x 7) kernel:

$$M_{sp} = \sigma\left(Conv_{7\times7}([M_{avg}; M_{max}])\right) \tag{10}$$

The spatially refined feature map is computed as:

$$F_{sp} = F_{ch} \otimes M_{sp} \tag{11}$$

In Eq. (9) [36], Eq. (10) [36], and Eq. (11) [36], $AvgPool(\cdot)$ refers to the average pooling operation, while $MaxPool(\cdot)$ represents the maximum pooling operation. The symbols M_{avg} and M_{max} denote the spatial descriptors generated from average pooling and max pooling, respectively. The notation $Conv_{7\times7}$ indicates a convolution operation with a 7×7 kernel size, and M_{sp} represents the resulting spatial attention map.

5. Cross-Scale Attention Mechanism

To capture multiscale contextual information, the spatially refined feature map is processed using convolutional filters of different receptive fields:

$$F_1 = Conv_{1\times1}(\mathrm{F_{sp}}),$$
$$F_2 = Conv_{3\times3}(F_{sp}),$$
$$F_3 = Conv_{3\times3}^{d=2}(F_{sp}) \tag{12}$$

These features are concatenated and fused:

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

$$F_{cs} = Conv_{1\times1}([F_1; F_2; F_3]) \quad (13)$$

A gating vector is computed using global average pooling and sigmoid activation:

$$g = \sigma(FC(GAP(F_{cs}))) \quad (14)$$

The final attention-refined feature map is obtained as:

$$F_{ha} = F_{cs} \otimes g \quad (15)$$

Algorithm 1: Training Pipeline of the Proposed HAREN Model for PCOS Detection

Input:

Grayscale ovarian ultrasound image dataset

$$\mathcal{D} = \{(I_i, y_i)\}_{i=1}^{N},$$

where $I_i \in \mathbb{R}^{224\times224\times1}$ and $y_i \in \{0, 1\}$ denotes nonPCOS and PCOS classes

Output:

Trained HAREN model $\mathcal{M}_{HAREN}$

Algorithm Steps

1. Initialize Parameters
   Initialize learning rate $\alpha = 0.001$,
   batch size B, number of epochs E,
   dropout rate $p = 0.3$, and L2 regularization coefficient $\lambda = 10^{-4}$.
2. Dataset Preprocessing
   Resize all images to 224×224
   Normalize pixel values to the range [0, 1].
   Apply data augmentation (rotation, flipping, zooming).
3. GrayscaletoRGB Conversion
   For each input image I_{gray} , generate
   $$\mathrm{I}_{rgb} = \mathrm{Conv}_{1\times1}(\mathrm{I}_{gray})$$
4. Initialize Backbone Networks
   Load pretrained ResNet50, DenseNet121, and EfficientNetB0 models with ImageNet weights.
   Remove fully connected layers from each backbone.
5. Hybrid Attention Feature Extraction
   For each backbone $\mathcal{B}_{\mathcal{K}}(\cdot)$:
   Extract convolutional features F_k
   Apply Channel Attention
   Apply Spatial Attention
   Apply CrossScale Attention
   Obtain refined feature maps F_{ha}^{k}
6. Global Feature Pooling
   Apply Global Average Pooling to each F_{ha}^{k} to obtain feature vectors F_k.
7. Ensemble Feature Fusion
   Concatenate feature vectors:
   F_{ens} = $[\mathrm{f}_{ResNet}; \mathrm{f}_{DenseNet}; \mathrm{f}_{EffNet}]$
   Apply dropout regularization.
8. Classification
   Pass F_{ens} through a fully connected layer with softmax activation to predict class probabilities.
9. Loss Computation
   Compute categorical crossentropy loss:
   $$\mathcal{L} = -\sum_{c=1}^{2} \mathrm{y}_c \log(\widehat{y_c})$$
10. Model Optimization
    Update network parameters using Adam optimizer and backpropagation.
11. Model Evaluation
    Compute Accuracy, Precision, Recall, F1Score, and AUC.
12. Explainability Analysis
    Generate GradCAM, GradCAM++, LayerCAM, LIME, and Saliency Maps using backbonespecific feature maps.
13. Save Model and Artifacts
    Store trained model weights, evaluation metrics, and XAI visualizations.

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

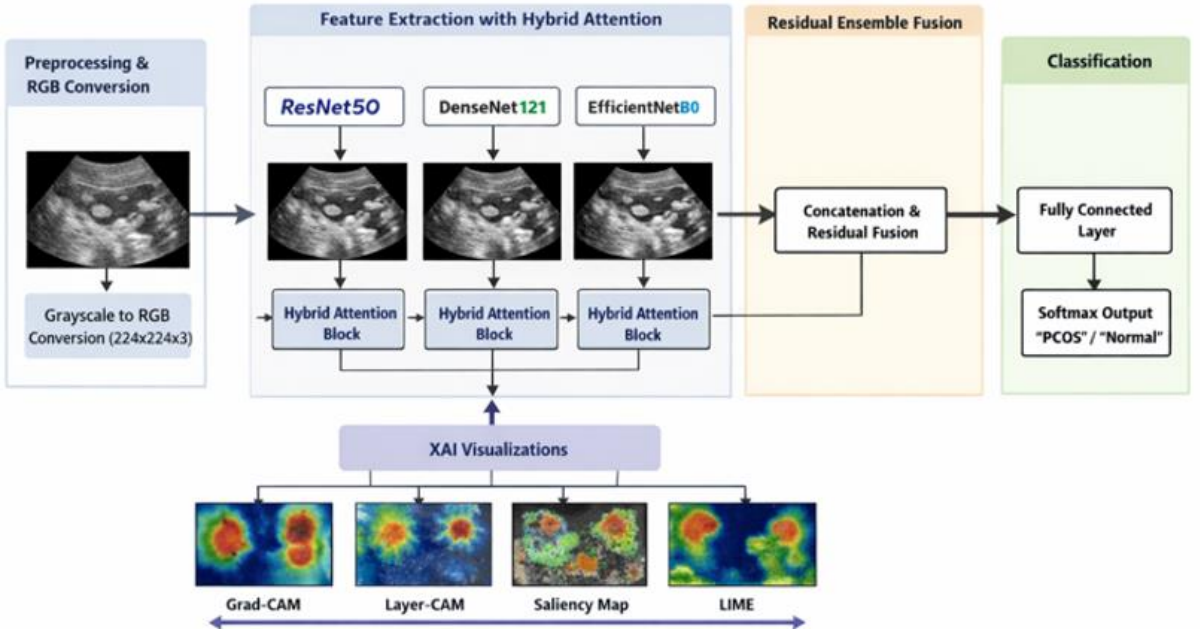


Fig.2. Hybrid Attention Residual Ensemble Network (HAREN) Methodology

In Eq. (12) [37], Eq. (13) [37], Eq. (14) [37], and Eq. (15) [37], $Conv_{1\times1}, Conv_{3\times3}$ and $Conv_{3\times3}^{d=2}$ denote convolutional filters, where the last represents a 3×3 convolution with a dilation rate of 2. The terms F_1,F_2 and F_3refer to the corresponding feature maps, while F_{cs} denotes the cross-scale feature representation derived from them. The variable g represents a gating vector, and F_{ha} indicates the final hybrid attention–refined feature map.

6. Ensemble Feature Fusion

For each backbone, global average pooling converts the refined feature maps into feature vectors:

$$f_k = GAP(F_{ha}^k) \quad (16)$$

The ensemble feature representation is formed by concatenation:

$$F_{ens} = [f_1; f_2; f_3] \quad (17)$$

Dropout regularization is applied to reduce overfitting. In Eq. (16) [37] to Eq. (17) [37], $f_1; f_2; f_3$ denotes the fused features representation from each backbone. F_{ens} denotes ensemble concatenated feature representation.

7. Classification and Loss Function

The final classification layer uses a softmax activation to predict class probabilities for PCOS and NOT PCOS classes:

$$\hat{y} = \text{Softmax}(WF_{ens} + b) \quad (18)$$

The network is trained using categorical crossentropy loss:

$$\mathcal{L} = -\sum_{c=1}^{2} y_c \log(\hat{y}_c) \quad (19)$$

In Eq. (18) [38] to Eq. (19) [38], W denotes weights, b denotes the bias, $\hat{y}$ denotes probability for predicted class, $\mathcal{L}$ denotes categorical ccross-entropyloss, and y_c and $\hat{y_c}$ represent the ground truth and predicted probabilities, respectively. The mathematical formulation clarifies how hybrid attention, residual learning, and ensemble fusion are integrated within the HAREN framework. This structured design enhances feature discrimination, stabilizes training, and improves classification performance while maintaining interpretability suitable for clinical applications.

Algorithm 1 gives the training pipeline of the proposed HAREN network. The HAREN model training pipeline for PCOS detection begins by preprocessing grayscale ovarian ultrasound images by resizing to 224x224, normalizing to, and augmenting with rotations, flips, and zooms, and then converting them to RGB by duplicating channels. Pretrained backbones (ResNet50, DenseNet121, EfficientNetB0) are loaded, with top layers removed for feature extraction. Each backbone's convolutional features undergo hybrid attention: channel attention refines inter-channel relationships, spatial attention highlights key regions, and cross-scale attention fuses multi-level details into refined maps. Global average pooling yields compact feature vectors from these maps, which are then concatenated into an ensemble vector and regularized with dropout. This fused vector feeds a fully connected layer with softmax for binary classification, optimized via categorical cross-entropy loss and Adam backpropagation over multiple epochs. Evaluation metrics are computed post-training. Explainability is enhanced through various XAI

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

techniques. Fig. 2 shows the diagram for the methodology used in the HAREN network.

III. Result

A. Experimental Configuration

All experiments were conducted using a T4 GPU, with models trained for 20 epochs using the Adam optimizer. A batch size of 32 and a learning rate of 0.001 were employed. The total training time for the complete HAREN framework was approximately several hours, depending on the backbone configuration, demonstrating practical computational feasibility for offline clinical deployment. The detailed training configuration is summarized in Table 2.

Table 2. Training configuration.

Hyper parameter	Value
Input image size	224 × 224 × 1
Learning rate	0.001
Batch size	32
Number of epochs	20
Dropout rate	0.3
Loss function	Categorical Cross-Entropy
Optimizer	Adam

B. Baseline and Comparative Models

To comprehensively evaluate the effectiveness of the proposed Hybrid Attention Residual Ensemble Network (HAREN), multiple baseline and comparative deep learning models were implemented and tested under the same experimental conditions.

The baseline models include a single backbone CNN ResNet50, which serves as a strong residual learning baseline. The ResNet50 augmented with attention was also implemented and assessed to evaluate the influence of hybrid attention mechanisms independently. In the end, the introduced HAREN model, which combines EfficientNetB0, DenseNet121, and ResNet50 using hybrid attention blocks and ensemble feature fusion, is implemented and evaluated. This comparative setup enables systematic analysis of single backbone, attention-based, and ensemble-based architectures, highlighting the advantages of the proposed approach.

C. Evaluation Metrics

The proposed HAREN Model's performance was assessed using measures such as Recall, F1 score, Accuracy, Precision, and Area under the Receiver Operating Characteristic Curve (AUCROC). Eq. (20) [39] to Eq. (23) [39] are used to calculate the values of these evaluation metrics. Here, TP indicates true positives, TN indicates true negatives, FP indicates false positives, and FN indicates false negatives. Accuracy measures the overall correctness of classification:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (20)$$

Precision evaluates how many predicted PCOS cases are truly positive:

$$Precision = \frac{TP}{TP+FP} \quad (21)$$

Recall (Sensitivity) measures the proportion of correctly identified PCOS cases:

$$Recall = \frac{TP}{TP+FN} \quad (22)$$

The F1score provides a harmonic balance between Precision and Recall:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (23)$$

The AUCROC metric is also used, as it evaluates classification performance across all decision thresholds and is particularly important in medical diagnosis, where class imbalance may exist [40]. In all equations, TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively. An explicit explainability pipeline is integrated into the HAREN framework using Grad CAM, Grad CAM++, Layer CAM, LIME, and Saliency Maps. Feature maps from the final convolutional layers of each backbone network are utilized to generate class-specific activation maps, which are subsequently aggregated to visualize diagnostically relevant regions. This integration enables post hoc

Table 3. Performance comparison (%).

Model	Accuracy	Precision	Recall	F1 score	AUC
ResNet50 (Vanilla)	88.8	81.2	81.2	81.2	88.9
DenseNet121 (Vanilla)	93.2	92.0	91.9	91.9	93.4
EfficientNetB0 (Vanilla)	94.0	92.3	92.8	92.8	94.5
ResNet50 + Hybrid Attention	95.7	95.2	95.2	94.8	94.8
DenseNet121 + Hybrid Attention	97.3	96.9	96.8	96.8	97.4
EfficientNetB0 + Hybrid Attention	97.9	97.5	97.6	97.6	98.1
HAREN (Proposed)	99.9	98.9	98.9	98.9	99.9

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

interpretability without altering the predictive architecture.

D. Results Achieved

The model was trained using the training configuration given in Table 1. The results are reported as Recall, F1-score, Accuracy, Precision, and AUC.

1. Quantitative Results

The quantitative performance achieved by all models is stored in Table 3. The results clearly show a consistent improvement in performance with the evolution of the models from a single backbone CNN to the proposed hybrid attention-based ensemble framework. To establish a comprehensive baseline, ResNet50, DenseNet121, and EfficientNetB0 were evaluated in their vanilla configurations. As reported in Table 2, DenseNet121 and EfficientNetB0 outperform the vanilla ResNet50 model, achieving accuracies of 93.2% and 94.0%, respectively. This improvement can be attributed to DenseNet121's dense feature reuse capability and EfficientNetB0's compound scaling strategy, which enables efficient multiscale feature representation.

Hybrid attention modules were independently integrated into each backbone to evaluate their effectiveness. The results demonstrate consistent performance improvements across all architectures. ResNet50 with hybrid attention improves accuracy from 88.8% to 95.7%, while DenseNet121 and EfficientNetB0 with hybrid attention achieve accuracies of 97.3% and 97.9%, respectively. These findings confirm that hybrid attention effectively enhances discriminative feature learning by jointly modeling channel-wise importance and spatial dependencies, irrespective of the underlying backbone architecture. The proposed HAREN framework achieves the highest overall performance, attaining 99.9% accuracy and AUC. Although DenseNet121 and EfficientNetB0 with hybrid attention demonstrate strong individual performance, neither matches the ensemble's results. This observation clearly indicates that the performance gains are not solely due to stronger backbones or attention mechanisms but arise from the complementary fusion of heterogeneous feature representations. By combining deep hierarchical features from ResNet50, fine-grained texture modeling from DenseNet121, and efficient multiscale representations from EfficientNetB0 through residual ensemble learning, HAREN achieves superior robustness and generalization.

The comprehensive ablation analysis given in Fig. 3 validates each architectural choice within the proposed framework. The performance table showcases the superior efficacy of the proposed HAREN model for PCOS detection on ovarian ultrasound images, achieving an impressive 98.9% across precision, recall,

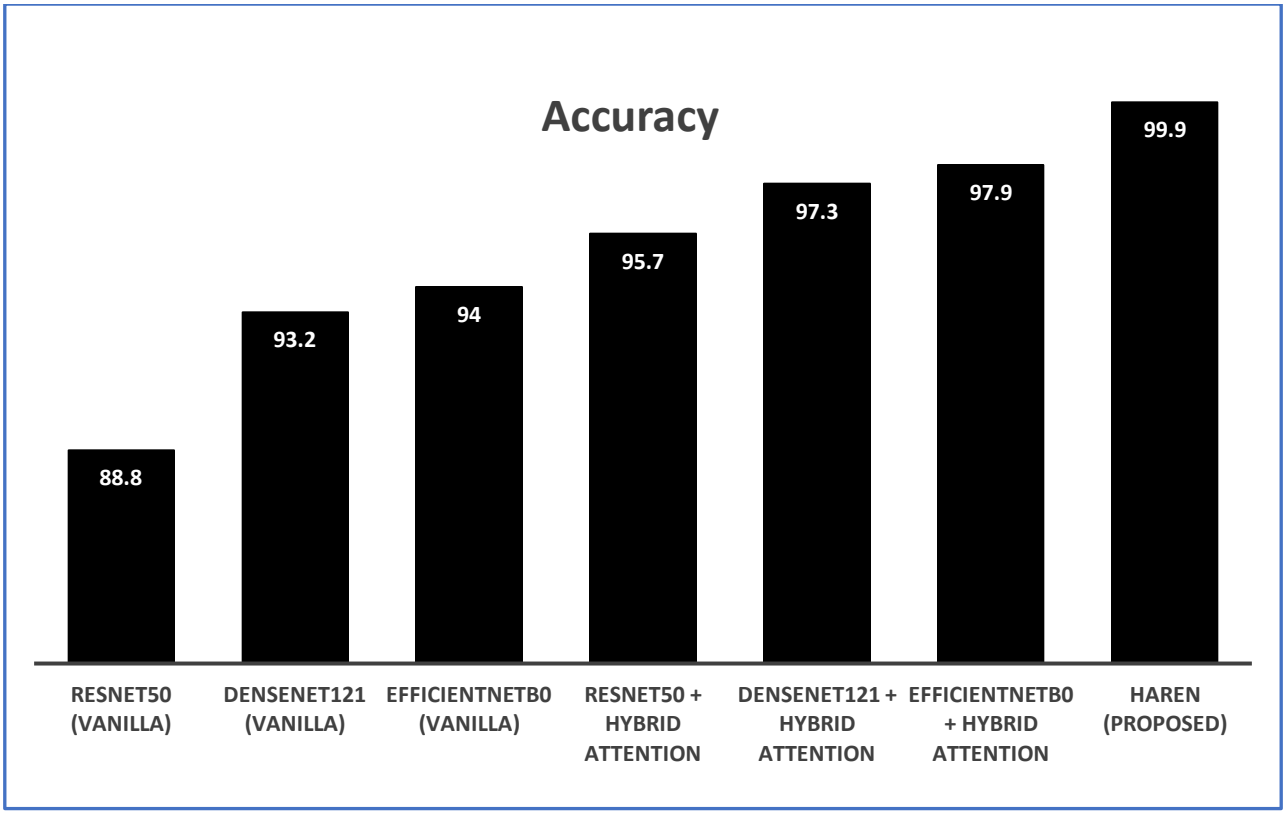


Fig. 3. Comprehensive ablation study: Effect of Backbone choice, Hybrid Attention, and Ensemble learning

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

and F1-score outperforming all baselines by wide margins. Vanilla backbones provide a solid foundation: ResNet50 achieves a uniform 81.2% accuracy, while DenseNet121 (91.9% F1) and EfficientNetB0 (92.8% F1) demonstrate stronger feature extraction from ImageNet pretraining. Adding hybrid attention yields substantial gains with ResNet50 jumps to 94.8-95.2%,

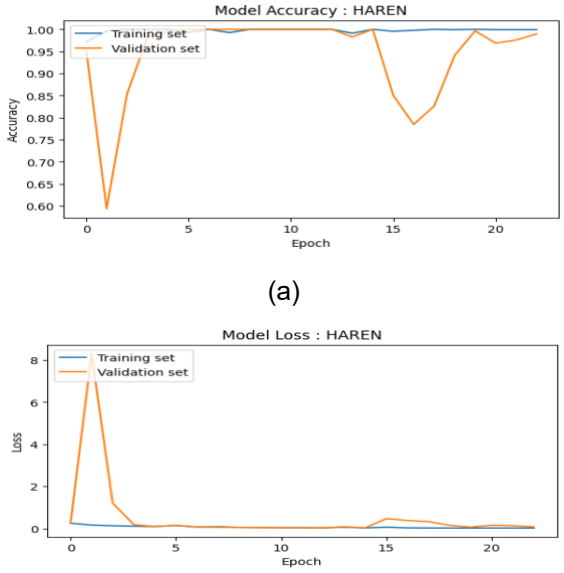


Fig. 4. HAREN performance (a) Accuracy (b) Loss visualization

DenseNet121 to 96.8%, and EfficientNetB0 to 97.6% F1, highlighting how channel, spatial, and cross-scale mechanisms sharpen discriminative features in noisy grayscale data. HAREN's ensemble fusion of this attention-refined backbones push metrics near perfection at 98.9%, balancing precision and recall to minimize false negatives critical for clinical diagnosis. These results underscore the role of attention in handling ultrasound variability like artifacts and class imbalance. Compared to vanilla models, HAREN offers ~7-17% absolute gains, validating the pipeline's preprocessing, pseudo-RGB conversion, and XAI integration. Such metrics suggest real-world deployability, potentially rivaling expert radiologists while enabling explainable AI for trust in biomedical applications.

2. Clinical Relevance of Quantitative Results

While the proposed HAREN framework demonstrates progressive performance improvements across architectural enhancements, the clinical relevance of these gains must be interpreted through the balance between sensitivity and specificity. In the context of PCOS diagnosis, high sensitivity (recall) is particularly important, as false negative cases may delay diagnosis and subsequent management of metabolic, reproductive, and endocrine complications. At the same time, excessive false positives can lead to unnecessary follow-up investigations, increased patient anxiety, and potential overtreatment.

The proposed model achieves a balanced precision–recall profile, indicating that improvements in sensitivity are not obtained at the expense of increased false positive rates. This balance is clinically significant, as it suggests that HAREN reliably identifies PCOS cases while maintaining diagnostic confidence in negative predictions. Such behavior aligns with real-world clinical expectations, where screening tools must minimize missed diagnoses without overwhelming clinicians with excessive false alarms.

3. Accuracy and loss curves

Fig. 4 illustrates the training and validation performance of the proposed HAREN model. As shown in Fig. 4(a), both training and validation accuracy increase steadily with increasing number of epochs. The training accuracy reaches a very high value in the later epochs, while the validation accuracy closely follows the same

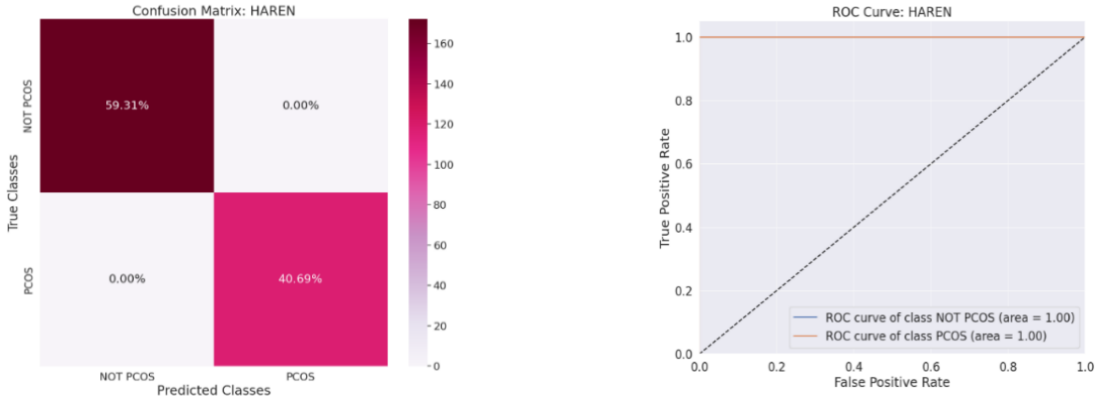


Fig. 5. HAREN performance (a) Confusion Matrix (b) ROC curve

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

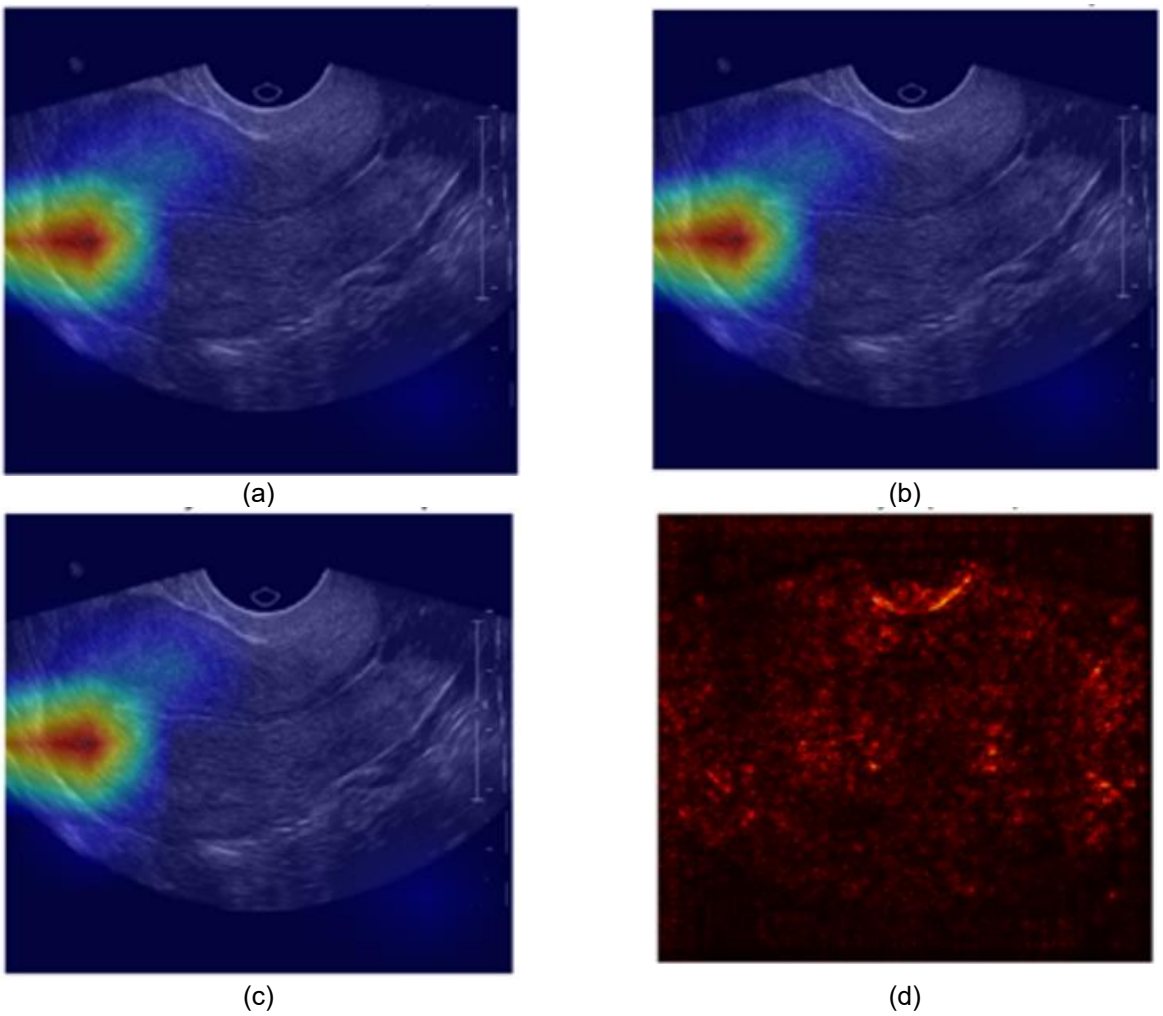


Fig. 6. XAI visualization (a) Grad CAM (b) Grad CAM++ (c) Layer CAM (d) Saliency Maps

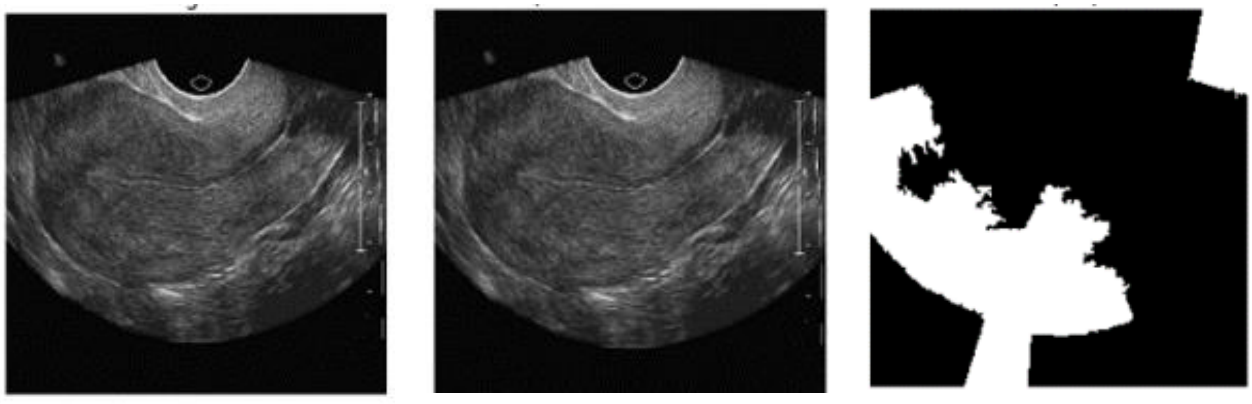


Fig. 7. Lime Visualization (a) Original (b) Positive Contribution (c) Mask (superpixels)

trend with only a marginal gap. This close alignment between training and validation accuracy indicates that the proposed HAREN model learns discriminative features effectively and generalizes well to unseen ultrasound images. Very few sharp fluctuations and deviations among the curves support even optimization and controlled overfitting.

The loss curves for training and validation are depicted in Fig. 4(b). Both training and validation loss

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

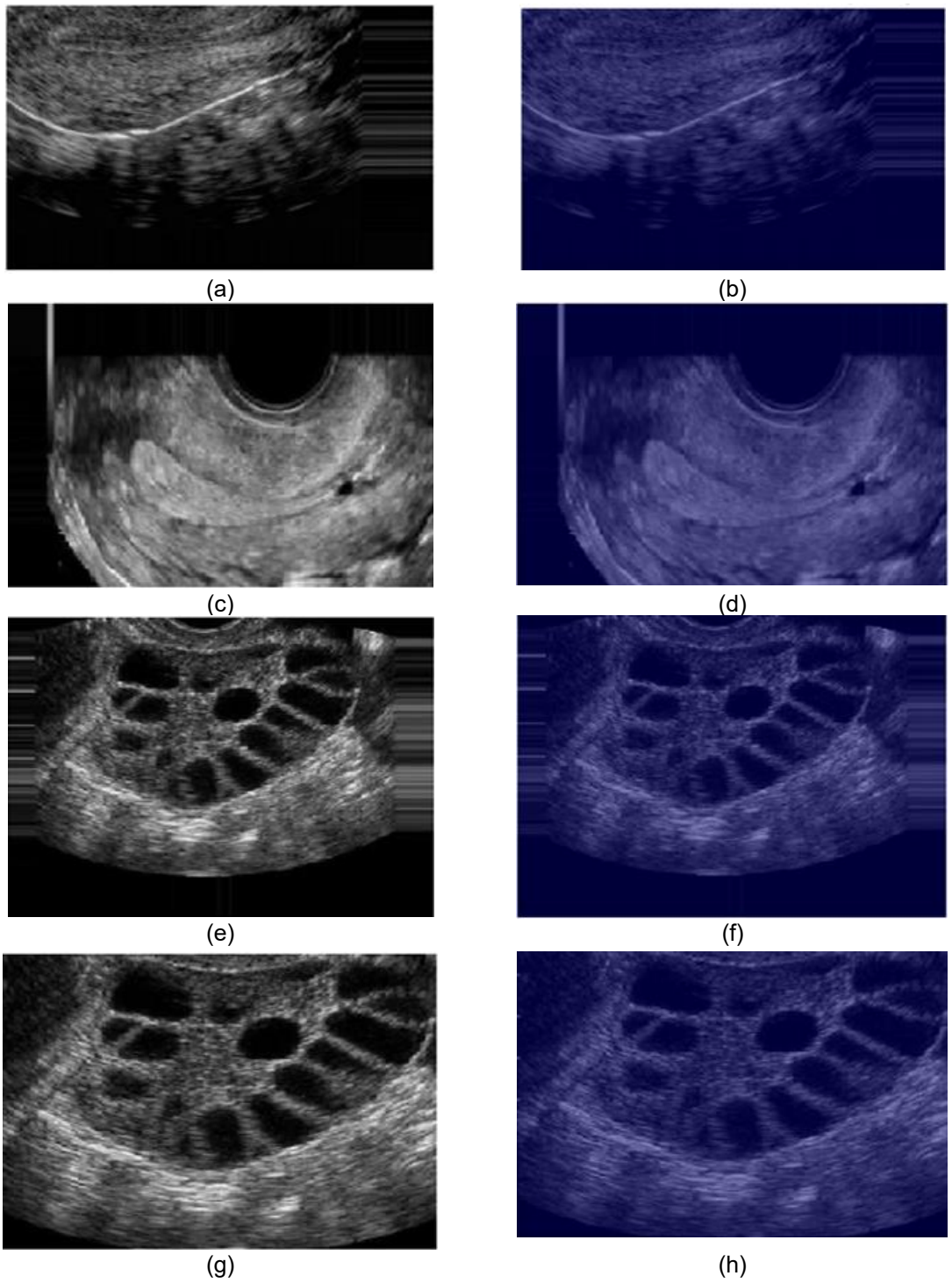


Fig. 8. Explainability based analysis of misclassified ultrasound images (a) true: NOT PCOS Predicted: PCOS (b) Grad Cam of Predicted PCOS in (a) (c) true: NOT PCOS Predicted: PCOS (d) Grad Cam of Predicted PCOS in (c) (e) true: PCOS Predicted: NOT PCOS (f) Grad Cam of Predicted NOT PCOS in (e) (g) true: PCOS Predicted: NOT PCOS (h) Grad Cam of Predicted NOT PCOS in (g)

continually decrease throughout the training and validation process. The slightly higher validation loss indicates a promising generalization rather than overfitting. Overall, the accuracy and loss outputs proved that the introduced HAREN framework attains

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

good training, convergence, and generalization performance.

4. Confusion Matrix and ROC Analysis.

Fig. 5(a) shows a confusion matrix generated by the proposed HAREN framework. The confusion matrix confirms the robust performance in distinguishing normal and infected PCOS samples through its very high true positive and true negative rates. The ROC curves shown in Fig. 5(b) further validate the robustness of the model. An average ROCAUC of 99.33% demonstrates near-perfect class separation, which is critical for reliable medical diagnosis.

5. Explainability Analysis

Numerous explainable artificial intelligence (XAI) techniques, such as LIME, Layer CAM, Saliency Maps, Grad CAM, and Grad CAM++, were used to enhance clinical trust. These techniques are applied to visualize the regions that influence the model's predictions. Fig. 6 shows the visualization of XAI methods used. Fig. 7 shows the visualization results achieved with LIME.

To assess the clinical relevance of the explainability visualizations, the Grad CAM and LIME maps were analyzed in the context of established radiological criteria for Polycystic Ovary Syndrome (PCOS). According to widely adopted diagnostic guidelines, including the Rotterdam criteria, PCOS is commonly associated with multiple peripherally arranged follicles (typically ≥12), increased stromal echogenicity, and enlarged ovarian volume on ultrasound imaging. The explainability maps in Fig. 6 and Fig. 7 demonstrate that the proposed model consistently attends to anatomically meaningful ovarian regions. In PCOS cases, high activation areas are predominantly localized around clusters of anechoic follicular structures and the surrounding stromal region. This behavior aligns with radiological practice, where the number, distribution, and morphology of follicles, as well as stromal texture, play a central role in diagnosis. The model's attention to these regions suggests that it implicitly learns clinically relevant sonographic patterns rather than relying on spurious background features.

In NOT PCOS cases, the attention maps exhibit a more diffuse activation pattern with reduced emphasis on follicular clusters, reflecting the absence of hallmark PCOS features. LIME explanations further corroborate this behavior by highlighting local image regions that contribute negatively to PCOS classification, often corresponding to uniform stromal texture and normal follicular distribution. Importantly, these observations indicate a strong alignment between the model's decision-making process and human expert reasoning used in ultrasound-based PCOS assessment. Although the model does not explicitly compute follicle counts or ovarian volume, the explainability analysis confirms that its predictions are guided by clinically meaningful features routinely evaluated by radiologists. This enhances the interpretability and trustworthiness of the proposed framework for potential clinical decision support applications.

Fig. 8 provides an explainability-based assessment of representative misclassified cases. False-positive predictions are primarily associated with attention activation over diffuse textural patterns and imaging artifacts that visually resemble polycystic characteristics. Conversely, false negative cases exhibit weakened or spatially diffused attention over diagnostically relevant follicular regions, particularly in low contrast or atypical PCOS presentations. These observations suggest that misclassifications arise from intrinsic ultrasound ambiguity rather than unstable model behavior, underscoring the importance of explainability-driven evaluation.

IV. Discussion

A. Vanilla Backbone Models

The vanilla backbone models provide a base for evaluating the performance of the enhanced architecture. Table 3 shows the performance achieved by the vanilla models used. ResNet50 achieves an accuracy of 88.8%. It also achieves 81.2% recall, precision, and F1 score, and an AUC of 88.9%. This indicates limited sensitivity under noisy ultrasound conditions. DenseNet121 improves performance to 93.2%. 92.0%, 91.9%, 91.9%, and 93.4% values of accuracy, precision, recall, F1-score, and AUC, respectively. This is due to the benefit of dense feature reuse. EfficientNetB0 achieved 97.9% accuracy, 97.5% precision, 97.6% recall, 97.6% F1-score, and an AUC of 98.1%. This demonstrates good multi-scale feature extraction. However, the relatively low recall and AUC across vanilla models highlight the need for feature-refinement mechanisms for reliable PCOS diagnosis.

B. Hybrid Attention Models

Hybrid attention mechanisms considerably enhanced the performance of backbone networks. The accuracy of the ResNet50 is improved from 88.8% to 95.7%. AUC is also improved up to 94.8%. This shows significant suppression of unrelated background regions. DenseNet121 attains an improved accuracy of 97.3% and an AUC of 97.4% with attention. It depicts good complementary interaction between attention and dense connectivity. EfficientNetB0 attained the best results of 97.9% accuracy and 98.1% AUC with hybrid attention. These results signify that hybrid attention improves good discriminative ability by enhancing

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

Table 4. Ablation Study on Residual Feature Fusion.

Model Variant	Accuracy	Precision	Recall	F1 score	AUC
HAREN (w/o Residual Fusion)	97.8	96.9	96.7	96.8	98.1
HAREN (Full, with Residual Fusion)	99.9	98.9	98.9	98.9	99.9

Table 5. Comparison with Existing Studies.

Study (Reference)	Model / Approach	Dataset Size	Dataset Source	Image Type	Accuracy (%)	F1Score (%)	AUC (%)
Zhao et al. [28]	YOLOv11based DL	933 cases	Private (clinical)	Grayscale	96.7	96.90	96.70
Alamoudi et al. [29]	DL + clinical fusion	NR	Private (clinical)	Grayscale	84.81	72.73	84.81
Moral et al. [30]	Hybrid / CystNet	3856 images	Public (Kaggle)	Grayscale	96.54	95.75	95.92
Bedi et al. [31]	Attentionbased ResUNet	3800 images	Public (Kaggle)	Grayscale	97.69	98.59	97.69
Pratibha et al. [32]	ResNet50 + SE blocks (Dataset 1)	1924 images	Public	Grayscale	98.25	98.15	–
	ResNet50 + SE blocks (Dataset 2)	4648 images	Public (Kaggle)	Grayscale	94.56	92.93	–
Vijaykumar et al. [33]	Hybrid DL model	3800 images	Public (PCOSgen)	Grayscale	91.50	90.80	94.00
HAREN (Proposed)	Hybrid Attention + Explainability	**3400+ images**	**Public + Private**	**Grayscale**	**99.93**	**98.92**	**99.93**

recall and AUC, particularly, which are crucial for clinical screening.

C. Ensemble Learning and Residual Feature Fusion

The introduced HAREN network achieved the highest overall performance, with accuracy, precision, recall, F1-score, and AUC of 99.9%, 98.9%, 98.9%, 99.9%, and 99.9%, respectively. Here, Multibackbone ensemble learning efficiently integrates complementary representations from DenseNet121, EfficientNetB0, and ResNet50. Ablation results shown in Table 3 confirm that residual feature fusion plays a significant role preserving discriminative backbone-specific features. It also stabilizes the ensemble learning. The balance between recall and precision further specifies its significance in clinical PCOS diagnosis.

D. Comparison with Existing literatures

A comparison between proposed approaches and the existing studies available in the literature is shown in Table 4 and Table 5. Zhao *et al.* [28] developed a YOLOv11-based DL model to detect PCOS in grayscale ultrasound images. They used a small private dataset of 933 patients. Their model attained 92.90%, 93.40%, and 97.90% of accuracy, F1 score, and AUC, respectively. The limitation was a small dataset and a single model. Alamoudi et al. [29] developed a DL network for clinical data fusion to detect PCOS from grayscale ultrasound images. They also used a private dataset. They achieved 84.81%, 72.73%, and 84.81% of accuracy, F1 score, and AUC, respectively. Moral et al. [30] developed a hybrid DL network where they incorporated multilevel thresholding with deep learning. They used a 3856-images dataset available on Kaggle. They achieved 96.54%, 95.75%, and 95.92% of accuracy, F1 score, and AUC, respectively. Bedi et al. [31] proposed a ResUNet model hybridized with an attention mechanism for detecting PCOS. They used a 3800 grayscale ultrasound image dataset from Kaggle. Their approach used a single backbone with attention. They achieved 97.69%, 98.59%, and 97.69% of accuracy, F1 score, and AUC, respectively. Pratibha et al. [32] proposed a hybrid of squeeze-and-excitation (SE) blocks and ResNet50 for PCOS detection. They used two different experimental setups. A publicly available

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

dataset of 1924 ultrasound images was used in this research.

They attained an accuracy of 98.25% with this dataset. They also applied their model to a larger dataset (4668 images) and achieved a reduced accuracy of 94.56%. Vijaykumar et al. [33] presented a hybrid DL network for PCOS detection. They hybridized clinical data with MobileNetV2. They evaluated their model on 3800 grayscale images available on Kaggle. Their model achieved 91.50% accuracy, 90.80% F1-score, and 94.00% AUC. All the above studies produced accuracies in the range of 84.8% to 98.5%. They also observed a trade-off between AUC and F1-score. The proposed HAREN network constantly outclasses these models across all performance measures. This confirms that hybrid attention, residual feature fusion, and multi-backbone ensemble learning together prove their best discriminative capability and clinical reliability for PCOS detection.

E. Limitations

This study uses a single dataset made of 11784 ultrasound images, which does not include patient details. This limits the subject diversity. This may undermine the generalization, even with augmentation. The proposed framework introduces hybrid attention with ensemble learning, which normally increases computational efforts. This can create some issues while deploying in a resource-constrained environment. Adding to that, the evaluation here is carried out with standard evaluation metrics. External validation or real-time clinical assessment is also not performed.

Future work includes the validation of the proposed network on a large multi center ultrasound dataset. This validation will be done to improve robustness and generalizability. We will also work to optimize our network by using lightweight models and pruning to reduce computational efforts and to provide real-time use. Additionally, improved XAI models and adaptive attention will also be investigated to improve interpretability, which in turn enhances the trust in patients. The proposed framework demonstrates the effectiveness of hybrid attention and ensemble learning for ultrasound-based diagnosis. From an engineering perspective, the architecture is adaptable to other medical imaging tasks. Clinically, the high recall and balanced precision support its use as a PCOS screening aid. From a medical informatics viewpoint, the framework facilitates scalable, automated diagnostic pipelines. Grad-CAM–based XAI was employed to visualize decision-relevant regions. The generated heatmaps show consistent focus on diagnostically meaningful ovarian regions, with hybrid attention producing more localized and coherent explanations. This enhances transparency, clinician trust, and clinical applicability.

V. Conclusion

This study proposes a Hybrid Attention Residual Ensemble Network (HAREN) for automated detection and classification of Polycystic Ovary Syndrome (PCOS) in grayscale ovarian ultrasound images. The model addresses challenges such as noise susceptibility, poor feature localization, and limited interpretability in conventional deep learning methods. HAREN integrates channel, spatial, and cross-scale attention mechanisms with residual learning and the ensemble fusion of pretrained CNNs such as ResNet50, DenseNet121, and EfficientNetB0, to extract complementary and clinically relevant ovarian features. The framework incorporates explainable AI (XAI) techniques to enhance diagnostic transparency and clinical trust. Evaluation on a public dataset of 11,784 ultrasound images (6,784 PCOS, 5,000 non-PCOS) showed that HAREN achieved 99.9% accuracy, 98.9% precision, 98.9% recall, 98.9% F1-score, and an AUC of 99.9%, outperforming baseline and single backbone attention models. Explainability analysis confirmed the model's focus on clinically meaningful ovarian regions. Overall, HAREN offers a robust, accurate, and interpretable solution for reliable PCOS diagnosis with strong potential for clinical decision-support integration.

Acknowledgment

The authors would like to express sincere gratitude to all who have helped us in this research work.

Funding

This research received no specific funding.

Data Availability

Data will be available upon request.

Declarations

Ethical Approval

This research was done to fulfil the requirements of the research scholar's PhD. It is also adhered to the PhD academic regulations established by the Faculty of Engineering Technology, Drs. Kiran and Pallavi Patel Global University.

Consent for Publication Participants.

Consent for publication was given by all participants.

Competing Interests

There is no competing interests.

References

[1] H. J. Teede *et al.*, "Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome," *Human Reproduction*, vol. 33, no. 9, pp. 1602–1618, Jul. 2018, https://doi.org/2010.1093/humrep/dey256.

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

[2] R. Azziz et al., "The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report," Fertility and Sterility, vol. 91, no. 2, pp. 456–488, Oct. 2008, https://doi.org/10.1016/j.fertnstert.2008.06.035.

[3] "Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome," *Fertility and Sterility*, vol. 81, no. 1, pp. 19–25, Jan. 2004, https://doi.org/10.1016/j.fertnstert.2003.10.004.

[4] Y. Kumar et al., "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda,"Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 7, pp. 8459–8486, Jan. 2022, https://doi.org/10.1007/s12652-021-03612-z.

[5] J. Kang et al.,"MRI-Based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, p. 2222, Mar. 2021, https://doi.org/10.3390/s21062222.

[6] X. Zhang et al.,"Attention-Based Multi-Model Ensemble for Automatic Cataract Detection in B-Scan Eye Ultrasound Images," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-10, https://doi.org/10.1109/IJCNN48605.202092076 96.

[7] S. C. Lingamaiah et al., "Attention-based deep learning model for clinical assessment of focal liver lesions using ultrasound imaging," *Biomedical Signal Processing and Control*, vol. 116, p. 109563, Jan. 2026, https://doi.org/10.1016/j.bspc.2026.109563.

[8] H. Jia, J. Zhang, K. Ma, X. Qiao, L. Ren, and X. Shi, " Application of convolutional neural networks in medical images: a bibliometric analysis," *Quant Imaging Med Surg.* vol 14, no. 5, pp. 3501-3518, May 1 2024, https://doi.org/10.21037/qims-23-1600.

[9] DR. Sarvamangala, RV. Kulkarni, " Convolutional neural networks in medical image understanding: a survey," *Evol Intell*. Vol. 15, no. 1, pp. 1-22, 2022, https://doi.org/10.1007/s12065-020-00540-3.

[10] Z. Fan *et al.*, "TMAN: A triple morphological feature attention network for fine-grained classification of breast ultrasound images," *Expert Systems with Applications*, vol. 234, 2023, Art. no. 120969, https://doi.org/10.1016/j.eswa.2023.120969.

[11] M. Zhao *et al.*, "An attention-based hybrid deep learning framework integrating brain connectivity and activity of resting-state functional MRI data," *Med Image Anal*. Vol. 78:102413, May 2022, https://doi.org/ 10.1016/j.media.2022.102413.

[12] R.G. Akindele *et al.*, " A hybrid attention-based deep learning framework for precise early diagnosis of Alzheimer's disease,". *Discov Appl Sci*, Vol 7, 856, 2025, https://doi.org/10.1007/s42452-025-07492-2

[13] Z. Chen, C. Zhang, Z. Li, J. Yang and H. Deng, " Automatic segmentation of ovarian follicles using deep neural network combined with edge information," *Front. Reprod. Health* Vol. 4,877216, 2022, https://doi.org/10.3389/frph.2022.877216.

[14] C. Kamala and J.M. Shivaram, "Segmentation of ovarian cyst using improved U-NET and hybrid deep learning model," *Multimed Tools Appl*, Vol. 83, pp. 42645–42679, 2024, https://doi.org/10.1007/s11042-023-16998-z

[15] M. Azmoodeh-Kalati *et al.*," Leveraging an ensemble of EfficientNetV1 and EfficientNetV2 models for classification and interpretation of breast cancer histopathology images," *Sci Rep* vol. 15, 21541, 2025, https://doi.org/10.1038/s41598-025-06853-6.

[16] SS. Koshy *et al.*,"HED-Net: a hybrid ensemble deep learning framework for breast ultrasound image classification," *Front. Artif. Intell., vol* 8:1672488, 2026 https://doi.org/10.3389/frai.2025.1672488.

[17] B. T. Z. Afif, W. Wiharto, and U. Salamah, "Classification of Ultrasound Images Using ResNet-50 with a Convolutional Block Attention Module(CBAM)", *j.electron.electromedical.eng. med.inform*, vol. 8, no. 1, pp. 284-303, Jan. 2026, https://doi.org/10.35882/jeeemi.v8i1.1406.

[18] S. Pawar, P. Dhane, D. Shelke, P. Dheple and N. Doshi, " PCOS Detection using Hybrid CNN-XGBoost Model - A Multimodal Data Approach,". *Biomed Pharmacol J*, Vol. 18, no. 3, 2025, https://dx.doi.org/10.13005/bpj/3252

[19] H. Liu, P. Zhang, J. Hu *et al.*, " Attention residual network for medical ultrasound image segmentation," *Sci Rep* Vol. 15, 22155, 2025, https://doi.org/10.1038/s41598-025-04086-1

[20] M.A. Aslam, A. Naveed, N. Ahmed *et al.*, " A hybrid attention network for accurate breast tumor segmentation in ultrasound images," *Sci Rep* vol 15, 39633, 2025, https://doi.org/10.1038/s41598-025-23213-6.

[21] E. Silambarasan, G. Nirmala, I. Mishra, "Polycystic ovary syndrome detection using optimized SVM and DenseNet," *Int. j. inf. tecnol.* Vol. 17, pp. 1039–1047, 2025, https://doi.org/10.1007/s41870-024-02143-y.

[22] N. Kaur, G. Gupta and P. Kaur, "Transfer-Based Deep Learning Technique for PCOS Detection Using Ultrasound Images," *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, Bengaluru, India, 2023, pp. 1-6, https://doi.org/2010.1109/NMITCON58196.2023.10276245

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

[23] K. S. Meghana, T. T, G. Shravinya and S. S. S. Siri, "XAI-Enabled Deep Convolutional Model for Polycystic Ovary Syndrome Detection from Ultrasound Imaging," *2025 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Cochin, India, 2025, pp. 254-261, https://doi.org/10.1109/RAICS66191.2025.11332581.

[24] P.B. Patil *et al.*, "Explainable ensemble-based machine learning model for polycystic ovary syndrome detection using hybrid feature selection," *Int. j. inf. tecnol.*, 2025, https://doi.org/10.1007/s41870-025-03044-4

[25] P. Jain *et al*, " Xplainable AI for deep learning model on PCOD analysis," pp. 131-152, January 2024, https://doi.org/10.1016/B978-0-323-95315-3.00012-7.

[26] V. Lakshmi, and B. Pushpa, "Explainable multimodal deep learning using cross-attention fusion of ultrasound and clinical features for PCOS classification,"*Discov Computing* vol. 29, no. 7, 2026. https://doi.org/10.1007/s10791-025-09901-x

[27] A. Choudhari, Pcos detection using ultrasound images. https://www.kaggle.com/datasets/anaghachoudhari/pcosdetectionusingultrasoundimages (2024)

[28] B. Zhao *et al.*, "A Deep Learning-Based Automatic Recognition Model for Polycystic Ovary Ultrasound Images," *Balkan Med J.* vol. 42, no. 5, pp. 419-428, sep 2025, doi: 10.4274/balkanmedj.galenos.2025.2025-5-114.

[29] A. Alamoudi *et al.*, "A deep learning fusion approach to diagnosis the polycystic ovary syndrome (PCOS)," Applied Computational Intelligence and Soft Computing, vol. 2023, pp. 1–15, Feb. 2023, https://doi.org/10.1155/2023/9686697.

[30] P. Moral et al., "CystNet: An AI driven model for PCOS detection using multilevel thresholding of ultrasound images," Scientific Reports, vol. 14, no. 1, p. 25012, Oct. 2024, https://doi.org/10.1038/s41598-024-75964-3

[31] P. Bedi et al., "An integrated adaptive bilateral filterbased framework and attention residual Unet for detecting polycystic ovary syndrome," Decision Analytics Journal, vol. 10, p. 100366, Nov. 2023, https://doi.org/10.1016/j.dajour.2023.100366.

[32] P. Pratibha et al.,"Deep Learning Technique for Interpretable Diagnosis of Polycystic Ovary Syndrome in Ultrasound Imaging," Aptisi Transactions on Technopreneurship, Vol. 7., 2025, 779792, https://doi.org/10.34306/att.v7i3.768.

[33] R. Vijayakumar et al., "PCOSVision A Hybrid Deep Learning Model for Polycystic Ovary Syndrome Detection using MobileNetV2 and Clinical Data," in Advances in computer science research, 2025, pp. 865–877. https://doi.org/10.2991/9789464637182_74.

[34] A. M. Naser, "Color to grayscale image conversion based dimensionality reduction with Stationary Wavelet transform," 2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA), Baghdad, Iraq, 2016, pp. 1-5, https://doi.org/10.1109/AIC-MITCSA.2016.7759946.

[35] B. Wong et al., "Rethinking Pre-Trained Feature Extractor Selection in Multiple Instance Learning for Whole Slide Image Classification," 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, USA, 2025, pp. 1-5, https://doi.org/10.1109/ISBI60581.2025.10981015.

[36] N. Das and S. Das, "Attention-UNet architectures with pretrained backbones for multi-class cardiac MR image segmentation," Current Problems in Cardiology, vol. 49, no. 1, p. 102129, Oct. 2023, https://doi.org/10.1016/j.cpcardiol.2023.102129.

[37] G. Li et al., "HAM: Hybrid attention module in deep convolutional neural networks for image classification," Pattern Recognition, vol. 129, p. 108785, May 2022, https://doi.org/10.1016/j.patcog.2022.108785.

[38] F. Chahkoutahi et al.,"Loss functions in classification: An comprehensive overview and comparative study," Applied Soft Computing, vol. 84, p. 113778, Aug. 2025, https://doi.org/10.1016/j.asoc.2025.113778.

[39] S. A. Hicks et al., "On evaluation metrics for medical applications of artificial intelligence," Scientific Reports, vol. 12, no. 1, p. 5979, Apr. 2022, https://doi.org/10.1038/s41598-022-09954-8.

[40] M. S. Mohosheu et al., "ROC Based Performance Evaluation of Machine Learning Classifiers for Multiclass Imbalanced Intrusion Detection Dataset," 2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering, Kuala Lumpur, Malaysia, 2023, pp. 1-6, https://doi.org/10.1109/ICRAIE59459.2023.10468177.

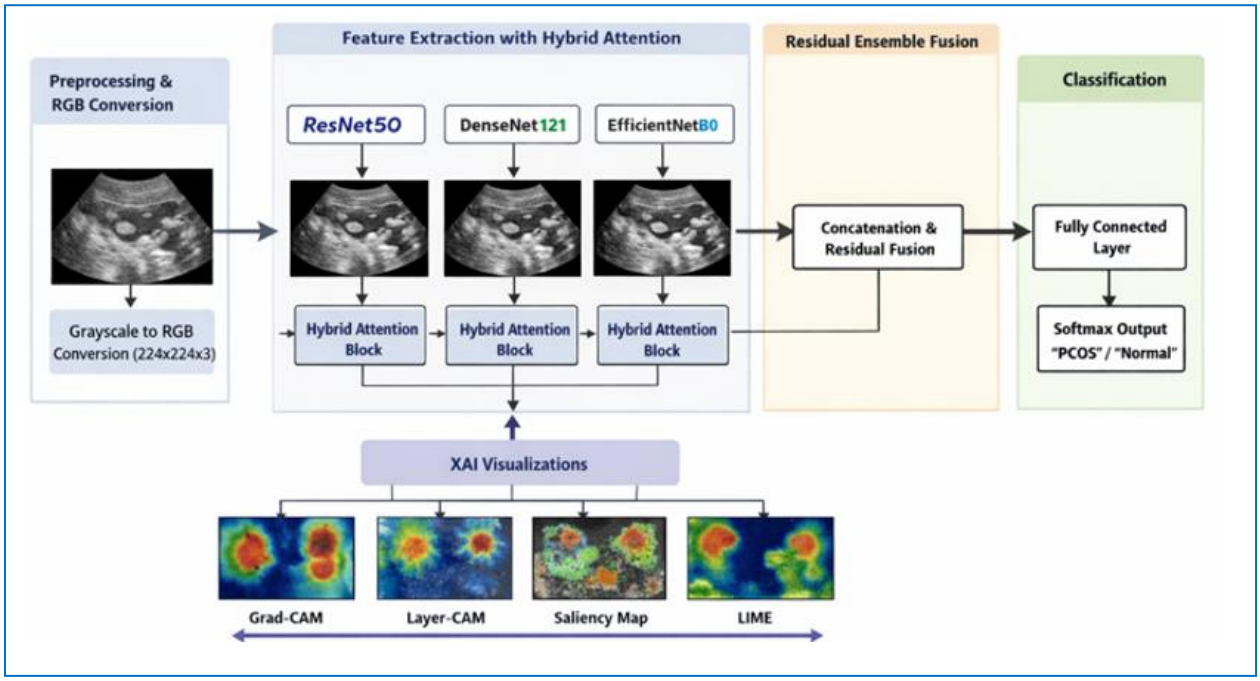
Author Biography

Pragati Patil is a Ph.D. scholar in computer engineering at Drs. Kiran and Pallavi Global University, Vadodara. She has 07 years of experience in teaching and research. Currently pursuing her Ph.D. in computer engineering at Drs. Kiran and Pallavi Global

Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441

University, Vadodara, she holds the position of assistant professor at the Department of CSE (AI&ML), Rajarambapu Institute of Technology, Rajaramnagar, Maharashtra. Her areas of interest in research and academic includes deep learning, machine learning, and image processing. She has published 14 research papers in national and international publications and conferences, making major contributions to her fields of interest.

Dr. Nandini M. Chaudhari (**B.E., M.E. CSE, Ph.D.)** has recently been working as Professor and Director, Krishna School of Emerging Technology with additional charge of Dean Faculty of Engineering and Technology at DRS Kiran and Pallavi Patel Global University, Vadodara, Gujarat (KPGU). She worked as a principal and Vice principal at previous colleges and has more than 33 years of experience. She has guided 3 research scholars and is currently guiding 5 Research Scholars in the subject of computer engineering under the faculty of Engineering and Technology at KPGU. Her 27 papers has been published in international journals indexed in SCOPUS/WOS/UGC listed journals and reputed journals. She has presented 18 papers at national and international conferences and worked as a reviewer and session chair at various conferences. Her research interests include image and video processing, AI and Machine learning, IoT, and Quantum Computing. She had published 2 patents and 1 Copyright. She had contributed to publishing 3 books and 3 book chapters with national/international publishers.



Manuscript received 30 December 2025; Revised 10 March 2026; Accepted 20 April 2026; Available online 8 May 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1441