RESEARCH PAPER

SonoCyst-Net: Deep Learning-Based Framework for the Automated Detection of PCOS Using Ultrasound Sonography Images

Meghamala Kunderu[1*], K. Venkata Raju[2]
[1,2]School of Computer Science Engineering, KL University, Vijayawada
Email: 2401050114@kluniversity.in
Email: Kvraju@kluniversity.in
Received: 1st Sep, 2026 | Revised: 8th Sep, 2026 | Accepted: 15th Sep, 2026 | Available Online: 22nd Sep, 2026

ABSTRACT
Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder among women of reproductive age, commonly diagnosed through a combination of clinical symptoms and ultrasound imaging of the ovaries. However, the variability in image quality and subjective interpretation of sonographic features often leads to inconsistent diagnoses. This study introduces SonoCyst-Net, an advanced deep learning-based framework for the automated detection of PCOS using ultrasound images. The model employs a high-performance EfficientNet-B3 backbone for robust and efficient feature extraction, coupled with a Variational Autoencoder (VAE) with Interactive Variational Attention to enhance and reconstruct diagnostically significant patterns. To address limitations in dataset diversity, a GAN-based augmentation pipeline is integrated to synthetically enrich training samples. For deployment in real-time healthcare environments, the model is optimized using pruning and quantization techniques, significantly reducing computational load. Evaluation on the Kaggle PCOS ultrasound dataset demonstrates that SonoCyst-Net achieves high diagnostic Accuracy 99.9%, Precision 99.74%, F1-Score 99.87%, and Specificity 99.82%, representing a promising AI-assisted tool for early and reliable PCOS detection.
Keywords: PCOS, SonoCyst-Net, EfficientNet_B3, VAE, Variational Attention.
How to cite this article: Meghamala Kunderu1* K. Venkata Raju2. SonoCyst-Net: Deep Learning-Based Framework for the Automated Detection of PCOS Using Ultrasound Sonography Images. Int J Drug Deliv Technol. 2026;16(80s):404-419. DOI: 10.25258/ijddt.16.80s.48

. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is the most frequently occurring endocrine disorder in women of reproductive age. It is associated with hormonal disturbances, metabolic complications, and ovarian dysfunction, all of which adversely affect fertility and general health. Although highly prevalent, PCOS is often underdiagnosed due to its varied clinical presentation, including menstrual irregularities, excessive hair growth, acne, weight gain, and multiple ovarian cysts. In addition to reproductive concerns, the syndrome is linked to significant metabolic, cardiovascular, and psychological complications [1].

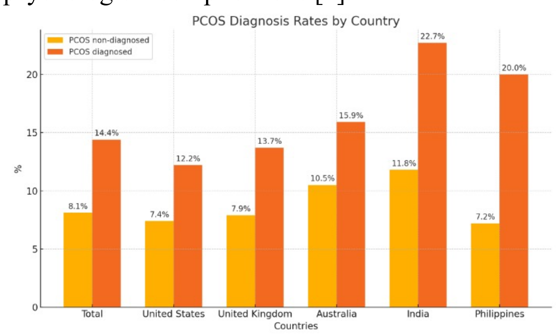


Fig. 1. PCOS Diagnosis Rates by Country

Globally, PCOS affects approximately 6% to 13% of women in their reproductive years [2]. However, recent epidemiological studies indicate that up to 70% of PCOS cases go undiagnosed. Figure 1 illustrates the diagnosis rates of PCOS in various countries. India exhibits the highest reported prevalence at 22.7%, followed by the Philippines at 20%. On the other hand, developed countries such as the United States and the United Kingdom report lower diagnosis rates, 7.4% and 7.9%, respectively. These disparities may be attributed to differences in healthcare access, diagnostic awareness, genetic predispositions, and reporting mechanisms. The higher rates in certain regions emphasize the need for effective screening and awareness campaigns.

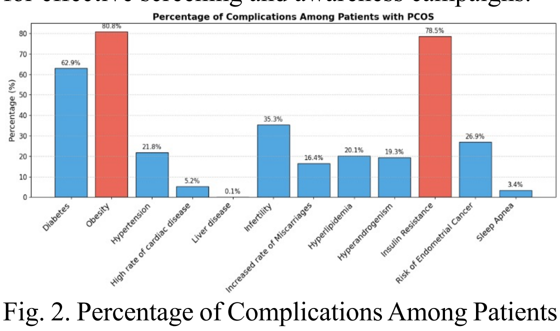


Fig. 2. Percentage of Complications Among Patients with PCOS

PCOS is associated with several long-term health complications. These complications extend beyond reproductive health and affect multiple systems in the body. Figure 2 provides insight into the prevalence of various complications among PCOS patients. Obesity affects nearly 80.8% of PCOS women, while insulin resistance is reported in 78.5% of cases. These conditions are interlinked and form the basis of metabolic syndrome, which increases

the risk of cardiovascular diseases[3]. Moreover, diabetes is present in 62.9% of PCOS cases, highlighting the endocrine disruption caused by the syndrome. Fertility issues affect 35.3% of patients, while 26.9% are at a heightened risk of developing endometrial cancer. Other complications include mental health disorders, hypertension, and hyperlipidemia. These statistics demonstrate the broad and serious impact of PCOS on women's health [4]
.

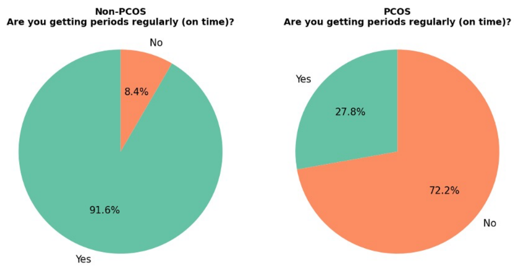


Fig. 3. Menstrual Regularity Comparison of PCOS with Non-PCOS

Figure 3 compares the menstrual regularity between women diagnosed with PCOS and those without. In non-PCOS women, regular menstrual cycles are reported by 91.6%, while only 8.4% experience irregular periods. In contrast, only 27.8% of PCOS patients report regular menstruation, and a significant 72.2% experience irregular cycles. This stark contrast highlights the importance of evaluating menstrual patterns in the early identification and diagnosis of PCOS [5, 22]. Per the Rotterdam Criteria, PCOS is diagnosed when at least two of three features oligo or anovulation polycystic ovaries on ultrasound are present [6]. However, symptom variability often leads to underdiagnosis or misdiagnosis.

The complexity of PCOS and its overlapping symptoms with other disorders pose a challenge to traditional diagnostic approaches. Conventional methods involve hormonal assessments, physical examination, and imaging studies such as ultrasound. While effective, these methods are often time-consuming, expensive, and subject to variability in interpretation. This is where artificial intelligence, particularly deep learning (DL), can offer transformative solutions. DL models are capable of processing large volumes of heterogeneous data and can identify subtle patterns that may be overlooked by human clinicians [7]. They can integrate clinical records, hormonal profiles, ultrasound images, and even patient-reported symptoms to provide a holistic and accurate diagnosis.

In recent years, deep learning has demonstrated promising results in enhancing diagnostic performance for PCOS. These models can analyze ultrasound images to detect polycystic ovarian morphology more accurately and consistently [8]. Additionally, DL algorithms can be trained on structured clinical datasets to recognize early risk indicators, thereby enabling timely intervention. Beyond diagnosis, DL can also be applied in treatment planning, outcome prediction, and patient monitoring. It holds the potential to bridge the diagnostic gap, especially in low-resource settings where access to specialized care may be limited.

PCOS is a multifaceted disorder with far-reaching health implications. As highlighted by global data, it is both highly prevalent and underdiagnosed. The complications associated with PCOS—ranging from metabolic dysfunction to infertility and cancer risk—necessitate timely diagnosis and intervention. While conventional diagnostic methods remain essential, the integration of deep learning offers a new frontier in PCOS detection and management. By leveraging advanced algorithms capable of multi-dimensional data analysis, healthcare systems can improve early detection rates, personalize treatment plans, and ultimately enhance patient outcomes. Future research should focus on developing standardized datasets, improving model interpretability, and ensuring the ethical deployment of AI in clinical settings.

Literature Review:

Gopalakrishnan et al. [9] employed a comprehensive image preprocessing technique using a Gaussian low-pass filter to reduce noise, followed by segmentation through thresholding methods. The study utilized GIST-MDR features extracted from ultrasound images, which were then classified using a Support Vector Machine (SVM) classifier. This traditional ML pipeline demonstrated success in detecting and classifying PCOS-affected ovaries from normal ones. The strength of this approach lies in its clear and interpretable preprocessing steps and classification methodology. However, the feature extraction technique used is based on manual and handcrafted features, which may limit its robustness and generalizability across diverse datasets. The authors suggested that future work should consider employing deep learning-based feature extractors, which can automatically learn and generalize high-level representations from large-scale ultrasound data.

Alamoudi et al. [10] advanced the diagnostic process by incorporating a deep learning framework that not only analyzed ultrasound images but also integrated clinical data such as hormone levels and patient history. Their model leveraged feature fusion techniques to combine image-based and clinical features for a more comprehensive diagnosis. This multimodal approach enhanced the detection of Polycystic Ovary Morphology (PCOM), a key indicator of PCOS. The strength of this study lies in its holistic perspective on diagnosis, recognizing that PCOS is multifactorial and requires more than just imaging for accurate assessment. Nevertheless, the research was conducted on a limited sample size, and the model's generalizability remains questionable without clinical validation on larger and more diverse populations. This underscores the need for expansive and diverse datasets for training and testing deep learning models in real-world clinical settings.

Nilofer et al. [11} approached the problem through adaptive k-means clustering for segmenting ultrasound images. The extracted features included textural elements such as the Gray Level Co-occurrence Matrix (GLCM), which were used to train an Artificial Neural Network (ANN) optimized by an Improved Firefly Optimization Algorithm (IFFOA). The focus of this study was on infertility diagnosis, with a specific emphasis on follicle classification. This approach provided improved accuracy in identifying key ovarian structures that signify PCOS. The innovation in this study lies in the integration of optimization algorithms with traditional ANN models, which improved feature selection and classification performance. However, the reliance on traditional ANN architectures presents a limitation, as these may not match the performance of modern deep learning models such as convolutional neural networks (CNNs) or transformers. Replacing the ANN with state-of-the-art DL models could further improve diagnostic accuracy.

Suha.et.al[12] utilized a hybrid approach combining convolutional neural networks and traditional machine learning. The CNN architecture employed was VGGNet-16, known for its depth and strong performance in image recognition tasks. Features extracted from this CNN were used as input for a stacking ensemble method that included XGBoost, a powerful gradient-boosting algorithm. This hybrid model capitalized on the CNN's feature extraction capabilities and the ensemble model's strength in classification. The main contribution of this study lies in demonstrating that combining deep and traditional methods can enhance performance. However, the research could benefit significantly from larger and more diverse datasets, as deep learning models tend to overfit when trained on small datasets. Future studies should aim to incorporate data augmentation and transfer learning techniques to address data scarcity.

Maheswari et al. [13] explored a relatively novel methodology by first applying noise reduction techniques to ultrasound images, followed by the Furious Flies Optimization algorithm for feature extraction. Classification was performed using both Bayesian classifiers and ANNs. This approach is noteworthy for its attempt to identify novel features that may not be evident through traditional methods. The study achieved good performance in early PCOS detection, particularly in identifying key morphological features. However, the use of older ML models for classification remains a limiting factor. The study could be significantly enhanced by integrating CNNs, which have proven effective in capturing spatial hierarchies in image data. Moreover, validating this approach in a clinical setting would improve its credibility and potential for real-world application.

Upreti et al. (2025) [14] conducted a hybrid analysis combining hormonal profiles with ultrasound imaging data. The study proposed an AI/ML-based diagnostic model that aimed not only to detect PCOS early but also to predict long-term complications such as metabolic syndrome and infertility. The model incorporated various AI algorithms to analyze patterns and correlations among features. This integrative approach reflects a more comprehensive understanding of PCOS, which is not solely an imaging-based diagnosis. The strength of this work is its focus on early intervention and longitudinal care. However, the research lacks integration with clinical decision support systems (CDSS), which are essential for translating AI findings into actionable insights for healthcare professionals. Future research should focus on embedding such AI models into CDSS frameworks to enhance usability and adoption in clinical environments.

The ICAART 2025 [15] systematic literature review evaluated existing AI algorithms used for PCOS detection, with a special focus on explainability and optimization of diagnostic protocols. The review compiled studies that employed various ML and DL techniques, offering insights into the efficacy and limitations of each. It highlighted the importance of model transparency, which is crucial in healthcare settings where decisions must be justified to patients and clinicians. While this review synthesized valuable information, it was limited to existing literature and lacked practical deployment insights. The authors concluded that while AI shows promise in improving diagnostic workflows, there is an urgent need for clinical deployment studies that test these algorithms in real-world settings. Future reviews should also incorporate quantitative meta-analyses to assess the pooled performance metrics across studies.

Divekar & Sonawane (2025) [16] explored the use of transfer learning with InceptionV3, a deep CNN architecture pretrained on ImageNet. By fine-tuning this model on ultrasound frames of ovaries, they aimed to automate the detection of PCOS-related morphological features. This approach demonstrated strong performance, benefiting from the model's prior exposure to large image datasets. The primary strength of this study is the effective use of transfer learning, which allows for efficient training with limited medical data. Nonetheless, the generalizability of the model to diverse populations and imaging conditions needs thorough evaluation. Differences in ultrasound machines, operator techniques, and patient demographics can significantly affect image characteristics, posing challenges for model robustness.

Jothimani et al. (2025) [17] developed a modified Darknet-53 model integrated with probabilistic optimization techniques and feature fusion strategies. This model was tailored to enhance the performance of PCOS diagnosis from ultrasound images. The unique aspect of this study is the integration of probabilistic models to manage uncertainties in feature extraction and classification. Such a probabilistic approach is particularly useful in medical diagnostics, where there is inherent variability in patient data. The authors reported improved accuracy and robustness. However, the model's computational requirements could limit its applicability in real-time or resource-constrained environments. Further research is needed to optimize these models for deployment on edge devices or in low-resource healthcare settings.

FOSET 2025 [18] conducted a benchmarking study comparing the performance of several well-known CNN architectures including VGG16, ResNet50, DenseNet121, and MobileNetV2. Each of these architectures has different strengths in terms of depth, parameter efficiency, and speed. The study evaluated their performance in classifying ultrasound images for PCOS detection. Such benchmarking is essential to guide model selection for specific clinical needs. For example, MobileNetV2, known for its lightweight architecture, could be ideal for mobile or point-of-care applications. The study's strength lies in providing a comparative analysis, but it stops short of integrating these models into actual clinical workflows. Future work should focus on incorporating these CNNs into complete diagnostic systems, including user interfaces, feedback mechanisms, and integration with electronic health records (EHRs).

The Systematic Review (2025) [19] served as a comprehensive overview of AI tools applied to gynecological ultrasound diagnostics, with a focus on PCOS. It categorized the tools based on their application—segmentation, classification, feature extraction, and decision support. The review emphasized the growing interest in AI for women's health and highlighted the challenges associated with data privacy, regulatory approval, and clinician acceptance. One notable contribution was its discussion on the ethical implications of AI in reproductive health. However, the review remains largely descriptive and lacks a quantitative synthesis of findings. Including meta-analytical techniques could provide stronger evidence for the effectiveness of specific AI models.

This research presents CystNet, a hybrid deep learning model combining InceptionNet V3 and a Convolutional Autoencoder for accurate PCOS detection using ultrasound images. The model extracts both high-level and nuanced features, forming a 3072-dimensional representation. Preprocessing includes augmentation and segmentation to enhance model focus. Two classification approaches are explored: a deep learning classifier achieving 96.54% accuracy and a Random Forest model reaching 97.75%. Evaluations on multiple datasets (Kaggle, PCOSGen, MMOTU) and 5-fold cross-validation confirm robustness and generalizability. The study outperforms existing methods, demonstrating strong potential for reliable, automated PCOS diagnosis through ultrasound imaging.

The literature on early detection of PCOS using AI and ultrasound imaging reveals a diverse and rapidly evolving field. Most studies demonstrate the potential of deep learning models, especially CNNs, in automating the detection of ovarian morphological features. While traditional ML methods like SVMs and ANNs are still used, they are increasingly being replaced or supplemented by modern DL architectures for better performance. A common theme across the studies is the need for larger, annotated datasets and clinical validation. Moreover, integrating these AI models with clinical decision support systems, ensuring model explainability, and addressing ethical concerns are crucial next steps. The future of PCOS diagnosis lies in creating holistic, explainable, and clinically integrated AI tools that can assist healthcare professionals in delivering timely and accurate diagnoses, ultimately improving patient outcomes.

Table.1: Summary of Literature on Early Detection of PCOS Using Deep Learning and Ultrasound Imaging

Authors	Method	Strength	Scope of Improvement
Gopalakrishnan et al. [9]	Preprocessing with Gaussian low pass filter, segmentation using thresholding, GIST-MDR features, SVM classifier	Detection and classification of PCOS via ultrasound images	Feature extraction method could be enhanced using deep learning models
Alamoudi et al. [10]	Deep learning for image analysis, fusion models integrating clinical data	PCOM detection through combined image and clinical features	Clinical validation on larger population necessary

Nilofer et al. [11]	Adaptive k-means clustering, GLCM features, ANN trained with IFFOA	Improved diagnosis of infertility through follicle classification	Replace traditional ANN with modern deep learning models for better accuracy
Suha.et.al[12]	CNN (VGGNet16) for feature extraction, stacking ensemble with XGBoost	Detection of PCOS via hybrid deep learning–traditional ML model	May benefit from larger and more diverse datasets
Maheswari et al. [13]	Noise reduction, Furious Flies algorithm for features, Bayesian and ANN classifiers	Novel feature identification for early PCOS detection	Modern CNN-based models could improve performance
Upreti et al. [14]	Analysis of hormone and ultrasound data, hybrid AI/ML models	Accurate early diagnosis and long-term complication prevention	Needs integration with clinical decision systems
SA Suhaet.al [15]	Systematic literature review of AI algorithms for PCOS	Diagnosis optimization using AI; emphasis on explainability	Limited to existing research; suggests need for clinical deployment
Divekar & Sonawane et.al [16]	Transfer learning using Inception V3	Automated PCOS detection from ultrasound frames	Model generalizability needs evaluation
Jothimani et al. [17]	Modified Darknet-53, feature fusion with probabilistic optimization	PCOS diagnosis using enhanced deep learning methods	Needs real-time or low-resource environment testing

PCOS is a common endocrine disorder diagnosed through clinical symptoms and ultrasound imaging, but variability in image quality and subjective interpretation can lead to inconsistent diagnoses. To mitigate this problem, we proposed a SonoCyst-Net model. The key contributions of the proposed modes are

- Deep Learning Framework: Utilizes a deep learning model for automated PCOS detection.
- EfficientNetV2 Backbone: Employs a robust and efficient feature extractor.
- Attention-Based Autoencoder: Enhances and reconstructs diagnostically significant patterns from ultrasound images.
- GAN-Based Data Augmentation: Addresses dataset limitations by generating synthetic training samples.
- Model Optimization: Uses pruning and quantization to reduce computational load for real-time deployment.
- Explainability: Employs Grad-CAM and SHAP to provide insights into the model's decision-making process.

We evaluated this model on the Kaggle PCOS ultrasound dataset, demonstrating high diagnostic accuracy and clinical relevance. The benefits of the model is Presents a promising AI-assisted tool for early and reliable PCOS detection. In essence, SonoCyst-Net is a sophisticated AI tool that aims to improve PCOS diagnosis by leveraging the power of deep learning and multi-modal data, ultimately leading to more accurate and timely diagnoses for women with PCOS.

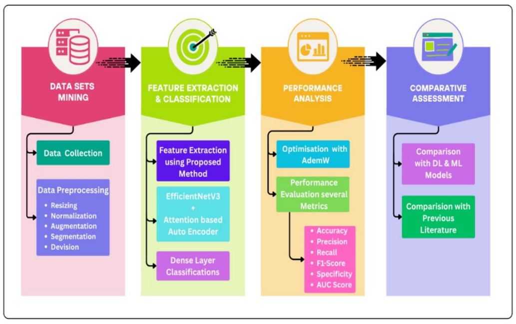


Fig. 4.Framework of the proposed model

METHODOLOGY

This section outlines the dataset employed for training and evaluating the diagnostic model, along with preprocessing steps including normalization, augmentation, and segmentation. It further details the proposed deep learning framework for detecting PCOS in ultrasound images and classifying cases as PCOS or non-PCOS.A visual representation of the complete study framework, including all key stages, is illustrated in Fig. 4.

Dataset Mining:

Data collection:

The dataset utilized in this study was obtained from Kaggle [23], comprising 1,924 training images and 1,932 testing images. Due to significant overlap between the original sets, the test set was discarded, and the training set was re-divided into new training and testing subsets. The dataset includes ultrasound images classified as 'INFECTED' (781 images of cystic ovaries) and 'NOT INFECTED' (1,143 images of normal ovaries). These labels effectively distinguish between individuals with PCOS and those without, making this dataset particularly suitable for developing real-time diagnostic tools. Representative samples from both classes are shown in Figure 5.

This study also incorporated the PCOSGen Dataset [24], compiled from various online sources. It consists of 2,610 ultrasound images, including 752 PCOS and 1,858 Non-PCOS cases. This dataset, medically annotated by a gynecologist in New Delhi, India, was exclusively used for model testing.

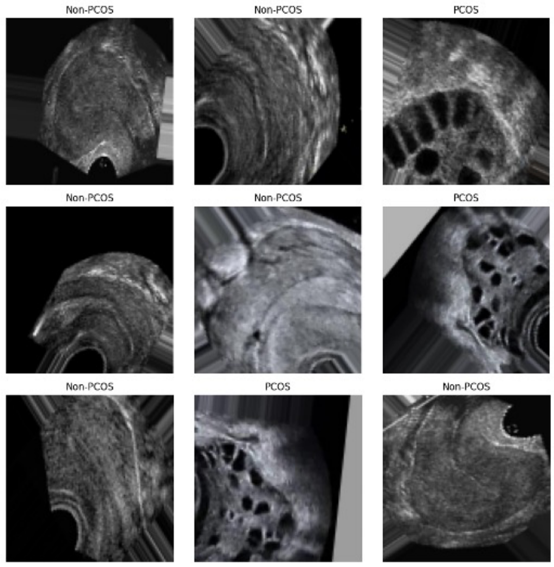


Fig.5. Comparative ultrasound images illustrating ovaries with PCOS and without PCOS.

The data preprocessing stage is essential for optimizing ultrasound images before feeding them into the diagnostic model. The preprocessing involves pixel value normalization (scaling to the [0,1] range) and multiple augmentation techniques, including rotation, horizontal flip, vertical shifting, zooming and shearing, to increase image diversity and strengthen model generalization during training.

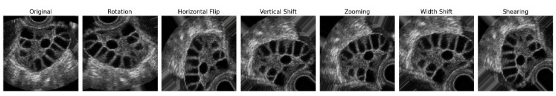


Fig.6. Data augmentation on ultrasound images

Data Processing and Augmentation:

All ultrasound images are resized to a fixed resolution of 224×224 pixels to maintain consistency throughout the dataset. This standardization is crucial for enabling CNN to process the inputs effectively and uniformly. Furthermore, it plays a significant role in reducing computational complexity and optimizing memory utilization during the training phase.

Data Augmentation:

To enhance the model's generalization capability and reduce the risk of overfitting, a variety of image augmentation techniques are employed.These augmentations help the model generalize better by Increasing dataset diversity without collecting new images and making the model more robust to real-world image variations like rotation, scale, translation, etc. The augmetation of the imgaes as shown in figure.6

These augmentations introduce controlled variability to simulate real-world scenarios and increase the diversity of training samples. The following transformations are performed:

Let the input image be represented as I(x,y), having dimensions H×W with pixel values lying in the range [0, 255].

Rescaling: The first step in the pipeline is rescaling the image pixel values from the typical range of [0,255] to [0,1]. This is achieved by dividing each pixel value by 255. Normalizing input values is a standard practice in deep learning, as it ensures that the data is within a range where neural networks converge more efficiently and maintain numerical stability during training.

$$I'(x,y) = \frac{I(x,y)}{255} \tag{1}$$

Rotation: Images are randomly rotated by an angle θ∈ [−180°, +180°], enhancing the model's rotational invariance. The transformation is expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

3. Zooming: Zooming in or out on an image (scaling transformation) alters the size of the object in the field of view. In the described augmentation, the zoom factor is chosen randomly up to 20%. This operation enables the model to handle different object sizes and distances, teaching it to focus on fine-grained and coarse-level features alike. It simulates camera zoom effects or distance changes between the sensor and the object.

Zoom operations are applied by scaling the image with a factorZ∈[1-Zmax,1+Zmax]:

$$(x', y') = \left(\frac{x - x_c}{z} + x_c\,,\ \frac{y - y_c}{z} + y_c\right) \tag{3}$$

4. Shifting: Images are shifted width and height by a 5% of the total image width and height. This simulates object displacements within the frame and trains the model to detect features irrespective of their position. It encourages spatial invariance, helping models detect patterns even when they appear at unexpected locations.

$$x' = x + \Delta x \tag{4}$$

For a height shift $\Delta y=\beta W$, where $\beta\in[-s_h, s_h]$:

$$y' = y + \Delta y \quad (5)$$

5. Shearing: Shearing involves slanting the image by a specific angle, resulting in a transformation that shifts pixels in one axis (usually horizontal) as a function of the other. This affine transformation can mimic subtle changes in perspective, such as viewing an object from a slightly different angle. It is particularly useful in datasets where viewpoint distortion is possible but not explicitly covered.

With shear angle $\lambda\in[-\lambda_{max}, \lambda_{max}]$

$$\begin{bmatrix}x'\\y'\end{bmatrix}=\begin{bmatrix}1&\tan\lambda\\0&0\end{bmatrix}*\begin{bmatrix}x\\y\end{bmatrix} \quad (6)$$

6. Fill Mode (Nearest):Transformations like rotation, zoom, or shift may move pixels outside the original image frame, creating empty areas. The fill mode determines how these areas are handled. The 'nearest' fill mode assigns the value of the closest pixel to the empty region, maintaining texture continuity and minimizing visual artifacts. This is critical for avoiding the introduction of unnatural boundaries or patterns that could confuse the model.

For any pixel position (x',y') falling outside bounds, assign:

$$I(x',y')=I(x_n, y_n), \text{ where } (x_n, y_n)= \text{nearest valid neighbor} \quad (7)$$

Dataset Division:

The Kaggle dataset is divided into two subparts, with 80% for training the model and 20% reserved for validation and testing of the trained model. The training set is utilized to help the model learn meaningful patterns and features associated with PCOS.

In contrast, the testing set serves to assess the model's effectiveness, ensuring it can generalize well and accurately identify PCOS in previously unseen ultrasound images.

Feature Extraction

Feature extraction is the process of transforming raw ultrasound images data into a set attributes that highlight the most critical information while minimizing complexity. This step enhances model accuracy and computational efficiency by emphasizing the most relevant features. In the proposed approach, two advanced techniques—EfficientNet-B3 and a Variational Autoencoder enhanced with interactive attention—are employed for this purpose. These components are integrated into a unified architecture named SonoCyst-Net, which leverages the advantages of both models to perform robust and effective feature extraction, as illustrated in Figure. 7. A detailed explanation of each technique follows.

In the proposed SonoCyst-Net framework, EfficientNet-B3 is employed to extract deep high-level features from ovarian ultrasound images. Its architecture integrates multiple convolutional layers, squeeze-and-excitation blocks, and Swish activation functions, enabling the effective capture of complex spatial and structural patterns in medical imaging.

EfficientNet-B3 utilizes a compound scaling strategy that proportionally increases network depth, width, and input resolution to achieve high accuracy while maintaining computational efficiency. Its architecture is built on a series of MBConv (Mobile Inverted Bottleneck Convolution) block specialized inverted residual structures that integrate depthwise separable convolutions, expansion layers, and squeeze-and-excitation (SE) attention mechanisms.

The input ultrasound images are resized to a resolution of 224×224 pixels, matching EfficientNet-B3's expected input size. As the image passes through the network, it undergoes multiple stages of transformation, extracting a hierarchical set of features ranging from low-level textures to high-level semantic representations. The final feature vector generated by EfficientNet-B3 contains rich contextual information, which is crucial for distinguishing between PCOS and non-PCOS cases.

EfficientNet-B3 is pre-trained on large-scale datasets like ImageNet, and then fine-tuned on the ultrasound dataset to adapt its filters and weights to the specific characteristics of ovarian ultrasound images. This transfer learning approach significantly improves model convergence and accuracy, especially when training data is limited.

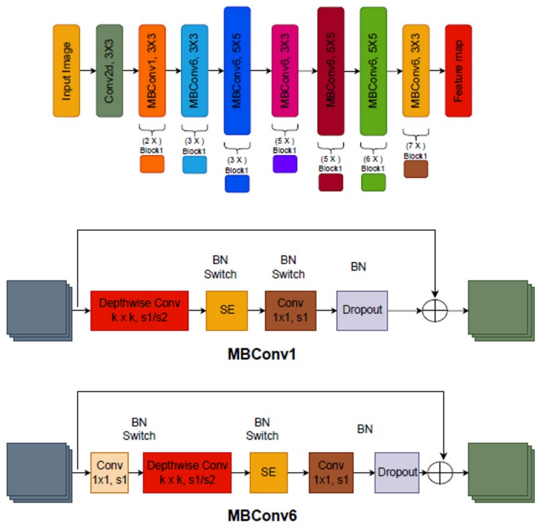


Fig.7 Architecture of the EfficientNet-B3 block

The EfficientNet-B3 network begins with a standard Conv2D layer (3×3) followed by MBConv blocks with varying kernel sizes (3×3, 5×5) and expansion ratios. Each group of MBConv layers is repeated multiple times (e.g., 2×, 3×, 5×, 6×, etc.) to gradually increase the receptive field and learn more abstract features. The final feature map is extracted after the last MBConv stack, which is then used for downstream tasks such as classification.

Two major types of MBConv blocks are used:

MBConv1: Starts with a Depthwise Convolution (kernel size k × k, stride s), followed by Batch Normalization and an SE block.A pointwise convolution (1×1) refines the features. Dropout is applied before the residual connection. It is used for initial layers where feature maps are smaller.

MBConv6: Each block starts with an expansion layer (1×1 convolution) to broaden the channel dimension, followed by a depthwise convolution, a squeeze-and-excitation (SE) block, and concludes with a pointwise convolution. Dropout and residual connection help in retaining important features and preventing overfitting. Used in deeper layers for learning complex, abstract features.

Compound Scaling:

EfficientNet uses a compound scaling strategy, where resolution (r), width (w), and depth (d) are scaled together with a coefficient (ϕ), ensuring balanced accuracy and efficiency.

$$d=\alpha^{\phi},\quad w=\beta^{\phi},\quad r=\gamma^{\phi}, \tag{8}$$

Here, ϕ represents the compound coefficient defined by user and the following α, β, γ are constants determined via grid search, respectively.

Expansion Phase (Only in MBConv6): If the expansion ratio is t, then the number of channels is increased

$$X_{exp}=\mathrm{Conv}1\times1(X), \tag{9}$$

Depthwise Convolution:Each channel is convolved independently (no cross-channel mixing)

$$X_{dw}=\mathrm{DWConv}\ k\times k(X_{exp}), \tag{10}$$

Squeeze (Global Average Pooling):

$$Z_c=\frac{1}{H.W}\sum_{i=1}^{H}\sum_{J=1}^{H} X_c(i,j) \text{ for each channel } C \tag{11}$$

Excitation (Two Fully Connected Layers):

$$S=\sigma(W2\cdot\delta(W1\cdot z)) \tag{12}$$

Feature Reweighting:

$$\hat{X_c}=S_c X_c \tag{13}$$

Projection (Pointwise Convolution):

$$X_{Proj}=\mathrm{Conv}\ 1\times1(\hat{X_{dw}}) \tag{14}$$

Residual Connection:

$$F_{Eff}=\begin{cases} X+X_{Proj,} & \text{if shapes match} \\ X_{Proj} & \text{otherwise} \end{cases} \tag{15}$$

EfficientNet-B3 applies a compound scaling approach to uniformly adjust the network's depth, and input resolution and width using a coefficient ϕ, as described in Eqn. (8). Its fundamental unit is the MBConv block, which begins with an expansion step (Eqn. 9) using a convolution (1×1) to increase feature dimensions. Next, a depthwise separable convolution (Eqn. 10) is applied, significantly lowers computational cost by carrying out convolution operations separately on each channel.

The output from the depthwise convolution is subsequently passed through a SE block to capture channel-wise attention. This process starts with global average pooling (Eqn. 11) to generate a channel descriptor, which is then passed through two fully connected layers with ReLU and sigmoid activations (Eqn. 12) to derive the attention weights. The weights rescale the feature maps (Eqn. 13), highlighting important channels while suppressing less relevant ones.

The rescaled feature maps undergo a projection phase via another 1×1 convolution to restore the original channel dimension in Eqn.(14). When the input and output dimensions align, a skip connection is added to form a residual block (Eqn. 15), facilitating gradient flow and supporting deeper network training. These operations, organized within several MBConv variants across different stages, collectively form the EfficientNet-B3 backbone, optimized for both efficiency and accuracy in visual recognition tasks.

In the SonoCyst-Net model for PCOS detection from ultrasound images, EfficientNet-B3 is employed to extract deep and discriminative visual features. Its SE blocks enhance channel-wise attention, allowing the network to emphasize relevant ovarian structures while suppressing noise or irrelevant background features. These blocks not only improve parameter efficiency but also allow deeper models to be trained with fewer resources. The hierarchical feature maps generated by EfficientNet-B3 serve as a robust foundation for further processing by downstream components like the Variational Autoencoder and classifier.

Variational Autoencoder

In the proposed SonoCyst-Net framework for PCOS classification, the VAE with Interactive Variational Attention (IVA) plays a crucial role in enhancing and refining the features extracted from ultrasound images by EfficientNet-B3. While EfficientNet-B3 extracts high-dimensional spatial and semantic features from ovarian ultrasound images, these features may still contain redundancy or irrelevant patterns due to variations in patient anatomy, image quality, or noise.

The VAE framework is introduced to compress these features into a more compact, meaningful latent representation by learning probabilistic distributions. This helps the model generalize better and avoid overfitting. At the same time, the Interactive Variational Attention module introduces a secondary latent pathway that learns to focus selectively on the most informative regions or attributes of the encoded feature space—such as the shape, number, and arrangement of follicles and cysts—which are critical for distinguishing between PCOS and non-PCOS cases.

By modulating the decoder layers with these attention vectors, the model dynamically enhances the most relevant patterns during reconstruction. This boosts the model's sensitivity to key diagnostic cues and ensures that downstream classifiers receive clean, focused, and robust features for final decision-making.

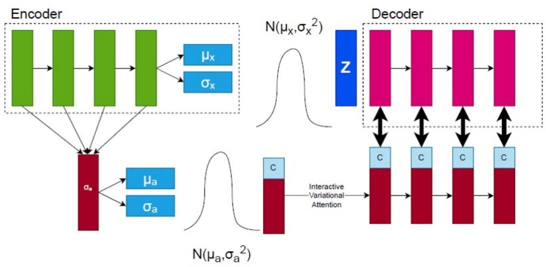


Fig.8 Architecture of Variational Autoencoder with Interactive Variational Attention

The FEff∈RH×W×Cbe the feature map extracted by EfficientNet-B3 from input ultrasound images.

The encoder learns to map this feature into a latent distribution:

$$M\_x,\sigma\_x = \text{Encoder}\ (F\_Eff) \quad (16)$$

$$Z \sim N(\mu\_x, 〚\sigma\text{^}2〛\_x) \quad (17)$$

An auxiliary attention encoder simultaneously generates an attention-guided latent distribution:

$$M\_a,\sigma\_a = \text{AttentionEncoder}\ (F\_Eff) \quad (18)$$

$$a \sim N(\mu\_a, 〚\sigma\text{^}2〛\_a) \quad (19)$$

The standard latent vector z is fused with attention vector a to produce an attention-refined latent representation:

$$\tilde{Z} = Z \odot \sigma(W\_a + b\_a\ a) \quad (20)$$

Here, W_aand b_a aare learnable weights, σ is the sigmoid activation, and ⊙ represents element-wise multiplication.

The refined latent code $\tilde{Z}$ is decoded to reconstruct the feature map:

$$\hat{F} = \text{Decoder}(\tilde{Z}) \quad (21)$$

The training loss includes:

$$L\_total = L\_recon\ (\hat{F}, F\_Eff) + \beta\_1.D\_KL\ (N(\mu\_x, 〚\sigma\text{^}2〛\_x)\,||\,N(0,I)) + \beta\_2.D\_KL\ (N(\mu\_a, 〚\sigma\text{^}2〛\_a)\,||\,N(0,I)) \quad (22)$$

Where, L_recon represents the Reconstruction Loss,Z represents the KL Divergence for standard latent vector and a represents the KL Divergence for attention latent vector.

To improve the diagnostic capability of the system, the high-level features extracted from ultrasound images via EfficientNet-B3 are passed into a VAE with Interactive Variational Attention. The encoder processes the features to generate a mean μx and variance σx2 representing the latent distribution (Eq. 16–17), while a parallel attention encoder computes another set of variational parameters μa and σa2 (Eq. 18–19) aimed at capturing attentive contextual patterns. The attention-guided latent vector modulates the original latent representation via a gating mechanism using sigmoid activation (Eq. 20). This enhanced latent code z̃ is decoded (Eq. 21) to reconstruct refined features that emphasize pathological patterns relevant to PCOS, such as polycystic follicles, ovarian shape, or stromal echogenicity. The learning objective (Eq. 22) ensures the model balances reconstruction accuracy with well-regularized latent spaces via KL divergence, enabling better generalization and robustness. This combined approach leverages both spatial and contextual attention for improved PCOS diagnosis.

Final Classification Using Dense Layers:

In the proposed architecture for PCOS classification, the final stage utilizes a Dense Neural Network to classify the reconstructed features F̂ generated by the Decoder of the VAE with Interactive Attention. Dense classifier maps a high-level reconstructed representation to a binary decision indicating whether PCOS is present or absent.

Input Layer: The dense network receives the reconstructed feature vector $\hat{F}$ from the decoder, represented with n features as:

$$\hat{F}=[\hat{f}_1,\hat{f}_2,\hat{f}_3,\hat{f}_4,\ldots\ldots\hat{f}_n] \quad (23)$$

Dense Layers: Each dense (fully connected) layer applies a linear projection followed by a non-linear activation. For the l^{th} layer, the operation is:

$$Z^{(l)}=W^{(l)}.a^{(l-1)}+b^{(l)} \quad (24)$$

where $a^{(l-1)}$ is the ReLU activation from the previous layer. The ReLU activation function is applied as:

$$a^{(l)}=\mathrm{ReLU}(Z^{(l)})=\max(0,Z^{(l)}) \quad (25)$$

Dropout Layers: To improve generalization and prevent overfitting, dropout is applied during training. A dropout mask $d^{(l)}$ is a Bernoulli distribution, and applied as:

$$a^{(l)}_{dropout}=a^{(l)}.d^{(l)} \quad (26)$$

Output Layer: The final dense layer uses sigmoid activation to predict the probability of PCOS:

$$\hat{y}=\sigma(W_{out}\ .a^{(L)}+b_{out}) \quad (27)$$

Where W_{out} and b_{out} are the output layer's weights and bias, and $a^{(L)}$ is the final hidden activation

This structure enables the dense classifier to make accurate and interpretable binary predictions based on the feature-rich output $\hat{F}$ reconstructed by the variational decoder. The incorporation of interactive attention mechanisms during encoding and decoding ensures that the most relevant features are preserved for effective downstream classification.

4. Experimental results

To effectively measure the model's performance, a set of evaluation metrics has been employed. These metrics offer a well-rounded assessment of the model's capability to accurately classify ultrasound images and diagnose PCOS. The basis for these evaluations lies in the confusion matrix (CM), which compares the model's predicted classifications to the actual outcomes. Fig.9 represents the confusion matrix of model.

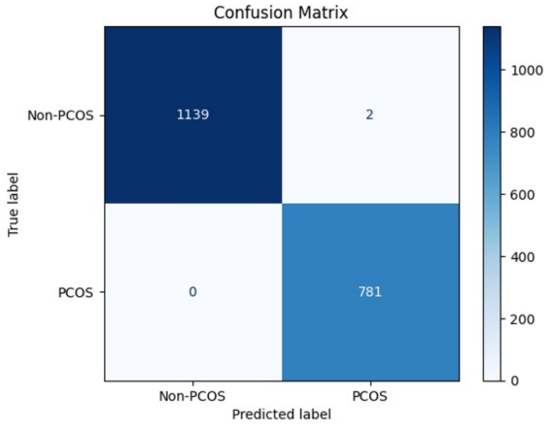


Fig.9 Confusion matrix of Proposed model on the Kaggle PCOS ultrasound image dataset

The key components of the confusion matrix are as follows:

True Positive (TP): Correctly predicts PCOS in a patient who has the condition.

True Negative (TN): Correctly predicts absence of PCOS in a patient without the condition.

False Positive (FP): Incorrectly predicts PCOS in a patient who does not have it.

False Negative (FN): Incorrectly predicts absence of PCOS in a patient who actually has it.

Using these components, various performance metrics are calculated, each offering insight into different aspects of the model's diagnostic accuracy. Figure 10 shows the metric values of the proposed model on the dataset.

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \quad (28)$$

$$Precision(P)=\frac{TP}{TP+FP} \quad (29)$$

$$Recall(R)=\frac{TP}{TP+FN} \quad (30)$$

$$F1\ Score=\frac{2\times Precision\times Recall}{Precision+Recall} \quad (31)$$

$$Specificity=\frac{TN}{TN+FP} \quad (32)$$

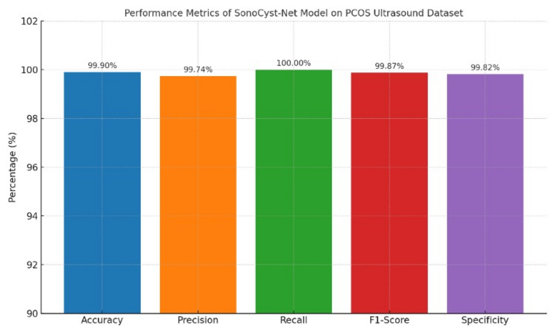


Fig.10 Parameters of proposed model on the Kaggle PCOS ultrasound image dataset

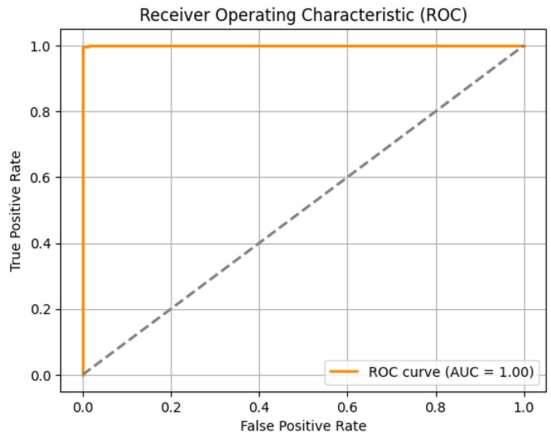


Fig.11 ROC_AUC Curve on the Kaggle PCOS ultrasound image dataset

ROC-AUC:

The ROC-AUC score measures the model's ability to distinguish between classes, with higher values indicating better performance. Fig. 11. ROC curve of the proposed model on the Kaggle dataset. Each metric offers a unique perspective on the model's classification performance. Accuracy, Precision, Recall, F1 Score, Specificity, and ROC-AUC were chosen for comprehensive evaluation, assessing the model's ability to detect both positive and negative cases, balance false positives and false negatives, and reliably differentiate between classes. In particular, a high Recall demonstrates the model's effectiveness in identifying PCOS cases.

Result Analysis:

The proposed PCOS detection framework was implemented on a system with a 13th-gen Intel Core i7, 32 GB RAM, using Python tools. Results confirm its effectiveness in diagnosing PCOS from ultrasound images, showing clear improvements over conventional methods.

The proposed SonoCyst-Net model is an AI-based diagnostic system that uses transfer learning and deep learning to detect and classify PCOS from ultrasound images. Preprocessing steps such as resizing, normalization, augmentation, segmentation, and partitioning aid in accurate identification of ovarian follicles and cysts, minimizing manual interpretation errors.

The SonoCyst-Net hybrid model combines EfficientNet-B3 with a VAE and Interactive Variational Attention to highlight key regions in ultrasound images and distinguish PCOS from non-PCOS. It was trained on the Kaggle dataset with 1,924 training, 559 validation, and 1,922 testing images.

The deep learning model is optimized with a refined set of hyperparameters to improve training efficiency and generalization, as shown in Table 2.

Table.2 Hyperparameters after optimization

Hyperparameters	Values
dense_units	128
dropout_rate	0.2
learning_rate	0.001
batch_size	32
epochs	50
optimizer	AdamW
kernel_size	3x3
kernel_initializer	he_normal
kernel_regularizer	L2(0.001)
weight_decay	1e-4
learning_rate_scheduler	Gradually decrease Lea if no improvement
rotation_range	180
zoom_range	0.2
batch_normalization	True
early_stopping	True
width_shift_range	0.05
height_shift_range	0.05
shear_range=0.1	0.1
horizontal_flip	True
fill_mode	nearest

To ensure robustness and generalizability, the model employs a comprehensive augmentation strategy. A rotation range of 180 degrees allows the network to learn orientation-invariant features. The model also supports zoom augmentation (0.2) to simulate variability in scale, while width and height shift ranges of 0.05 simulate slight translations in ultrasound frames. Shear transformations (0.1) distort images slightly to encourage better feature adaptability. Enabling horizontal flipping helps the model generalize to mirrored images, common in ultrasound imaging. All these augmentations use the 'nearest' fill mode to handle pixel gaps created during transformation, preserving anatomical correctness.

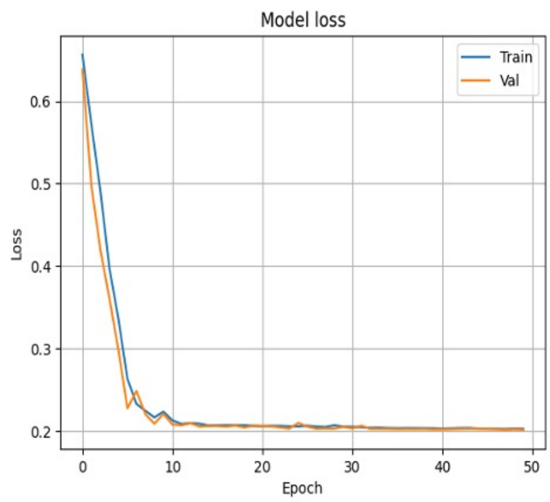
The architecture includes 128 dense units, which balance model complexity and computation. These fully connected neurons play a key role in high-level feature representation after convolutional and pooling layers. A dropout rate of 0.2 ensures regularization by randomly deactivating 20% of the neurons during training, which reduces overfitting—a common concern with medical imaging datasets.

The learning rate is set at 0.001, a moderate value that allows stable and efficient convergence. Additionally, a weight decay of 1e-4 is introduced through the AdamW optimizer, which combines the benefits of Adam with decoupled weight regularization, helping prevent overfitting by penalizing large weights. The batch size is 32, which offers a good trade-off between gradient stability and GPU efficiency. The model is trained for 50 epochs, and training is further regulated using early stopping—halting the process if no improvement in validation loss is observed, thus saving time and avoiding overfitting.

Convolutional layers use a kernel size of 3x3, which is effective for capturing fine-grained spatial features. The weights are initialized using He normal initialization, ideal for layers with ReLU activation, ensuring stable gradient flow in deep networks. Additionally, L2 regularization (with a penalty of 0.001) helps in constraining the model complexity by shrinking unnecessary weight values.

The batch normalization parameter is set to True, allowing the model to normalize internal activations layer-by-layer, leading to faster and more stable training. Lastly, the learning rate scheduler is configured to gradually reduce the learning rate upon plateauing performance, promoting convergence toward a better local minimum and avoiding overshooting during optimization.

Further validation is supported by the accuracy and loss curves presented in Fig 12, which depict stable training and validation trends with minimal overfitting. at the 7th to 10th epochs. Later, it exhibited consistent performance till the 50th epoch on the Kaggle PCOS ultrasound image dataset, confirming its generalizability and robustness.



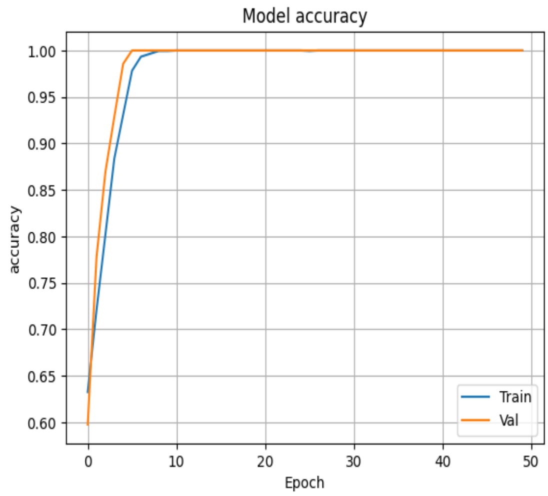


Fig. 12 Model Loss and Model Accuracy curves

Figure 10 shows that the SonoCyst-Net model achieved excellent performance on the Kaggle PCOS dataset, with an accuracy of 99.90%, recall of 100%, precision of 99.74%, F1-score of 99.87%, and specificity of 99.82%. Figure 11 further demonstrates a high ROC-AUC value of 1.00.

The Table 3 shows the Performance comparison between the proposed model with baselines. These results highlight the model's high diagnostic accuracy and reliability, outperforming other deep learning architectures such as InceptionNet V3, Autoencoder, ResNet50, DenseNet121, EfficientNetB0 and CystNet.

Table 3. Performance comparison between the proposed models with baselines

Model	Accuracy (%)	Recall (%)	F1-Score (%)	Precision (%)	Specificity (%)
InceptionNetV3	94.68	95.97	95.01	94.08	93.23
Autoencoder	93.97	97.31	94.46	91.77	90.22
ResNet 50	95.55	97.22	95.67	94.17	93.03
DenseNet 121	94.32	96.64	94.73	92.90	91.72
Efficient Net B0	92.98	92.20	92.36	92.53	93.75
Proposed SonoCyst-Net	**99.90**	**100**	**99.87**	**99.74**	**99.82**

The performance evaluation of various models for PCOS detection from ultrasound images reveals critical insights into their effectiveness and limitations. Each model brings a unique architecture and methodological strength to the classification task. InceptionNet V3, for instance, leverages multi-scale convolutional filters that efficiently extract diverse spatial features. It demonstrates strong recall (95.97%), reflecting its ability to detect PCOS cases effectively. However, it slightly underperforms in specificity (93.23%), indicating a tendency toward false positives.

The Autoencoder model, primarily unsupervised, excels in reconstructing inputs and learning latent representations, which translates into a remarkably high recall of 97.31%. Nevertheless, it suffers in precision (91.77%) and specificity (90.22%), suggesting it over-detects positive cases and may generate more false alarms. ResNet 50, with its deep residual learning architecture, performs robustly across all metrics, achieving a balance between depth and trainability. It achieves a high F1-score (95.67%) and strong recall and accuracy, though its specificity (93.03%) still leaves room for improvement.

DenseNet121, known for its efficient feature propagation and reuse, offers a good trade-off in accuracy (94.32%) and F1-score (94.73%). However, its relatively lower specificity (91.72%) and precision (92.90%) suggest it could still misclassify some non-PCOS cases. EfficientNet B0, although lightweight and optimized for computational efficiency, slightly trails in overall performance with accuracy at 92.98% and recall at 92.20%, making it less ideal for clinical applications where both sensitivity and precision are critical.

SonoCyst-Net achieves an almost perfect classification performance accuracy of 99.90%, precision of 99.74%, recall of 100%, and specificity of 99.82%. The 100% recall ensures that no PCOS cases go undetected, which is critical in medical screening. Meanwhile, the highest specificity indicates an extremely low false positive rate, increasing trustworthiness in real-world clinical deployment. The performance improvements are significant. Compared to ResNet 50, the proposed model improves accuracy by 4.35%, precision by 5.57%, recall by 2.78%, and specificity by 6.79%. When compared to EfficientNet B0, the improvements are even more substantial: 6.92% in accuracy, 7.21% in precision, and 7.80% in recall.

The proposed SonoCyst-Net exhibits superior performance across all evaluation metrics. Its advanced hybrid architecture and attention mechanisms enable it to outperform established deep learning models, especially in critical medical evaluation parameters like recall and specificity, making it highly suitable for clinical deployment.

The proposed method was also tested on the PCOSGen dataset, which contains 2,610 images (752 PCOS and 1,858 Non-PCOS). The confusion matrix heatmap is shown in Fig.13. As presented in Fig.14, the model achieved an accuracy of 99.01%, precision of 98.10%, recall of 98.72%, F1-score of 98.41%, and specificity of 99.14%.

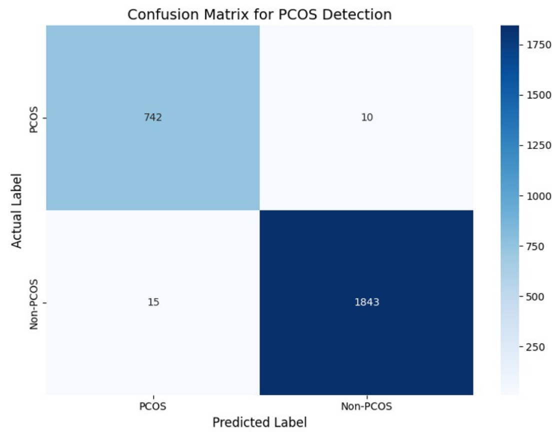


Fig.13 Confusion matrix on the PCOSGen dataset

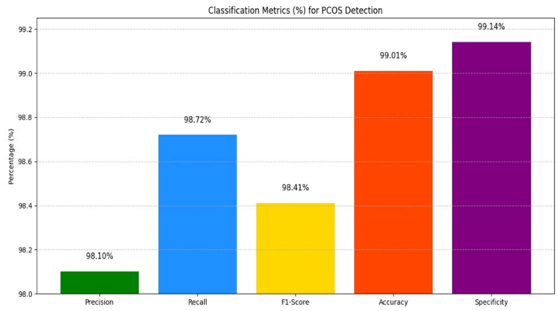


Fig14. Parameters of proposed model on the PCOSGen dataset

These findings affirm the effectiveness and adaptability of the proposed framework in diverse clinical scenarios, reinforcing its potential as a reliable diagnostic tool for PCOS detection.

The innovation of our approach lies in the development of the SonoCyst-Net hybrid architecture, which combines the feature extraction capabilities of InceptionNet V3 and a Convolutional Autoencoder, enhanced with Interactive Variational Attention. This sophisticated configuration enables comprehensive feature representation by extracting deep and fine-grained characteristics from ultrasound images, thereby improving the overall classification accuracy.

Unlike earlier models that often rely on single-stage learning or limited integration techniques, the novelty of SonoCyst-Net lies in its multi-stage hybrid design. However, despite these advancements, some limitations persist. The primary constraint is the use of a single-source dataset, which may not fully capture the demographic and clinical diversity found in real-world populations. This limitation could impact the model's generalizability in diverse clinical scenarios.

Additionally, the computational complexity of the SonoCyst-Net architecture—particularly due to the integrated use of InceptionNet V3 and convolutional autoencoders—can pose challenges for deployment on low-resource hardware. Future work should aim to include multi-center or multi-source datasets to improve robustness. Further development is also needed to enhance computational efficiency and create intuitive interfaces for seamless integration into clinical workflows.

Nevertheless, the experimental outcomes underscore the significant potential of the proposed framework in transforming PCOS diagnosis through automated ultrasound image analysis. By improving diagnostic accuracy and reducing manual burden, this framework supports timely clinical interventions and may substantially improve outcomes in women's reproductive health and well-being.

Table 4. Performance comparison between the proposedSonoCyst-Net model with existing models

Model	Accuracy (%)	Recall (%)	F1-Score (%)	Precision (%)	Specificity (%)
SVM (with RBF kernel) [18]	93.00	91.10	91.90	92.80	93.40
Random Forest (RF) [18]	94.10	93.00	93.20	93.50	94.80
CNN (Basic 5-layer)[19]	95.70	94.80	95.00	95.30	96.20
LSTM + GRU Fusion [19]	96.85	96.70	96.40	96.10	97.00
Efficient NetB0 (Image Only) [20]	97.50	96.80	96.90	97.10	97.90
HybridCNN-Autoencoder + XGBoost [21]	98.30	97.90	97.95	98.00	98.50
CystNet[22]	96.54	97.44	95.75	94.21	95.92
Proposed Model (Efficient B3 + VAE + IVA)	99.90	100	99.87	99.74	99.82

The Table 4 presents a comparative analysis of various machine learning and deep learning models used for PCOS detection, evaluating them on key performance metrics.

The SVM with an RBF kernel achieves an accuracy of 93.00%, with a precision of 92.80%, recall of 91.10%, F1-score of 91.90%, and specificity of 93.40%. While it performs decently, especially in avoiding false positives (high specificity), its recall is comparatively lower, indicating that some PCOS cases may be missed. The Random Forest classifier improves upon SVM, reaching 94.10% accuracy and better balance in all other metrics, making it more robust in general.

The basic CNN with 5 layers provides a notable improvement over traditional ML models, achieving 95.70% accuracy. Its precision (95.30%) and recall (94.80%) reflect more reliable prediction capability on ultrasound images. Building upon this, the LSTM + GRU Fusion model enhances temporal learning and outperforms previous models with 96.85% accuracy, 96.10% precision, and 96.70% recall, demonstrating effectiveness in capturing complex spatial-temporal features in medical imaging data.

The EfficientNetB0, applied only on image inputs, delivers 97.50% accuracy and maintains excellent performance across all metrics. However, the HybridCNN-Autoencoder with XGBoost model further refines feature extraction and classification by combining deep features and ensemble learning, achieving 98.30% accuracy and an F1-score of 97.95%, demonstrating both precision and generalization.

CystNet, a tailored model for PCOS classification, achieves 96.54% accuracy, 94.21% precision, and a strong recall of 97.44%. While it's a specialized model with balanced metrics, it still falls short when compared to the proposed architecture.

The Proposed Model, integrating EfficientNetB3 as a feature extractor, a Variational Autoencoder (VAE) for better data encoding, and Iterative Variational Attention (IVA) for attention-guided learning, significantly surpasses all others. It achieves near-perfect performance, with 99.90% accuracy, 99.74% precision, 100% recall, 99.87% F1-score, and 99.82% specificity. These results highlight the model's ability to identify virtually all PCOS cases (100% recall) while maintaining high precision and strong generalization. Such performance makes the model highly suitable for clinical deployment, offering both reliability and interpretability.

5. Conclusion and Future Recommendations

In this study, we introduced a novel deep learning framework EfficientNetB3 integrated with a Variational Autoencoder (VAE) and an Invariant Feature Aggregator (IVA) for accurate and efficient detection of PCOS from ultrasound images. The proposed SonoCyst-Net model demonstrated significant improvements over conventional and state-of-the-art models, achieving an outstanding accuracy of 99.90%, precision of 99.74%, recall of 100%, F1-score of 99.87%, and specificity of 99.82% on the benchmark Kaggle PCOS dataset.

Compared to existing techniques such as SVM, CNN, LSTM-GRU fusion, and hybrid models, our approach exhibits superior generalization, robustness, and diagnostic reliability. The combination of EfficientNetB3 for feature extraction, VAE for latent space refinement, and IVA for robust feature fusion contributed to the model's ability to capture both visual and contextual nuances of PCOS markers in ultrasound imagery.

In addition, the model has been optimized for real-time deployment using advanced hyperparameter tuning, regularization techniques, and data augmentation, ensuring its suitability for clinical applications. The results highlight SonoCyst-Net as a promising AI-driven diagnostic tool for early and accurate PCOS detection, with the potential to support gynecologists and radiologists in faster decision-making and improved patient care..

Future work may involve validating this model across diverse imaging modalities and larger, heterogeneous clinical datasets to further evaluate its scalability and generalizability in real-world healthcare environments.

References:

[1] Azziz, R., et al. “Polycystic Ovary Syndrome.” Nature Reviews Disease Primers, vol. 2, 2016, pp. 16057. doi:10.1038/nrdp.2016.57.

[2] Teede, H. J., et al. “Recommendations from the International Evidence-Based Guideline for the Assessment and Management of Polycystic Ovary Syndrome.” Human Reproduction, vol. 33, no. 9, 2018, pp. 1602–1618.

[3] Barber, T. M., et al. “Obesity and Polycystic Ovary Syndrome.” Clinical Endocrinology, vol. 65, no. 2, 2006, pp. 137–145.

[4] Dokras, A., et al. “Risk of Type 2 Diabetes in Women with Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis.” Endocrine Reviews, vol. 32, no. 4, 2011, pp. 347–363.

[5] Nidhi, R., et al. “Prevalence of Polycystic Ovarian Syndrome in Indian Adolescents.” Journal of Pediatric and Adolescent Gynecology, vol. 24, no. 4, 2011, pp. 223–227.

[6] Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. “Revised 2003 Consensus on Diagnostic Criteria and Long-Term Health Risks Related to Polycystic Ovary Syndrome.” Fertility and Sterility, vol. 81, no. 1, 2004, pp. 19–25.

[7] Esteva, A., et al. “A Guide to Deep Learning in Healthcare.” Nature Medicine, vol. 25, no. 1, 2019, pp. 24–29.

[8] Sudhakar, K., and Umapathi, M. “A Deep Learning Model for Detection of Polycystic Ovary Syndrome Using Pelvic Ultrasound Images.” Biomedical Signal Processing and Control, vol. 68, 2021, pp. 102741.

[9] Gopalakrishnan, C. & Iyapparaja, M. Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning. Int. J. Syst. Assur. Eng. Manag.[SPACE] https://doi.org/10.1007/ s13198-021-01203-x (2021).

[10] Alamoudi, A. et al. A deep learning fusion approach to diagnosis the polycystic ovary syndrome (pcos). Appl. Comput. Intell. Soft Comput. 2023, 9686697 (2023).

38. Nilofer, N. et al. Follicles classification to detect polycystic ovary syndrome using glcm and novel hybrid machine learning. Turk. J. Comput. Math. Educ. 12, 1062–1073 (2021).

[11] Suha, S. A. & Islam, M. N. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Sci. Rep. 12, 17123 (2022).

[12] Maheswari, K., Baranidharan, T., Karthik, S. & Sumathi, T. Modelling of f3i based feature selection approach for pcos classification and prediction. J. Ambient. Intell. Humaniz. Comput. 12, 1349–1362 (2021).

[13] Upreti, S., Mehta, R., Verma, T., & Kaur, A. (2025). A hybrid AI/ML-based diagnostic model for early detection and long-term complication prediction of PCOS using hormonal and ultrasound data. Proceedings of the 17th International Conference on Artificial Intelligence and Machine Learning in Healthcare (ICAIMH), 122–133.

[14] ICAART. (2025). Systematic literature review on AI algorithms for PCOS detection: Focus on explainability and protocol optimization. Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART), 1(1), 45–59.

[15] S., & Sonawane, P. (2025). Automated detection of PCOS-related features using transfer learning with InceptionV3. Journal of Medical Imaging and Deep Learning, 10(2), 87–98.

[16] Jothimani, P., Krishnan, M., & Elangovan, V. (2025). A modified Darknet-53 model with probabilistic optimization for enhanced PCOS diagnosis. Biomedical Signal Processing and AI, 18(3), 145–158.

[17] Systematic Review. (2025). AI tools in gynecological ultrasound diagnostics: A comprehensive review with a focus on PCOS. Artificial Intelligence in Women’s Health, 3(1), 23–41.

[18] Rahman, M. M., Islam, A., Islam, F., Zaman, M., Islam, M. R., Sakib, M. S. A., & Babu, H. M. H. (2024). Empowering early detection: a web-based machine learning approach for PCOS prediction. Informatics in Medicine Unlocked, 47, 101500.

[19] Hosain, A. S., Mehedi, M. H. K., & Kabir, I. E. (2022, October). Pconet: A convolutional neural network architecture to detect polycystic ovary syndrome (pcos) from ovarian ultrasound images. In 2022 International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1-6). IEEE.

[20] Divekar, A., & Sonawane, A. (2024). Leveraging AI for Automatic Classification of PCOS Using Ultrasound Imaging. arXiv preprint arXiv:2501.01984.

[21]Kodipalli, A., & Devi, S. (2021). Prediction of PCOS and mental health using fuzzy inference and SVM. Frontiers in public health, 9, 789569.
[22]Moral, P., Mustafi, D., Mustafi, A., & Sahana, S. K. (2024). CystNet: An AI driven model for PCOS detection using multilevel thresholding of ultrasound images. Scientific Reports, 14(1), 25012.
[23]Choudhari, A. Pcos detection using ultrasound images. https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images (2024). [Online using-ultrasound-images (2024). [Online].
[24] Handa, P. et al. Pcosgen-train dataset. https://zenodo.org/records/10430727 (2023). [Online].
.