I.J. Image, Graphics and Signal Processing, **2026, 1, 88-105**
Published Online on February 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2026.01.06

Automated PCOS Detection Using Fine-Grained Deep Feature Extraction and Explainable AI: A Transformer-Based Ensemble Approach

Ifra Bilal Shah*
National Institute of Technology/Computer Science and Engineering, Srinagar, 190006, India
E-mail: ifra_2023phacse009@nitsri.ac.in
ORCID iD: https://orcid.org/0009-0009-3593-8262
*Corresponding Author

Pramod Kumar Yadav
National Institute of Technology/Computer Science and Engineering, Srinagar, 190006, India
E-mail: pramod.kumar@nitsri.ac.in
ORCID iD: https://orcid.org/0000-0001-8693-8216

Received: 20 June, 2025; Revised: 20 October, 2025; Accepted: 18 December, 2025; Published: 08 February, 2026

Abstract: Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine condition affecting women of reproductive age, hallmarked by hormonal abnormalities, ovarian cysts, and metabolic issues. Early diagnosis is essential to prevent long-term effects such as infertility, diabetes, and cardiovascular issues. Conventional diagnostic approaches relying on manual interpretation of ultrasound images are time-consuming and error-prone. To overcome these limitations, we propose an automated diagnostic framework leveraging deep feature extraction and ensemble learning. Initially, ResNet50 is utilized as a convolutional feature extractor, and its extracted features are classified using ensemble of Random Forest (RF) and Gradient Boosting (GB) classifiers. Subsequently, we also employed the Swin Transformer which is a hierarchical vision transformer to extract deep features from ultrasound images, which were fed to Random Forest and Gradient Boosting classifiers. These features were handled separately from those of ResNet50, and no feature concatenation was done. Compared to the ResNet50-based ensemble model, which achieved a classification accuracy of 99.2%, the Swin Transformer–based ensemble model performed better by attaining the accuracy of 99.87%. Furthermore, Explainable AI approaches (Grad-CAM) were applied to both ResNet50-based model and Swin Transformer-based model to highlight key regions contributing to the predictions. This scalable and interpretable system offers encouraging potential for advancing PCOS detection and other medical imaging applications.

Index Terms: PCOS Detection, Swin Transformer, Resnet50, Ensemble Learning, Explainable AI

1. Introduction

Polycystic ovary syndrome (PCOS), also known as polycystic ovary disease (PCOD), is among the most common endocrine disorders affecting women of reproductive age, typically between 15 and 45 years. Despite its widespread occurrence, the underlying cause of PCOS remains incompletely understood. The disorder is closely linked to a range of hormonal and metabolic abnormalities, most notably irregular menstrual cycles, hyperandrogenism, and insulin resistance [1,2]. Because many of these clinical features overlap with those of other conditions, such as obesity and cardiovascular disease, PCOS is often underdiagnosed or misdiagnosed. When left untreated, the condition can have serious long-term consequences, as it represents a major cause of reproductive morbidity and is associated with an elevated risk of endometrial cancer.

Insulin resistance, which is frequently observed in women with PCOS, further contributes to disease complexity by substantially increasing the risk of developing type-2 diabetes. The clinical presentation of PCOS is heterogeneous and may include irregular or absent menstruation, ovarian cysts, excessive hair growth, weight gain, hair thinning, oily skin, acne, dyslipidemia, and abnormalities in skin pigmentation [3]. These manifestations often evolve over time and can significantly impair reproductive health. As reported by Lobo et al. [4], infertility affects 4.5% of married adolescents, 32% of individuals aged 35–40 years, and up to 70% of those above 40 years, underscoring the substantial reproductive burden associated with the disorder.

In addition to clinical and metabolic factors, ethnicity has also been suggested as a variable influencing PCOS prevalence, and several studies have reported differences in detection rates across ethnic populations. For example, Ahmed *et al.* [5] observed a higher reported prevalence of PCOS among Asian women (31.3%) compared with white American women (4.8%), as summarized in Table 1. However, such differences should be interpreted cautiously, as they are likely shaped by multiple confounding factors, including disparities in healthcare access, variations in diagnostic criteria, socioeconomic conditions, lifestyle behaviors, and differences in the prevalence of central obesity and insulin resistance, rather than ethnicity alone.

Table 1. Reported PCOS detection rates among women from various ethnic groups [5]

S. No	Ethnic Group	Detection Rate (%)
1	White Americans	4.8
2	African Americans	8
3	Spain	6.8
4	Mexican	13
5	Asian	31.3

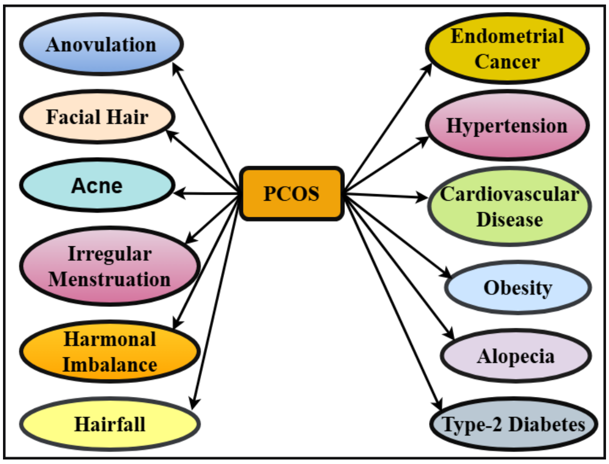


Fig. 1. Symptoms and basic effects of PCOS on women.

The remainder of this paper is arranged as follows: Section 2 reviews recent developments in machine learning and deep learning-based approaches for PCOS detection. Section 3 outlines the proposed methodology, which includes information about the dataset, pre-processing steps, deep feature extraction leveraging ResNet50 and Swin Transformer, and the ensemble learning strategy. Section 4 presents the experimental results and comparative analysis between CNN based models and transformer-based models, highlighting the performance metrics of the proposed framework. Section 5 addresses the integration of explainable AI techniques, particularly Grad-CAM and Attention Rollout, to provide interpretability to the model's predictions. Section 6 focuses on the clinical applicability and practical considerations of the proposed system, including its potential role in assisting radiologists and gynecologists in real-world diagnostic settings. Finally, Section 7 concludes the paper and outlines future options for enhancing diagnostic skills using multi-modal data and more advanced interpretability tools.

2. Related Work

The pie chart of "Estimated number of PCOS publications using ML and DL (2014–2024)" depicts the tremendous growth in research in the last ten years based on ML and DL applications for the management of PCOS. It gives a brief overview of how AI methodologies are increasingly adopted in dealing with all kinds of problems regarding the identification, prediction, and treatment related to PCOS.

In this domain, from 2014 to 2017, the research activity was minimal, collectively accounting for less than 7% of the total publications. In fact, the contributions in this domain were limited to 0.7% in the year 2014, 1.2% in 2015, 1.7% in 2016, and 2.9% in 2017 [6]. Such slow early progress can be attributed to the non-availability of large annotated datasets, limited computational resources, and resistance to the adoption of ML/DL approaches because of their complexity and perceived black-box nature.

The trend of adoption is very clear from 2018 onward. The year 2018 was the beginning of a steep rise and constituted 4.1% of total publications in that year. This was the time when computational infrastructure was hugely improving, the ML/DL architecture was gaining momentum, and the awareness of the potential of AI in healthcare was growing. The share stood at 5.3% in the following year, thus showing a gradual but significant rise in research interest.

Early works studying CNNs and other architectures of deep learning for medical applications had started to appear, forming fertile ground for further research in the domain under study.

Starting in 2020, the rate of growth clearly accelerated, with contributions reaching 9.7%. Part of this may be attributed to the globalization of focused efforts for data-driven healthcare solutions during the COVID-19 pandemic, alongside the rapid expansion of AI in medical areas. As ML/DL tools became easier to access and data more varied, more and more researchers began recognizing the full potential of these tools to address the diagnostic challenges of PCOS. The upward trend continued in 2021, with publications accounting for 12.1%, reflecting growing confidence in the effectiveness of AI-assisted healthcare solutions. This share increased further in 2022 to 15.1%, signaling a transition from exploratory studies to established and application-driven research activities.

In this stage, researchers increasingly employed pre-trained deep learning models like ResNet and YOLO and explainable AI techniques such as Grad-CAM and LIME to reduce the interpretability issues of deep learning models [7]. The further integration of ensemble learning, hybrid architectures, and transfer learning methods would help to significantly improve the prediction performance and make the model more robust, which obviously contributed to the expanding research landscape. The most significant growth can be seen in the last two years, namely, in 2023 and 2024, which, together, dominate the research output, contributing 19.4% and 27.8% of publications, respectively. This sharp increase reflects the maturation of AI-driven approaches in the study of PCOS, wherein there is a profound emphasis on advanced feature extraction, classification, and predictive modeling. The wide acceptance of ML and DL as part of the clinical decision-support system reflects a clear shift from experimental methodologies to practical deployable solutions in PCOS management [8].

Overall, the trend illustrated in Fig. 2 demonstrates how ML and DL have transformed PCOS research over the last decade. The pronounced growth in recent years suggests that these techniques are no longer supplementary tools but central components of modern PCOS diagnostics. While challenges such as limited dataset diversity and the need for enhanced model interpretability remain, ongoing efforts in optimization, multimodal learning, and explainable AI indicate a promising future for AI-driven PCOS diagnosis and management.

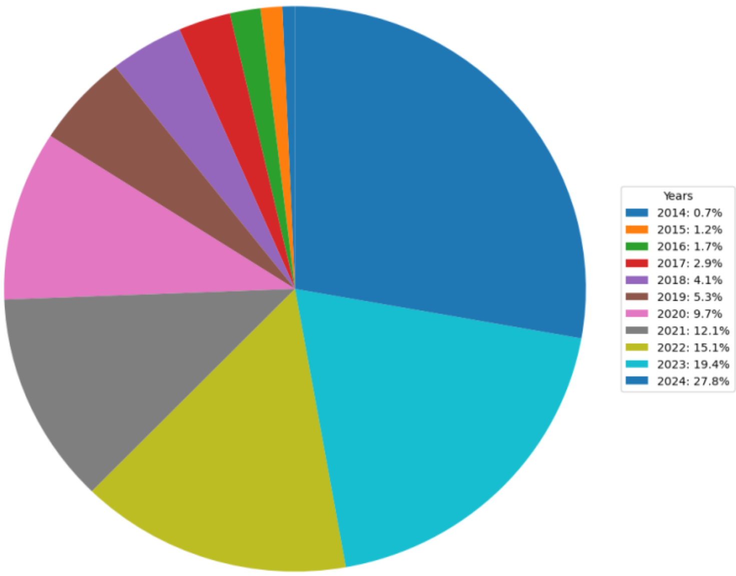


Fig. 2. Estimated number of PCOS publications using ML and DL (2014-2024)

2.1. Machine Learning-Based Classification Approaches

Sayma Alam Suha *et al.* [9] proposed an improved machine-learning approach to diagnose PCOS based on ultrasound scans of ovaries. This work assesses an enhanced machine learning categorization approach for the prediction of PCOS from 594 USG images of the ovaries. In the proposed approach, features from the ultrasound images are extracted using CNNs with refined algorithms and transfer learning. Classification of PCOS and non-PCOS ovaries is performed based on the stacking ensemble machine learning. In this approach, reduced feature sets and the traditional models are used as base models, and the bagging or boosting models as meta-models. As one can conclude, analyzing this method compared to other approaches of ML, the accuracy is enhanced, whereas the training time is reduced significantly. Notably, the highest performance is obtained after using the described CNN architecture as a feature extractor with the help of the “VGGNet16”model and applying the “XGBoost” stacking ensemble model as a meta-learner to reach the highest classification accuracy of 99.89%. A PCOS diagnostic tool designed by B Rachana *et al.* [10] could take much less time required in the manual examination for tracing the follicle and defining the characteristics of the geometry of the same. In this case, the use of a KNN classifier has been the suggested method, which has demonstrated its ability to achieve classification accuracies of over 97%. By increasing the efficiency and reliable diagnosis of PCOS, this classifier increases the possibility of early diagnosis and decreases the risk of a fatal outcome. C. Gopalakrishnan *et al*. [11] only target the sick and propose an image-based automated system for diagnosing and binning PCOS using Ultrasound. There are two types of input imagery pre-processing: Gaussian low pass filter for segmentation and multi-layer thresholding. The computation of features is done with the help of the GIST-MDR approach, and the classification of the instances of the PCOS is done with the help of the modern methods of supervised machine learning. According to the evaluation elements, the proposed system is more effective, which confirms the success of the approach. An overall accuracy of 93.82% for the Support Vector Machine, 89.7% for Random Forest, 91.05% for Linear Discriminant Analysis, and 88.26% for the Naive Bayes algorithms with the best features have been achieved by the feature selection model GIST-MDR. Wagh *et al.* [12] proposed the enhancement of early diagnosis of PCOD by developing machine learning technologies. Indeed, CatBoost was found to be the best model for the tabular data, where it produced an F1-score of 88.68%; for the ultrasound scan images, the best model was identified as YOLOv2 with an F1-score of 85.86%. Furthermore, the research also suggested creating an application through which doctors can enter tabular data and ultrasound images to predict PCOD and help in the diagnosis process to be more efficient and effective while also storing the data for further use. Shazia Nasim *et al.* [13] proposed an efficient predictive model for PCOS diagnosis depending on clinical and physical indices of 541 patients with and without the disorder. To identify the most relevant predictors for PCOS, the authors introduced a feature selection method known as CS-PCOS, which is based on chi-squared optimization. The study compares ten different machine learning algorithms, but Gaussian Naive Bayes outperformed all other models with 100% accuracy, perfect precision, Recall, and F1 Score. The results are further tested with k-fold cross-validation, to reduce variability. A study by Subha *et al.* [14] suggested a hybrid model that blends the swarm intelligence (SI) feature selection method with the machine learning algorithms. The model makes use of classifiers like Random Forest and XGBoost. The model with the Particle Swarm Optimization method and Random Forest classifier achieved the highest accuracy of 92.24%. In order to improve the diagnosis of PCOS, Poonam *et al.* [15] have presented a robust framework utilizing advanced machine learning techniques. The study utilized a hybrid PODBoost classifier, which is a combination of bagging and boosting techniques hence achieved a remarkable accuracy of 97.42%. Some improvements made in the study are SMOTE-Tomek Links (SMTL) algorithms and Grey Wolf optimization (GWO) performance feature selection techniques to reduce the dimensionality of the data while enhancing computational efficiency. Furthermore, for policy recommendation, the enhancement of Explainable AI (XAI) tools such as LIME was utilized for modeling interpretability. S. Sreejith *et al.* [16] developed a clinical decision support system (CDSS) to aid PCOS diagnosis by improving machine learning accuracy through feature selection. The study removes irrelevant features that can reduce model performance. Using a wrapper approach with the red deer algorithm (RDA), which balances exploration and exploitation, the most relevant features are identified. These are then evaluated using a Random Forest classifier with 50 decision trees, achieving 89.81% accuracy. Khanna *et al.* [17] presented a novel method for diagnosing PCOS by utilizing explainable AI and ML. The study used a dataset of 541 patients and employed feature selection techniques such as Mutual information and Harris Hawk optimization to identify important characteristics like follicular count and hormone levels. The developed ensemble model stack-3 outperformed other classifiers and achieved the highest accuracy of 98%. To enhance interpretability, XAI tools, including SHAP and LIME, were incorporated, offering insights into the significance of features and reasoning behind predictions.

2.2. Deep Learning-Based Techniques

Abrar Alamoudi *et al.* [18] proposed a PCOS diagnosis model for CAD systems using two datasets: ovarian ultrasound scans and clinical parameters like blood pressure, pulse rate, biochemical profile, and PCOS-specific complaints. Three deep learning frameworks are evaluated for AI-based CAD deployment. The study focuses on improving PCOS diagnosis using ultrasound images to enhance accuracy and reduce false positives. Fine-tuning the Inception architecture with ultrasound images achieves 84.81% accuracy. The research also explores deep learning on image and clinical features and two fusion models, emphasizing the significance of clinical data in PCOS diagnosis. This automated model can assist physicians, accelerate patient evaluations, and reduce early diagnosis omissions. Other researchers like Pradeep Bedi *et al.* [19] have also developed a hybrid model for the enhanced diagnosis of PCOS, which is a vital marker in predicting the onset of female gynecological cancer. They used a model known as the attention residual u-net (AResUNet) and the adaptive bilateral filter for image pre-processing. AResUNet also provides structural flexibility for both 2D and multi-modal images in parallel, and adaptive bilateral filtering to eliminate noises. The findings of the proposed hybrid model show a substantial improvement in the detection of PCOS compared to the existing methods. In particular, the AResUNet variant demonstrated spectacular results when using two-dimensional and multi-modal images. In recent years, computational Intelligence has drawn significant attention due to its potential to improve the early detection of infertility in women, especially in the diagnosis of PCOS, which is a major cause of infertility. Sowmiya *et al.* [20] developed a novel PCOS diagnostic method using the F-Net architecture, ensuring high prediction accuracy and computational speed for clinical use. The model employs advanced image pre-processing for effective feature extraction of ovarian follicles. Using a rich dataset, it outperforms other methods in classification efficacy, accuracy, and discrimination. The study also addresses model interpretability with Grad-CAM, enabling clinicians to pinpoint key areas influencing predictions. F-Net facilitates faster, less human-dependent PCOS diagnosis with improved outcomes. Table. 2 shows the average accuracy of existing models in the PCOS diagnosis.

Table 2. Average Accuracy of Existing Models

Author	Data Type	Algorithm	Highest Accuracy (%)
Sayma Alam et al. [9]	images	CNN with ensemble learning	99.8
B Rachana et al. [10]	images	KNN classifier	97
C. Gopalakrishnan et al. [11]	images	SVM, RF, NB	93.82
Pratik Wagh et al. [12]	images and textual data	CatBoost and YOLOv2	88.68
Shazia Nasim et al. [13]	textual data	Chi-square and GNB	100
Subha et al. [14]	Textual data	PSO and RF	92.24
Poonam et al. [15]	Textual data	Grey Wolf optimization	97.42
S.Sreejith et al. [16]	Textual data	Red Deer and RF	89.81
Khanna et al. [17]	Textual data	STACK-3, XAI	98
Abrar Almoudi et al. [18]	images and textual data	CNN, VGG-16, ResNet50	82.4
Pradeep Bedi et al. [19]	Multi-modal 2D images	ResUNet	97.7
Sowmiya et al. [20]	Images	YOLOv8, ResNet152v2	97.5

2.3. Research Gaps and Contributions

Despite much advancement in PCOS diagnosis using machine learning and deep learning, several research gaps still remains. Most existing studies are constrained by the use of single-modality data such as clinical parameters or ultrasound images rather than combining both for comprehensive diagnosis. While deep learning models such as CNNs have in-creased detection accuracy, they often lack the ability to capture long-range dependencies, which is crucial in identifying minute patterns in ultrasound images. Transformer-based architectures, which could potentially solve these issues, are still underexplored in PCOS detection. Additionally, the lack of interpretability in many high-performing models undermines clinical trust and acceptance. Few studies have leveraged explainable AI techniques like Grad-CAM or SHAP to visualize decision-making processes in a clinically interpretable way. Given the complex and multifaceted nature of PCOS, accurate and early diagnosis is essential.

To address this challenge, the research proposes a sophisticated deep learning-based framework that leverages the strengths of both convolutional and transformer-based models. Specifically, the feature extraction capabilities of the pre-trained ResNet50 are utilized to analyze ovarian ultrasound images, capturing important diagnostic patterns [21]. To assess the effectiveness of various architectures, we conducted a comparative analysis between CNN-based and transformer-based models to ascertain which approach provides superior feature extraction for PCOS diagnosis. To further enhance classification performance, ensemble learning methods—Random Forest (RF) and Gradient Boosting (GB)—are employed, combining their predictive strengths. Additionally, to explore more robust feature extraction, a Swin Transformer is also utilized. With its hierarchical representation and shifted window mechanism, the Swin Transformer is proficient at capturing both local and global spatial features, leading to improved accuracy over traditional CNN-based methods. This demonstrates the efficacy of transformer-based models in medical image analysis. Overall, the proposed framework pro-vides a scalable, automated, and accurate diagnostic tool for clinical settings, enabling early and trustworthy detection of PCOS.

3. Methodology

The proposed framework for automated PCOS diagnosis is built around a two-branch architecture, where each branch separately processes ovary ultrasound images using a different deep feature extractor such as ResNet50 and Swin Trans-former. In each model branch, preprocessed images are fed into the respective model to extract deep features. These features are then fed to the ensemble of two traditional machine learning classifiers: Random Forest (RF) and Gradient Boosting (GB). The predictions from RF and GB are averaged to calculate the final class prediction for each feature ex-tractor. This ensemble approach is used separately for each model thus enabling performance comparison. Furthermore, Grad-CAM is applied to the ResNet50-based pipeline to produce visual explanations for predictions thereby enhancing model interpretability. This dual-branch framework ensures both robustness and explainability in the diagnostic process. Fig 3 depicts the workflow of the proposed system for the detection of PCOS using ultrasound images.

3.1. Dataset Description

The dataset used in this study consists of 3,846 ovarian ultrasound images collected from a publicly available Kaggle repository and organized into two diagnostic categories: PCOS-affected and non-PCOS cases. From the total samples, 1,562 images correspond to PCOS-affected ovaries, exhibiting clinically relevant morphological features such as ovarian enlargement and the presence of multiple peripherally arranged follicles. The remaining 2,284 images represent normal ovarian anatomy, with no visible pathological abnormalities. This results in an inherent class distribution of approximately 40.6% PCOS-affected and 59.4% non-PCOS samples, reflecting a moderately imbalanced dataset that is representative of real-world clinical prevalence.

To ensure a robust and unbiased evaluation, the dataset was divided into training and testing subsets using stratified sampling, thereby preserving the original class distribution in both partitions. A total of 1,924 images were assigned to the training set, while 1,922 images were reserved for independent testing, maintaining the same PCOS-to-non-PCOS ratio observed in the complete dataset. Importantly, no class rebalancing techniques, including oversampling, SMOTE, or synthetic data generation, were applied prior to data partitioning, thereby preventing information leakage and ensuring the integrity of the evaluation protocol.

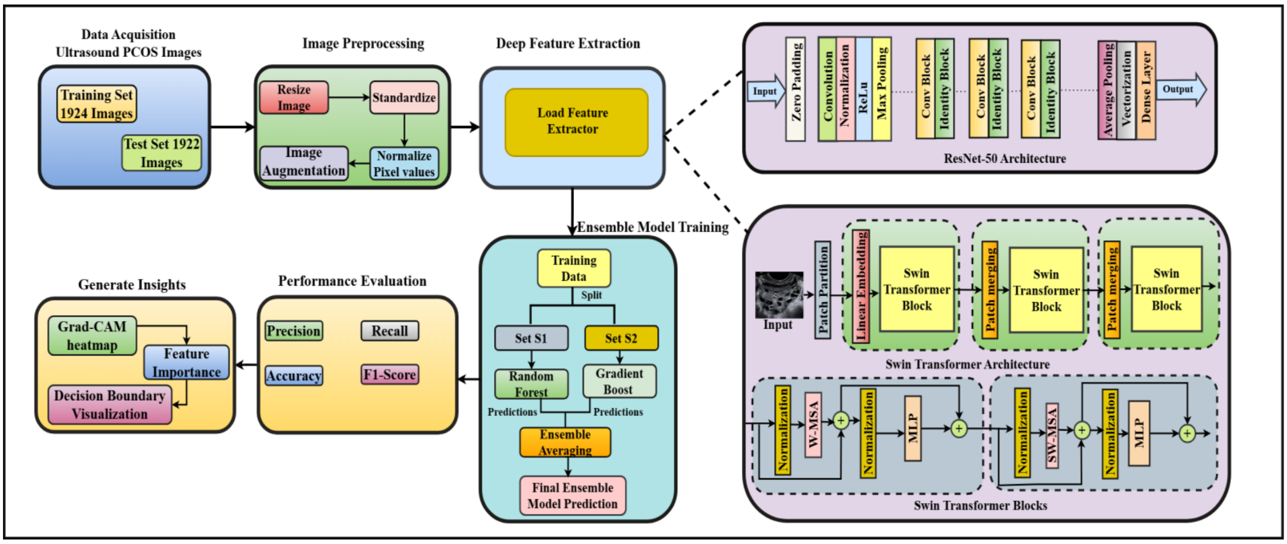


Fig. 3. Block diagram of the proposed PCOS detection system using ultrasound images.

3.2. Data Augmentation and Pre-processing

To improve model performance and enhance generalization, particularly in medical ultrasound imaging tasks where annotated datasets are often limited, a combination of image pre-processing and data augmentation was employed using the *ImageDataGenerator* utility from the TensorFlow Keras framework. All input images were normalized by rescaling pixel intensities from the original range of [0, 255] to [−1, 1], a preprocessing strategy that is well suited for pre-trained architectures such as ResNet50 and the Swin Transformer, ensuring compatibility with their learned weight distributions.

Data augmentation was strictly confined to the training set in order to prevent any form of data leakage between training and evaluation phases. Several geometric transformations were applied to introduce controlled variability and to mitigate over-fitting during model training. These transformations included random rotations within ±20◦, horizontal and vertical translations of up to ±10% of the image dimensions, zoom operations within the range [0.9, 1.1], shear trans-formations of up to ±10%, and random horizontal flipping. Collectively, these operations simulate plausible anatomical variations and acquisition inconsistencies commonly observed in real-world ovarian ultrasound examinations.

Table 3. Hyperparameter settings and model configuration for the proposed PCOS classification framework.

Parameter	Value
Input image size	224 × 224 × 3
Batch size (*B*)	32
Optimizer	Adam
Learning rate (η)	0.0001
Loss function	Binary Cross-Entropy
Epochs (*E*)	50
Early stopping patience	Enabled
Random seed	42 (for reproducibility)
Feature Extraction Configuration	
ResNet50 feature dimension	2048
Swin Transformer feature dimension	1024
Ensemble Classification Configuration	
Ensemble classifiers	Random Forest + Gradient Boosting
Number of estimators	100 (each classifier)
Ensemble fusion rule	Arithmetic averaging of predictions

By increasing the effective diversity of the training samples, the adopted augmentation strategy encourages the models to learn more robust and invariant feature representations. In contrast, the test set was left completely un-augmented, ensuring that all reported performance metrics reflect genuine generalization to previously unseen data rather than adaptation to artificially modified samples.

3.3. Implementation Details

The proposed framework was implemented using TensorFlow 2.12 with Keras API for deep feature extraction and Python 3.10.

For both ResNet50 and Swin Transformer pipelines, the following settings were used as shown in Table 3.

3.4. Deep Feature Extraction using ResNet50

ResNet50, a pre-trained model on ImageNet, is employed for feature extraction. Residual Network 50, abbreviated as ResNet50, is a deep convolutional neural network that consists of 50 layers [22]. ResNet50 captures local spatial features such as edge patterns, textures, contours, and fine-grained details at multiple levels of abstraction. A typical convolutional operation is expressed as:

$$a_f = \left(I.W_f\right) + \beta \tag{1}$$

Where I is the input image, W_f is the filter weight, β is the bias, * denotes the convolutional operation, and a_f is the output feature map. He *et al.* [23] introduced Deep Residual Learning for image recognition in 2015. ResNet50 is widely accepted for its residual connections, which mitigate the vanishing gradient issue and enable the training of extremely deep networks. A residual block of ResNet50 can be expressed as:

$$R = \emptyset\left(I, \{W_j\}\right) + I \tag{2}$$

Where R is the residual output, $\emptyset$ is the transformation function, I is the input feature map, and W_j represents the learned weights associated with the transformation.

ResNet50 consists of five stages of convolutional layers followed by global average pooling and a fully connected layer that is designed for robust image classification. The global average pooling (GAP) layer operation is usually given as:

$$\acute{y} = \frac{1}{h.w} \sum_{r=1}^{h} \sum_{c=1}^{w} x_{rc} \tag{3}$$

Where $\acute{y}$ is the output of the GAP layer, h and w are the respective height and width of the feature map, and x_{rc} is the value at point *(r,c)* in the feature map. Mathematically, the fully connected layer can be expressed as shown in Equation

$$Y = WI + \beta \tag{4}$$

Where W is the weight matrix, I is the input vector, β is the bias, and Y is the output of the fully connected layer.

The first input layer processes $224 \times 224 \times 3$ images, pre-processed for pixel adjustments. A 7×7 convolution with 64 filters and stride 2 is followed by a 3×3 max pooling layer, reducing spatial dimensions. The network has 16 residual blocks across four stages, with filters increasing from 64 to 512. Skip connections mitigate vanishing gradients, ensuring efficient 50-layer training. A global average pooling layer condenses spatial data, and feeds features into a dense layer for classification. ResNet50's depth-efficiency balance makes it a standard for feature extraction and transfer learning.

Because ResNet50's innovative design balances depth and efficiency, it is considered a standard model for feature extraction and transfer learning.

3.5. Deep Feature Extraction using Swin Transformer

The Swin Transformer introduced by Liu et al. (2021) is a hierarchical vision transformer that processes images effectively while maintaining ideal computing efficiency. Swin transformer is employed to capture multi-scale contextual and spatial relationships across the image, integrating both local (within windows) and global (across windows) context. It enhances the conventional vision transformer by using a hierarchical feature representation and leveraging a shifted window mechanism for self-attention. Due to these innovations, Swin Transformers are well suited for tasks like PCOS detection from ovary ultrasound images. These advancements also address the limitations of traditional vision transformers, such as the quadratic complexity in self-attention.

Before splitting ultrasound images into non-overlapping patches, the model processes the image by resizing and normalizing it. Each of these patches is then handled as a token, and feature vectors are produced using a linear embedding layer. Mathematically, the patch embedding operation can be represented as:

$$\mathrm{Embed}(\mathrm{P_i}) = \mathrm{W_P P_i} + \beta_\mathrm{P} \tag{5}$$

where W_P is a learnable weight matrix and β_P is the bias term. This process converts the image into a series of patch tokens, $T = [t_1, t_2, \ldots t_n]$, allowing the model to process the image in smaller, more manageable chunks. As a part of the Swin Transformer's core attention mechanism, each patch token interacts with other patch tokens inside a local window to depict spatial relationships. Three matrices - Query (Q), Key (K), and Value (V) are used to compute the self-attention operation. Mathematically, the attention mechanism is defined as:

$$Attention(Q_{patch}, K_{patch}, V_{patch}) = Softmax\left(\frac{Q_{patch}K_{patch}^{T}}{\sqrt{D_k}} + \beta_p\right)V_{patch} \tag{6}$$

where Q_{patch}, K_{patch}, and V_{patch} are the Query, Key, and Value matrices, respectively, and β_p is the relative position bias that improves spatial modeling. The term $\frac{Q_{patch}K_{patch}^{T}}{\sqrt{D_k}}$ ensures that the attention values are scaled properly, preventing unnecessarily large number's that can disrupt the learning process.

Algorithm 1. PCOS Detection and Explanation (ResNet50 as Feature Extractor)

Require: Ovary ultrasound images $\{I_i\}_i^N = 1$ labeled as Infected or Not-infected
Ensure: Prediction labels and Grad-CAM heatmap

Pre-process all input images:
for each image I_i **do**
    Normalize and resize to (224 × 224)
    $I_{resized}$= Resize(I_i, (224 × 224))
    Extract features using ResNet50
    $F_i = \emptyset(I_{resized})$
end for
Create training and validation datasets:
Split $\{F_i\}$ into training and validation sets:
(F_{train}, F_{val}, y_{train}, y_{val}) = Split (F, y)
Train classifiers:
Train Random Forest (RF):
RF = Train(Random Forest(n = 100), F_{train}, y_{train})
Train Gradient Boosting (GB):
GB = Train(Gradient Boosting(m = 100), F_{train}, y_{train})
Ensemble predictions:
for each feature vector F_test from test images **do**
    Predict using the ensemble model:
    $P_{ensemble} = \frac{1}{2}\,(RF(F_{test}) + GB(F_{test}))$
    Determine the final prediction label:
    $y_{pred} = \arg\max(P_{ensemble})$
end for
Grad-CAM explanations:
for each test image I_{test} **do**
    Generate the Grad-CAM heatmap:
    H = Grad-CAM(φ, I_{test})
    Output the heatmap and prediction label
end for

This function allows the model to learn intricate patterns in ovary ultrasound images that indicate PCOS-infected or PCOS-not-infected characteristics by focusing on pertinent features inside each window. A primary feature of the Swin Transformer is its shifted window system, which alternates window partitioning between successive layers. This shifting facilitates cross-window interactions, enabling the model to capture long-range dependencies across the image. The shift can be mathematically represented as: The weight update in the Swin Transformer follows a gradient-based optimization strategy, which can be mathematically represented as:

$$W_{new} = W_{old} + \Delta W_{grad} \tag{7}$$

where W_{new} reflects the newly updated weight matrix, W_{old} is the weight matrix before applying any change, and ΔW_{grad} is the gradient-based update to the weights. In the case of PCOS detection, this design enhances the model's ability to identify intricate, spatially distributed features, such as cysts or follicular patterns, by encouraging richer interactions between distant regions in the image.

Algorithm 2. PCOS Detection and Explanation (Swin Transfomer as Feature Extractor)

Require: Ovary ultrasound images $\{I_i\}_i^N = 1$ labeled as Infected or Not-infected
Ensure: Prediction labels and Grad-CAM heatmap
1. **Pre-process all input images**:
2. **for** each image I_i **do**
3. Normalize and resize to (224 × 224)
4. $I_{resized}$= Resize(I_i, (224 × 224))
5. Extract features using Swin Transformer
6. **For** each resized Image $I_{resized}$
7. $F_{i=}\emptyset_{Swin}(I_{resized)}; H, W, L, d, M)$
8. **end for**
9. **end for**
10. **Create training and validation datasets:**
11. Split $\{F_{i\}}$ into training and validation sets:
12. (F_{train}, F_{val}, y_{train}, y_{val}) = Split (F, y)
13. **Train classifiers:**
14. Train Random Forest (RF):
15. *RF* = Train(Random Forest(n = 100), $F_{train}, y_{train})$
16. Train Gradient Boosting (GB):
17. *GB* = Train(Gradient Boosting(m = 100), $F_{train}, y_{train})$
18. **Ensemble predictions:**
19. **for** each feature vector F_test from test imag**es do**
20. Predict using the ensemble model:
21. $P_{ensemble} = \frac{1}{2}\ (RF(F_{test}) + GB(F_{test}))$
22. Determine the final prediction label:
23 $y_{pred} = \arg\max(P_{ensemble})$
24. **end for**
25: **Explainability (XAI):**
26: **for** each test image I_{test} **do**
27: Generate Attention Rollout map using $\emptyset_{Swin}$
28: **end for**

As the model evolves through deeper levels, patch merging operations improve the number of channels while reducing spatial resolution. By integrating adjacent patches, this operation enables the model to refine its feature maps and efficiently learn both local and global patterns. The patch merging operation can be mathematically represented as:

$$T_{merged} = concat(T_{even}, T_{old})W_{merge} \tag{8}$$

where T_{even} and T_{old} are the even and odd indexed patches, respectively. This operation combines local features, thereby helping the model maintain computational efficiency while gradually improving its representations.

The hierarchical structure of the Swin Transformer enables it to learn features at different scales, building smaller and more abstract versions of the input image. This is particularly advantageous for medical imaging, where it is essential to capture both larger structures (such as the morphology of ovaries) and finer details (such as individual cysts). The model extracts features at varying levels of abstraction by increasing the number of channels while reducing the spatial resolution at each iteration. Mathematically, this process can be defined as:

$$\mathcal{F}_i = \emptyset_i(\mathcal{F}_{i-1}) \tag{9}$$

where $\mathcal{F}_i$ is the feature map at the i^{th} layer, $\emptyset_i$ represents the stage-specific transformation function.

Once the hierarchical feature maps have been developed, the model generates a global feature vector $\mathcal{F}_{global}$, providing a summary of the entire image. This $\mathcal{F}_{global}$ is then fed to the classification head, which usually consists of fully connected layers, to determine whether the image corresponds to a PCOS or non-PCOS case. The classification step can be expressed as:

$$\mathcal{F}_{global} = MLP\left(\mathcal{F}_{final}\right) \tag{10}$$

where $\mathcal{F}_{final}$ is the final output from different layers of the Swin Transformer, and the Multi-Layer Perceptron (MLP) transforms the feature vector into class labels. By integrating local and global features obtained from self-attention and hierarchical aggregation, along with the computational efficiency of the shifted window mechanism, the Swin Transformer proves to be a robust and scalable method for PCOS identification in high-resolution ovary ultrasound images. These mathematical principles underpin the model's ability to detect minute yet significant patterns, leading to accurate and effective diagnosis.

3.6. Ensemble Learning

The proposed framework categorizes ovary ultrasound images as infected (PCOS-positive) or not-infected (PCOS-negative) using an ensemble learning technique. Deep features extracted via ResNet50/Swin Transformer are fed into an ensemble of machine learning classifiers—Random Forest (RF) and Gradient Boosting (GB). The RF classifier, set with 100 estimators, handles high-dimensional feature sets by constructing multiple decision trees on bootstrapped subsets of training data. Majority voting aggregates predictions, reducing over-fitting and improving generalizability. The Gradient Boosting classifier, also with 100 estimators, uses sequential learning where weak learners iteratively correct errors of previous models. By emphasizing misclassified samples in each iteration, Gradient Boosting minimizes prediction error, ensuring robustness. The complementary nature of RF and GB is leveraged through ensemble averaging, enhancing overall accuracy. This approach achieves exceptional test accuracy while ensuring robustness. The ensemble framework is particularly well-suited for medical imaging, combining the predictive capabilities of machine learning classifiers with the feature extraction power of ResNet50.

4. Results and Discussion

4.1. Results with ResNet50 as a Backbone for Feature Extraction

The efficacy of the proposed ensemble framework is assessed using robust metrics to ensure its reliability in dividing ultrasound images into infected and not-infected PCOS categories. The model's impressive validation accuracy of 99.22% demonstrates its effective generalization to unseen data during validation. This high accuracy is further confirmed by detailed evaluation metrics, including precision, recall, and F1-score, as shown in table 4. The confusion matrix, as shown in Fig. 4, provides a clear visualization of the model's performance with ResNet50 as a feature extractor, showing the true positives, true negatives, false positives, and false negatives. The graph presented in Fig. 5 displays the training and validation accuracies of RF and GB models across several training steps.

Table 4. Classification Performance Metrics

Metric	Class 0 (Non-PCOS)	Class 1 (PCOS)
Precision (%)	98	
Recall (%) F1-	100	100
Score (%)	99	99 98

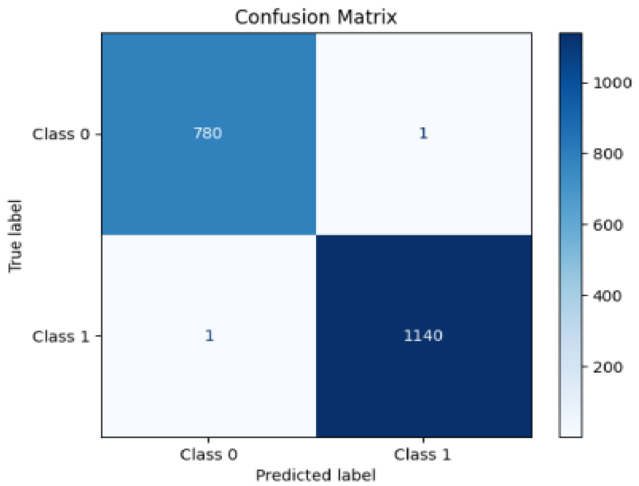


Fig. 4. Confusion matrix using ResNet50 as Feature Extractor

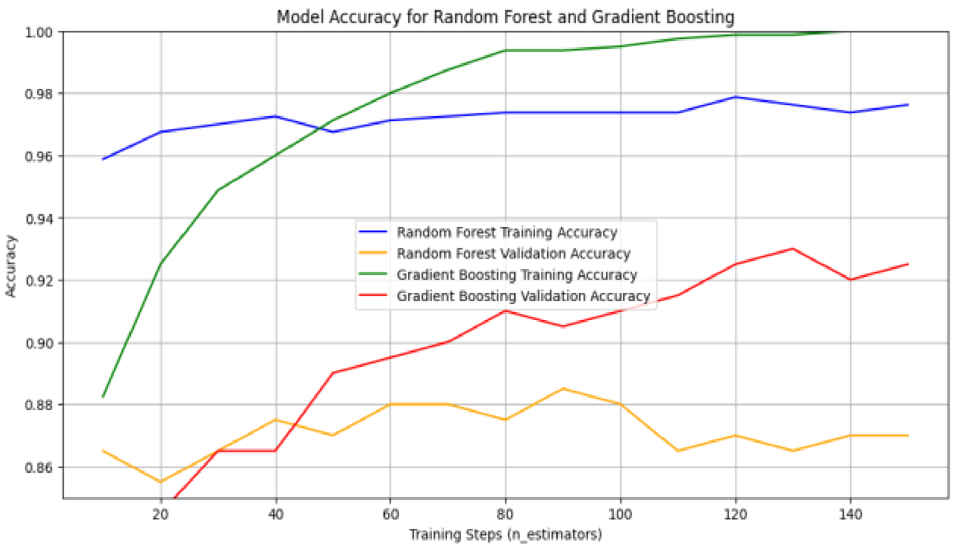


Fig. 5. Training and validation accuracies of RF and GB models using ResNet50 as Feature Extractor

4.2. Results with Swin Transformer as a Backbone for Feature Extraction

The proposed system, which combines the Swin Transformer (ST) for feature extraction with Random Forest and Gradient Boosting Classifiers, outperformed the ResNet50-based approach. Based on the experimental evaluation, the hierarchical feature extraction ability of the Swin Transformer, when combined with the ensemble learning strengths of RF and GB, resulted in an overall accuracy of 99.8%, which is higher than the 99.2% previously achieved with ResNet50. The high accuracy is further confirmed by detailed evaluation metrics, including precision, recall, and F1-score, as shown in Table 5. The confusion matrices shown in Fig. 6 and classification reports for both approaches further demonstrate that the adoption of the Swin Transformer significantly improved precision, recall, and F1-score. The Swin Transformer efficiently captured both global contextual information and fine-grained information (cystic structures) from ovary ultrasound images, leading to better classifier performance. The drastic improvement in accuracy and other metrics indicates the Swin Transformer's ability to address the shortcomings of the Convolutional architecture,

Table 5. Classification Performance Metrics

Metric	Class 0 (Non-PCOS)	Class 1 (PCOS)
Precision (%) Recall (%) F1-Score (%)	99 100 99	100 98.9 99.9

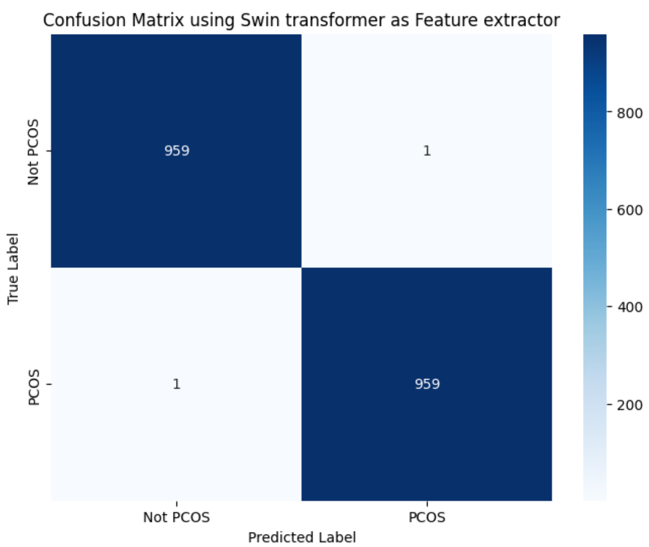


Fig. 6. Confusion matrix using Swin Transformer as Feature Extractor

as seen in ResNet50. One of the advantages of the shifted windowing approach in the Swin Transformer is its scalable feature representation, which enables the model to differentiate fine details in ultrasound images relevant for PCOS detection. Fig. 7 illustrates the training and validation accuracies of the Random Forest (RF) and Gradient Boosting (GB) models, demonstrating the effectiveness of using the Swin Transformer as a feature extractor for PCOS detection. This development signifies the progress made in the field of transformer-based architectures for medical imaging, enabling a more effective and dependable technique for PCOS diagnosis.

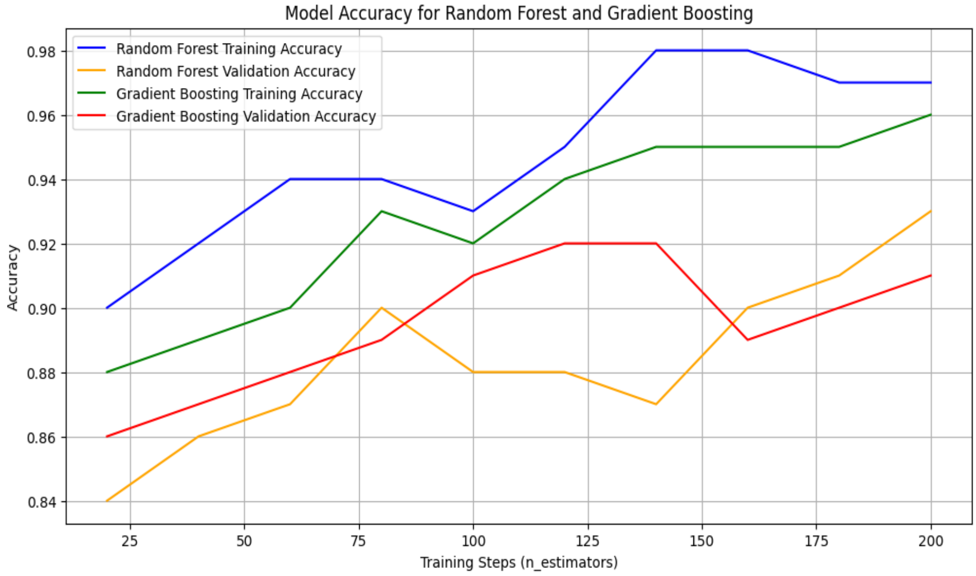


Fig. 7. Training and validation accuracies of RF and GB models using Swin Transformer as Feature Extractor

4.3. Comparison of existing literature with the Proposed Framework

A comparison between existing studies and the proposed framework is presented in Table 6. Previous work in this domain has explored a variety of modeling approaches and pre-processing strategies, leading to a wide range of reported accuracies. For example, Hosain *et al.* [25] combined convolutional neural networks with manually extracted features and reported an accuracy of 85.3%. While their approach demonstrated the feasibility of automated PCOS detection, its performance was limited by the small size of the dataset and the absence of data augmentation, which is particularly important in ultrasound-based analysis. Similarly, Moral *et al.* [15] achieved an accuracy of 87.1%; however, their frame-work lacked deeper integration between feature extraction and classification, as well as comprehensive pre-processing, which may have constrained its ability to generalize.

More advanced architectures have reported improved results. Sowmiya *et al.* [20] employed an F-Net–based architecture and achieved an accuracy of 97.5%, highlighting the benefits of refined network design. Building on these developments, the proposed framework further enhances performance by combining robust deep feature extraction with ensemble learning. Using ResNet50 features together with Random Forest and Gradient Boosting classifiers, the proposed method achieved an accuracy of 99.2%, supported by systematic data augmentation and pre-processing. Moreover, replacing the convolutional backbone with a Swin Transformer led to a further improvement, yielding an accuracy of 99.89%. This gain suggests that transformer-based feature representations are particularly effective in capturing the complex structural patterns present in ovarian ultrasound images.

Table 6. Comparison of the Proposed Framework with Existing Techniques for PCOS Detection.

Reference	**Dataset**	**Methodology**	**Accuracy(%)**
Hosain et al. [25]	2,500 images (Kaggle)	CNN-based	85.3
Moral et al. [15]	3,000 images (Public USG dataset)	CystNet: Multilevel thresholding	87.1
Umapathy et al. [20]	1,148 images (Public USG dataset)	F-Net Architecture	97.5
Proposed Framework (ours)	3,846 images (Kaggle)	ResNet50 + RF + GB	99.2
Proposed Framework (ours)	3,846 images (Kaggle)	Swin Transformer + RF + GB	99.8

5. Explainable AI (XAI)

The black-box nature of deep learning models presents challenges to XAI, which is crucial in healthcare diagnostics for trust, transparency, and interpretability. While AI systems achieve high accuracy in disease detection, such as PCOS diagnosis via ultrasound, their lack of interpretability limits clinical adoption [26]. XAI bridges this gap by explaining predictions, highlighting key features, and enabling clinicians to validate results. Interpretability fosters trust as clinicians assess alignment with medical knowledge [27]. XAI also helps researchers detect biases, refine models, and ensure adherence to ethical standards [28]. Beyond transparency, it aids in identifying weaknesses and improving models [29]. Integrating XAI transforms AI from a black box into a trusted partner, enhancing safety and patient care by aligning technology with clinical expertise.

5.1. Gradient-weighted Class Activation Mapping (Grad-CAM)

In the proposed framework, an XAI approach was utilized to enhance the understanding and transparency of the polycystic ovarian syndrome detection model, addressing significant aspects of artificial intelligence in healthcare. This research employs the Grad-CAM method to visualize and identify the most significant regions within ovary ultrasound images that contributed to the model's prediction. The results obtained by using the Grad-CAM method for PCOS diagnosis demonstrate how well it works in identifying significant regions in ovary ultrasound images and improving the interpretability of the model. A trained deep learning model uses the grey-scale ovary ultrasound image as an input, shown in Fig. 8 (a). Though the deep learning model predicts the presence or absence of PCOS, its decision-making process still remains unknown due to the inherent "black-box" aspect of deep learning. Grad-CAM addresses this by displaying the regions of the image having the biggest impact on the model's prediction. These critical areas are visible when a Grad-CAM heatmap, as shown in Fig. 8 (b), is overlaid on the original image. Grad-CAM operates by utilizing the gradients of the target class S_c (e.g., 'PCOS infected' or 'PCOS not infected') with respect to the activations of the last convolutional layers of the neural network [30]. The gradients indicate how much these spatial locations in the feature maps contribute to the prediction.

The gradient of the class c score S_c for a particular CNN layer with respect to the activation function F_K of the k^{th} feature map is provided by:

$$\frac{\partial S_c}{\partial F_k} \tag{11}$$

The importance weight of the k^{th} feature map, β^k is obtained by averaging the gradients over all spatial locations (i, j) as:

$$\beta^k = \frac{1}{H \times W} \sum_i \sum_j \frac{\partial S_c}{\partial F_{ij}^k} \tag{12}$$

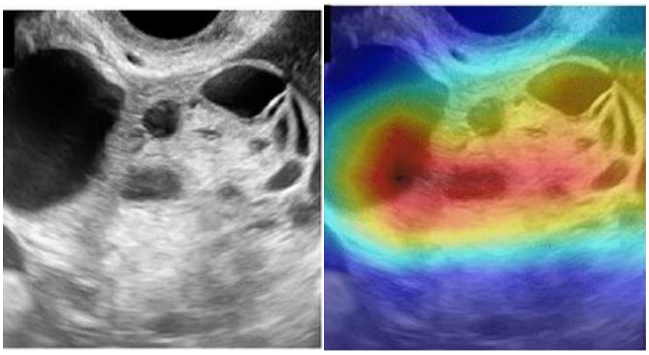
where H and W represent the height and width of the feature map.

The Grad-CAM heatmap $H_{\text{Grad-CAM}}^c$ is then calculated by combining feature maps in a weighted manner, followed by a ReLU operation to retain only the positive influences: The Grad-CAM heatmap is computed as:

$$H_{Gradcam=Relu(\sum_k \beta^k F^k)}^c \tag{13}$$

By averaging these gradients and weighting the corresponding feature map channels, a heatmap is generated that highlights the region of interest in the input image. This heatmap is then normalized and placed back on the original image to provide clear visibility of areas pertinent to the conclusion made by the model. The Grad-CAM heatmap, when overlaid on the ultrasound images, provides a useful visual representation of what the model is focusing on. The Grad-CAM heatmap, when overlaid on the ultrasound images, provides a useful visual representation of what the model is focusing on, allowing researchers and clinicians to check whether the prediction corresponds to clinically relevant features, such as the presence of cysts or abnormally arranged structures in ovaries [31]. As shown in Fig. 8 (b), the Grad-CAM heatmap provides key insights into the model's decision-making process for PCOS diagnosis. The red and yellow patches represent regions of high activation, highlighting significant features that greatly influenced the model's predictions. These highlighted locations are likely associated with cysts or structural abnormalities. In contrast, cooler (blue) regions indicate areas with low activation, which are less relevant to the model's decision-making process. Such interpretability is crucial for clinical applications, ensuring a balance between high accuracy and the consistency of the AI system with current medical practices [32]. Thus, the proposed PCOS detection framework, which incorporates Grad-CAM, addresses a crucial gap by combining high-accuracy AI models with explainability. This study emphasizes the importance of integrating XAI techniques to ensure that AI models not only achieve technical proficiency but also adhere to ethical and practical requirements in real-world medical applications.

The black-box nature of deep learning models presents significant challenges in healthcare diagnostics, where transparency, trust, and interpretability are essential for clinical adoption. Although artificial intelligence systems have demonstrated high accuracy in disease detection tasks, such as the automated diagnosis of Polycystic Ovarian Syndrome (PCOS) from ultrasound images, their limited interpretability restricts their acceptance in real-world medical practice [26]. Explainable Artificial Intelligence (XAI) addresses this limitation by providing insights into model predictions, identifying influential regions and features, and enabling clinicians to validate model decisions against established medical knowledge [27]. Furthermore, XAI facilitates the identification of potential biases, supports the ethical deployment of models, and aids in refining model architectures for improved reliability and robustness [28, 29]. Integrating XAI transforms deep learning models from opaque decision-making systems into transparent and trustworthy clinical decision-support tools.

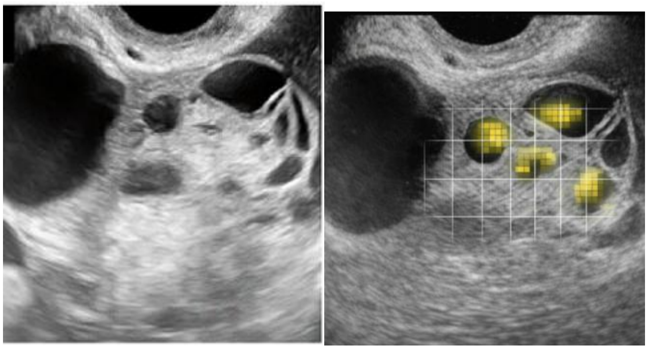


(a) (b)

Fig. 8. Explainable AI Visualization with ResNet50 as a Feature Extractor (a) Original ultrasound image of a PCOS-affected ovary. (b) Grad-CAM heatmap highlighting the regions of interest used for model prediction.

5.2. Attention Rollout for Swin Transformer-Based Feature Extraction

To enhance explainability and address reviewer concerns regarding transformer-based interpretability, the Attention Rollout technique was applied on the Swin Transformer backbone used in the proposed framework. Unlike ResNet50, which relies on convolutional operations to extract localized spatial features, the Swin Transformer leverages hierarchical, patch-based self-attention that captures both local and long-range contextual dependencies through shifted window attention. Consequently, interpretability in Swin Transformers differs fundamentally from convolutional networks, as the model attends to inter-patch relationships rather than localized receptive fields, as shown in Fig. 9.



(a) (b)

Fig. 9. Explainable AI Visualization with Attention Rollout (a) Original ultrasound image of a PCOS-affected ovary. (b) Attention Rollout map highlighting the regions of interest used for model prediction.

In the Swin Transformer-based pipeline, the input ultrasound image $\mathbf{I} \in \mathbb{R}^{3\times H_0\times W_0}$ partitioned into non-overlapping patches and processed through multiple transformer stages. For interpretability, the attention maps from all layers were utilized to compute the rolled-out attention, capturing cumulative contributions of each patch to the final prediction. . Let the attention map from layer l and head h be:

$$A^{(l,h)} \in \mathbb{R}^{N\times N} \tag{14}$$

where N denotes the number of patches. The multi-head attention maps are averaged over heads:

$$\tilde{A}^{(l)} = \frac{1}{H}\sum_{h=1}^{H} A^{(l,h)} \tag{15}$$

H being the number of heads. The Attention Rollout is then recursively computed across layers to account for residual connections:

$$R^{(l)} = \tilde{A}^{(l)} R^{(l-1)}, \quad R^{(0)} = I_N \tag{16}$$

where I_N is the identity matrix representing the initial attention and $R^{(l)}$ is the cumulative attention at layer l. The final rolled-out attention $R^{(L)}$ captures the overall contribution of each patch to the class prediction.

The spatially interpretable saliency map is obtained by reshaping $R^{(L)}$ to the patch grid of the last stage and up-sampling using bilinear interpolation to match the input image resolution:

$$L^{c}_{\text{Attention}} = \text{Upsample}(R^{c}) \tag{17}$$

where R^{c} corresponds to the rolled-out attention for the target class c.

Compared to ResNet50-based Grad-CAM, which highlights regions primarily based on localized convolutional activations, Swin Transformer-based attention maps integrate both local and global dependencies across the image. This allows the model to capture complex spatial relationships, such as follicular arrangements and stromal abnormalities, providing a more context-aware explanation Fig. 10. Thus, explainable AI derived from Swin Transformers not only identifies salient ovarian regions but also visualizes their inter-patch interactions, offering richer and clinically meaningful interpretability that is unattainable with traditional convolutional feature-based XAI approaches.

5.3. Expert Radiologist Assessment

In addition to quantitative performance evaluation, the interpretability of the visualizations was independently assessed by an experienced gynecological radiologist. The expert looked at the saliency maps from both the ResNet50 and Swin Transformer backbones and evaluated their correspondence with the recognized sonographic markers of PCOS, such as the appearance of follicular cysts, ovarian enlargement, and abnormal follicular distribution. According to the radiologist, the focus maps created by the Swin Transformer showed more anatomical consistency and contextual understanding, with the most important areas being able to identified more frequent clinically relevant ovarian structures. This qualitative evaluation is in line with the clinical feasibility of the model explanations and thus, from the diagnostic point of view, it is in agreement with the explanatory PCOS detection framework that has been proposed, which is more reliable.

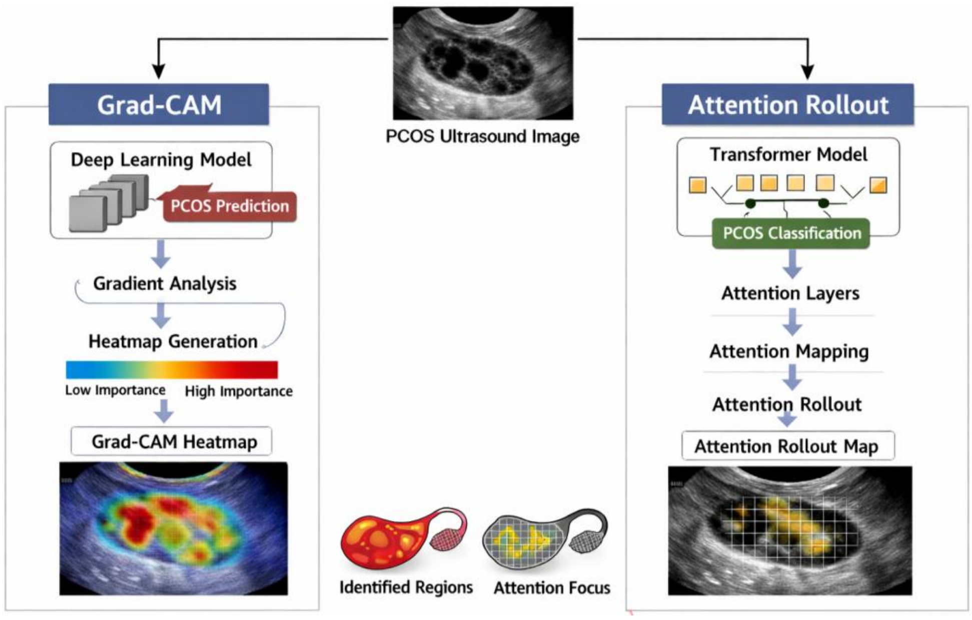


Fig. 10. Explainable AI Visualization for PCOS Diagnosis using Grad-CAM and Attention Rollout.

6. Clinical Applicability and Practical Considerations

Although the proposed PCOS diagnosis framework performs well on the selected ultrasound image dataset, there are a few real-world implementation aspects to consider before applying the framework to clinical settings. One critical aspect to consider is the variability introduced by human observers when interpreting ultrasound images. Since human observers possess varying levels of expertise, their interpretations may vary, thereby creating variability during the training and testing phases. Another critical aspect to consider is the variability introduced by the imaging equipment used during ultrasound image acquisition.

In terms of its implementation, the computational performance of the proposed approach, including the processing time of the model, its memory demand, as well as the accessibility of the required hardware, needs to be seriously considered to readily incorporate the proposed approach as a natural extension of the existing academic setting of PCOS diagnosis. Moreover, the potential clinical implications of a wrongly predicted result, including both positive and negative predictions, urgently demand a serious discussion on the significance of the proposed approach in the clinical setting of PCOS diagnosis in the future.

7. Conclusion and Future Research Directions

To address the shortcomings in existing methods, this work presents an advanced ensemble-based approach for the automation of the detection of Polycystic Ovarian Syndrome in ovarian ultrasound images by leveraging the strengths of the Swin Transformer architecture in feature extraction, in addition to its complementarity with the Random Forest classifier and the Gradient Boosting classifier. Although ResNet50 has long been recognized for its efficacy in learning high-level discriminate features, its convolutional neural network architecture largely focuses on the importance of local spatial patterns. However, in contrast to conventional convolutional neural network architecture, the Swin Transformer architecture has its layers designed in a hierarchical manner with shifted windows by using self-attention mechanisms, which are apt in modeling both local patterns as well as long-term dependencies.

The proposed method attained an accuracy of 99.89%, which is an improvement over the ResNet50 ensemble model and all previous state-of-the-art approaches. Although the difference in accuracy may seem small, improvements of this nature can be quite significant for medical diagnosis applications, where small margins of difference can translate into large differences in outcomes. The new method conclusively proves the superiority of the Swin Transformer model over ResNet50 in capturing features related to diagnoses and can be used for improving the detection of PCOS.

The generalization capability of the proposed model can be strengthened by the application of systematic approaches to data augmentation, thereby reducing issues with over-fitting and addressing any variance in ultrasound image acquisition. For an extensive evaluation of model performances, various parameters, such as precision, recall, and confusion matrices, in addition to accuracy, can be utilized. This allows an equal measure of emphasis to be placed upon different parameters. Additionally, to provide explainable insights into the proposed model approaches, such as those using the concept of Grad-CAM, can be applied. The result will be an indication of regions in the ultrasound images of ovaries influencing the proposed model's predictions, in compliance with PCOS characteristics.

The effective result obtained by the Swin Transformer-based ensemble model confirms its usability for clinical use, especially as a decision-support system that can help in the early and accurate diagnosis of PCOS, thereby minimizing the need for the interpretation of results. This study has made a good foundation for the development of an automatic system for the detection of PCOS by incorporating the latest developments in deep learning models into the ensemble method. In addition, future work may also further improve the proposed framework by considering multimodal data, including hormonal patterns, clinical information, and physical inputs, along with ultrasound images to achieve a holistic diagnostic process. Additionally, the combination of other explainable AI methods, like SHAP, LIME, and ELI5, may help to attain a further understanding regarding the workings of the model, thereby further improving the framework's effectiveness for PCOS diagnosis and Women's Health.

Acknowledgment

We sincerely thank Dr. Rehana Afshana, a radiologist at Superb Diagnostic Clinic, for her expert evaluation of the Explainable AI visualizations and for assessing their correspondence with established diagnostic indicators of Polycystic Ovary Syndrome, thereby supporting the clinical relevance of the proposed framework.

References

[1] Stefano Palomba and et al. Complications and challenges associated with polycystic ovary syndrome: current perspectives. International Journal of Women's Health, pages 745–763, 2015.

[2] Robert L. Rosenfieldand David A. Ehrmann. The pathogenesis of polycystic ovary syndrome (pcos): the hypothesis of pcos as functional ovarian hyperandrogenism revisited. Endocrine Reviews, 37(5):467–520, 2016.

[3] Ricardo Azziz and et al. Polycystic ovary syndrome. Nature Reviews Disease Primers, 2:1–18, 2016.

[4] Rogerio A Lobo and Enrico Carmina. The importance of diagnosing the polycystic ovary syndrome. Annals of internal medicine, 132(12):989–993, 2000.

[5] Samia Ahmed, Md Sazzadur Rahman, Ismate Jahan, M Shamim Kaiser, ASM Sanwar Hosen, Deepak Ghimire, and Seong-Heum Kim. A review on the detection techniques of polycystic ovary syndrome using machine learning. IEEE Access, 11:86522–86543, 2023.

[6] Yihui Xu, Ziyan Cao, Ting Chen, and Jiawei Ren. Trends in metabolic dysfunction in polycystic ovary syndrome: a bibliometric analysis. Frontiers in Endocrinology, 14:1245719, 2023.

[7] RummanAhmad,LameesAMaghrabi,IshfaqAhmadKhaja,LouaiAMaghrabi,andMusheerAhmad. Smote-based automated pcos prediction using lightweight deep learning models. Diagnostics, 14(19):2225, 2024.

[8] Bharti Panjwani, Jyoti Yadav, Vijay Mohan, Neha Agarwal, and Saurabh Agarwal. Optimized machine learning for the early detection of polycystic ovary syndrome in women. Sensors, 25(4):1166, 2025.

[9] Sayma Alam Suha and Muhammad Nazrul Islam. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. ScientificReports, 12(1):17123, 2022.

[10] B. Rachana and et al. Detection of polycystic ovarian syndrome using follicle recognition technique. Global Tran- sitions Proceedings, 2(2):304–308, 2021.

[11] C. Gopalakrishnan and M. Iyapparaja. Multilevel thresholding based follicle detection and classificationof polycys- tic ovary syndrome from the ultrasound images using machine learning. International Journal of System Assurance Engineering and Management, pages 1–8, 2021.
[12] Pratik Wagh, Megha Panjwani, and S. Amrutha. Early detection of pcod using machine learning techniques. In ArtificialIntelligence and Speech Technology, pages 9–20. CRC Press, 2021.
[13] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas. A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics. IEEE Access, 10:97610–97624, 2022.
[14] R. Subha, B. R. Nayana, R. Radhakrishnan, and P. Sumalatha. Computational intelligence for early detection of infertility in women. Engineering Applications of ArtificialIntelligence, 127:107400, 2024.
[15] Poonam Moral, Debjani Mustafi, and Sudip Kumar Sahana. Podboost: an explainable ai model for polycystic ovarian syndrome detection using grey wolf-based feature selection approach. Neural Computing and Applications, 36(30):18627–18644, 2024.
[16] S. Sreejith, H. K. Nehemiah, and A. Kannan. A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier. Healthcare Analytics, 2:100102, 2022.
[17] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, and G. K. Hegde. A distinctive explainable machine learning framework for detection of polycystic ovary syndrome. Applied System Innovation, 6(2):32, 2023.
[18] Abrar Alamoudi, Irfan Ullah Khan, Nida Aslam, Nourah Alqahtani, Hind S Alsaif, Omran Al Dandan, Mohammed Al Gadeeb, and Ridha Al Bahrani. A deep learning fusion approach to diagnosis the polycystic ovary syndrome (pcos). Applied Computational Intelligence and Soft Computing, 2023(1):9686697, 2023.
[19] Pradeep Bedi et al. An integrated adaptive bilateral filter-based framework and attention residual u-net for detecting polycystic ovary syndrome. Decision Analytics Journal, 10:100366, 2024.
[20] S. Umapathy, O. Alhajlah, F. Almutairi, S. Aslam, and A. RK. F-net: Follicles net an efficienttool for the diagnosis of polycystic ovarian syndrome using deep learning techniques. PLoS One, 19(8):Article e0307571, 2024.
[21] Dimitris Perdios and et al. Cnn-based image reconstruction method for ultrafast ultrasound imaging. IEEE Trans- actions on Ultrasonics, Ferroelectrics, and Frequency Control, 69(4):1154–1168, 2021.
[22] Sheldon Mascarenhas and Mukul Agarwal. A comparison between vgg16, vgg19 and resnet50 architecture frame- works for image classification. In 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON), volume 1, pages 96–99. IEEE, 2021.
[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
[24] Ze Liu, Yutong Lin, Yue Cao, Han Hu, YixuanWei, ZhengZhang, StephenLin, andBainingGuo. Swintransformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision, pages 10012–10022, 2021.
[25] AKM Salman Hosain, Md Humaion Kabir Mehedi, and Irteza Enan Kabir. 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), pages 1–6. IEEE, 2022.
[26] A. B. Arrieta, N. D´ıaz-Rodr´ıguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, and F. Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information Fusion, 58:82–115, 2020.
[27] Pavel Hamet and Johanne Tremblay. Artificialintelligence in medicine. metabolism, 69:S36–S40, 2017.
[28] H. Hagras. Toward human-understandable, explainable ai. Computer, 51(9):28–36, 2018.
[29] Subhashis Suara, Aayush Jha, Pratik Sinha, and Arif Ahmed Sekh. Is grad-cam explainable in medical images? In International Conference on Computer Vision and Image Processing, pages 124–135. Springer, 2023.
[30] Mengying Xiao, Liyuan Zhang, Weili Shi, Jianhua Liu, Wei He, and Zhengang Jiang. A visualization method based on the grad-cam for medical image segmentation model. In 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), pages 242–247. IEEE, 2021.
[31] W. Lv, Y. Song, R. Fu, X. Lin, Y. Su, X. Jin, and G. Huang. Deep learning algorithm for automated detection of polycystic ovary syndrome using scleral images. Frontiers in Endocrinology, 12:789878, 2022.
[32] Jong-Chih Chien et al. The usefulness of gradient-weighted cam in assisting medical diagnoses. Applied Sciences, 12(15):7748, 2022.
[33] Mengcheng Cai, Zhexin Ni, Zike Yuan, Jin Yu, Danying Zhang, Ruipin Yao, Ling Zhou, and Chaoqin Yu. Past and present a bibliometric study on polycystic ovary syndrome. Journal of Ovarian Research, 16(1):42, 2023.

Authors’ Profiles

Ifra Bilal Shah is currently pursuing a Ph.D. in computer science and engineering from the National Institute of Technology Srinagar, Jammu and Kashmir, India. She received the Master of Computer Science and Technology degree from the Department of Computer Science and Technology, Sharda University, Greater Noida, India, and the Bachelor of Technology degree in computer science and engineering from the Islamic University of Science and Technology, Awantipora, Jammu and Kashmir, India. Her major field of study is computer science and engineering, with a focus on machine learning and deep learning methodologies.

She is presently engaged in doctoral research at the National Institute of Technology Srinagar. Her research primarily focuses on the application of machine learning and deep learning techniques for automated diagnosis of Polycystic Ovary Syndrome (PCOS) using medical imaging and clinical data, along with research interests in audio steganography. She has authored one research article published in a Scopus-indexed journal and one paper presented at an IEEE international conference. Ms. Shah's current research interests include medical image analysis, explainable artificial intelligence, and deep learning–based diagnostic systems.

Pramod Kumar Yadav is currently working as an Assistant Professor in the Department of Computer Science and Engineering at the National Institute of Technology Srinagar, Jammu and Kashmir, India. He received the Ph.D. degree in computer science and engineering from Jamia Millia Islamia, New Delhi, India, in 2018. His major field of study includes computer science and engineering with an emphasis on data-driven and large-scale computing systems.

He has extensive academic and research experience in teaching and research. He has published several research articles in reputed international journals and conferences and has authored seven textbooks in various domains of computer engineering. He is the inventor of three granted international patents and two published national patents. In recognition of his outstanding academic performance and dedication to teaching and research, he has received the "Outstanding Faculty Award" and the "Award of Excellence." His current research interests include big data analytics, data mining, query optimization, cloud computing, machine learning, and the Internet of Things. Dr. Yadav has actively contributed to academic research and innovation through scholarly publications, patents, and academic leadership in the field of computer science and engineering.

How to cite this paper: Ifra Bilal Shah, Pramod Kumar Yadav, "Automated PCOS Detection Using Fine-Grained Deep Feature Extraction and Explainable AI: A Transformer-Based Ensemble Approach", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.1, pp. 88-105, 2026. DOI:10.5815/ijigsp.2026.01.06