



## OPEN Automated high precision PCOS detection through a segment anything model on super resolution ultrasound ovary images

S. Reka<sup>1</sup>, T. Suriya Praba<sup>1</sup>✉, Mukesh Prasanna<sup>1</sup>, Vanipenta Naga Nithin Reddy<sup>1</sup> & Rengarajan Amirtharajan<sup>2</sup>

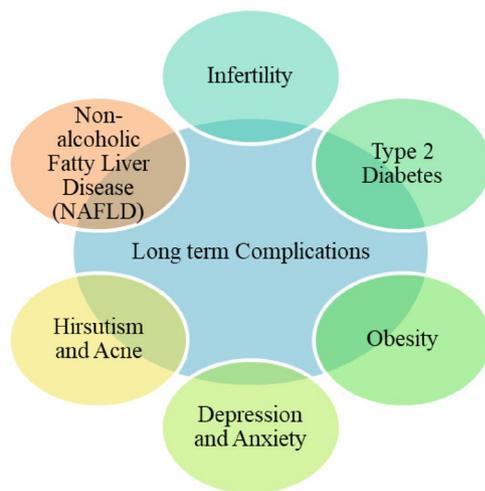
PCOS (Poly-Cystic Ovary Syndrome) is a multifaceted disorder that often affects the ovarian morphology of women of their reproductive age, resulting in the development of numerous cysts on the ovaries. Ultrasound imaging typically diagnoses PCOS, which helps clinicians assess the size, shape, and existence of cysts in the ovaries. Nevertheless, manual ultrasound image analysis is often challenging and time-consuming, resulting in inter-observer variability. To effectively treat PCOS and prevent its long-term effects, prompt and accurate diagnosis is crucial. In such cases, a prediction model based on deep learning can help physicians by streamlining the diagnosis procedure, reducing time and potential errors. This article proposes a novel integrated approach, QEI-SAM (Quality Enhanced Image – Segment Anything Model), for enhancing image quality and ovarian cyst segmentation for accurate prediction. GAN (Generative Adversarial Networks) and CNN (Convolutional Neural Networks) are the most recent cutting-edge innovations that have supported the system in attaining the expected result. The proposed QEI-SAM model used Enhanced Super Resolution Generative Adversarial Networks (ESRGAN) for image enhancement to increase the resolution, sharpening the edges and restoring the finer structure of the ultrasound ovary images and achieved a better SSIM of 0.938, PSNR value of 38.60 and LPIPS value of 0.0859. Then, it incorporates the Segment Anything Model (SAM) to segment ovarian cysts and achieve the highest Dice coefficient of 0.9501 and IoU score of 0.9050. Furthermore, Convolutional Neural Network – ResNet 50, ResNet 101, VGG 16, VGG 19, AlexNet and Inception v3 have been implemented to diagnose PCOS promptly. Finally, VGG 19 has achieved the highest accuracy of 99.31%.

**Keywords** Poly-cystic ovary syndrome, Enhanced super resolution generative adversarial networks, Segment anything model, Convolutional neural network

Poly-Cystic Ovarian Syndrome (PCOS) is one of the most frequent endocrine illnesses affecting reproductive-aged women globally, with a frequency ranging from 8 to 13% depending on diagnostic criteria and demographic investigated<sup>1</sup>. PCOS is a disorder marked by irregular menstruation, hyperandrogenism, and poly-cystic ovarian morphology, as seen on ultrasound imaging<sup>2</sup>. Ovarian enlargement and the development of several tiny cysts inside the ovaries or on the surface of the eggs are the symptoms associated with this illness. The immature eggs inside the cysts are not released during ovulation. In addition, Women with PCOS have aberrant hormonal and metabolic conditions, which may increase their risk of getting cancer. Prolonged hormone stimulation can cause women to develop breast, ovarian, and endometrial cancers. Figure 1 shows the long-term consequences of PCOS.

Ultrasound imaging is the primary technique for evaluating ovarian morphology, which is essential for diagnosing and treating PCOS. According to the Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group (2004), the 2003 criteria define PCOS as the presence of 12 or more follicles in each ovary, measuring 2–9 mm in diameter and/or an elevation in ovarian volume (> 10 mL). Also, the presence of two of the three features: poly-cystic ovaries on ultrasound, clinical and/or biochemical signs of hyperandrogenism, and anovulation<sup>3</sup>. Ovarian morphology is the most commonly identifiable feature of ultrasonography among these

<sup>1</sup>School of Computing, SASTRA Deemed University, Thirumalaisamudram, Thanjavur 613401, India. <sup>2</sup>School of Electrical and Electronics Engineering, SASTRA Deemed University, Thirumalaisamudram, Thanjavur 613401, India. ✉email: suriyaprabha@cse.sastra.edu



**Fig. 1.** Long-term complications of PCOS.

criteria, and it is frequently used in clinical practice to diagnose PCOS. Although ultrasonography is widely used to diagnose PCOS, the manual interpretation of ovarian morphology poses some challenges. Numerous factors, including the quality of the equipment, inter-operator variability, and the parameters used for image acquisition, can affect the detection of cysts and follicles on ultrasound images.

Numerous approaches to image preprocessing, including CLAHE (Contrast Limited Adaptive Histogram Equalization)<sup>13</sup> and deep learning techniques such as (SRCNN) Super-Resolution Convolutional Neural Network<sup>19</sup>, have been used and produced noticeable results in medical imaging. CLAHE is a conventional method that increases the contrast in images by redistributing pixel intensities according to the image's histogram. It independently adjusts the contrast of local image regions. Another technique, SRCNN, is a deep learning model focusing on single-image super-resolutions. It uses convolutional neural networks to learn an end-to-end mapping between low-resolution and high-resolution images. Various techniques have encouraged prospects for enhancing the reliability of PCOS identification from ultrasonography ovary images<sup>4,5,21–26,31</sup>. Whilst many solutions have been presented for PCOS detection with ultrasound imaging, specific issues often limit their performance. This work finds and fixes the following gaps based on recent research on PCOS identification. Preprocessing is crucial in medical image analysis to improve image quality, eliminate noise, and increase model accuracy.

Nevertheless, less focus has been given to resolution enhancement<sup>14,23,25</sup>, which is essential for improving feature extraction by fine-tuning image details. Enhancing resolution can greatly help identify subtle patterns in ultrasound pictures, increasing the accuracy of PCOS prediction. Additionally, the segmentation process is necessary for precise feature extraction to detect important markers like follicle count, ovarian volume, cystic structures and diagnosis. Using deep learning techniques for the segmentation process in PCOS prediction has contributed less<sup>5,14,20,25,26</sup>.

To address these issues, the PCOS detection model requires automatic image enhancement such as contrast enhancement, noise reduction and sharpening to improve the quality and clarity of ultrasound images, making them more suitable for accurate analysis. Furthermore, segmentation models precisely define ovarian structures and identify certain regions of interest, making it easier to perform quantitative analysis to detect abnormalities associated with PCOS. These methods aim to reduce inter-operator variability, streamline PCOS diagnosis, and improve clinical workflow efficiency. Following is the key contribution of this article:

1. Ultrasound images exhibit poor quality; thus, preprocessing is required to increase their effectiveness. In this article, ESRGAN is employed to increase the resolution, sharpen the edges, and restore the finer structure of the ultrasound ovary images.
2. We believe this is the first adaptation of SAM (Segment Anything Model) in ultrasound ovary image segmentation to identify the cyst's existence in PCOS detection.
3. An in-depth evaluation and comparative analysis of the image enhancement model (ESRGAN) and segmentation model (SAM) using various performance metrics.
4. The Application of Convolutional Neural Network architectures such as ResNet 50, ResNet 101, AlexNet, VGG 16, VGG 19, and the Inception v3 model to predict PCOS using an enhanced ultrasound image produced by ESRGAN and a mask generated by SAM.

This article is organised as follows. “[Related works](#)” section presents the existing works on image preprocessing, segmentation and classification related to PCOS detection. “[Methods and materials](#)” section explains the methods and methodologies used in this research work. The proposed model is illustrated in “[Proposed model](#)”. Dataset description and implementation details are described in “[Experiments](#)” section. The result of the proposed model is discussed in “[Results and discussion](#)” section, and the conclusion of the research work is presented in “[Conclusions](#)”.

## Related works

### Preprocessing and segmentation background

Preprocessing and segmentation are essential to properly analyse, interpret, and make clinical decisions using medical image data. By improving image quality, standardising datasets and extracting relevant characteristics, these techniques enhance medical imaging research and improve patient care. Preprocessing and segmenting medical image datasets using deep learning has emerged as a promising approach. It enhances the precision and efficiency of preprocessing and segmentation operations by automating the extraction of significant features from unstructured medical data using neural networks. Egger et al. provide a review of deep learning techniques in medical applications. Various methods and challenges are discussed, including data scarcity, time/cost investment in feature extraction, and disease diagnosis with numerous datasets<sup>6</sup>. Akkus et al., The author provides an overview of deep learning-based ultrasonic applications that enhance clinical workflow by enhancing ultrasound image acquisition, real-time image quality evaluation, object detection and disease diagnosis<sup>7</sup>. The two-dimensional Fractional Fourier Transform (2D-FrFT) derived a denoising filter to eliminate distortions from PCOS ultrasound images in the time-frequency domain. To analyse the optimal fractional operator parameter of the 2D-FrFT, the VGG-16 deep learning model is used. Metrics like PSNR, SSIM, and RMSE are used to evaluate the improved image quality<sup>8</sup>.

Chandrasiri et al. Utilise ESRGAN to convert low-resolution images into high-resolution images. The generator in this model uses two different channel attentions (SENet and ECA-Net) to increase the output image resolution. NIQE and LPIPS metrics are used to evaluate the quality level of the generated images<sup>9</sup>. SAM-IE (Segment Anything Model for Image Enhancement) was used by C. Wang et al. to improve medical images. Their method entailed generating attention maps by merging the original image with binary and contour masks produced by SAM to enhance the performance of the diagnostic model. The authors tested four sets of medical picture data to verify the SAM-IE model's efficiency<sup>10</sup>. C. Wang et al. presented a framework for medical image annotation using SAM. This framework comprises two sub-modules to generate annotations automatically and assists with manual annotation of medical images<sup>11</sup>. Three stages of an integrated MP-YOLO architecture for segmenting and visualising ultrasonic images were proposed by Wang et al. The YOLO method is used to identify the follicular region and extract the salient features. The authors have contrasted the outcomes of automated ROI generated by integrated MP-YOLO with expert-annotated ROI.

With an IoU (Intersection over Union) score of 94.63%, the outcome shows that automated ROI has obtained the highest results<sup>12</sup>. Alwakid et al. used an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) and Contrast-Limited Adaptive Histogram Equalization (CLAHE) to improve the model's learning capacity. The authors experimented on the Asia Pacific Tele Ophthalmology Society (APTOS) dataset; the Inception v3 model achieved 98.7% accuracy compared to other models<sup>13</sup>. Nazarudin et al. proposed a hybrid segmentation technique (Otsu's thresholding and Chan-Vese method) for identifying follicles in the ultrasound ovary images. Otsu thresholding is used to create a binary mask and boundary of the follicle using the Chan-Vese method. The performance of the proposed hybrid model is compared with the traditional Chan-Vese method. The results showed a significant improvement in Dice score, Jaccard Index and sensitivity<sup>14</sup>.

Marinov et al. presented the various challenges of medical image segmentation, such as the time-intensive nature of manual annotation, which requires expert knowledge. Also, noise and artefacts in the data, variations in scanner types, and variations in population demographics are additional obstacles to obtaining reliable and accurate segmentation<sup>15</sup>. Xiao et al. reviewed transformer-based segmentation models such as U-Net, Swin transformer and vision transformer (ViT). After analysing the numerous research works, it is summarised that the U-net-based transformer model has achieved remarkable results with various medical datasets<sup>16</sup>.

Kirillov et al., SAM (Segment Anything Model) was introduced to improve the segmentation performance by generating high-quality masks. The authors used SAM to generate 1 billion masks on 11 M licensed images<sup>17</sup>. Qian et al. proposed a hybrid NAS (Neural Architecture Search) for ultrasound ovary image segmentation. Transfer learning is applied by incorporating the pre-trained architecture instead of starting from scratch and developing the high-performance model. The authors have experimented on two large ultrasound image datasets, echinococcosis with 9-class and ovary datasets for segmentation<sup>18</sup>.

Sawant et al. proposed a hybrid filter combining the Wiener, Median, Noise adaptive Fuzzy switching median filter and Adaptive median filter for image denoising. Then, denoised images are passed to a super-resolution convolutional Neural Network (SRCNN) to enhance the image quality. The performance of the proposed model has been evaluated using performance metrics like PSNR, SSIM, and Universal Quality Index (UQI), and a significant improvement in PSNR value was observed from 40 to 50%<sup>19</sup>. Gopalakrishnan et al., Various filters were applied to reduce speckle noise in ultrasound images, such as the Lee filter, Kuan filter, Frost filter, Median filter, Gaussian filter, and Wiener filter. The performance of the denoised image was evaluated using various performance metrics, and it was observed that the Frost filter outperformed the other filters well. The result obtained from modified Otsu thresholding was considered as an initial mask. A combination of Modified Otsu with active contour has been employed to detect the exact number of follicles from PCOS ultrasound ovary image<sup>20</sup>.

### Classification model

Bedi et al. used adaptive bilateral filtering for image denoising and proposed AResUNet (Attention-based Residual Unet) for segmentation. They have experimented with 2D and multi-modal ultrasound images and achieved 98% accuracy<sup>21</sup>. Reka et al. used ultrasound images and Hormonal datasets as input for PCOS prediction. A Generative Adversarial Network (GAN) is employed to increase the dataset size. The authors experimented with ultrasound ovary images using the Residual and Inception network, and various machine-learning approaches were implemented to classify PCOS using hormonal data. The performance of the proposed model is evaluated using accuracy, precision, recall and F1 Score<sup>22</sup>. Alamoudi et al. employed AHE - Adaptive

Histogram Equalization for image enhancement. The authors have used learning models such as MobileNet, DenseNet 121, DenseNet 201, VGG 16, VGG 19, and Inception v3 to extract the important features and classify the PCOS data<sup>23</sup>.

Chitra et al. developed a hybrid CNN model – a combination of ResNet, Alexnet, VGG 16 and Inception v3 for classifying PCOS ultrasound images. They achieved the highest accuracy of 95% than individual models<sup>24</sup>. Salman Hosain et al. proposed a PCONet - PCOS detection model using the transfer learning method and fine-tuned pre-trained Inception V3 model. Experimentation was conducted on ultrasound images, and 98.12% of the results were achieved with the proposed PCONet model compared to fine-tuned Inception v3<sup>25</sup>. Gopalakrishnan et al. used Gaussian low pass filter for image preprocessing and multilevel thresholding for image segmentation. The authors used a supervised machine learning model for classification and presented the Multifactor Dimensional Reduction (MDR) technique for feature extraction. Support Vector Machine (SVM) outperformed with an accuracy of 93.82% compared to other tested classifiers, including Random Forest, Naïve Bayes, and Linear Discriminant<sup>26</sup>.

Classical image enhancement and segmentation techniques have been the main focus of PCOS detection in previous research works<sup>5,14,20,23</sup> to improve classifier model performance. Hitherto, the focus has been on metabolic markers instead of using ovarian ultrasound images to determine PCOS. Ultrasound ovary images provide more precise and accurate diagnostic results than metabolic indications, a numerical value that might not always provide exact solutions. PCOS diagnosis requires visual information from ultrasound imaging, which shows crucial ovarian characteristics such as the size and existence of cysts. There is also a dearth of research on the segmentation of ultrasound ovary images. These deficiencies highlight the need for additional research to solve these issues and improve methods for PCOS diagnosis.

This research aims to address these gaps with ultrasonography ovarian imaging by providing a reliable and timely diagnosis of Poly-Cystic Ovarian Syndrome. The suggested approach uses ESRGAN for ultrasound ovarian image enhancement and comparison analysis. Additionally, SAM (Segment Anything Model) is employed for cyst segmentation. Finally, Convolutional Neural Networks such as ResNet 50, ResNet 101, VGG 16, VGG 19, Alexnet, and Inception v3 have been implemented to diagnose PCOS.

## Methods and materials

### Ultrasound ovary image enhancement using ESRGAN

ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) is essential for improving ultrasound ovarian image quality to predict PCOS. By producing high-fidelity images that preserve precise information, it successfully handles issues related to ultrasonography. With a generator and a discriminator, ESRGAN uses a dual-network design.

From a low-resolution input image,  $z$ , the generator network aims to produce a high-resolution image,  $G(z)$ . A function can represent it  $G: R \rightarrow X$ , where  $X$ , is the space of high-resolution images and  $R$  is the space of low-resolution images. Decreasing a reconstruction loss, often determined by comparing the output image pixel-by-pixel to ground truth high-resolution images, can map input images  $z$  to high-resolution images  $G(z)$ .

The discriminator network aims to distinguish between the produced high-resolution images  $G(z)$  and the real high-resolution images. A function  $D: X \rightarrow [0, 1]$  is often used to represent the discriminator, where  $D(x)$  represents a probability that the image  $x$  is real. It is designed to maximise the probability of assigning low scores to newly generated images and high scores to real ones. To enhance the quality of generated images, ESRGAN uses an adversarial training technique in which the discriminator and generator networks dynamically interact. The adversarial training process is represented in Eqs. (1),

$$\min_G \max_D E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_{data}(z)} [\log (1 - D(G(z)))] \quad (1)$$

Where  $G(z)$  indicates generated high-resolution images,  $P_{data}(x)$  denotes real high-resolution image distribution and  $P_{data}(z)$  represents the low-resolution input image distribution.

### Segmentation using SAM (Segment anything Model)

Segmentation masks for medical images can be generated using pre-trained SAM. It creates segmentation masks for each discernible region in a medical image and organises them into a comprehensive list<sup>10</sup>. This article uses SAM to identify the cyst region and extract the most pertinent features. SAM is designed for flexible segmentation and works well in identifying cyst regions without requiring explicit instance segmentation, making it well-suited for medical imaging tasks. Also, unlike models like UNet, which require extensive training on domain-specific datasets, SAM can be used directly without pre-training. The important problem in medical datasets, especially in cyst detection, can be biased due to demographic factors, age, community, and dietary habits. Traditional models like UNet might struggle to generalise across different datasets, while SAM, being foundation-model-based, is designed to work across diverse data distributions. Unlike traditional segmentation models requiring retraining for each dataset, SAM can segment cysts effectively across different datasets without requiring extensive fine-tuning. It consists of three components: an image encoder, a prompt encoder, and a mask decoder.

Image Encoder: SAM takes advantage of the scalability and powerful pre-training approaches using a Masked Auto Encoder (MAE) pre-trained Vision Transformer (ViT), finely intended to handle high-resolution inputs with few modifications. An image encoder is used before prompting the model, and each image is processed only once.

**Prompt Encoder:** SAM utilises two types of prompts: dense (masks) and sparse (points, boxes). It represents points and boxes by combining positional encodings with learnt embeddings for each prompt type. Convolutions are used to embed dense prompts (masks), combined element-wise with image embedding.

**Mask Decoder:** The prompt embedding, output, and picture embedding are effectively mapped to a mask using a mask decoder. A dynamic mask prediction block is used with a modified Transformer decoder block. To update all embeddings, the redesigned decoder block of SAM employs prompt self-attention and cross-attention in two directions: prompt-to-image embedding and image embedding to prompt. The image embedding is upsampled by SAM and mapped to a dynamic linear classifier through a multilayer perceptron. After executing two blocks, the classifier computes the mask foreground probability at each image point.

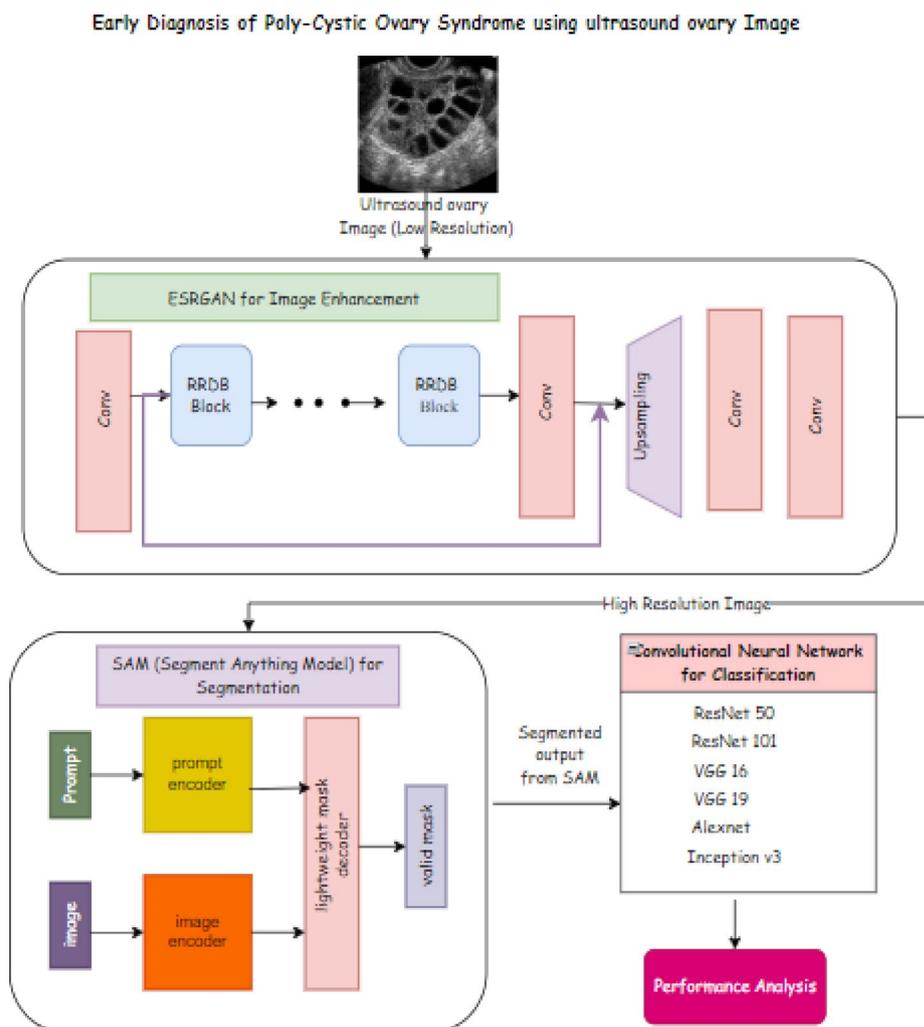
### Classification model using convolutional neural network

A popular deep learning model for image classification tasks is the convolutional neural network (CNN). CNN is built to automatically and adaptably identify feature spatial hierarchies from raw image data. It consists of several layers, including fully connected, pooling, and convolutional. In this article, ResNet 50, ResNet 101, Alexnet, Inception v3, VGG 16 and VGG 19 models are considered for classifying PCOS and Non-PCOS images.

### Proposed model

Multiple ovarian cysts, irregular menstrual cycles, increased testosterone levels and insulin intolerance characterise Poly-cystic Ovary Syndrome (PCOS). Its exact origin is still elusive. Many ailments, such as infertility, obesity, diabetes, and cardiovascular disease, are associated with PCOS<sup>23,25</sup>. Early PCOS diagnosis is crucial for better results and prevention of difficulties. To help physicians identify PCOS, leveraging deep learning models can be highly beneficial in the diagnostic process using ultrasound ovary images<sup>21,23,24</sup>. The proposed work has been divided into three major parts to develop an effective treatment strategy (Fig. 2).

1. Image Enhancement using ESRGAN (Enhanced Super-Resolution Generative Adversarial Network).
2. Segment Anything Model (SAM) to obtain Region of Interest (RoI).



**Fig. 2.** Proposed architecture of early diagnosis of PCOS.

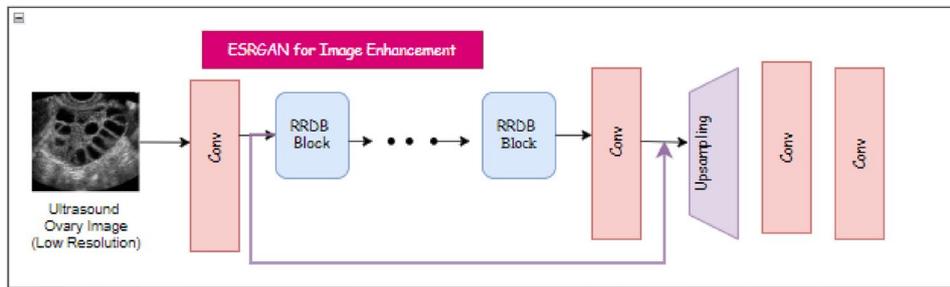


Fig. 3. ESRGAN architecture.

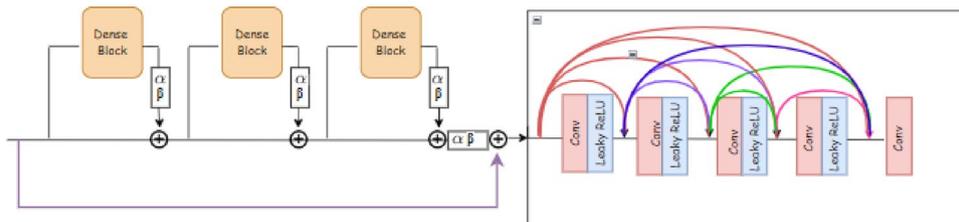


Fig. 4. Residual in residual dense block (RRDB).

3. Classification Model.

**Preprocessing and image enhancement using ESRGAN (enhanced super-resolution generative adversarial network)**

Preprocessing ultrasound ovary images is crucial for promptly identifying Poly-cystic Ovary Syndrome (PCOS). In this article, CLAHE<sup>13</sup> has been used as a preprocessing technique by enhancing local contrast in ultrasound ovary images, preventing noise amplification and improving the ability to discern important ovarian structures like follicles. It efficiently manages uneven illumination and enhances image clarity to improve segmentation and feature extraction. In addition, strengthening the ultrasound ovary image can highlight significant details and structures and increase the visibility of anomalies linked to PCOS. These techniques help identify cysts, follicular changes and other PCOS indicators early by increasing contrast, reducing artifacts, and highlighting the pertinent region of interest. This article uses ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)<sup>27</sup> to enhance the ultrasound ovary image by increasing the resolution. This network uses the generator to produce high-resolution images by converting the low-resolution input images and discriminator to differentiate the high-resolution images made by the generator from the original images.

Figure 3 illustrate the architecture of ESRGAN. It comprises several convolutional filters, residual-in-residual blocks (RRDB), and an upsample activation map. To identify the low-level features such as edges, textures, corners and intensity gradients, a convolutional layer with 3 × 3 convolution operation is applied. These inputs are passed to the RRDB block to extract the most pertinent features. Figure 4 depicts the architecture of Residual in Residual Dense Block (RRDB). The RRDB contains three instances of the dense\_block to extract hierarchical features, residual connections to enhance feature reuse by stabilising the model training and skip connection, which passes the important information directly. Each dense block consists of five convolutional filters, for instance, a normalisation function with two parameters, Alpha and Beta, to scale and shift the normalised output followed by an activation function (Leaky ReLU) to avoid dead neurons. Each convolutional layer in the dense\_block function uses 64 filters. The alpha and Beta parameters are set to 0.2. The upscaling factor of 4 is used to scale the low-resolution images, and the residual scalar value is set at zero. The upsampling preserves an ultrasound ovary image’s fine textures and features. Finally, 3 × 3 convolutions are applied to convert the upscaled features to RGB images, which uses Tanh activation to limit pixel values between - 1 and 1. The operation of each specific layer in dense\_block is represented in Eqs. (2),

$$Y_{HR\_image} = \beta \cdot IN (conv (X_{LR\_input})) + \alpha \tag{2}$$

where,  $conv (X_{LR\_input})$  represents the convolution operation on a low-resolution image  $X_{LR\_input}$ ,  $IN$  indicates the Instance Normalization and  $\alpha, \beta$  represents the parameters used to scale and shift the normalised output.

Instance Normalisation (IN) of an input feature map  $X_{LR\_input}$  is given by,

$$X_{norm} = \frac{X_{LR\_input} - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{3}$$

where  $\mu$  and  $\sigma$  indicate each feature map's mean and standard deviation,  $\epsilon$  is the value for stabilising the normalisation process.

The scale and shift parameters  $\alpha$  and  $\beta$  are applied to the normalised output.  $X_{norm}$ , which is represented in Eqs. (4),

$$Y = \alpha X_{norm} + \beta \quad (4)$$

Three different approaches are being used to improve the quality of the images: CLAHE<sup>13</sup>, SRCNN<sup>19</sup> and ESRGAN<sup>9,13</sup>. Metrics such as Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR) are used to identify which of these three models performs effectively. Algorithm 1 shows the process of the image enhancement technique.

---

Input:  $X_{LR\_input}$  : Low-resolution ultrasound ovary image

Output:  $Y_{HR\_image}$  : Enhanced high-resolution image

$M_{best}$  : Value w.r.t PSNR, SSIM and LPIPS

1. Preprocessing:

$$X_{norm} = \frac{X_{LR\_input} - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad // \text{normalise the input image}$$

2.  $M = \{g, c, s\}$  // g, c, s represents ESRGAN ( $X_{LR\_input}$ ), CLAHE ( $X_{LR\_input}$ ), SRCNN ( $X_{LR\_input}$ )

$Q = \{P, S, L\}$  // P, S, L indicates PSNR, SSIM, LPIPS values

3. for k, n in M.items(): // k indicates the image enhancement technique  
enhanced = n( $X_{LR\_input}$ ) // n indicates the method to perform enhancement

$$PSNR\_V = 10 \cdot \log_{10} \frac{MAX^2}{MSE(X_{LR\_input}, \text{enhanced})}$$

$$SSIM\_V = \frac{(2\mu_{I_1} \mu_{I_2} + B_1)(2\sigma_{I_1 I_2} + B_2)}{(\mu_{I_1}^2 + \mu_{I_2}^2 + B_1)(\sigma_{I_1}^2 + \sigma_{I_2}^2 + B_2)}$$

$$LPIPS\_V = v(I_1, I_2) = \sum_i w_i * v_i(I_1, I_2), \text{ where,} \\ v_i(I_1, I_2) = \frac{1}{p_i} \sum_n \| f_{j(I_1)_n} - f_{j(I_2)_n} \|_2$$

Q[“P”] += [(k, PSNR\_V)]

Q[“S”] += [(k, SSIM\_V)]

Q[“L”] += [(k, LPIPS\_V)]

4. Determine the  $M_{best}$  to obtain  $Y_{HR\_image}$

if ( $P_g > P_c$ ) & ( $P_g > P_s$ ) & ( $S_g > S_c$ ) & ( $S_g > S_s$ ) & ( $L_g < L_c$ ) & ( $L_g < L_s$ ), then

$M_{best} = \text{ESRGAN}$  //  $P_g, P_c, P_s, S_g, S_c, S_s, L_g, L_c, L_s$  refers to Q w.r.t M

else if ( $P_c > P_s$ ) & ( $S_c > S_s$ ) & ( $L_c < L_s$ ) then

$M_{best} = \text{CLAHE}$

else

$M_{best} = \text{SRCNN}$

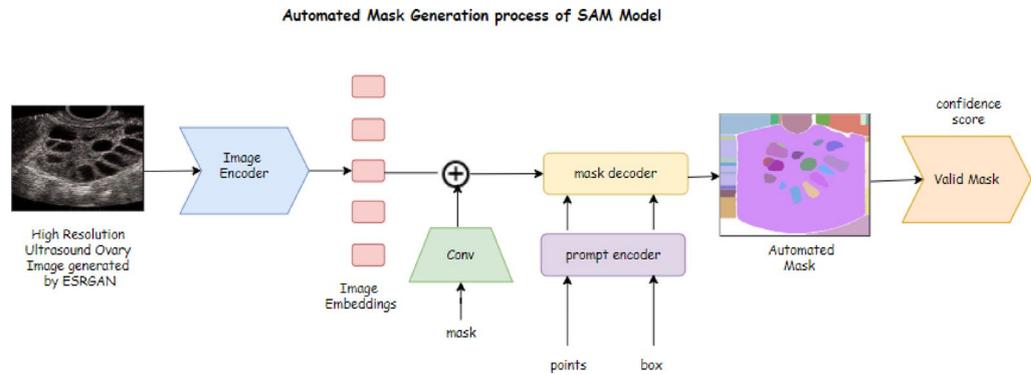
---

**Algorithm 1.** Enhancing low resolution ultrasound ovary image.

---

### SAM (segment anything model) for segmentation

For the early diagnosis of PCOS, segmentation of ultrasonography ovary images is crucial. Accurate segmentation allows for early diagnosis, minimises diagnostic errors, and enhances treatment outcomes. In this article, the Segment Anything Model generates automatic annotation and segments ovarian cysts by developing an automated mask. The Segment Anything Model (SAM) performed well and provides significant benefits in ultrasound image segmentation<sup>28</sup>, such as automated segmentation, improved accuracy and generalisation to new tasks, reduced task-specific expertise, and the ability to handle speckle noise. Figure 5 depicts the architecture of the Segment Anything Model. The enhanced high-resolution images obtained from ESRGAN are used as an input to segment the ovarian cysts. The image encoder and prompt encoder represent the input and prompt image features. The model decoder combines the image results and the prompt encoder to create the mask. The performance of the generated mask is evaluated using IoU (Intersection over Union) and Dice score. Algorithm 2 explains the process of automated mask generation using SAM.



**Fig. 5.** Automated mask generation using segment anything model.

Input:  $Y_{HR\_image}$ : Enhanced high-resolution image,  $I_{prompt}$ : input prompt

Output: Segmented image with object boundary

1. Image Encoder:

$$R_{Y_{(HR\_image)}} = f_i(Y_{HR\_image}) // R_{Y_{(HR\_image)}} \text{ is the feature representation of input image,}$$

$$// f_i \text{ represents the image encoder function}$$

2. Prompt Encoder:

$$P_{I_{prompt}} = f_p(I_{prompt}) // P_{I_{prompt}} \text{ is the feature representation of the input prompt,}$$

$$// f_p \text{ represents prompt encoder function}$$

3. Mask Decoder:

$$M_D = f_m(R_{Y_{(HR\_image)}}, P_{I_{prompt}}) // f_m \text{ indicates mask decoder function}$$

4. SAM Model:

$$SAM_M = f_m(f_i(Y_{HR\_image}), f_p(I_{prompt})) // SAM_M \text{ indicates final mask produced by SAM}$$

5. Evaluate the performance of SAM Metrics (IoU, Dice) to evaluate SAM result

IoU (Intersection over Union):

$$IoU = \frac{TP}{TP+FP+FN}$$

Dice coefficient:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}$$

**Algorithm 2.** Automated mask generation using segment anything model.

### Classification using ESRGAN and SAM results

In this article, a novel integrated approach (QEI-SAM) for the classification of Poly-Cystic Ovarian Syndrome (PCOS) using several prominent Convolutional Neural Network (CNN) classifiers, such as AlexNet, ResNet 50, ResNet 101, VGG 16, VGG 19 and Inception v3 is proposed. With their deep residual learning and skip connections, ResNet-50 and ResNet-101 effectively capture hierarchical information while avoiding vanishing gradients, which makes them appropriate for classifying and segmenting numerous cystic follicles and enlarged ovaries. However, deeper models such as ResNet-101 need a high-dimensional dataset to prevent overfitting. Inception v3 is used due to its factorised convolutions and multi-scale feature extraction, effectively capturing follicular variations at various scales. VGG-16 and VGG-19, recognised for their stacked  $3 \times 3$  convolutions, excel at detecting fine-grained textures such as small follicular cysts, making them helpful for segmentation-based PCOS identification. The lightweight AlexNet model lacks the depth required to identify complex ovarian anomalies used for comparison. The performance of the proposed model is evaluated using various performance metrics, including accuracy, precision, recall, F1 score and AUC score. Algorithm 3 presents the steps involved in the classification of PCOS.

---

Input:  $E_{HR\_image}$ : Enhanced high-resolution image by ESRGAN  
 $SAM_{Seg}$ : Segmented results of enhanced images by SAM

Output: Classification of PCOS images

1. Load enhanced ultrasound image  $E_{HR\_image}$  and segmented cysts result from  $SAM_{Seg}$
2. Build a classifier model
 

Alexnet:  $(T, K)(u, v) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} T(u+i, v+j) K(i, j)$  //convolution operation  
 $g(u) = \max(0, u)$  //ReLU Activation  
 $f = W \cdot u + b$  //Fully connected layer

ResNet:  $R_l = F(R_{l-1}) + R_{l-1}$  //  $F(R_{l-1})$  is the output after applying convolution with different filter size, Batch Normalization, ReLU activation function

VGG:  $I_j = ReLU(\sum_{i \in G_j} T_i * K_{ij} + b_j)$  //convolution operation  
 $V_l = d_l(w_l V_{l-1} + b_l)$  // output of convolution and fully connected layer

Inception v3:  $x = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w H_{conv}(i, j)$  //Global Average Pooling  
 $y = W_{FC} \cdot x + b_{FC}$  // fully connected layer  
 $S(c) = \frac{g^{y_c}}{\sum_{k=1}^K g^{y_c}}$  //softmax layer
3. Evaluate the performance of the classifier and choose the best model with the highest accuracy using
 
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

---

**Algorithm 3.** PCOS classification model.

---

## Experiments

### Dataset

This research conducted experiments on a publicly available dataset<sup>32</sup>: ultrasound ovarian images with and without PCOS from the Kaggle repository to determine the efficiency of the proposed QEI-SAM in improving the accuracy of PCOS classification. This dataset contains two subfolders: train and test with two categories: infected (781 with PCOS) and non-infected (1143 Healthy ovaries). It consists of 1932 ultrasound ovary images. The size of the initial low-resolution image is  $300 \times 300$ , and ESRGAN was trained with this low-resolution image and produced the high-resolution image of  $900 \times 900$ .

### Implementation details and performance metrics

The experiments were performed on the DELL Precision 3660, Intel core i5-12500 Processor and 64GB Memory, and NVIDIA T1000 4GB Discrete Graphics. All models were implemented using Jupyter Notebook. The experimental parameter settings used during implementation are shown in Table 1.

The performance analysis of the classification model with and without ESRGAN and SAM results is evaluated using metrics including AUC (Area under Curve), accuracy, precision, recall and F1 score. Additionally, PSNR (Peak Signal-to-Noise Ratio), Learned Perceptual Image Patch Similarity (LPIPS) and Structural Similarity Index (SSIM) are used to evaluate the quality of enhanced images produced by ESRGAN. Also, to assess the segmented region identified by the SAM -Segment Anything Model, the IoU (intersection over union) and Dice scores are used.

## Results and discussion

### ESRGAN and SAM performance on image enhancement and segmentation

In this article, image enhancement and segmentation of ultrasound ovary image is accomplished with the support of ESRGAN & SAM, as depicted in Fig. 6. The initial row depicts the sample's original low-resolution image. The enhanced image produced by ESRGAN is presented in the second row. The binary and contour masks made by the Segment Anything Model (SAM) for each sample image are displayed in the third row. In this paper, CLAHE<sup>13</sup>, SRCNN<sup>19</sup> and ESRGAN<sup>9,13</sup> are employed as image enhancement techniques. Enhanced image quality is evaluated using performance metrics such as PSNR, SSIM and LPIPS. The results are shown in Table 2. From the results, it is clear that the enhanced image produced by ESRGAN achieved the highest PSNR and SSIM values of 38.60 and 0.9383 and a low LPIPS value of 0.0859 compared with CLAHE and SRCNN results.

Further, an ANOVA test was carried out to find the statistical difference between ESRGAN, SRCNN, and CLAHE by comparing the means of the three groups. Subsequently, the post-hoc pairwise test was conducted to determine the specific group's significant difference. With PSNR, the F-statistic is 10920.99, and the p-value is  $2.07e-11$  ( $< 0.05$ ); for SSIM, it achieves an F-statistic of 840013.40 and a p-value of  $4.55e-17$  ( $< 0.05$ ). Further, LPIPS obtained 134152.79 of F-statistic and a p-value of  $1.11e-14$  ( $< 0.05$ ). Since all metrics' F-statistic value is

Model	Key parameters	Value
ResNet 50 & 101	Activation function	ReLU
	Regularisation	L1 (0.01)
	Optimizer	Adam
	Learning rate	0.0001
	Loss function	Sparse categorical crossentropy
VGG 19	Activation function	ReLU
	Regularisation	L1 (0.01)
	Optimizer	Adam
	Learning rate	0.00001
	Loss function	Sparse categorical crossentropy
Alexnet	Activation function	ReLU
	Regularisation	L1 (0.01)
	Optimizer	Adam
	Learning rate	0.000001
	Loss function	Sparse categorical crossentropy
Inception v3	Activation function	ReLU
	Regularisation	L1 (0.01)
	Optimizer	Adam
	Learning rate	0.000001
	Loss function	Sparse categorical crossentropy

**Table 1.** Experimental parameter settings.

higher and the  $p$ -value is  $< 0.05$ , a significant difference exists between groups in all metrics. Thus, a post-hoc pairwise  $t$ -test was conducted between three models, CLAHE, SRCNN and ESRGAN, to find which model outperforms the others. The results are given in Table 3.

The pairwise test results show that ESRGAN significantly outperformed both CLAHE and SRCNN, while CLAHE and SRCNN performed similarly regarding PSNR. Further, with SSIM, ESRGAN significantly works better than CLAHE and SRCNN. Additionally, CLAHE also significantly outperforms SRCNN. Regarding LPIPS, ESRGAN and CLAHE perform similarly, significantly outperforming SRCNN. Across all metrics, SRCNN is consistently outperformed by both ESRGAN and CLAHE. These results statistically confirm that ESRGAN is the best-performing model regarding reconstruction quality, structural similarity, and perceptual quality.

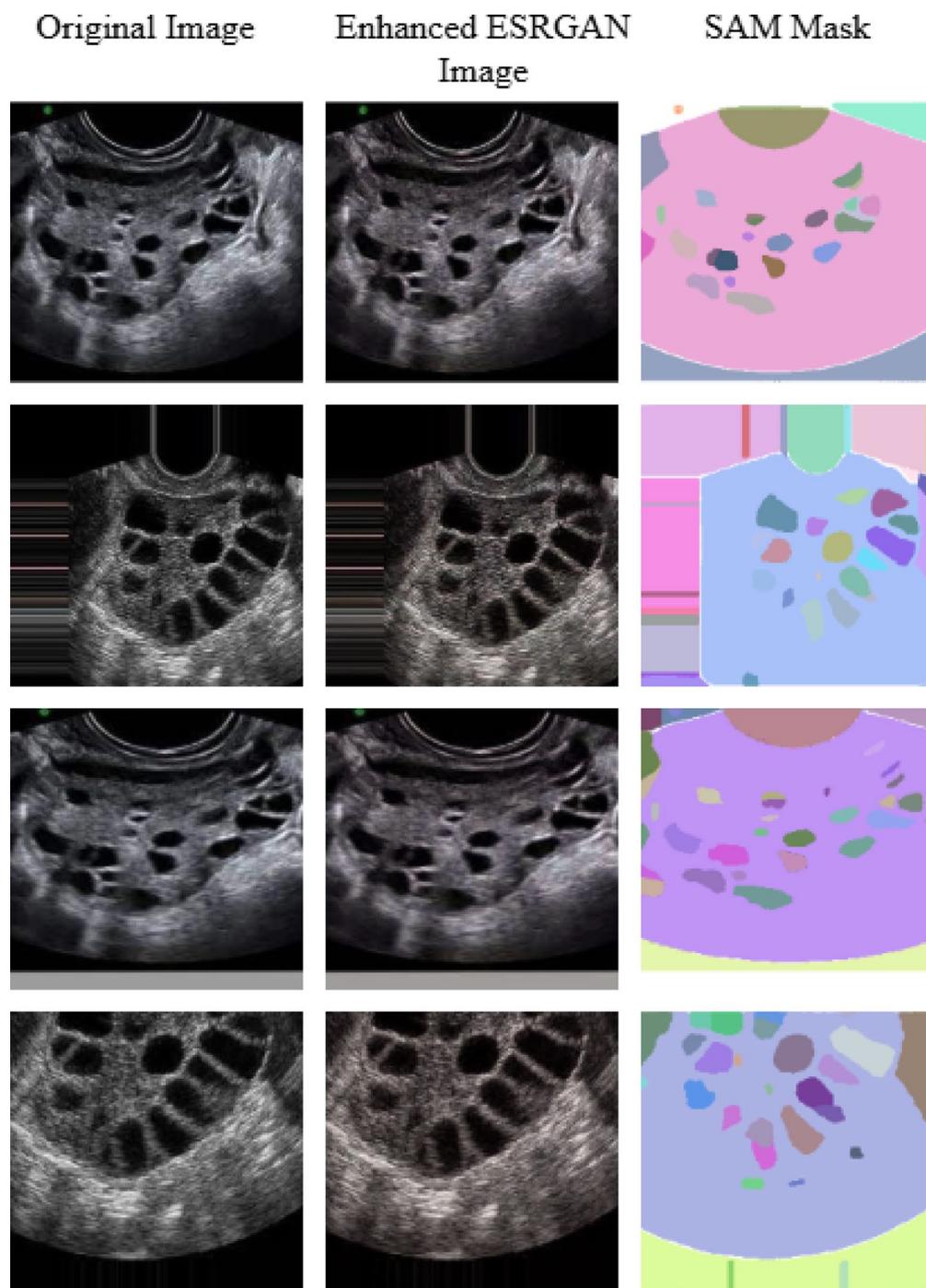
To effectively highlight the cysts region and boundaries in an ultrasound ovary image, SAM (Segment Anything Model) is used. Performance metrics such as IoU and Dice are used to evaluate the performance of SAM segmentation results. A comparative analysis of SAM results with different segmentation models, including U-Net, U-Net++, Cascaded U-Net, PSPNet and Deeplab v3, is also conducted. Table 4 depicts the performance analysis of various segmentation techniques. The result shows that SAM achieved the highest IoU and Dice values of 0.9602 and 0.9501 compared to other models. Further, ANOVA and post-hoc pairwise tests were conducted to determine the statistical significance between the models. With the ANOVA test, the  $F$ -statistic is 2525.7, which is higher, and the  $p$ -value is extremely small ( $1.07e-17 < 0.05$ ), far below the significance threshold of 0.05. This indicates a statistically significant difference in the performance among the models. Thus, a post-hoc pairwise test was conducted to find the statistically significant model. The results are presented in Table 5.

It is observed that SAM has significantly outperformed all the other models. While compared with U-Net, it achieved the  $p$ -value  $< 0.0001$ . Even though U-Net++ has produced remarkable results in the medical dataset, SAM performs significantly better and obtained a  $p$ -value of  $< 0.0001$ . Further, the  $p$ -values of 0.0002, 0.0001, and 0.0005 were obtained by SAM with cascaded PSPnet and Deeplab v3. Thus, SAM consistently performs better with statistically significant differences ( $p$ -values  $< 0.05$  in all pairwise comparisons) than all other models.

### Performance of classifier model

To assess the effectiveness of the proposed QEI-SAM in accurately classifying PCOS ultrasound images. This article examines the classification performance on popular classification architectures, including Alexnet, ResNet 50, ResNet 101, VGG 16, VGG 19, and Inception v3, before and after employing QEI-SAM. The classification results of the classifier model without and with QEI-SAM are displayed in Tables 6 and 7.

Table 6 shows the classification result of individual classifiers without using the images produced by QEI-SAM. It is observed from the results that VGG 19 achieved an accuracy of 85.44%, 88% precision and Recall, F1 score of 85% and 82% AUC score, among other models. Additionally, to visually represent the differences in classification results without QEI-SAM, the ROC (Receiver Operating Characteristic) curve was plotted, as depicted in Fig. 7. It is noticed that the true positive rate of VGG 19 is better than other models. Table 7 illustrates that ultrasound ovary images enhanced with QEI-SAM performed significantly better than the original images on several classifier models. All classification metrics show notable improvements. The results show that VGG 19 outperformed other models with 99.31% accuracy, 99% precision, recall, and F1 score, respectively. Figure 8



**Fig. 6.** Representation of automated mask generation of SAM with enhanced image produced by ESRGAN.

Model	PSNR	SSIM	LPIPS
CLAHE	28.09	0.5993	0.0862
SRCNN	28.20	0.1522	0.3586
ESRGAN	38.60	0.9383	0.0859

**Table 2.** Performance analysis of image enhancement techniques - CLAHE, SRCNN and ESRGAN.

Metrics	Comparative models	p-value
PSNR	CLAHE & SRCNN	0.2878
	CLAHE & ESRGAN	2.15e-02 (<0.05)
	SRCNN & ESRGAN	1.98e-02 (<0.05)
SSIM	CLAHE & SRCNN	3.24e-02 (<0.05)
	CLAHE & ESRGAN	2.87e-02 (<0.05)
	SRCNN & ESRGAN	1.92e-02 (<0.05)
LPIPS	CLAHE & SRCNN	2.55e-02 (<0.05)
	CLAHE & ESRGAN	0.2302
	SRCNN & ESRGAN	2.14e-02 (<0.05)

**Table 3.** The post-hoc pairwise statistical analysis between CLAHE, SRCNN and ESRGAN.

Model	IoU	Dice
U-Net	0.1756	0.2988
U-Net++	0.8648	0.9275
Cascaded U-Net	0.5479	0.7079
PSPNet	0.6500	0.7879
Deeplab v3	0.8629	0.8908
SAM	0.9602	0.9501

**Table 4.** Performance analysis of segmentation models using an enhanced image produced by ESRGAN.

Base Model	Comparative models	t-statistic	p-value
U-Net	U-Net++	-2.85	0.0043
	Cascaded U-net	-2.98	0.0031
	PSPnet	-3.10	0.0021
	Deeplab v3	-3.78	0.0005
	SAM	-5.25	<0.0001
U-Net++	Cascaded U-net	-1.85	0.0674
	PSPnet	-2.57	0.0110
	Deeplab v3	-2.95	0.0032
	SAM	-4.75	<0.0001
Cascaded U-net	PSPnet	-1.95	0.0527
	Deeplab v3	-2.79	0.0053
	SAM	-4.10	0.0002
PSPnet	Deeplab v3	-2.31	0.0213
	SAM	-4.35	0.0001
Deeplab v3	SAM	-3.85	0.0005

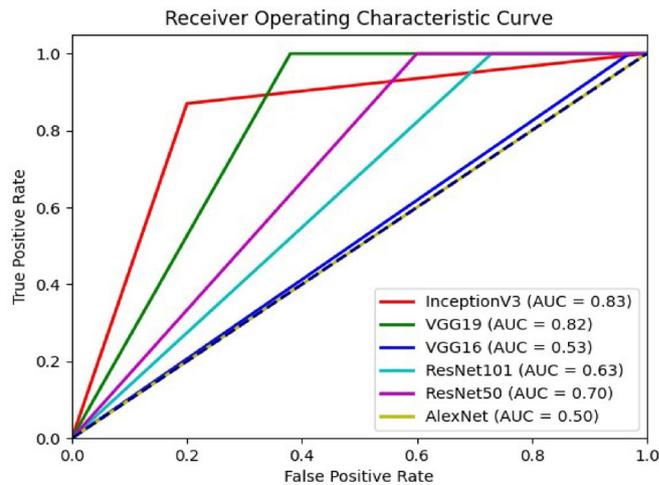
**Table 5.** The post-hoc pairwise statistical analysis of segmentation models.

Model	Accuracy	Precision	Recall	F1 score	AUC score
ResNet 50	76.08	83	76	73	70
ResNet 101	70.88	80	71	65	63
VGG 16	62.39	77	62	50	53
VGG 19	85.44	88	88	85	82
Alexnet	60.31	36	60	45	50
Inception v3	83.71	83	83	83	83

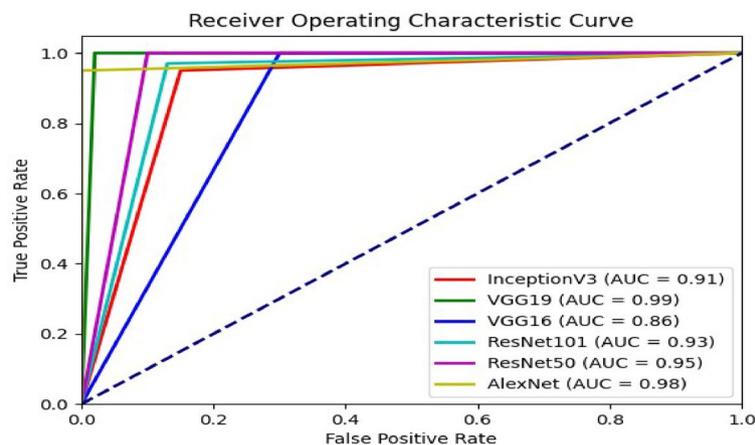
**Table 6.** Performance analysis of individual classifier model without QEI-SAM.

Model	Accuracy	Precision	Recall	F1 score	AUC score
ResNet 50	96.36	97	96	96	95
ResNet 101	93.58	94	94	94	93
VGG 16	88.91	91	89	88	86
VGG 19	99.31	99	99	99	99
Alexnet	97.05	97	97	97	98
Inception v3	91.50	92	92	91	91

**Table 7.** Performance analysis of classification models with ESRGAN & SAM results.



**Fig. 7.** ROC curve of individual classifier without QEI-SAM.



**Fig. 8.** ROC plot of an individual model with ESRGAN and SAM.

shows the RoC plot of the proposed QEI-SAM and that the true positive rate of VGG 19 is higher than that of other models.

The performance of the proposed work is compared with the existing work, as depicted in Table 8. Moral et al.<sup>29</sup> utilised various techniques to enhance the image quality and to obtain RoI, including image resizing, watershed and multilevel thresholding. CystNet was introduced for feature extraction. The improved images with CystNet were considered for model training. The neural network -Dense Layer (FC) and Machine Learning models, including KNN, AdaB, NB, and RF, were implemented, and Random Forest obtained the highest accuracy of 97.75%. Bernatin et al.<sup>24</sup> employed a filter-based univariate attribute selection approach to extract pertinent features and normal scaling for preprocessing. Five classifiers were implemented: Alexnet, Inception V3, Resnet50, VGG16 and hybrid model. The hybrid model has obtained the highest accuracy of 95% than other models. Shanmugavadivel et al.<sup>30</sup> used CNN as a feature extraction technique and employed classifiers

References	Model	Image enhancement technique	Segmentation model/feature extraction	Accuracy %
Moral et al. <sup>29</sup>	Dense Layer (FC), ML: KNN, AdaB, NB, RF	Image resizing, Normalization, Watershed technique, Multilevel thresholding, Morphological Processing	CystNet	97.75
Shanmugavadivel et al. <sup>30</sup>	LR, NB, SVM, CNN, VGG 16	Morphological Processing	CNN for feature extraction	98.29
Bernatin et al. <sup>24</sup>	Alexnet, Inception V3, Resnet50, VGG16 and Hybrid Models	Image resizing	Filtering-based uni variate attribute selection	95
Proposed model (QEI-SAM)	Six CNN models: ResNet 50, ResNet 101, VGG 16, VGG 19, Alexnet, Inception v3	ESRGAN(Enhanced Super Resolution Generative Adversarial Networks)	SAM (Segment Anything Model)	99.31

**Table 8.** Result comparison of proposed QEI-SAM result vs. existing work.

Model	Accuracy	Precision	Recall	F1 score	AUC score
ESRGAN + ResNet 50	90.76	91	91	91	90
ESRGAN + Resnet 101	90.36	91	90	90	89
ESRGAN + VGG 16	68.28	75	68	62	60
ESRGAN + VGG 19	90.81	92	91	91	88
ESRGAN + Alexnet	60.31	36	60	45	50
ESRGAN + Inception v3	86.30	86	86	86	85

**Table 9.** Ablation experimentation results of ESRGAN with classifiers.

including LR, NB, SVM, CNN and VGG 16 to classify PCOS. VGG 16 has achieved a notable accuracy of 98.29%, outperforming machine learning classifiers. The proposed QEI-SAM incorporates ESRGAN for image enhancement and SAM for segmentation. Also, various CNN models such as VGG 16, VGG 19, Alexnet, ResNet 50, ResNet 101 and Inception v3 were implemented, and the model's performance was evaluated. VGG 19 surpassed other models and achieved 99.31% accuracy, higher than the existing results.

### Ablation experiments

This article describes three different ablation experiments to evaluate the proposed model's performance in each stage.

- Performance analysis of the classifier model without preprocessing and segmentation is displayed in Table 6. Due to noise in ultrasound images and inadequate preprocessing, the model has produced less accuracy of 85.44%.
- Further, to improve the performance of the model, CLAHE (only contrast enhancement, fails to increase resolution), SRCNN (obtained finer details, confine to identifying specific patterns from the image), whereas ERGAN effectively preserves anatomical structures and produces edges and textures that are more realistic than SRCNN. The performance analysis of these methods is given in Table 2. Since ESRGAN outperformed the other two methods, classifier models were trained with high-resolution images produced with ESRGAN, which provides a better accuracy of 90.81% with the VGG 19 model, as depicted in Table 9. However, capturing the complicated correlations between features in enhanced ultrasound images produced by ESRGAN is challenging for the model. Consequently, it is difficult for the model to locate crucial information needed for accurate PCOS identification, like cysts boundaries and contour masks.
- To further improve the overall accuracy of the PCOS detection model, the Segment Anything Model (SAM) is integrated with ESRGAN-enhanced ultrasound image for automatic annotations and mask generation to segment ovarian cysts. By focusing solely on the ovary, the model can learn features specific to PCOS, such as cysts size, count and distribution, resulting in the highest accuracy of 99.31% compared to other segmentation models, as shown in Table 10.

### Conclusions

This article demonstrates data-driven early identification of PCOS in women to treat and manage the illness appropriately. The proposed QEI-SAM improves the model's classification performance by enhancing and highlighting the target region and contour in the image. Furthermore, QEI-SAM has identified the precise features that will serve as primary indicators for diagnosing PCOS, hence assisting physicians in quickly and accurately identifying diseases and being cost-effective for patients who have undergone several tests to diagnose their illnesses. In this article, as an initial step, comparative analyses of three image enhancement techniques such as CLAHE, SRCNN and ESRGAN, were conducted, and the quality of the enhanced images was evaluated using PSNR, SSIM and LPIPS. As a result, the best-enhanced images produced by ESRGAN with the highest PSNR value of 38.60, SSIM of 0.938 and LPIPS value of 0.0859 are considered. Subsequently, the Segment Anything Model (SAM) was employed to segment ovarian cysts from enhanced images, and their performance was evaluated and compared with various segmentation techniques. As a result, SAM achieved the highest IoU of 0.9602 and the Dice coefficient of 0.9501, surpassing other models. Finally, Convolutional Neural Network

Model + image enhancement	Segmentation model	Accuracy	Precision	Recall	F1 score	AUC score
ResNet 50 + ESRGAN +	U-net	91.86	92	92	92	90
	U-net++	94.84	93	93	93	94
	Cascaded U-net	90.34	91	91	91	89
	PSPnet	92.38	92	92	91	92
	Deeplab v3	93.89	93	93	93	92
	SAM	96.36	97	96	96	95
ResNet101 + ESRGAN +	U-net	91.57	91	91	90	91
	U-net++	91.87	93	93	93	90
	Cascaded U-net	90.75	90	90	91	90
	PSPnet	92.16	92	92	93	92
	Deeplab v3	92.83	93	93	94	91
	SAM	93.58	94	94	94	93
VGG 16 + ESRGAN +	U-net	86.14	90	88	88	86
	U-net++	87.52	89	90	89	86
	Cascaded U-net	85.28	87	88	87	85
	PSPnet	86.71	89	88	88	85
	Deeplab v3	86.92	88	89	86	85
	SAM	88.91	91	89	88	86
VGG 19 + ESRGAN +	U-net	94.81	95	95	94	93
	U-net++	96.74	96	95	96	96
	Cascaded U-net	92.87	93	93	92	91
	PSPnet	94.89	94	95	94	94
	Deeplab v3	95.78	95	94	95	95
	SAM	99.31	99	99	99	99
Alexnet + ESRGAN +	U-net	84.31	62	60	65	83
	U-net++	92.56	93	91	91	92
	Cascaded U-net	86.54	85	85	84	87
	PSPnet	90.76	86	85	84	90
	Deeplab v3	93.49	92	92	91	94
	SAM	97.05	97	97	97	98
Inception v3 + ESRGAN +	U-net	86.30	85	85	86	85
	U-net++	90.62	89	88	89	90
	Cascaded U-net	86.87	85	85	86	85
	PSPnet	90.56	88	89	89	90
	Deeplab v3	89.67	90	90	88	90
	SAM	91.50	92	92	91	91

**Table 10.** Ablation experimentation results of classifiers with ESRGAN and segmentation models.

models, including ResNet 50, ResNet 101, Alexnet, VGG 16, VGG 19 and Inception v3 have been implemented and evaluated using various performance metrics. Table 7 shows that enhanced and segmented images produced by the proposed QEI-SAM performed well with VGG 19 and achieved the highest accuracy of 99.31% compared to other models. SAM guarantees precise and flexible segmentation of ovarian structures, while ESRGAN improves ultrasound image resolution and contrast. This synergy provides generalizability without requiring much retraining and increases diagnostic accuracy for PCOS identification. To enhance the clinical applicability, future work will concentrate on making the approach more efficient, enhancing structural consistency, and validating it on larger and more diverse datasets.

### Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Received: 26 September 2024; Accepted: 8 May 2025

Published online: 15 May 2025

### References

1. Teede, H. J. et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum. Reprod.* **33**(9), 1602–1618. <https://doi.org/10.1093/humrep/dey256> (2018).

2. Azziz, R. et al. The androgen excess and PCOS Society criteria for the poly-cystic ovary syndrome: The complete task force report. *Fertil. Steril.* **91**(2). <https://doi.org/10.1016/j.fertnstert.2008.06.035> (2009).
3. Fauser, B. C. J. M. Revised 2003 consensus on diagnostic criteria and long-term health risks related to poly-cystic ovary syndrome. *Fertil. Steril.* **81**(1), 19–25. <https://doi.org/10.1016/j.fertnstert.2003.10.004> (2004).
4. GhoshRoy, Debasmita, P. A., Alvi & Santosh, K. C. AI tools for assessing human fertility using risk factors: A state-of-the-art review. *J. Med. Syst.* **47**(1), 91 (2023).
5. Rachana, B. et al. Detection of poly-cystic ovarian syndrome using follicle recognition technique. *Glob. Trans. Proc.* **2**(2), 304–308. <https://doi.org/10.1016/j.gltp.2021.08.010> (2021).
6. Egger, J. et al. Medical deep learning—a systematic meta-review. *Comput. Methods Programs Biomed.* **221**, 106874 (2022).
7. Akkus, Z. et al. A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow. *J. Am. Coll. Radiol.* **16**(9), 1318–1328 (2019).
8. Jha, M., Gupta, R. & Saxena, R. Noise cancellation of poly-cystic ovarian syndrome ultrasound images using robust two-dimensional fractional fourier transform filter and VGG-16 model. *Int. J. Inform. Technol.* **16** (4), 2497–2504. <https://doi.org/10.1007/s41870-024-01773-6> (2024).
9. Liu, J. & Chandrasiri, N. P. CA-ESRGAN: Super-resolution image synthesis using channel attention-based ESRGAN. *IEEE Access* **12**(January), 25740–25748. <https://doi.org/10.1109/ACCESS.2024.3363172> (2024).
10. Wang, C., Chen, H., Zhou, X., Wang, M. & Zhang, Q. SAM-IE: SAM-based image enhancement for facilitating medical image diagnosis with segmentation foundation model. *Expert Syst. Appl.* **249**(PC), 123795. <https://doi.org/10.1016/j.eswa.2024.123795> (2024).
11. Wang, C. et al. SAMMed: A medical image annotation framework based on large vision model. (2023). <http://arxiv.org/abs/2307.05617>.
12. Wang, C. J., He, C. S., Yan, R. X. & Liu, Y. C. Application of MP-YOLO for segmentation and visualization of ovarian ultrasound imaging. in *2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, ECBIOS 2023* 130–132. (2023). <https://doi.org/10.1109/ECBIOS57802.2023.10218459>.
13. Alwakid, G., Gouda, W. & Humayun, M. Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement. *Healthc. (Switzerland)* **11**(6), 1–17. <https://doi.org/10.3390/healthcare11060863> (2023).
14. Nazarudin, A. A. et al. Performance analysis of a novel hybrid segmentation method for poly-cystic ovarian syndrome monitoring. *Diagnostics* **13**(4). <https://doi.org/10.3390/diagnostics13040750> (2023).
15. Marinov, Z., Jäger, P. F., Egger, J., Kleesiek, J. & Stiefelbogen, R. Deep interactive segmentation of medical images: A systematic review and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* (2024).
16. Xiao, H., Li, L., Liu, Q., Zhu, X. & Zhang, Q. Transformers in medical image segmentation: A review. *Biomed. Signal Process. Control* **84**, 104791 (2023).
17. Kirillov, A. et al. *Segment Anything* 4015–4026 (n.d).
18. Qian, J. et al. HASA: Hybrid architecture search with aggregation strategy for echinococcosis classification and ovary segmentation in ultrasound images. *Expert Syst. Appl.* **202**(October 2021), 117242. <https://doi.org/10.1016/j.eswa.2022.117242> (2022).
19. Sawant, A. & Kulkarni, S. Ultrasound image enhancement using super resolution. *Biomedical Eng. Adv.* **3**(May), 100039. <https://doi.org/10.1016/j.bea.2022.100039> (2022).
20. Gopalakrishnan, C. & Iyapparaja, M. Active contour with modified Otsu method for automatic detection of poly-cystic ovary syndrome from ultrasound image of ovary. *Multimedia Tools Appl.* **79**(23–24), 17169–17192. <https://doi.org/10.1007/s11042-019-07762-3> (2020).
21. Bedi, P., Goyal, S. B., Rajawat, A. S. & Kumar, M. An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting poly-cystic ovary syndrome. *Dec. Anal. J.* **10**(June 2023). <https://doi.org/10.1016/j.dajour.2023.100366> (2024).
22. Reka, S., Mohan, K. & Praba, T. S. Prompt diagnosis of poly-cystic ovary syndrome using ultrasonography—a machine learning approach. *AIP Conf. Proc.* **3031**(1). <https://doi.org/10.1063/5.0194703> (2024).
23. Alamoudi, A. et al. A deep learning fusion approach to diagnosis the Poly-cystic ovary syndrome (PCOS). *Appl. Comput. Intell. Soft Comput.* **2023** <https://doi.org/10.1155/2023/9686697> (2023).
24. Chitra, P. et al. Classification of ultrasound PCOS image using deep learning based hybrid models. in *Proc. 2023 2nd Int. Conf. Electron. Renew. Syst. ICEARS 2023* 1389–1394 (2023). <https://doi.org/10.1109/ICEARS56392.2023.10085400>.
25. Salman Hosain, A. K. M., Mehedi, M. H. K. & Kabir, I. E. Technologies, I. C. E. E. T. PCONet: A convolutional neural network architecture to detect poly-cystic ovary syndrome (PCOS) from ovarian ultrasound images. in *8th International Conference on Engineering and Emerging 2022*, October, 1–6. (2022). <https://doi.org/10.1109/ICEET56468.2022.10007353>.
26. Gopalakrishnan, C. & Iyapparaja, M. Multilevel thresholding based follicle detection and classification of poly-cystic ovary syndrome from the ultrasound images using machine learning. *Int. J. Syst. Assur. Eng. Manage.* <https://doi.org/10.1007/s13198-021-01203-x> (2021).
27. Wang, X. et al. and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. in *Proceedings of the European conference on computer vision (ECCV) workshops* (2018).
28. Ning, G., Liang, H., Jiang, Z., Zhang, H. & Liao, H. The potential of segment anything (SAM) for universal intelligent ultrasound image guidance. *Biosci. Trends* **17**(3), 230–233. <https://doi.org/10.5582/bst.2023.01119> (2023).
29. Moral, P. et al. CystNet: An AI driven model for PCOS detection using multilevel thresholding of ultrasound images. *Sci. Rep.* **14**, 25012. <https://doi.org/10.1038/s41598-024-75964-3> (2024).
30. Shanmugavadivel, K. & R, M. S. M. D. T. Optimised polycystic ovarian disease prognosis and classification using AI based computational approaches on multi-modality data. *BMC Med. Inf. Decis. Mak.* **24**, 281. <https://doi.org/10.1186/s12911-024-02688-9> (2024).
31. Reka, S., Suriya, P. T. & Mohan, K. Expeditious prognosis of PCOS with ultrasonography images—a convolutional neural network approach. *Commun. Comput. Inf. Sci.* **1929**, 367–376. [https://doi.org/10.1007/978-3-031-48774-3\\_26](https://doi.org/10.1007/978-3-031-48774-3_26) (2024).
32. Dataset. <https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images>.

## Author contributions

Dr.T.P.—Investigation, Writing—review & editing, R.S. Implementation and Writing—original draft, M.P., V.N. N.R. and R.A.—Implementation Support.

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article contains no studies with human participants performed by authors.

### Additional information

**Correspondence** and requests for materials should be addressed to T.S.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025