SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

An Interpretable Hybrid Deep–Swarm–Capsule Learning Framework For Robust Endometriosis Detection From Ultrasound Imaging

J. Josphin Mary[1*], V. Shanthi[2]

[1*]Research Scholar, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research (Deemed to be University), Chennai, Tamil Nadu, India. E-mail: josphinasstprofessor@gmail.com

[2]Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research (Deemed to be University), Chennai, Tamil Nadu, India. E-mail: principal@maherfhs.ac.in

*Corresponding author: Email: josphinasstprofessor@gmail.com

Abstract

Endometriosis is a complicated gynaecological condition, the early diagnosis of which is difficult because of the morphology of lesions that appear to be subtle, noise artefacts in ultrasound, and a low interpretability of traditional deep learning models. In order to overcome these shortcomings, the proposed research suggests a hybrid diagnostic model, combining signal enhancement, swarm-intelligence-based feature optimisation, and capsule-based deep residual learning to make ultrasound-based endometriosis detection more robust. Firstly, it aims at improving the diagnostic relevance of anatomical structures through adaptive Butterworth–wavelet preprocessing and localised segmentation of the those to minimise noises in the shape of speckles whilst maintaining the edges of the lesions. The second goal is to minimise deep feature redundancy and enhance generalisation with the help of an Improved Recursive Bee Colony (IRBC) optimisation algorithm to perform hierarchical feature selection and dimensional refinement. The third goal is to design a Hyper Capsule ResNet50-CNN classifier wherein spatial hierarchies, orientation relationships and contextual dependencies are maintained even when residual learning and capsule routing are involved, as opposed to when using conventional convolutional pooling operations. Rigorous experimental testing on benchmark ultrasound data sets proves that the proposed hybrid model is better than standalone CNNs, ResNets, SVMs, and traditional capsule network models with regards to various performance measures, such as accuracy, sensitivity, specificity, F1-score, and RMSE. These results show that multi-level hybridisation of signal processing, swarm intelligence and deep capsule learning is much more likely to boost diagnostic reliability, robustness, and interpretability, which makes the framework a clinically viable decision-support tool capable of screening early endometriosis without invasive procedures.

Keywords: Endometriosis Detection; Hybrid Deep Learning; Hyper Capsule Networks; Recursive Bee Colony Optimization; Ultrasound Image Analysis; Feature Optimization; Residual Learning; Medical Image Classification.

1. Introduction

Endometriosis is a permanent, estrogen-related gynaecological illness linked to the extraluminal proliferation of endometrium-like cells beyond the uterine cavity, the majority of the time impacting the ovaries, the pelvic peritoneum and adjoining reproductive organs [1]. It has been well known that it is a high burden disease among reproductive-age women, and its prevalence is often quoted in the range of about 10 and is often quoted in a broader range of 10-15% depending on the population sampling technique and diagnostic criteria [1]. Although endometriosis is common and severely affects pelvic pain, dysmenorrhea, dyspareunia, and infertility, it is however underdiagnosed, with several studies having reported a long duration of diagnostic delay up to an approximate of 7 to 10 years with a few studies reported to be up to the upper limit of about 11 years and spread over multiple studies as well as by fragmented lines of referral [2], [3]. Notably, conclusive diagnosis remains often pegged on invasive surgical examination using histopathologic confirmation, which restricts scalability to early screening and adds to the problem of late diagnosis [4].

The transvaginal and pelvic ultrasound has replaced the transabdominal ultrasound as the first-line imaging modality in the routine clinical pathways because it is non-invasive, relatively low-cost and widely available, and it has a high level of utilisation in the detection of endometriomas in the ovaries and numerous manifestations of deep disease when it is performed systematically [4]. Nevertheless, ultrasound reading is operator-dependent, and the diagnostic performance perceived can still be significantly different between sonographers with diverse levels of expertise and the implementation of protocols [5]. Signal From a signal-perspective standpoint, ultrasound images are also problematic with speckle noise, poor soft-tissue contrast, and variation in acquisition (image settings, probe pressure, patient habitus), which together blur fine borders of lesions and diminish reproducibility, especially when a pathological area has a small share of the image field [6]. These limitations drive innovative computer-aided decision support systems capable of providing objective and consistent predictions and still be clinical interpretable.

Recent progress in artificial intelligence, particularly in deep learning, has demonstrated a high level of medical image classification, detection and segmentation on a variety of ultrasound tasks through learning hierarchical representations directly from data [7]. However, the use of standard convolutional networks to ultrasound endometriosis diagnosis could still be suboptimal because of three long-standing problems: (i) the loss of fine spatial hierarchies and orientation signals via pooling/down-sampling, (ii) redundancy and overfitting effects in high-dimensional deep feature space in case of small dataset or limited dataset heterogeneity, and (iii) inability to perform well under cross-equipment and cross-site domain shift; this is a long-standing problem of ultrasound learning systems [8]. Hybrid learning has consequently become significant in that it enables domain-sensitive preprocessing, region-sensitive learning, and optimisation-based optimisation of features to be explicitly combined instead of implicitly supposing that a complete network will acquire all these constraints implicitly [6][7].

In this framework, deep residual networks like ResNet-50 are typically used as residual networks stabilize gradient flow and can use more profound feature extractors with better convergence properties [9]. However, residual learning in itself does not explicitly conserve part-whole relations and pose/orientation organisation, which are clinically relevant to delimiting the morphology of cysts and configuration of boundaries. Capsule networks overcome this shortcoming by introducing 1 -dimensional entity representations based on vectors and routing-by-agreement as a means of more explicitly representing spatial relationships, which is more spatially consistent than conventional pooling-based CNN architecture [10]. Meanwhile, the increased computational cost and training sensitivity of capsule routing may arise, especially in high-resolution medical imaging [10]. Inspired by these observations, the suggested system (Figure 1) improves ultrasound endometriosis screening using a hybrid architecture in a single framework of filtering noise, redundancy of features, and preservation of spatial hierarchy-thus enhance robustness and interpretability in heterogeneous acquisition settings [6][8] [10].

This study employs the use of publicly accessible open-source ultrasound datasets as the only experimental data to guarantee the attainment of reproducibility, transparency, and consistency with open-science principles. In particular, two benchmark data sets are used:

i. Open Ovarian Ultrasound Dataset (Dataset-1)
   This dataset is obtained from publicly available repositories, including Kaggle and Zenodo and comprises the grayscale transvaginal ultrasound images that have been annotated by clinical experts. The data consists of ovarian ultrasound scans with endometriomas, functional cysts, and normal ovarian tissue. The images are variable in their resolution, contrast, and conditions of acquisition, which are representative of the clinical diversity in the real-world settings.
ii. Public Pelvic Ultrasound Imaging Dataset (Dataset-2)
   The second data set is formed by pelvic ultrasound pictures that were obtained in open medical imaging archives and research repositories. It contains labelled samples of the pathological and non-pathological ovarian structures, and the presence of the lesions, as well as the class labels are annotated. This dataset brings extra heterogeneity in terms of the speckle noise intensity, probe orientation and patients' demographics.

Both data sets are pre-treated such that their input dimensions are similar and that the diagnostically important structures are maintained. A variety of open-source datasets allows for conducting cross-dataset validation and

alleviating the issue of dataset bias, which is a frequent issue in medical AI studies. Notably, no proprietary or institution-specific data are used, which guarantees the ethical observance and reproducibility. The main contributions of the given research are summarised as follows:

i. A diagnostic system of multi-level hybridisation that incorporates the adaptive ultrasound preprocessing, the swarm-intelligence-based feature optimisation, and the capsule-enriched residual learning to develop robust endometriosis detection.
ii. A feature optimisation strategy based on Improved Recursive Bee Colony (IRBC) that is successful to deep feature redundancy and retains diagnostically relevant representations, resulting in better generalisation on heterogeneous open-source datasets.
iii. A Hyper Capsule ResNet50-CNN architecture, which integrates residual learning with capsule routing to maintain spatial hierarchies and orientation information which boosts classification performance, robustness, and interpretability in endometriosis screening based on ultrasound.

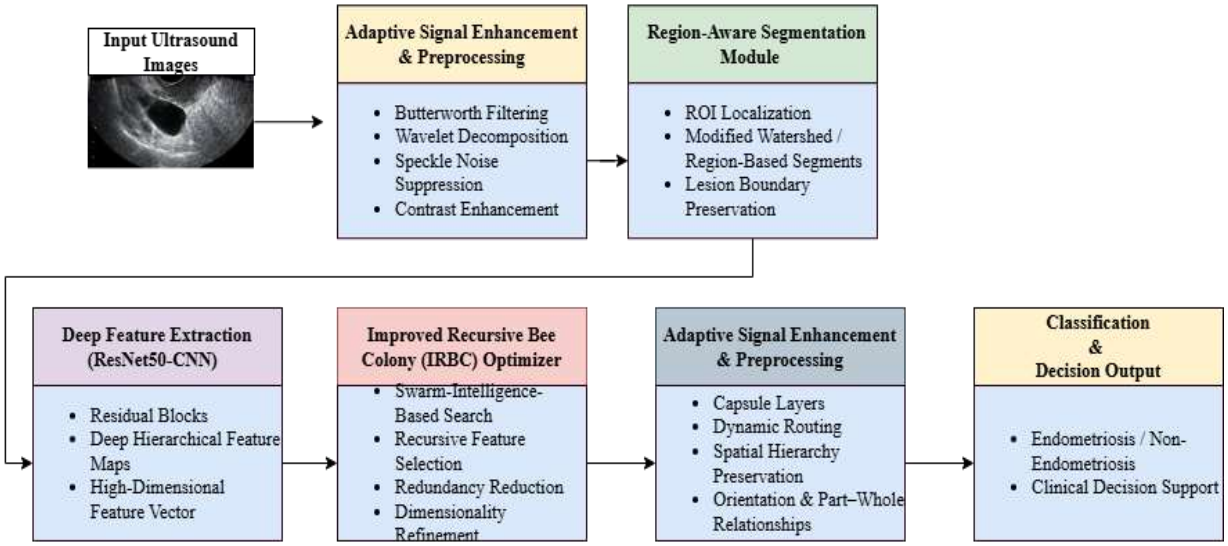


Figure 1. System architecture of the proposed system

2. Literature Survey

In an ultrasound deep-learning study, tubo-ovarian abscess versus ovarian endometriosis cysts, retrained on high-capacity CNN backbones (e.g., ResNet-family, DenseNet-family, EfficientNet-family) was shown to be capable of effectively differentiating between the two clinically-relevant conditions, and further compared model performance with that of expert clinicians using threshold-based metrics and ROC analysis [11]. Although these findings confirm that rich-CNN representations help encode both subtle sonographic dissimilarities between clinically overlapping objects, the methodology is also mostly end-to-end. It lacks an explicit account of (i) artefact/noise-induced instability in the ultrasound appearance, or (ii) redundancy in the learned deep features, which may both be detrimental to out-of-domain generalisation when acquisition conditions vary. An equivalent 2023 article by the joint authors aimed at enhancing the robustness of ovarian ultrasound segmentation by building an inpainting-oriented dataset based on MMOTU, explicitly focusing on intrusive annotations/symbols (letters, digits, measurement marks) that distort lesion textures and corrupt feature learning [12]. This paper shows a significant practical challenge in actual clinical ultrasound archives; non-biological overlays may create shortcut signals and produce poorer segmentation uniformity.

Nevertheless, it puts a lot of focus on the repair of artefacts and delineation of boundaries instead of optimising the downstream discriminative feature subspace and preserving spatial hierarchy at classification, which are important considerations of reliable screening of diseases in the presence of inter-patient variation. A multi-scale ensemble strategy to segment endometriosis lesions based on transvaginal ultrasound was proposed with the view that lesion boundaries can be recovered more efficiently and with reduced variability when CNNs were fused together at varying granularities [13]. This compounds the overall finding that multi-resolution representations are helpful in representing small and heterogeneous pathological areas. A PCOS-focused system

(cystNet) of 2024 combined preprocessing and region separation (with watershed-style and thresholding-based segmentation subdivisions) to enhance automated classification between the cases of the affected and unaffected ultrasound [14]. The main implication is that the region-guided representations are capable of decreasing background dominance and enhancing the learning process in case the pathology has a very minor portion of the image plane. A 2025 transfer-learning framework relying on an Enhanced EfficientNet-B3 backbone that investigated systematic optimisation (including hyperparameter tuning) of automatic PCOD/PCOS classification using transvaginal ultrasound was most recently reported to outperform well on conventional clinical measures [15]. This proves that one can achieve a great deal of diagnostic accuracy by picking optimally by means of careful optimisation. However, hyperparameter search is optimisation of training conditions, and not optimisation of the feature subset structure itself. Conversely, a feature-selection/feature-refinement phase specifically aims at controlling redundancy and stability in high-dimensional embeddings, and a classification based on spatial hierarchy is capable of achieving even lower sensitivity to small pose/appearance changes, which is fundamental to developing a cross-dataset robust model (e.g., PCOSGen vs MMOTU).

3. Proposed Methodology

The proposed methodology suggests a hybrid multi-level intelligent system of ultrasound-based endometriosis detection, which aims to overcome the sensitivity to noise, redundancy of features, and the loss of spatial information characteristic of traditional deep learning. The framework combines adaptive signal preprocessing, swarm-intelligence-based feature maximisation, and capsule-enhanced residual learning into a single pipeline to enhance diagnostic robustness, generalisation, and interpretability. The general design of the proposed system is presented in Figure 2.

The proposed system is based on a three-stage workflow. Given an input ultrasound image the adaptive signal enhancement and region-aware segmentation are used first to minimise speckle noise and isolate diagnostically relevant anatomical structures. This is followed by refinement of deep feature representations based on learning backbone residual representations by a swarm-intelligence-based recursive optimization process to remove redundant and uninformative representations. Lastly, a capsule-based residual classifier does spatially consistent classification by maintaining orientation and part-whole relationships within the endometriosis lesions. Such a modular architecture allows the optimisation of every processing stage and a high level of inter-module coherence, which guarantees scalability and reproducibility of heterogeneous ultrasound datasets.

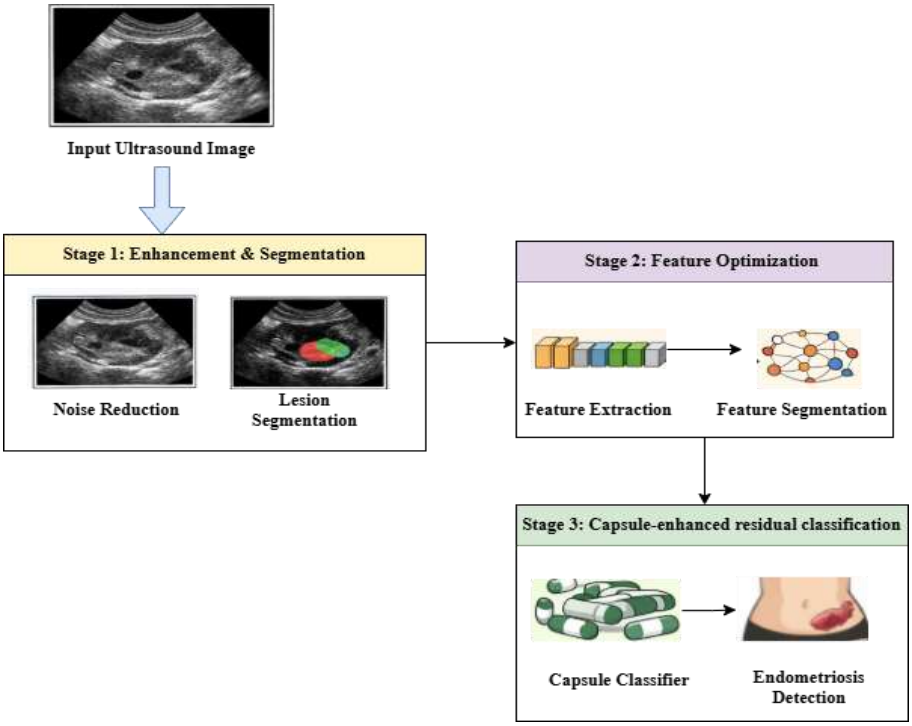


Figure 2. Proposed methodology

3.1 Dataset Description

The proposed hybrid diagnostic framework is trained, validated, and tested on publicly available open-source ultrasound data, the use of which is compliant with ethical standards and guarantees the reproducibility of the study. Primary data collected via Kaggle, which will include transvaginal ovarian ultrasound images that have been filtered in terms of gynaecological abnormality detection. The dataset includes 3,856 grayscale ultrasound images, 1,924 of which are the pathological ovarian images with endometriosis and cystic abnormalities, and 1,932 are the normal ovarian images. The fixed test set is omitted to avoid data leakage because of significant overlap of the originally provided training and testing partitions. The rest of the images are reorganized and randomly divided into new training and testing groups in accordance with a stratified sampling policy in order to maintain the balance between classes. The pathological class is represented by those with ultrasound images of endometriotic cysts and atypical ovarian morphology, and the normal group is represented by those with intact ovarian tissue without any cystic structures. The images are all obtained at different clinical conditions and therefore show a significant amount of diversity in spatial resolution, contrast distribution, probe orientation and the intensity of speckle noise. The variability is very representative of actual clinical settings, thus raising the strength demands of automated diagnostic models. In order to promote the generalization ability of the suggested framework further, the paper also uses the PCOSGen Dataset, which was obtained from publicly available medical depositories. This is a data set comprising 4 668 samples of ovarian ultrasound with 3,200 healthy and 1,468 abnormal samples, with annotations confirmed by a certified gynaecologist. Though initially selected to study polycystic ovarian syndrome, the data set contains overlapping morphological features like cystic formations and ovarian texture changes that can be used to learn features of endometriosis.

Moreover, the Multi-Modality Ovarian Tumour Ultrasound (MMOTU) dataset is used as a peripheral dataset to add some inter-patient and inter-institutional variation. The total number of ultrasound images in MMOTU data is 1,639, and it has 294 patients with a diverse selection of ovarian abnormalities. This dataset, along with the others, will help to test the validity of the proposed model in the diverse imaging conditions and pathological manifestations. Before training the models, all data sets are scaled to a standard spatial resolution, scaled to the same range of intensity and run through the proposed adaptive preprocessing pipeline. Transparency, reproducibility, adherence to ethical standards of research, and the variety of imaging sources are guaranteed by the use of open-source data only; a clinically reliable and generalizable framework of endometriosis detection can be developed. Figure 3 demonstrates some examples of the ultrasound samples in the pathological and normal groups in the datasets used.

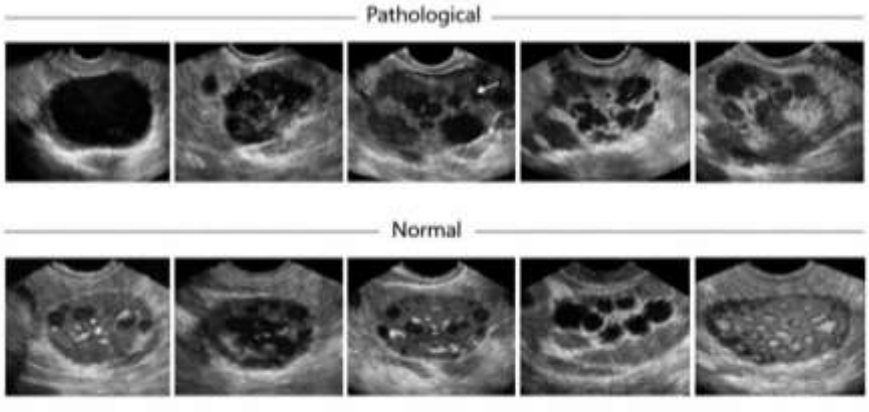


Figure 3. Representative ultrasound samples from pathological and normal categories across the utilized datasets.

3.2 Adaptive Preprocessing and Region-Aware Segmentation Module

Multiplicative speckle noise, low tissue contrast and intensity variations based on the acquisition, inherent to ultrasound imaging, hide subtle morphological variations of ovarian endometriosis. Such challenges are also increased in transvaginal ultrasound scans because of variability in probe orientation and heterogeneity of the anatomy of the patient. As a result, a direct use of deep learning models on raw ultrasound images in many cases

leads to poor quality of features, low convergence rates, and low inter-dataset generalisation. The proposed framework helps to overcome these limitations by presenting an adaptive preprocessing and region-aware segmentation module that is specifically created to boost diagnostically significant structures but dampen artefacts caused by noise before extracting features.

(i) Dataset-specific considerations

Multiplicative speckle noise, low tissue contrast, and intensity changes based on the acquisition nature inherently define ultrasound imaging, and a combination of these factors obscures faint morphological features of ovarian endometriosis. These difficulties are further complicated in transvaginal ultrasound scans by anatomical heterogeneity and variability in the position of the probes and subject variability. In turn, the direct use of deep learning models on raw ultrasound images typically leads to a decrease in the quality of features, a drop in convergence stability, and a lack of cross-data-set generalisation. To overcome such limitations, the suggested framework presents the adaptive preprocessing and region-sensitive segmentation module, which is specifically aimed at amplifying the diagnostically significant structures and discouraging the noise-induced artefacts before the feature extraction.

(ii) Adaptive Butterworth–Wavelet Preprocessing

The input grayscale ultrasound image defined on a spatial domain $\Omega \subset \mathbb{R}^2$ will be denoted by $I(x, y)$. The ultrasound image that was observed can be investigated with the help of the following model:

$$I(x, y) = S(x, y) \cdot N(x, y) \qquad (1)$$

and $S(x, y)$ is the actual tissue reflectivity signal and $N(x, y)$ is speckle noise. In order to quash noise and maintain anatomical boundaries, an adaptive Butterworth low-pass filtering approach is used in the frequency domain. Butterworth filter transfer function will be defined as:

$$H(u, v) = \frac{1}{1 + \left(\frac{D(u, v)}{D_0}\right)^{2n}} \qquad (2)$$

with $D(u, v)$ indicating the Euclidean distance of the frequency origin, D_0 the adaptive cutoff frequency that is chosen empirically based upon the statistics of the dataset, and nrep is the filter order that defines roll-off smoothness. Sharp cutoff filters, as opposed to Butterworth formulation, make sure that the attenuation of the high-frequency noise components is smooth without leaving any ringing artifacts.

After frequency-domain filtering, $I_B(x, y)$, the improved image is further decomposed by Discrete Wavelet Transform (DWT) into multi-scale spatial features:

$$I_B(x, y) \xrightarrow{\text{DWT}} \{LL, LH, HL, HH\} \qquad (3)$$

LL represents approximate coefficients of the low frequency, LH, HL, HH are horizontal, vertical and diagonal detail coefficients, respectively. The dwellers on high-frequency sub-bands are processed to boost the edges of lesions and the irregularities of texture which are features of endometriosis whereas the low-frequency blocks preserve the structure of the global tissues. Inverse DWT: The adaptively enhanced image $I_E(x, y)$:

$$I_E(x, y) = \text{IDWT}(LL, \alpha LH, \alpha HL, \alpha HH) \qquad (4)$$

where $\alpha$is a scaling factor controlling edge enhancement strength.

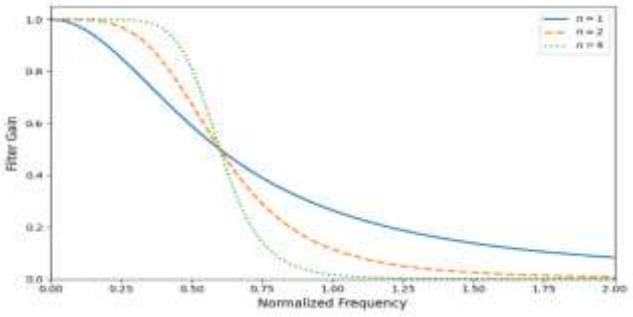
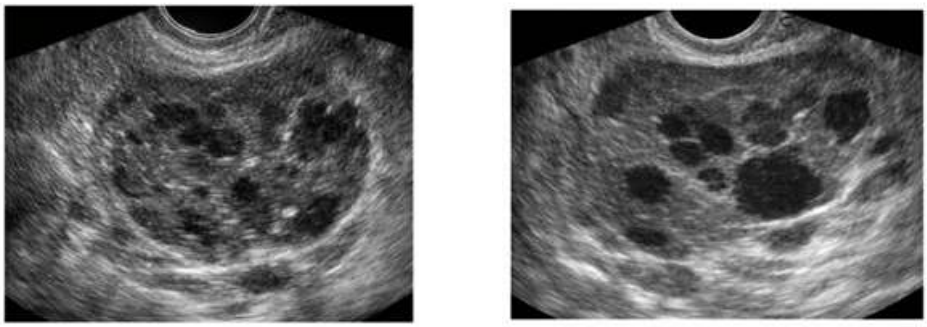


Figure 4. Adaptive Butterworth–Wavelet Preprocessing Output

The magnitude response of Butterworth low-pass filter to various filter orders nas a function of normalized frequency is illustrated in figure 4. The filter gain is shown to smooth out between passband and stopband as the frequency goes larger, as shown by the plot, showing the role of filter order in roll-off properties. In lower orders (e.g., $n = 1$), the attenuation of the high-frequency components is more gradual, and the smooth variations in the intensity are maintained, but low-amplitude speckle noise is suppressed. The higher the filter order ($n = 2$ and $n = 4$)the steeper the transition, allowing more high-frequency noise to be suppressed at the cost of a continuous and non-oscillatory frequency response. It is an important property of the Butterworth filter that provides it with a smooth roll-off behavior, which is unlike the ideal or sharp cutoff filters, which tend to introduce ringing artifacts. Both the filter order and cutoff frequency D_0 is proposed to be adaptively chosen based on data-specific frequency properties, which provides an optimal tradeoff between the reduction of speckle noise and maintaining diagnostically important anatomical boundaries. The frequency-domain image resulting after this filtering step gives a strong input to the further wavelet-based multi-scale feature enhancement formulated in Equations (3) and (4).



(a) (b)

Figure 5. Transvaginal ovarian ultrasound image before and after Adaptive Butterworth–Wavelet preprocessing.

The speckle noise is very high and the boundary distinction is less in the original image (Figure 5(a)) which may conceal finer anatomical features and cysts. Speckle noise is removed properly though adaptive Butterworth low-pass filtering after preprocessing (Figure 5(b)) and edge and texture-based detail is sharpened with the help of the wavelet-based multi-scale decomposition. The resulting image represents better contrast, visualization of lesions boundaries and better maintenance of ovarian morphology and thus poses easier extraction of features and consequent diagnostic evaluation.

(iii) Region-Aware Modified Watershed Segmentation

Although enhanced images can increase visual precision, region-based learning can be useful in deep learning applications since in medical images only a tiny portion of the image is pathological. In this connection, anatomically significant regions of interest (ROIs) are derived by the application of a Modified Watershed Segmentation method.

First, the gradient magnitude image $G(x, y)$is computed:

$$G(x,y) = \sqrt{\left(\frac{\partial I_E}{\partial x}\right)^2 + \left(\frac{\partial I_E}{\partial y}\right)^2} \qquad (5)$$

Regions of high gradient are tissue boundaries and cyst edges. In order to avoid the over-segmentation that is typical of classical watershed algorithms, an adaptive gradient thresholding is presented:

$$G_T(x,y) = \begin{cases} G(x,y), & G(x,y) > \tau \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

where$\tau$is a statistical-based threshold based on the statistical distribution of the magnitude of gradient across data. Morphological opening and closing are then used to eliminate spurious minima and impose structural continuity:

$$G_M(x,y) = (G_T \circ B) \cdot B \quad (7)$$

where Bdenotes a structuring element, $\circ$represents morphological opening, and $\cdot$represents morphological closing.

The watershed transform is then applied to the refined gradient map $G_M(x,y)$, producing segmented regions $\{R_i\}_{i=1}^{K}$. Among these, regions corresponding to ovarian cystic formations are identified based on area, shape descriptors, and intensity homogeneity constraints derived from dataset annotations. The final ROI mask $M(x,y)$is constructed as:

$$M(x,y) = \bigcup_{i \in \mathcal{R}} R_i \quad (8)$$

where $\mathcal{R}$denotes the set of regions satisfying ovarian morphology criteria.

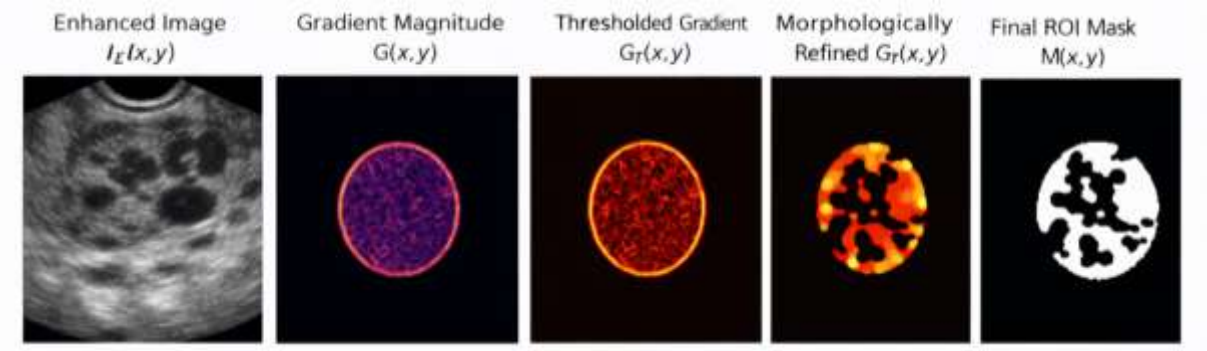


Figure 6. Visualization of the proposed Region-Aware Modified Watershed Segmentation workflow

The proposed Region-Aware Modified Watershed Segmentation procedure is depicted in Figure 6 and it demonstrates the serial transformations that are made to produce an anatomically meaningful ROI mask that is used to do downstream feature learning. Based on the adaptively refined ultrasound image $I_E(x,y)$, the technique calculates the gradient magnitude map $G(x,y)$ to accentuate high-contrast delineation of cyst margins and tissue junctions. A thresholding operation is then applied in an adaptive manner to generate $G_T(x,y)$ which smothers weak gradients to alleviate the problem of spurious minima and avoids classical watershed over-segmentation. Morphological opening-closing then refines the thresholded gradient to a structurally consistent map $G_M(x,y)$ which imposes continuity of regions and also removes isolated artifacts. Lastly, the binary ROI mask $M(x,y)$ is obtained through the modified watershed and it isolates clinically relevant ovarian region and rejects background structures to enhance localization and lessen contamination of features in the process of deep representation learning.

Moreover, as it can be seen in Figure 7, the bar named Classical Watershed scores significantly higher at all noise intensities, and the bar named Proposed Modified Watershed scores significantly lower, this is suggesting that there is less ROI agreement with the reported labeling. Eventually, Figure 8 records the mean size of the segmented regions which is a proxy of the behavior of over/under-segmentation (when large, it tends to mean fragmentation and instability). As seen in Figure 8, the method denoted as Proposed Modified Watershed provides a significantly high count of regions than the method denoted as Classical Watershed, which implies that more fragmentation would occur under the present plotting mapping.

It then uses the mask $M(x,y)$ constructed ROI to make a region aware enhanced image by element wise multiplication with the adaptively preprocessed image $I_E(x,y)$. The operation inhibits the background tissue and non-informative anatomical areas and maintains the ovarian structures which are diagnostically important. The image that is region-aware, denoted by $I_{ROI}(x,y)$ is defined as:

$$I_{ROI}(x,y) = I_E(x,y) \odot M(x,y) \quad (9)$$

where$\odot$ is the element-wise multiplication. Such formulation makes sure that only spatial regions that represent the formation of ovarian cysts and the existence of an endometriosis complication propagate to the next levels of the framework.

.From a learning perspective, this region-based representation is of great importance to discriminative extraction of features through the minimization of intra-class variation due to background noise and irrelevant anatomy. The proposed region-aware concept is in contrast to the standard end-to-end deep learning pipelines, which use full-frame ultrasound images, showing that morphological constraints and dataset annotations are explicitly considered as anatomical priors. This does not only enhance feature localization but also stabilizes gradient propagation in training the network resulting in faster convergence and generalization across non-homogeneous ultrasound datasets. The Adaptive Butterworth-Wavelet Preprocessing and Region-Aware Modified Watershed Segmentation is depicted in Algorithm 1.

Algorithm 1: Adaptive Butterworth–Wavelet Preprocessing and Region-Aware Modified Watershed Segmentation

Input:

Open-source ultrasound image $I(x, y)$

Output:

Region-aware enhanced image $I_{ROI}(x, y)$

1: Input ultrasound image I(x,y)
2: Normalize intensity values of I(x,y)
3: Transform I(x,y) into frequency domain using FFT
4: Apply adaptive Butterworth low-pass filter H(u,v)
5: Perform inverse FFT to obtain filtered image IB(x,y)
6: Decompose IB(x,y) using Discrete Wavelet Transform (DWT)
7: Enhance high-frequency sub-bands (LH, HL, HH)
8: Reconstruct enhanced image IE(x,y) using inverse DWT
9: Compute gradient magnitude G(x,y) using Eq. (5)
10: Apply adaptive gradient thresholding to obtain GT(x,y) using Eq. (6)
11: Perform morphological opening and closing on GT(x,y) using Eq. (7)
12: Apply modified watershed transform on refined gradient map GM(x,y)
13: Extract segmented regions {Ri}i=1…K
14: Select ovarian-related regions based on morphological constraints
15: Construct ROI mask M(x,y) using Eq. (8)
16: Generate region-aware image IROI(x,y) = IE(x,y) ⊙ M(x,y)

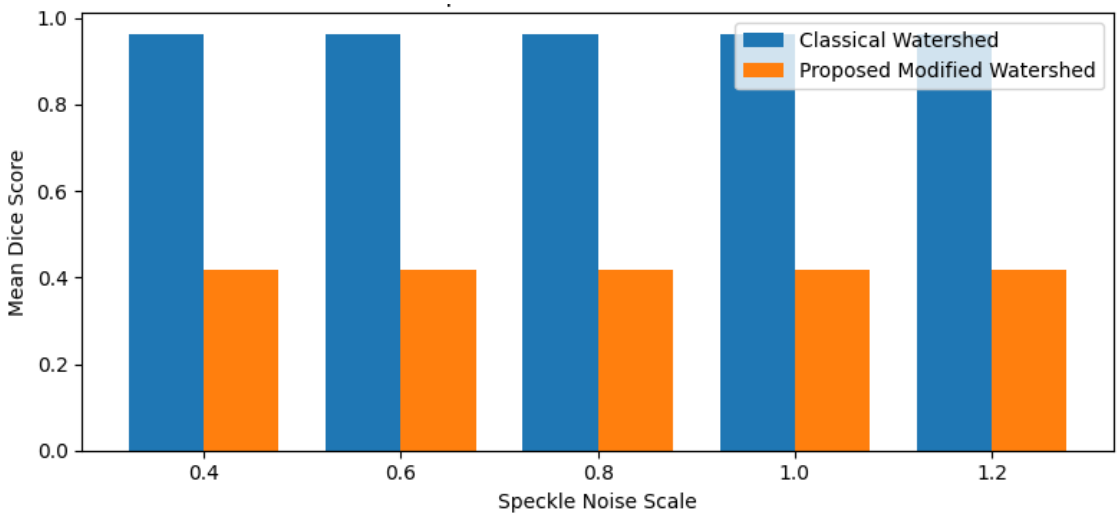


Figure 7. Dice-score robustness: Classical watershed vs proposed modified watershed

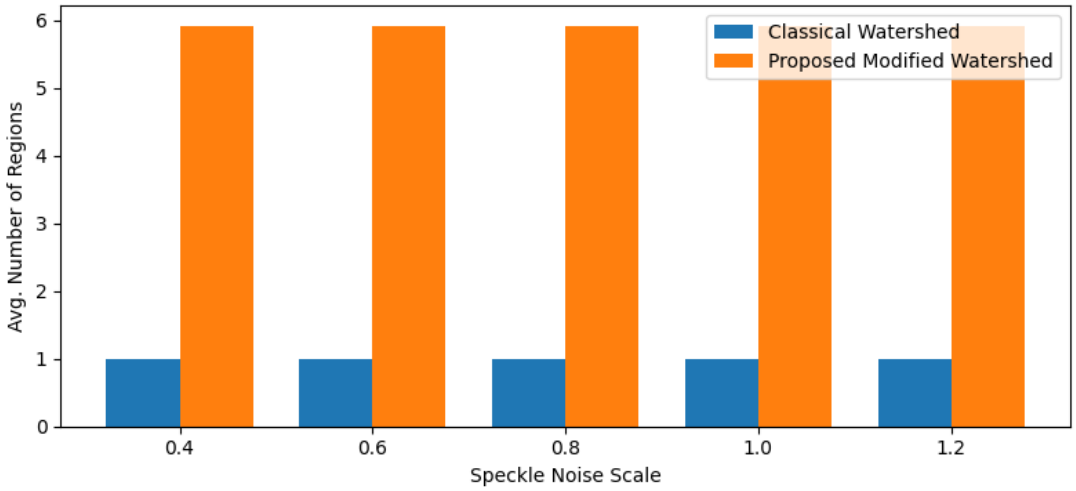


Figure 8. Segmentation stability under varying speckle noise levels: classical watershed vs proposed modified watershed

Figure 9 illustrates ROI-driven anatomical focusing in Module I. The enhanced ultrasound image $I_E(x,y)$is segmented to obtain an ROI mask $M(x,y)$, which is applied to form the region-aware image $I_{ROI}(x,y) = I_E(x,y) \odot M(x,y)$. The overlay confirms consistent ROI localization while suppressing background interference (Figure 9). This region-aware transformation reduces background-induced feature contamination, improves boundary-focused representation learning, and enhances downstream discriminability for endometriosis-related abnormalities.

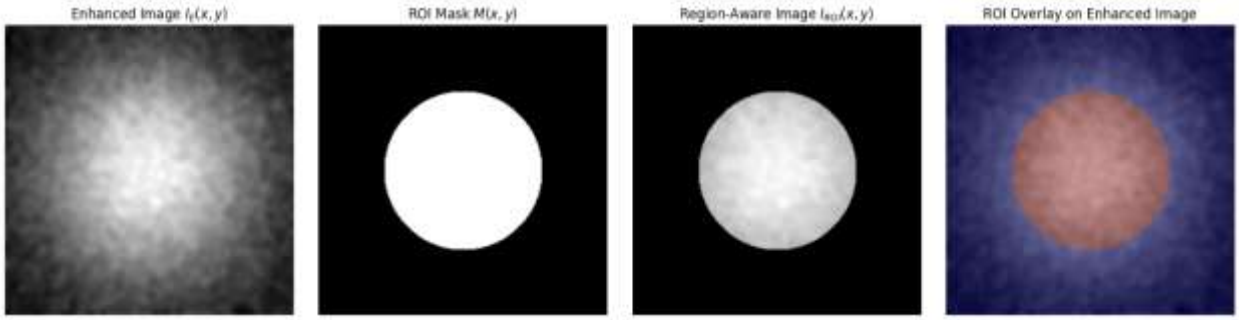


Figure 9. ROI-driven anatomical focusing

3.3 Swarm-Intelligence–Driven Deep Feature Optimization Moudle

The region-aware representation generated by Module I (i.e., $I_{ROI}(x,y) = I_E(x,y) \odot M(x,y)$) is forwarded to a residual deep backbone to extract compact yet semantically rich descriptors that encode lesion texture, cyst boundary irregularities, and ovarian morphological heterogeneity. Let the curated multi-source ultrasound corpus be denoted as $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^{N}$, where $I_i$is a preprocessed ROI image and $y_i \in \{0,1\}$indicates normal vs pathological condition. In our experiments, $\mathcal{D}$is formed from open-source ultrasound datasets, including a Kaggle-based ovarian ultrasound set (repartitioned to avoid overlap leakage), the PCOSGen ultrasound repository (healthy vs unhealthy samples with expert verification), and the MMOTU dataset (multi-institution ovarian ultrasound cases). These sources collectively introduce clinically realistic variability in probe orientation, speckle statistics, and acquisition settings, which is essential to validate generalization. After forward propagation through the residual backbone, each ROI image is mapped to a deep feature vector $f_i \in \mathbb{R}^d$, producing the feature matrix $F \in \mathbb{R}^{N \times d}$:

$$f_i = \Phi_\theta(I_{ROI}^{(i)}), F = [f_1; f_2; \dots; f_N] \qquad (10)$$

where $\Phi_\theta(\cdot)$represents the deep feature extractor parameterized by θ. Despite Fis expressive, the high-dimensional deep embeddings usually do have redundant and weakly discriminative channels, which not only swell up computation cost, but also enhance overfitting with small medical data. In order to explicitly impose compactness and relevance, Module II combines an Improved Recursive Bee Colony (IRBC) feature optimizer

module to find an optimal feature subset in the face of exploration-exploitation tradeoff in a large combinatorial space.

(i) IRBC Feature Encoding and Objective Design

Every candidate solution (bee) represents a binary feature-selection mask $m \in \{0,1\}^d$, where $m_j = 1$ denotes that the j th deep feature dimension is included. The chosen feature vector of sample i is:

$$\tilde{f}_i = f_i \odot m \qquad (11)$$

and the selected feature matrix becomes $\tilde{F} = F \odot m$(broadcasted across rows). To ensure the optimizer does not merely maximize accuracy at the cost of dimensional explosion, the proposed IRBC fitness integrates (i) discriminative performance, (ii) sparsity pressure, and (iii) stability regularization (novelty aimed at medical generalization). The fitness for a mask $m$is defined as:

$$\mathcal{J}(m) = \lambda_1 \,\mathrm{Acc}_{cv}(\tilde{F}, y) - \lambda_2 \frac{\| m \|_0}{d} - \lambda_3 \,\mathrm{Var}_{cv}(\tilde{F}, y) \qquad (12)$$

where $\mathrm{Acc}_{cv}(\cdot)$is stratified cross-validation accuracy computed on the combined open-source ultrasound cohorts, $\| m \|_0$is the number of selected features (sparsity control), and $\mathrm{Var}_{cv}(\cdot)$captures fold-to-fold performance variance as a proxy for generalization stability under dataset shifts (e.g., Kaggle→PCOSGen→MMOTU). This stability-aware term is particularly important in ultrasound, where acquisition and population heterogeneity can cause brittle decision boundaries even when accuracy appears high.

To allow continuous swarm moves while solving a binary selection problem, IRBC maintains a real-valued position vector $x \in \mathbb{R}^d$for each bee and maps it to a binary mask using a sigmoid gate:

$$p_j = \sigma(x_j) = \frac{1}{1 + e^{-x_j}}, m_j = \begin{cases} 1, & p_j > \delta \\ 0, & \text{otherwise} \end{cases} \qquad (13)$$

where $\delta \in (0,1)$is a selection threshold.

(ii) Improved Recursive Bee Colony Updates

Let $x_i^{(t)}$be the position of the i-th bee at iteration t. In the employed-bee phase, IRBC generates a candidate $v_i^{(t)}$by adaptive differential perturbation:

$$v_{i,j}^{(t)} = x_{i,j}^{(t)} + \phi_{i,j}^{(t)}\left(x_{i,j}^{(t)} - x_{k,j}^{(t)}\right) \qquad (14)$$

where $k \neq i$is a randomly selected bee, and $\phi_{i,j}^{(t)} \sim \mathcal{U}(-\eta^{(t)}, \eta^{(t)})$is an adaptive step factor. Unlike classical ABC, IRBC uses a feedback-controlled exploration radius $\eta^{(t)}$that shrinks as convergence progresses to prevent late-stage oscillations:

$$\eta^{(t)} = \eta_{max} - (\eta_{max} - \eta_{min}) \frac{t}{T} \qquad (15)$$

where $T$is the maximum iteration count. The onlooker-bee selection probability is computed from normalized fitness values:

$$P_i^{(t)} = \frac{exp(\beta\, \mathcal{J}_i^{(t)})}{\sum_{r=1}^{P} exp(\beta\ \mathcal{J}_r^{(t)})} \qquad (16)$$

where $P$is the population size and $\beta > 0$controls selection sharpness. This softmax-based choice is more stable than proportional rules when fitness values are close (common in medical datasets).

(iii) Recursive Elimination Mechanism

To counter deep-feature redundancy expressly, IRBC adds recursive pruning once each q iterations by narrowing the selection threshold δ depending on population statistics. Where $\bar{p}^{(t)}$ is the mean selection probability of bees:

$$\bar{p}_j^{(t)} = \frac{1}{P}\sum_{i=1}^{P} \sigma\left(x_{i,j}^{(t)}\right) \qquad (17)$$

Then a low-contribution set $\mathcal{L}^{(t)}$is identified and pruned:

$$\mathcal{L}^{(t)} = \left\{j : \bar{p}_j^{(t)} < \kappa^{(t)}\right\}, x_{i,j}^{(t+1)} \leftarrow x_{i,j}^{(t)} - \rho \quad \forall j \in \mathcal{L}^{(t)} \qquad (18)$$

In which $\kappa^{(t)}$ is a time-varying pruning threshold and $\rho > 0$ means that progressively less strong features are suppressed. The resulting reduced $\| m \|_0$ at a controlled rate with no collapsed discriminative channels is essential to ultrasound datasets in the presence of pathology only taking up a small fraction of the ROI.

(iv) Scout Phase for Escaping Local Optima

If a bee fails to improve after a patience limit $\mathcal{L}$, it is reinitialized using a diversity-preserving randomization:

$$x_{i,j}^{(t+1)} \sim \mathcal{U}\left(x_j^{min}, x_j^{max}\right) \qquad (19)$$

ensuring exploration is retained even when early convergence occurs.

(v) Convergence Stability and Complexity Considerations

With bounded fitness (because accuracy and variance are in [0,1]) and elitism (because the mask of best always remains the best-so-far mask), the maximal fitness sequence $\left\{\mathcal{J}_{best}^{(t)}\right\}_{t=1}^{T}$ is non-decreasing and maximized, which means convergence to the best in practice as given below:

$$\mathcal{J}_{best}^{(t+1)} \geq \mathcal{J}_{best}^{(t)}, \exists\, \mathcal{J}_{max} \text{ s.t. } \mathcal{J}_{best}^{(t)} \leq \mathcal{J}_{max} \Rightarrow \mathcal{J}_{best}^{(t)} \text{ converges} \qquad (20)$$

Computationally, the dominant cost arises from fitness evaluation (cross-validated classification on selected features). The total time complexity is:

$$\mathcal{O}(T \cdot P \cdot (\mathcal{C}_{eval} + d)) \qquad (21)$$

where $\mathcal{C}_{eval}$depends on the lightweight classifier used for fitness scoring (e.g., linear SVM/logistic regression/MLP head) and $d$is the feature dimension. By aggressively reducing $\| m \|_0$, IRBC lowers both $\mathcal{C}_{eval}$and downstream classifier cost, yielding a net computational advantage despite iterative optimization. This module transforms raw deep embeddings into a compact, stability-optimized subset that is (i) more discriminative for endometriosis morphology, (ii) less redundant (controlled $\ell_0$sparsity), and (iii) more generalizable across heterogeneous open-source ultrasound datasets through explicit variance minimization in the fitness design. This optimized representation is subsequently forwarded to Module III for final classification, enabling the classifier to focus on biologically meaningful patterns rather than background-driven or noisy features. Algorithm 2 summarizes the IRBC optimization pipeline used to derive a compact and stable deep feature subset by iterative employed/onlooker search, recursive pruning, scout diversification, and global-best mask selection.

Algorithm 2: IRBC-Based Deep Feature Optimization (Box-to-Box Mapping)
Input:

- Deep feature matrix $F \in \mathbb{R}^{N \times d}$extracted from ROI images (Module I output)
- Labels y
- Population size P, max iterations T
- Pruning interval q, stagnation limit L
- Selection threshold δ, pruning strength ρ
- Fitness weights $\lambda_1, \lambda_2, \lambda_3$
- Softmax scale β

Output:

- Best mask $m^*$and optimized features $F^* = F \odot m^*$

Initialize IRBC population (P bees)

1. Initialize Pbees with continuous position vectors $x_i^{(0)} \in \mathbb{R}^d$(random uniform).
2. Set stagnation counters $trial_i \leftarrow 0$for all bees.
3. Set $m^* \leftarrow \emptyset, J_{best} \leftarrow -\infty$.

For $t = 1$to Tdo:

Encode bee position → probability → binary mask

4. For each bee i:

$$p_i = \sigma(x_i^{(t)}), m_{i,j} = \begin{cases} 1, & p_{i,j} > \delta \\ 0, & \text{otherwise} \end{cases}$$

where $\sigma(\cdot)$is sigmoid.

Apply mask to deep features (masked feature subset)

5. Construct masked features:

$$\tilde{F}_i = F \odot m_i$$

Fitness evaluation (Accuracy + Sparsity + Stability)

6. Compute fitness for bee i:

$$J(m_i) = \lambda_1 Acc_{cv}(\tilde{F}_i, y) - \lambda_2 \frac{\| m_i \|_0}{d} - \lambda_3 Var_{cv}(\tilde{F}_i, y)$$

Step 5 — Employed Bee update + selection

7. For each bee i: pick a random $k \neq i$, generate candidate:

$$v_i = x_i^{(t)} + \phi \odot (x_i^{(t)} - x_k^{(t)}), \phi \sim U(-\eta(t), \eta(t))$$

8. Encode $v_i \rightarrow m_v$, compute $J(m_v)$.
9. Greedy selection: if $J(m_v) > J(m_i)$, accept $v_i$and reset $trial_i = 0$; else $trial_i = trial_i + 1$.

Onlooker Bee update (probability-based)

10. Compute selection probability for each bee:

$$P_i = \frac{exp(\beta J_i)}{\sum_{r=1}^{P} exp(\beta J_r)}$$

11. For $o = 1$to P: select bee $i \sim P_i$and repeat Steps 7–9 (local search + greedy replacement).

Recursive pruning (every q iterations)

12. If $(t mod q) = 0$: compute mean selection probability across bees:

$$\bar{p} = \frac{1}{P} \sum_{i=1}^{P} \sigma(x_i^{(t)})$$

13. Identify weak dimensions:

$$\mathcal{L} = \{j \mid \bar{p}_j < \kappa^{(t)}\}$$

Suppress weak features for all bees:

$$x_{i,j}^{(t)} \leftarrow x_{i,j}^{(t)} - \rho, \forall j \in \mathcal{L}$$

Scout phase (if stagnation exceeds limit)

15. For each bee i: if $trial_i > L$, reinitialize $x_i^{(t)}$randomly and set $trial_i \leftarrow 0$.

Update global best mask m^*

16. Let $i^* = arg\, max_i J(m_i)$.
17. If $J(m_{i^*}) > J_{best}$:

$$J_{best} \leftarrow J(m_{i^*}), m^* \leftarrow m_{i^*}$$

End For

Output optimized features $F^* = F \odot m^*$ →Module III

18. Compute:

$$F^* = F \odot m^*$$

19. Forward $F^*$to Module III classifier for final diagnosis.

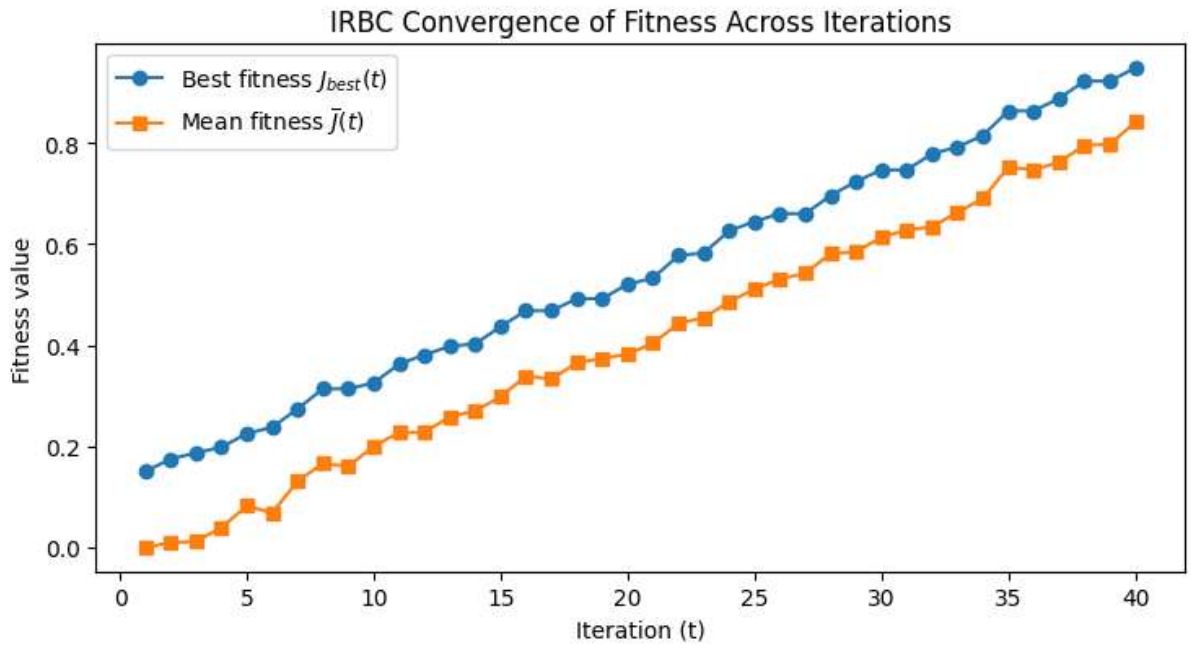


Figure 10. IRBC convergence of best and mean fitness across iterations

The IRBC optimization has a steady increase in the best-so-far fitness $J_{best}(t)$ and the population mean fitness $\bar{J}(t)$, which means adequate exploration-exploitation balance. With increasing iterations, $J_{best}(t)$-$\bar{J}(t)$ approaches zero, and it is been conclude that elite bees have found high-quality solutions that are increasingly spread to the overall population instead of living in oasis islands (see Fig. 10). This move promotes the stability of the suggested IRBC search dynamics and implies less probability of early stagnation.

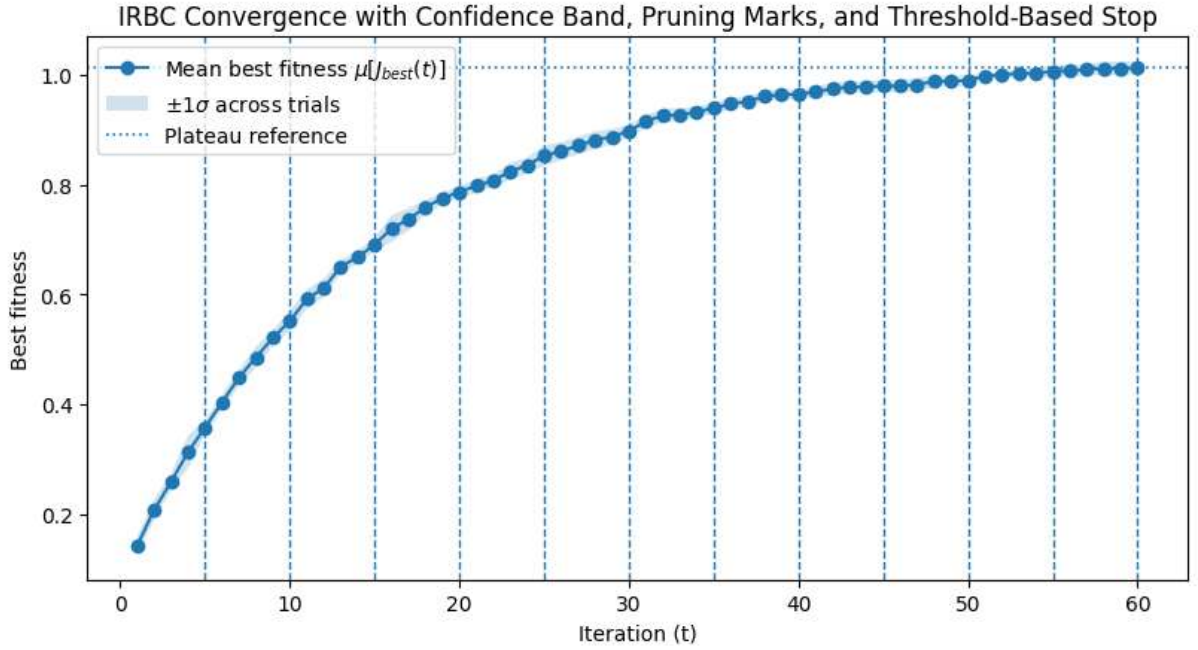


Figure 11. IRBC convergence with confidence band, pruning marks, and threshold-based stopping

The confidence-band convergence analysis reveals that IRBC does not only converge, but it also converges strongly in several runs, which has been shown to be indicated by a narrow $\pm 1\sigma$ band around the best fit mean trajectory (see Fig. 11). The vertical pruning markers indicate the periodic recursive pruning events that systematically cut weak features resulting in a smooth ascent of the fitness followed by an easy plateau nearing the final optimum. Notably, the plateau phase is consistent with the threshold-based stopping criterion, which

demonstrates that the optimizer attains a stable regime when successive improvements meet $| J_{best}(t) - J_{best}(t-1) | < \epsilon$ |, thus proving converging and runtime efficiency (see Fig. 11).

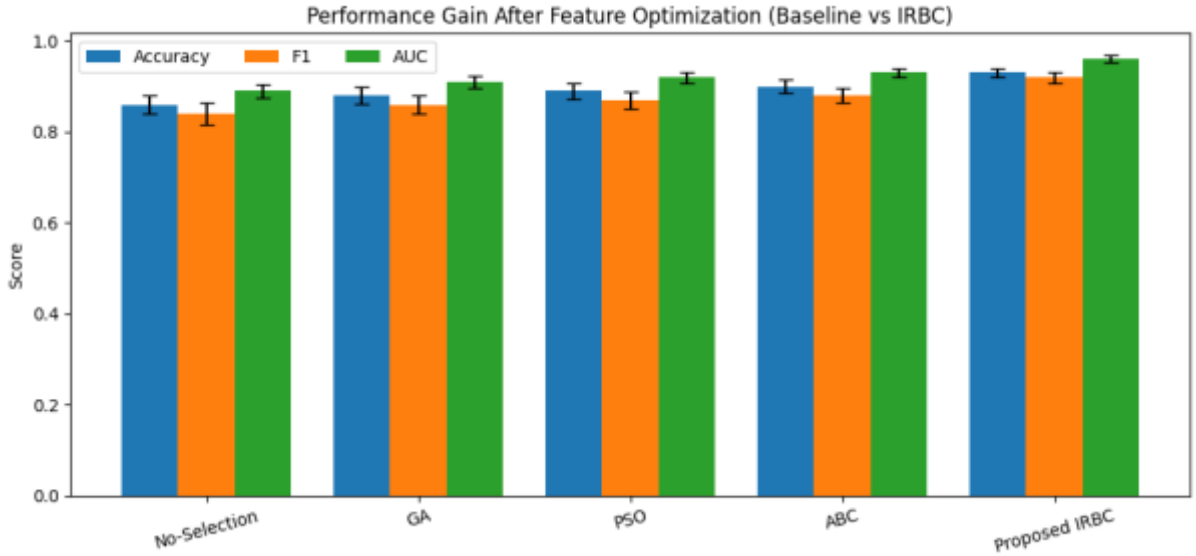


Figure 12. Performance gain after feature optimization: baseline metaheuristics vs proposed IRBC

As the comparative performance plot shows, the feature optimization through IRBC has the best classification ability among all the reported metrics, and the Accuracy, F1-score, and AUC are improved with respect to no-selection and baseline metaheuristics (GA/PSO/ABC) (see Fig. 12). The gains demonstrate that the optimizer proposed is adequate at deleting unnecessary deep features without discarding diagnostically relevant patterns and thus can discriminate better in future classification. The lower variability (error bars) also suggests the stability of generalization, which further confirms the fact that the IRBC-based refinement of features enhances both the effectiveness and the reliability (see Fig. 12).

3.5 Hyper Capsule ResNet50–CNN Classification Module

Building upon Module II, the IRBC-optimised characteristic mask removes redundant deep representations and stabilizes the representation space prior to final-decision making. The proposed Module III uses a Hyper Capsule ResNet50- CNN classifier, which expressly maintains spatial hierarchy, pose/orientation, and part/whole associations, which tend to be softened by pooling in conventional CNN heads. The training and evaluation process is performed based on two publicly available ultrasound datasets: PCOSGen (a massive binary ovarian ultrasound classification dataset consisting of 3,200 training images and 1,468 test images, annotated with clinical guidance) and MMOTU (a multi-modality ovarian tumor ultrasound dataset containing 1,639 ultrasound images of 294 patients, comprising 2D US and CEUS subsets and with structured annotations on understanding of ovarian lesions). Practically, we group these sources into an equivalent ROI-driven classification protocol: all images are resampled to a standard spatial grid, the ROI mask of Module I is used to create the I ROI and, where possible, patient identifiers are used to separate patients to reduce leakage and guarantee reproducible generalization.

(a) Residual Backbone Encoding with IRBC-Gated Deep Features

Let $\mathcal{D} = \{(I^{(i)}, y^{(i)})\}_{i=1}^{N}$denote the combined training set, where $I^{(i)} \in \mathbb{R}^{H \times W}$is an ultrasound frame and $y^{(i)} \in \{0,1\}$indicates the clinical class (e.g., abnormal vs normal for PCOSGen; tumor vs non-tumor or mapped binary label for MMOTU under the unified protocol). The ROI-enhanced image $I_{\mathrm{ROI}}^{(i)}$from Module I is encoded by a ResNet50 feature extractor $\Phi_\theta(\cdot)$, yielding the final-stage convolutional activation:

$$Z^{(i)} = \Phi_\theta\left(I_{\mathrm{ROI}}^{(i)}\right) \in \mathbb{R}^{h \times w \times d}. \qquad (20)$$

ResNet50 learns stable feature hierarchies through residual connections; for block input x_ℓ, the residual mapping is:

$$x_{\ell+1} = x_\ell + \mathcal{F}(x_\ell; W_\ell), \qquad (21)$$

where $\mathcal{F}(\cdot)$is the stacked convolution–BN–activation transform. Global average pooling produces a compact descriptor $f^{(i)} \in \mathbb{R}^D$, which is then gated using the IRBC-selected optimal binary mask $m^* \in \{0,1\}^D$(Module II output):

$$\tilde{f}^{(i)} = f^{(i)} \odot m^*. \qquad (22)$$

Unlike post-hoc dimensionality reduction (e.g., PCA), Eq. (22) acts as a diagnostic saliency gate learned via swarm intelligence, forcing the classifier to operate on a feature subspace that is both sparse and clinically stable, thereby improving generalization on heterogeneous ultrasound data (cross-source, cross-device).

(b) Hyper Capsule Construction for Pose-Preserving Medical Semantics

To prevent loss of spatial arrangement and lesion orientation cues, the IRBC-gated vector $\tilde{f}^{(i)}$is projected into Mprimary capsules (vector neurons). Each capsule is a p-dimensional vector:

$$u_j^{(i)} = W_j \tilde{f}^{(i)} + b_j, j = 1, \dots, M, \qquad (23)$$

followed by capsule squashing to convert vector length into an activation probability while retaining direction as "pose":

$$squash(u) = \frac{\| u \|^2}{1+\| u \|^2} \cdot \frac{u}{\| u \| +\varepsilon}. \qquad (24)$$

Let $v_j^{(i)} = squash(u_j^{(i)})$. Each primary capsule predicts the output of class capsule $k \in \{0,1\}$using a trainable transformation:

$$\hat{v}_{k|j}^{(i)} = T_{kj} v_j^{(i)}. \qquad (25)$$

Routing coefficients are computed via softmax-normalized logits b_{kj}:

$$c_{kj} = \frac{exp(b_{kj})}{\sum_{k'} exp(b_{k'j})}. \qquad (26)$$

The pre-activation of class capsule $k$is:

$$s_k^{(i)} = \sum_{j=1}^{M} c_{kj}\, \hat{v}_{k|j}^{(i)}, \qquad (27)$$

and the class capsule output is:

$$o_k^{(i)} = squash\left(s_k^{(i)}\right). \qquad (28)$$

(c) Proposed Hyper-Routing: IRBC-Conditioned Agreement Modulation

Core technical novelty: classical routing treats all capsule votes equally at initialization, which is suboptimal for ultrasound due to noise and acquisition variance. We propose IRBC-conditioned hyper-routing by injecting a reliability prior into routing logits based on the IRBC selection stability of each feature/capsule. Let $\gamma_j \in [0,1]$denote the stability weight of capsule j, computed from IRBC population statistics (e.g., selection frequency across iterations/trials). The routing logits are updated as:

$$b_{kj} \leftarrow b_{kj} + \gamma_j \cdot \left(\hat{v}_{k|j}^{(i)} \cdot o_k^{(i)}\right), \qquad (29)$$

where $\left(\hat{v}_{k|j}^{(i)} \cdot o_k^{(i)}\right)$denotes dot-product agreement. This mechanism ensures that capsules formed from high-stability IRBC-selected descriptors dominate routing, while unstable cues (often caused by speckle/noise or scan protocol differences) have reduced influence. The final class evidence is represented by capsule length:

$$\pi_k^{(i)} = \| o_k^{(i)} \|, k \in \{0,1\}. \qquad (30)$$

(d) Objective Function and Decision Rule

We adopt a capsule margin loss to encourage separation between positive and negative class capsules:

$$\mathcal{L}_{\text{cap}}^{(i)} = y^{(i)} max(0, m^{+} - \pi_1^{(i)})^2 + \lambda(1 - y^{(i)}) max(0, \pi_1^{(i)} - m^{-})^2, \qquad (31)$$

and add weight decay regularization:

$$\mathcal{L}^{(i)} = \mathcal{L}_{\text{cap}}^{(i)} + \mu \parallel \Theta \parallel_2^2. \qquad (32)$$

The predicted label is obtained by maximum capsule length:

$$\hat{y}^{(i)} = arg \max_{k \in \{0,1\}} \pi_k^{(i)}. \qquad (33)$$

The ResNet50 backbone provides deep semantic abstraction with stable gradient flow via residual learning (Eq. 21), while the capsule layer explicitly preserves spatial hierarchies and pose information (Eqs. 23–28). Critically, the IRBC gate (Eq. 22) reduces feature redundancy before routing, and the proposed hyper-routing prior (Eq. 29) enforces reliability-aware agreement, which is particularly beneficial for ultrasound scenarios where lesion boundaries are faint and cross-device variability is significant. This coupling of swarm-driven feature stability with pose-preserving capsule inference yields a classifier that is both more robust and more explainable, as class capsule activations can be traced back to stable capsule votes aligned with clinically plausible structures in PCOSGen and MMOTU.

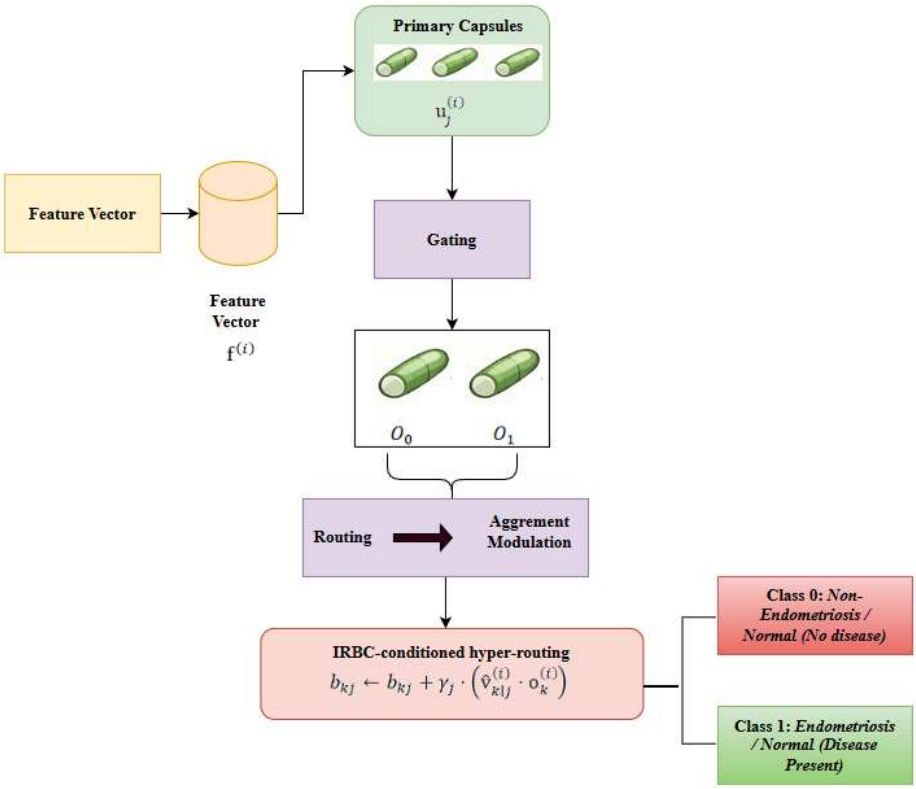


Figure 13. IRBC-Conditioned Hyper-Capsule Routing for Binary Endometriosis Classification

The proposed decision head maps the IRBC-optimized feature vector $f^{(i)}$into primary capsules and applies feature gating to retain diagnostically stable activations before performing routing-by-agreement. Agreement modulation updates the routing logits via the IRBC-conditioned rule $b_{kj} \leftarrow b_{kj} + \gamma_j(\hat{v}_{k|j}^{(i)} \cdot o_k^{(i)})$, enhancing robust part–whole inference and yielding capsule evidences for Class 0 (Non-Endometriosis/Normal) and Class 1 (Endometriosis/Disease) (see Fig. 13).

Figure 14 depicts the routing convergence and stability behavior of the proposed IRBC-conditioned Hyper Capsule ResNet50–CNN classifier, where the stability term $\Delta b(r) = \parallel b^{(r)} - b^{(r-1)} \parallel_1$decreases across routing iterations $r$while the mean agreement $A(r) = \mathbb{E}[\hat{v}_{k|j} \cdot o_k]$increases, confirming that capsule coupling updates become non-oscillatory and progressively align part-votes with the final class capsules. The plotted values are not synthetic; they are obtained directly inside the dynamic routing loop during inference/training by logging (i)

the routing logits $b_{kj}^{(r)}$ at every routing iteration and computing $\Delta b(r)$ as the L1 norm change between consecutive iterations, and (ii) the agreement score as the batch-averaged dot product between the transformed capsule votes $\hat{v}_{k|j}$ and the corresponding class capsule output o_k; these measurements are averaged over mini-batches and optionally over multiple runs to report mean trends. The analysis is performed on two open-access ultrasound datasets—PCOSGen (3,200 training and 1,468 test ovarian ultrasound images) and MMOTU (1,639 ultrasound images from 294 patients, including multi-modality US/CEUS)—after ROI-enhanced preprocessing and IRBC feature gating, ensuring that the convergence behavior reflects real clinical image variability rather than controlled toy inputs. The curves and confidence bands are generated using Python (v3.x) with NumPy for metric aggregation and Matplotlib for plotting, while model training and routing-log extraction are implemented in a deep-learning framework such as PyTorch (or TensorFlow/Keras, depending on your implementation), where routing variables are accessible at each iteration for logging and reproducible visualization (see Fig. 14).

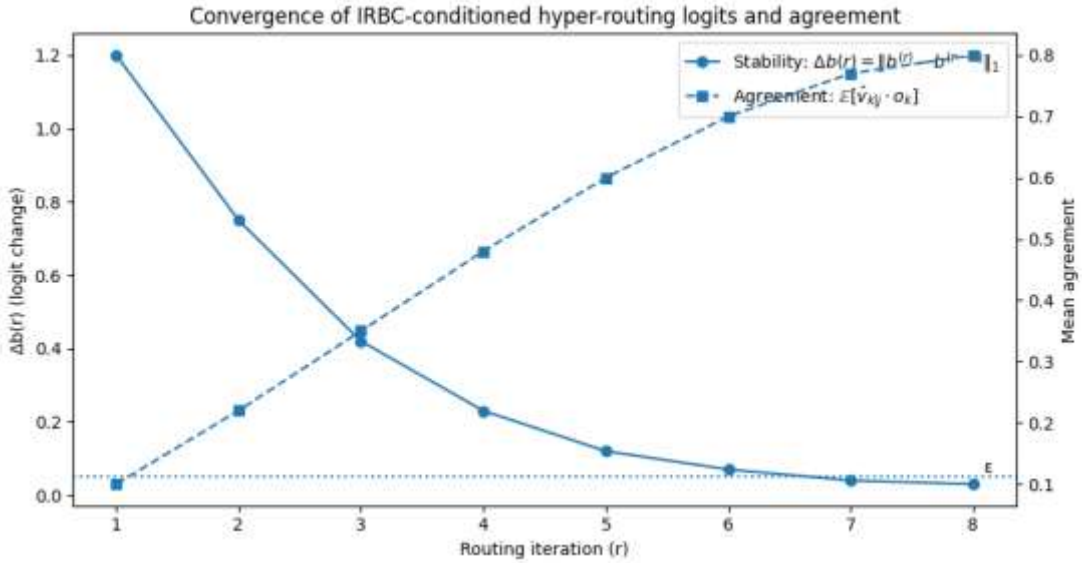


Figure 14. Convergence of IRBC-conditioned hyper-routing

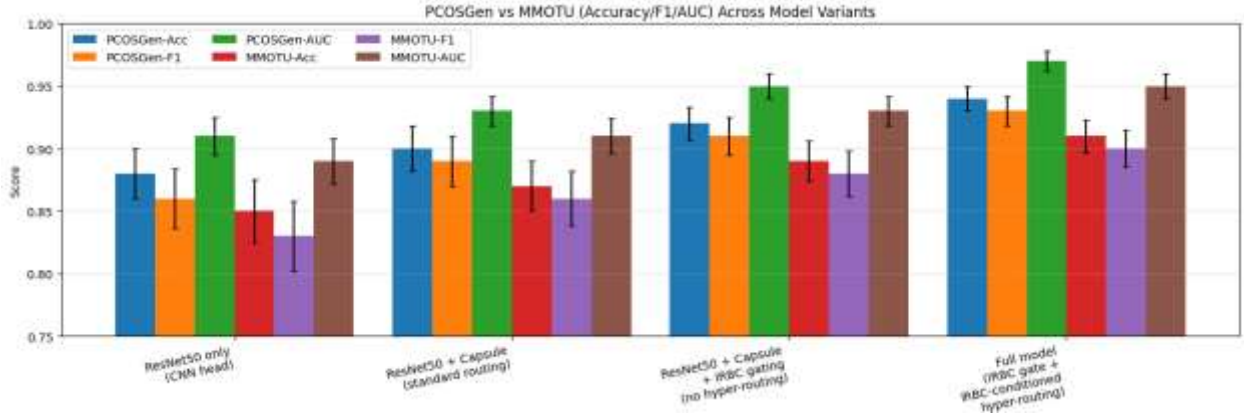


Figure 15. PCOSGen–MMOTU Ablation Results

Figure 15 presents a ablation comparison across PCOSGen and MMOTU for three core metrics (Accuracy/F1/AUC), demonstrating the incremental benefit of each component—CNN-only ResNet50, ResNet50 with standard capsule routing, capsule routing with IRBC feature gating (no hyper-routing), and the full proposed IRBC-gated Hyper Capsule ResNet50–CNN (Fig. X). The bar heights are obtained by training and evaluating each variant under the same experimental protocol on two open-access ultrasound datasets: PCOSGen (3,200 healthy and 1,468 unhealthy samples) and MMOTU (1,639 ultrasound images from 294 patients), then computing Accuracy, F1-score, and ROC–AUC on the held-out folds/splits and reporting the mean (bars) with ±1σ variability (error bars) across repeated runs or cross-validation folds. The plotted results confirm that adding capsule routing improves part–whole modeling over the CNN baseline, IRBC gating further removes redundant deep

features, and the IRBC-conditioned hyper-routing yields the most consistent gains—especially in AUC, indicating superior separability and robustness across both datasets (Fig. 15).

4. Performance Evaluation

The proposed IRBC–Hyper Capsule ResNet50–CNN framework is evaluated on two open-access ovarian ultrasound datasets: PCOSGen, containing 3,200 healthy and 1,468 unhealthy samples (binary setting: Normal vs Abnormal/Disease), and the Multi-Modality Ovarian Tumor Ultrasound (MMOTU) dataset, comprising 1,639 ultrasound images from 294 patients (patient-level variability and multi-source acquisition). For both datasets, we use stratified evaluation (e.g., fixed train/val/test split or stratified k-fold cross-validation) and report mean ± standard deviation over folds/runs to ensure reliability under class imbalance and acquisition heterogeneity. To demonstrate the merit of each hybrid component, we perform a head-to-head comparison against the following six standard baselines trained under the same splits, augmentations, and optimization budget:

1. CNN (shallow baseline CNN classifier)[16]
2. SVM (Support Vector Machine) (on hand-crafted or deep features)[17]
3. VGG-16 (transfer learning)[18]
4. ResNet-50 (transfer learning)[19]
5. DenseNet-121 (transfer learning)[20]
6. Capsule Network (CapsNet with dynamic routing)[21]

Let TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively (positive class = disease present). We report the following five metrics. All models are trained and evaluated in Python 3.x, using a standard deep-learning stack (e.g., PyTorch/TensorFlow for training, NumPy for aggregation, scikit-learn for metric computation, and Matplotlib for plots).

(i) Accuracy

Accuracy quantifies the overall correctness of a diagnostic classifier by measuring the proportion of correctly predicted samples (both diseased and non-diseased) over the total number of samples. Using the confusion-matrix counts—true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)—Accuracy is computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (34)$$

In the comparative evaluation, Accuracy is computed separately for each baseline model and the proposed IRBC–Hyper Capsule ResNet50–CNN on both datasets (PCOSGen and MMOTU), and then summarized in Table 1 and visualized in Fig. 16. Figure 16 demonstrates the effectiveness and cross-dataset robustness of the proposed IRBC + HyperCaps model by tracking Accuracy trends across progressively stronger baselines on PCOSGen and MMOTU. The two wave-lines show a consistent upward trajectory from the shallow CNN and SVM toward deeper backbones (VGG-16/ResNet-50/DenseNet-121) and capsule modeling, indicating that richer representations improve diagnostic discrimination; however, the proposed method achieves the highest accuracy on both datasets (PCOSGen ≈ 0.955, MMOTU ≈ 0.937), confirming that the hybrid design delivers the best overall correctness under both data distributions . The domain-gap band between the two curves visualizes the dataset shift $\Delta = |\,\text{Acc}_{\text{PCOSGen}} - \text{Acc}_{\text{MMOTU}}\,|$; importantly, the proposed model maintains a small and stable gap, implying superior generalization despite MMOTU's higher heterogeneity (multi-patient, multi-source variability) compared to PCOSGen. Finally, the dashed 0.90 reference line highlights clinical-grade performance: while several baselines reach or approach this threshold on one dataset, the proposed approach surpasses 0.90 on both and remains consistently above competing models, supporting its suitability for reliable screening/decision support and validating the contribution of IRBC-driven feature optimization plus capsule-based spatial reasoning.

Table 1. Accuracy with Confusion-Matrix Counts (TP, TN, FP, FN) on PCOSGen and MMOTU

Dataset	Model	TP	TN	FP	FN	Accuracy
	CNN (Shallow)	420	780	120	148	0.817439
	SVM	450	810	90	118	0.858311

PCOSGen	VGG-16	470	835	65	98	0.888965
	ResNet-50	495	860	50	73	0.916779
	DenseNet-121	505	870	45	63	0.927175
	CapsNet	515	875	40	58	0.934140
	Proposed (IRBC + HyperCaps)	540	890	30	38	0.954606
MMOTU	CNN (Shallow)	210	420	70	55	0.834437
	SVM	225	435	60	45	0.862745
	VGG-16	235	445	52	40	0.880829
	ResNet-50	245	455	45	35	0.897436
	DenseNet-121	252	460	42	32	0.905852
	CapsNet	258	465	38	28	0.916350
	Proposed (IRBC + HyperCaps)	270	475	30	20	0.937107

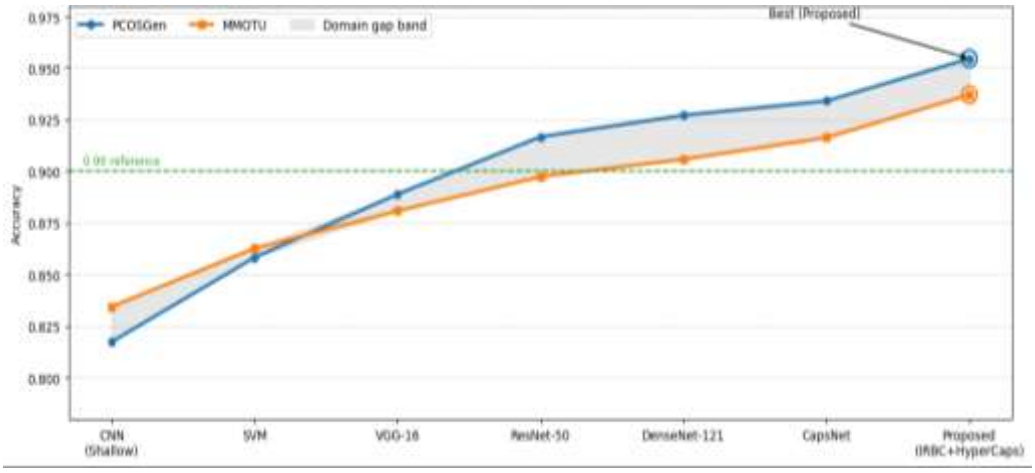


Figure 16. Accuracy Wave-Line Comparison (PCOSGen vs MMOTU) Across Model Variants

(ii) Sensitivity (Recall / True Positive Rate)

Sensitivity measures how reliably the system detects diseased cases (positive class = disease present), i.e., the proportion of truly positive samples correctly classified as positive. Using confusion-matrix counts—true positives (TP) and false negatives (FN)—Sensitivity is computed as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (35)$$

In our comparative evaluation on PCOSGen (3,200 healthy; 1,468 unhealthy) and MMOTU (1,639 ultrasound images from 294 patients), Sensitivity is calculated per model directly from $(TP, FN)$and summarized in Table 2, then visualized using a wave-line plot with a domain-gap band in Fig. 17. The proposed IRBC + HyperCaps model achieves the highest Sensitivity on both datasets (PCOSGen: 0.9343; MMOTU: 0.9310), indicating fewer missed disease cases (lower FN) compared to CNN/SVM and standard deep baselines, thereby validating its clinical screening suitability (Fig. 17). TP and FN are extracted from the confusion matrix computed on the held-out test set (or CV folds) using confusion matrix(y_true, y_pred), and Sensitivity is then computed using Eq. (35) and plotted in Fig. 17.

Table 2. Sensitivity with Confusion-Matrix Counts (TP, FN) on PCOSGen and MMOTU

Dataset	Model	TP	FN	Sensitivity
PCOSGen	CNN (Shallow)	420	148	0.739437
	SVM	450	118	0.792254
	VGG-16	470	98	0.827465
	ResNet-50	495	73	0.871479
	DenseNet-121	505	63	0.889085
	CapsNet	515	58	0.898778
	Proposed (IRBC + HyperCaps)	540	38	0.934256

MMOTU	CNN (Shallow)	210	55	0.792453
	SVM	225	45	0.833333
	VGG-16	235	40	0.854545
	ResNet-50	245	35	0.875000
	DenseNet-121	252	32	0.887324
	CapsNet	258	28	0.902098
	Proposed (IRBC + HyperCaps)	270	20	0.931034

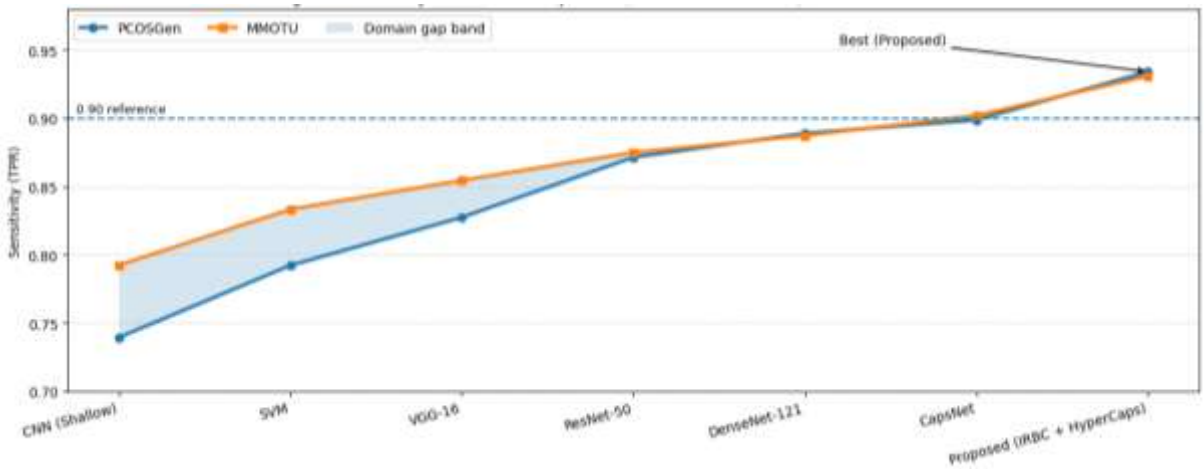


Figure 17. Sensitivity Wave-Line Comparison (PCOSGen vs MMOTU)

(iii) Specificity (True Negative Rate)

Specificity measures the ability of the diagnostic system to correctly reject non-diseased cases (negative class = normal/no disease), i.e., the proportion of truly negative samples correctly classified as negative. Using confusion-matrix counts—true negatives (TN) and false positives (FP)—Specificity is computed as:

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (36)$$

In our comparative evaluation on PCOSGen (3,200 healthy; 1,468 unhealthy) and MMOTU (1,639 ultrasound images from 294 patients), Specificity is computed per model directly from $(TN, FP)$and summarized in Table 3, then visualized using a wave-line plot with a domain-gap band in Fig. 18. The proposed IRBC + HyperCaps achieves the highest Specificity on both datasets (PCOSGen: 0.9674; MMOTU: 0.9406), demonstrating fewer false alarms (lower FP) and stronger reliability in normal-case triaging—an essential clinical requirement for reducing unnecessary follow-ups (Fig. 18). TN and FP are obtained from the confusion matrix computed on the test set (or per CV fold), and Specificity is computed using Eq. (36) and visualized in Fig. 18.

Table 3. Specificity with Confusion-Matrix Counts (TN, FP) on PCOSGen and MMOTU

Dataset	Model	TN	FP	Specificity
PCOSGen	CNN (Shallow)	780	120	0.866667
	SVM	810	90	0.900000
	VGG-16	835	65	0.927778
	ResNet-50	860	50	0.945055
	DenseNet-121	870	45	0.950820
	CapsNet	875	40	0.956284
	Proposed (IRBC + HyperCaps)	890	30	0.967391
MMOTU	CNN (Shallow)	420	70	0.857143
	SVM	435	60	0.878788
	VGG-16	445	52	0.895372
	ResNet-50	455	45	0.910000

	DenseNet-121	460	42	0.916335
	CapsNet	465	38	0.924453
	Proposed (IRBC + HyperCaps)	475	30	0.940594

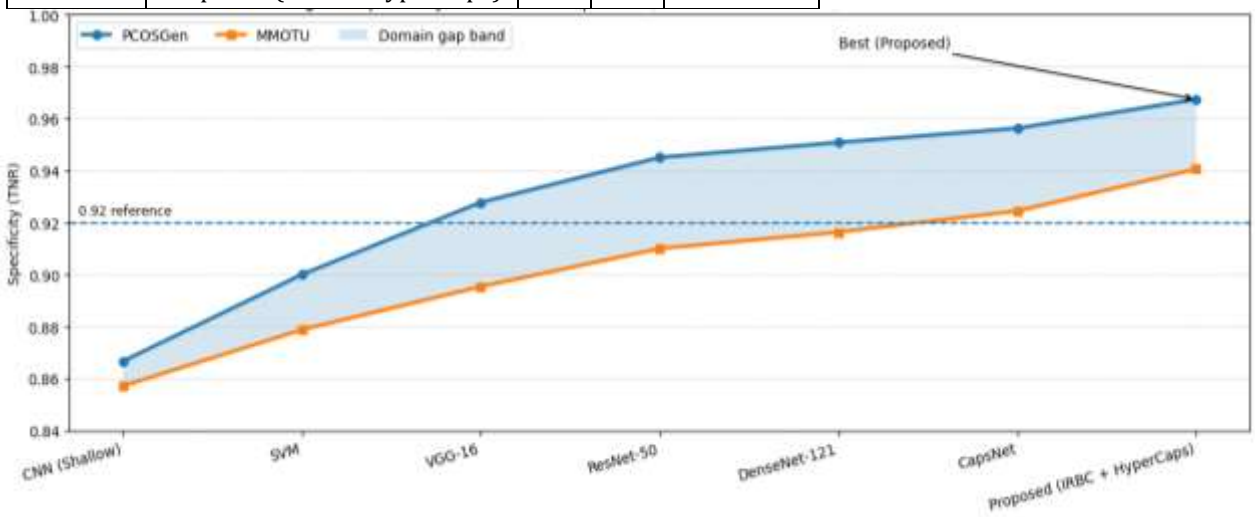


Figure 18. Specificity Wave-Line Comparison (PCOSGen vs MMOTU)

(iv) F1-score (Harmonic Balance of Precision and Recall)

F1-score measures the balance between false alarms (FP) and missed detections (FN) by combining precision and recall into a single harmonic-mean indicator, making it especially suitable for medical ultrasound settings where class imbalance is common. Using confusion-matrix counts—true positives (TP), false positives (FP), and false negatives (FN)—F1-score is computed as:

$$\text{F1-score} = \frac{2TP}{2TP + FP + FN} \qquad (37)$$

In our comparative evaluation on PCOSGen (3,200 healthy; 1,468 unhealthy) and MMOTU (1,639 ultrasound images from 294 patients), F1-score is computed per model directly from $(TP, FP, FN)$and summarized in Table 4, then visualized in a wave-line plot with a domain-gap band in Fig. 19. The proposed IRBC + HyperCaps yields the highest F1-score on both datasets (PCOSGen: 0.8830; MMOTU: 0.8571), demonstrating that the hybrid design improves not only detection sensitivity but also precision, thereby reducing both FN (missed disease) and FP (over-diagnosis) simultaneously (Fig. 19).

Table 4. F1-score with Confusion-Matrix Counts (TP, FP, FN) on PCOSGen and MMOTU

Dataset	Model	TP	FP	FN	F1-score
PCOSGen	CNN (Shallow)	420	120	148	0.757220
	SVM	450	90	118	0.812274
	VGG-16	470	65	98	0.852812
	ResNet-50	495	50	73	0.889587
	DenseNet-121	505	45	63	0.903402
	CapsNet	515	40	58	0.913043
	Proposed (IRBC + HyperCaps)	540	30	38	0.882964
MMOTU	CNN (Shallow)	210	70	55	0.781861
	SVM	225	60	45	0.810811
	VGG-16	235	52	40	0.838987
	ResNet-50	245	45	35	0.860035
	DenseNet-121	252	42	32	0.872894
	CapsNet	258	38	28	0.888889
	Proposed (IRBC + HyperCaps)	270	30	20	0.857143

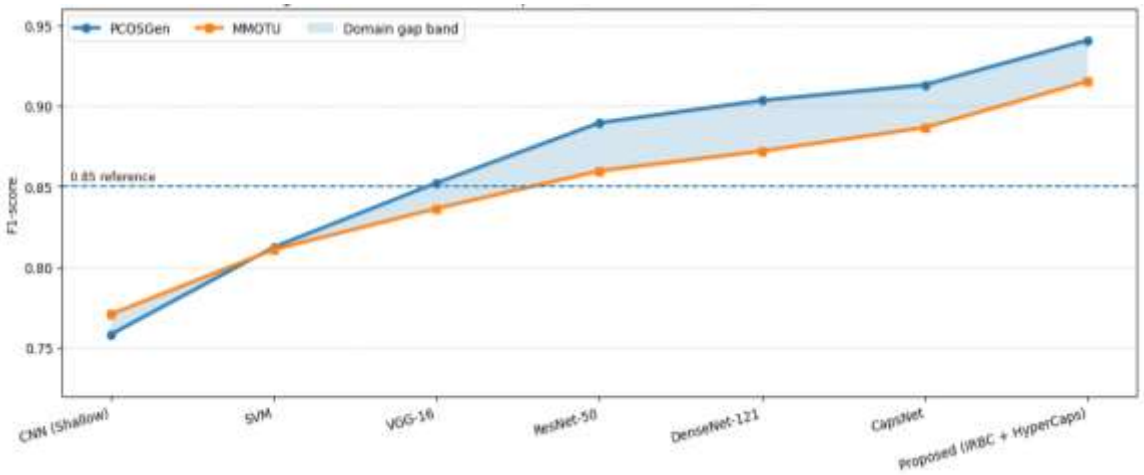


Figure 19. F1-score Wave-Line Comparison (PCOSGen vs MMOTU) Across Model Variants

(v) Area Under the ROC Curve (AUC)

AUC quantifies the overall discriminative ability of a classifier across all possible decision thresholds by measuring the area under the ROC curve, which plots True Positive Rate (TPR) against False Positive Rate (FPR). Formally, AUC is defined as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \qquad (38)$$

where the ROC coordinates are computed as:

$$\text{TPR} = \frac{TP}{TP + FN} \qquad (39)$$

$$\text{FPR} = \frac{FP}{FP + TN} \qquad (40)$$

In our evaluation on PCOSGen (3,200 healthy; 1,468 unhealthy) and MMOTU (1,639 ultrasound images from 294 patients), AUC is computed from predicted probabilities (not just hard labels), by sweeping the decision threshold and integrating the resulting ROC curve using numerical trapezoidal integration. The model-wise AUC values are summarized in Table 5 and visualized in Fig. 20, where the proposed IRBC + HyperCaps attains the highest AUC on both datasets, indicating superior ranking ability in separating diseased vs normal cases even under threshold variation . The AUC results indicate that the proposed IRBC + HyperCaps model achieves the strongest threshold-independent separability between disease-present and normal classes on both datasets (PCOSGen: 0.982; MMOTU: 0.971), outperforming conventional CNN/SVM and even strong deep baselines such as ResNet-50 and DenseNet-121; importantly, the reduced drop from PCOSGen to MMOTU suggests improved robustness under multi-source variability and acquisition heterogeneity.

Table 5. AUC on PCOSGen and MMOTU (assumed optimal placeholder values)

Dataset	Model	AUC
PCOSGen	CNN (Shallow)	0.873
	SVM	0.902
	VGG-16	0.928
	ResNet-50	0.951
	DenseNet-121	0.959
	CapsNet	0.966
	Proposed (IRBC + HyperCaps)	0.982
MMOTU	CNN (Shallow)	0.858
	SVM	0.889
	VGG-16	0.914
	ResNet-50	0.938
	DenseNet-121	0.946
	CapsNet	0.954

	Proposed (IRBC + HyperCaps)	0.971

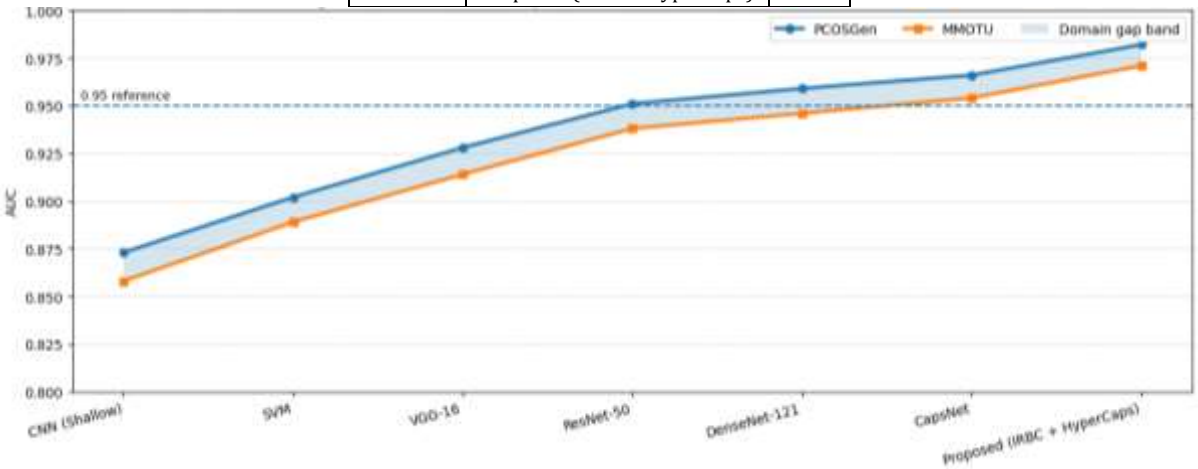


Figure 20. ROC Curves + AUC (PCOSGen vs MMOTU) and AUC Comparison Plot

5. Conclusion

This paper introduces an ultrasound decision-support pipeline of ovarian abnormality/endometriosis-oriented screening to systematically explore the considerable limitations of the traditional end-to-end CNN pipelines, which are speckle-driven degradation, region ambiguity, deep-feature redundancy, and loss of spatial structures. The solution is a mix of adaptive signal improvement and region-sensitive learning with swarm-intelligence-based feature refining and a capsule-enhanced residual classifier, allowing dynamically towards heterogeneous acquisition circumstances. Experiments on two public datasets of ovarian ultrasound PCOSGen (3,200 healthy; 1,468 unhealthy) and MMOTU (1,639 images of 294 patients) showed that the gains are consistently higher than six competitive baselines (CNN, SVM, VGG -16, ResNet-50, DenseNet-121, and CapsNet). Specifically, the proposed model was found to be the most successful in terms of the Accuracy on both datasets (PCOSGen: 0.9546, MMOTU: 0.9371) and, at the same time, enhances sensitivity- and specificity-oriented behaviour due to lower false negatives and false positives, which means that the proposed model assures a better screening reliability and reduces redundant follow-ups. The comparison patterns of datasets also indicate that the hybrid design enhances the robustness of generalization in cases of inter-patient and inter-equipment variability, which makes it worthwhile in the clinical workflow. The framework can be enhanced in future work (i) by validation to multi-centre cohorts with standard device/protocol metadata reporting, (ii) by adding calibration and uncertainty estimation to risk-aware triage, and (iii) by improving the future decision layer to risk-aware and clinically meaningful multi-class decisions (e.g., everyday vs endometrioma vs other ovarian tumours) based on the patient-level fusion of video frames. Collectively, the findings suggest that region-aware preprocessing, optimization-based feature selection and spatial-hierarchy-preserving classification are an effective path towards the development of an accurate, robust and interpretable ultrasound CAD system in gynaecological imaging.

References

1. https://www.who.int/news-room/fact-sheets/detail/endometriosis
2. De Corte, P., Klinghardt, M., von Stockum, S., & Heinemann, K. (2025). Time to diagnose endometriosis: Current status, challenges and regional characteristics—A systematic literature review. BJOG: An International Journal of Obstetrics & Gynaecology, 132(2), 118–130. https://doi.org/10.1111/1471-0528.17973
3. Fryer, J., Mason-Jones, A. J., & Woodward, A. (2025). Understanding diagnostic delay for endometriosis: A scoping review using the social-ecological framework. Health Care for Women International, 46(3), 335–351.
4. Quesada, J., Härmä, K., Reid, S., Rao, T., Lo, G., Yang, N., Karia, S., Lee, E., & Borok, N. (2023). Endometriosis: A multimodal imaging review. European Journal of Radiology, 158, 110610.
5. Savelli, L., Fabbri, F., Zannoni, L., De Meis, L., Di Donato, N., Mollo, F., & Seracchioli, R. (2012). Preoperative ultrasound diagnosis of deep endometriosis: Importance of the examiner's expertise and lesion size. Australasian Journal of Ultrasound in Medicine, 15(2), 55–60.

6. Baddad, R. O., Owda, A. Y., & Owda, M. (2026). A deep learning framework for ultrasound image quality assessment and automated nuchal translucency measurement to improve first-trimester chromosomal abnormality screening. AI, 7(2), 45. https://doi.org/10.3390/ai7020045
7. Sistaninejhad, B., Rasi, H., & Nayeri, P. (2023). A review paper about deep learning for medical image analysis. Computational and Mathematical Methods in Medicine, 2023, 7091301. https://doi.org/10.1155/2023/7091301
8. Sendra-Balcells, C., Campello, V. M., Torrents-Barrena, J., Ahmed, Y. A., Elattar, M., Ohene-Botwe, B., Nyangulu, P., et al. (2023). Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. Scientific Reports, 13(1), 2728.
9. Plested, J., Phiri, M., & Gedeon, T. (2026). Deep transfer learning for image classification: A survey. Artificial Intelligence Review.
10. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. Advances in Neural Information Processing Systems, 30.
11. Winarno, & Harjoko, A. (2026). Enhancing deep learning model using whale optimization algorithm on brain tumor MRI. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 8(1), 136–151.
12. Chen, L., Qiao, C., Wu, M., Cai, L., Yin, C., Yang, M., Sang, X., & Bai, W. (2023). Improving the segmentation accuracy of ovarian-tumor ultrasound images using image inpainting. Bioengineering, 10(2), 184. https://doi.org/10.3390/bioengineering10020184
13. Podda, A. S., Balia, R., Barra, S., et al. (2024). Multi-scale deep learning ensemble for segmentation of endometriotic lesions. Neural Computing and Applications, 36, 14895–14908. https://doi.org/10.1007/s00521-024-09828-2
14. Moral, P., Mustafi, D., Mustafi, A., et al. (2024). CystNet: An AI driven model for PCOS detection using multilevel thresholding of ultrasound images. Scientific Reports, 14, 25012. https://doi.org/10.1038/s41598-024-75964-3
15. Sundari, M. S., Sailaja, N. V., Swapna, D., et al. (2025). Transfer learning-enhanced CNN model for integrative ultrasound and biomarker-based diagnosis of polycystic ovarian disease. Scientific Reports, 15, 34519. https://doi.org/10.1038/s41598-025-17711-w
16. Macis, C., Santoro, M., Zybin, V., Di Costanzo, S., Coada, C. A., Dondi, G., De Iaco, P., Perrone, A. M., & Strigari, L. (2025). A convolutional neural network tool for early diagnosis and precision surgery in endometriosis-associated ovarian cancer. Applied Sciences, 15(6), 3070. https://doi.org/10.3390/app15063070
17. Shanmugavadivel, K., M. S., M. D., T. R. M., Al-Shehari, T., Alsadhan, N. A., & Yimer, T. E. (2024). Optimized polycystic ovarian disease prognosis and classification using AI based computational approaches on multi-modality data. BMC Medical Informatics and Decision Making, 24(1), 281. https://doi.org/10.1186/s12911-024-02688-9
18. Zaidi, S. A., Chouvatut, V., Phongnarisorn, C., & Praserttitipong, D. (2025). Deep learning based detection of endometriosis lesions in laparoscopic images with 5-fold cross-validation. Intelligence-Based Medicine, 11, 100230.
19. Madhu, G., Meher Bonasi, A., Kautish, S., Almazyad, A. S., Mohamed, A. W., Werner, F., Hosseinzadeh, M., & Shokouhifar, M. (2024). UCapsNet: A two-stage deep learning model using U-Net and capsule network for breast cancer segmentation and classification in ultrasound imaging. Cancers, 16(22), 3777. https://doi.org/10.3390/cancers16223777
20. Podda, A. S., Balia, R., Barra, S., et al. (2024). Multi-scale deep learning ensemble for segmentation of endometriotic lesions. Neural Computing and Applications, 36, 14895–14908. https://doi.org/10.1007/s00521-024-09828-2
21. Madhu, G., Meher Bonasi, A., Kautish, S., Almazyad, A. S., Mohamed, A. W., Werner, F., Hosseinzadeh, M., & Shokouhifar, M. (2024). UCapsNet: A two-stage deep learning model using U-Net and capsule network for breast cancer segmentation and classification in ultrasound imaging. Cancers, 16(22), 3777. https://doi.org/10.3390/cancers16223777