RESEARCH ARTICLE

Asian Research Association

INTERNATIONAL RESEARCH JOURNAL OF
MULTIDISCIPLINARY TECHNOVATION

DOI: 10.54392/irjmt2434

An Intelligent Computer Aided Diagnosis System for Classification of Ovarian Masses using Machine Learning Approach

Smital D. Patil [a,*], Pramod J. Deore [a], Vaishali Bhagwat Patil [b]

[a] Department of Electronics and Telecommunication, R.C. Patel Institute of Technology, Shirpur, Maharashtra-425405, India
[b] Department of Computer Science, RCPET's Institute of Management Research, Maharashtra-425405, India
*Corresponding Author Email: smitalpatil55@gmail.com
DOI: https://doi.org/10.54392/irjmt2434
Received: 19-08-2023; Revised: 26-03-2024; Accepted: 07-04-2024; Published: 16-04-2024

Abstract: Ovarian cancer, a difficult and often asymptomatic malignancy, remains a substantial global health concern in women. An ovary is a female reproductive organ, which lies on each side of the uterus and used to store eggs. Computer-aided diagnosis (CAD) is an approach that involves using computer algorithms and machine learning techniques to assist medical professionals in diagnosing ovarian malignancies, benign tumors or Poly-cystic ovaries (PCOS). The need for models that can effectively predict benign ovarian tumors and ovarian cancer has led to the use of machine learning techniques. Our research objective is to propose a machine learning-based system for accurate and early ovarian mass detection utilizing novel annotated ovarian masses. We have used an actual patient database whose input features were extracted from 187 transvaginal ultrasound images from database. The input image is preprocessed using the Block Matching 3D filter. The process involves employing binary and watershed segmentation techniques, followed by the integration of Gabor, Gray-Level Co-Occurrence Matrix (GLCM), Tamura, and edge feature extraction methods. K-Nearest Neighbors (KNN) and Random Forest (RF) are two classifiers used for classification. Based on our results, we are able to demonstrate that binary segmentation with RF classifiers is more accurate (above 86%) than KNN classifiers (under 84%).

Keywords: Computer-Aided Diagnosis (CAD), Transvaginal Ultrasound, Block Matching 3D Filter, Binary Segmentation, Watershed, KNN and RF

1. Introduction

A vital part of the female reproductive system is the ovary. The ovaries are two small, almond-shaped organs located in the pelvic cavity on either side of uterus. In addition to ova, the ovary produces estrogen, progesterone, and other female sex hormones for reproduction. The ovaries are susceptible to various health conditions, including ovarian cysts (fluid-filled sacs in ovary), polycystic ovary syndrome (PCOS), and ovarian cancer. Regular gynecological check-ups and screenings are essential to monitor the health of the ovaries and detect any abnormalities at an early stage. Figure 1 shows normal ovary, ovarian cyst and PCOS.

- Ovarian Cyst: In the ovaries, cysts are fluid-filled sacs which may form within the ovaries or around the ovaries.
- Ovarian Tumor: Ovarian tumors are abnormal growths of tissue within the ovaries.
- Ovarian Mass: An ovarian mass is a general term that encompasses both cysts and tumors found in or on the ovaries.
- Polycystic Ovary (PCO): A hormonal condition known as polycystic ovarian syndrome (PCO) affects the ovaries and causes the development of numerous tiny cysts.

Ovarian cancer is one of the top causes of cancer-related mortality among women, and early detection and effective treatment are crucial to survival. Ovarian cancer is categorized as a gynecological cancer, a prevalent cause of mortality, and the seventh most frequently diagnosed cancer among women. The survival rate for most patients typically ranges between 26 to 42 percent five years post-diagnosis [1]. In recent years, computer vision and machine learning techniques have developed as promising tools for enhancing the accuracy and efficiency of ovarian mass detection. This research aims to investigate the application of Random Forest and KNN classifiers with Binary and Watershed segmentation separately to improve the detection of ovarian masses.

Ovarian cancer pertains to a collection of conditions that arise within the ovaries and disseminate to the fallopian tubes and peritoneum [2].

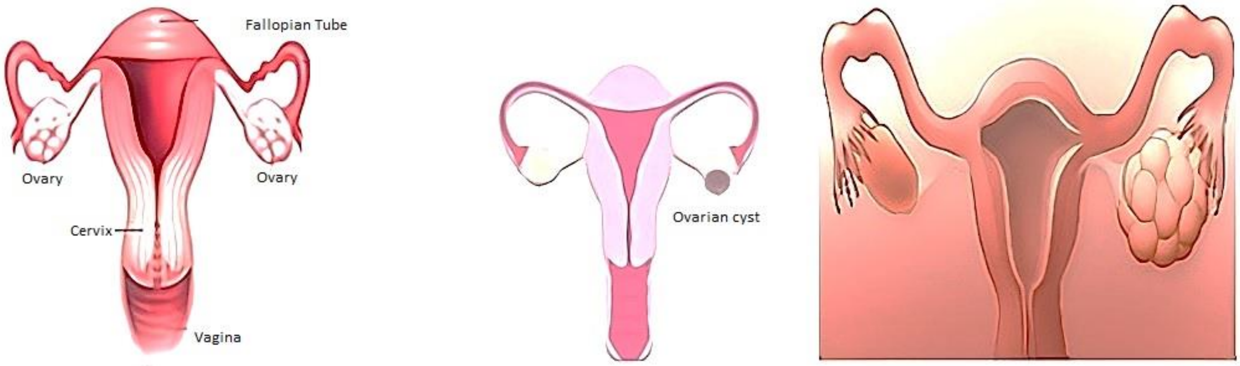

Figure 1. Normal ovary, ovarian cyst and polycystic ovary

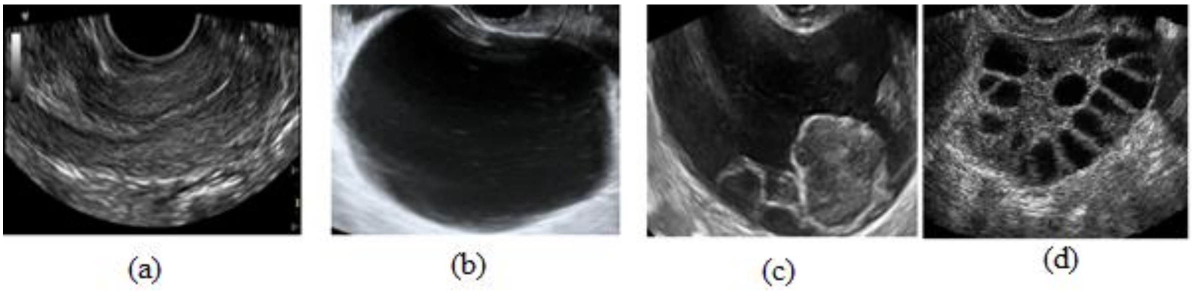

Figure 2. Sample images of (a) Normal ovary (b) benign (c) Malignant (d) PCOD

Masses can be either malignant (cancerous) or benign (non-cancerous). Ovarian masses cannot be detected in their early stages because there are no symptoms. This condition becomes more prevalent as women get older. Having an ovarian mass, general pelvic or abdominal symptoms occur and are typically only recognized during the developed stage, and there is currently no effective screening tool.

According to estimates, there are 204,000 new instances of ovarian mass worldwide each year, which result in about 125,000 fatalities. The best chance for lowering the death rate and providing long-term control over the disease is to find ovarian mass in its earliest stages. Women between the ages of 50 and 70 are mostly affected by it. More than 90% of patients who were diagnosed in the earlier stages survived in the past five years, compared to just around 70% of patients who were detected in stage II. As a result, ovarian mass must essentially be found in its earlier stages. Figure 2 shows sample ultrasound images of normal ovary and various types of cysts.

Compared to traditional abdominal ultrasound, transvaginal ultrasound provides high-resolution images and allows for better pelvic visualization. Various gynecological conditions, such as ovarian cysts, tumors, fibroids, and polycystic ovary syndrome (PCOS) can be assessed by this technique. Hence, we have used transvaginal ultrasound images in our research work for better diagnosis. Therefore, the transvaginal approach is the best. The information gathered by ultrasonography includes ovary size, aberrant ovarian lesions, pelvic or abdominal fluid, and blood flow in the ovary. These results are taken into consideration as diagnostic indicators to find ovarian mass in its early stages. Machine learning algorithms are shown to have the capacity to anticipate complicated diseases. The goal of our research is to assess various well-known Machine Learning (ML) methods that automatically classify ovarian masses into benign, malignant and PCOS.

Dataset: For this research work, we have used 187 benign and malignant tumor datasets of transvaginal ultrasonography images from the source: https://osf.io/n9abq/ [3]. Out of which 112 are benign and 75 are malignant. Polycystic ovary images are obtained from the Kaggle dataset.

2. Related work

Faust *et al.,* [4] review the use of textural features for computer-assisted diagnosis (CAD) systems in ultrasound (US) images of the prostate, breast, liver, thyroid, and ovaries for cancer diagnosis. The usefulness of texture characteristics as a method for extracting diagnostically valuable information from US images is demonstrated in this study.

The machine-learning algorithm is evaluated and the spatial domain algorithm is employed for feature extraction in four stages: LL, HL, LH, and HH. The authors extracted these features to reduce dimensionality and subsequently used an SVM classifier to classify the five different stages of ovarian cancer [5].

The survey conducted by Huang *et al.* [6] provides a comprehensive overview of the recent advancements and machine-learning algorithms that are applied in ultrasound CAD systems. The paper encompasses various topics, including the utilization of different machine learning techniques, feature extraction methods, and the performance evaluation of these systems across various medical applications. The survey discusses a wide range of machine learning algorithms including traditional algorithms like Random Forest, k-nearest Neighbors (k-NN), and Support Vector Machines (SVM) as well as more advanced techniques like Convolutional Neural Networks (CNNs), Deep Learning, and Ensemble methods. The paper also presents the pros and cons of each algorithm, emphasizing their suitability for specific diagnostic tasks.

The study by Kiruthika *et al.* [7] aims to develop an automated ovarian classification system using texture and intensity-based features extracted from medical images. The proposed approach leverages advanced image analysis techniques to classify ovarian masses based on their distinct textural and intensity patterns.

The study by Zhang and Han presents a novel approach for the recognition of ovarian tumors in obstetric ultrasound imaging using an advanced machine learning technique with a logistic regression classifier [8]. By integrating sophisticated feature extraction and machine learning methods, the proposed system exhibits promising results in enhancing the accuracy and efficiency of ovarian tumor detection.

Srivastava *et al.* [9] used a fine-tuned VGG-16 deep-learning network to develop an automated system for detecting ovarian cysts in ultrasound images. The proposed approach leverages transfer learning to enrich the performance of the deep learning model for accurate cyst detection. Using ultrasound scans, this algorithm can assess whether an ovarian cyst is there or not with an accuracy of 92.11%.

A novel method for classifying ovarian cysts using watershed segmentation and contour analysis in ultrasound images is presented by Nabilah *et al*. [10]. The integration of image processing techniques and machine learning classification demonstrates promising results in accurately identifying and distinguishing different types of ovarian cysts.

Image segmentation is a critical step in computer vision and image analysis, aimed at segregating an image into meaningful regions or entities. The watershed transformation algorithm is a popular technique used for image segmentation. Belaid and Mourou represent an in-depth exploration of the watershed transformation algorithm for image segmentation [11]. It discusses the principles, methodology, and applications of the algorithm in various image analysis tasks.

Numerous studies have investigated the significance of feature extraction and selection in PCOS detection. Gabor wavelets, Gray-Level Co-occurrence Matrix (GLCM), and texture analysis have been commonly utilized to extract relevant features from ultrasound images. Authors in have explored novel feature extraction techniques to improve detection accuracy with a classification accuracy of 99.89% [12].

Ravishankar *et al.* [13] introduced OCD-FCNN, an automated system designed for identifying and categorizing ovarian cysts through a Convolutional Neural Network with fuzzy rule-based mechanisms. This system demonstrated an impressive accuracy of 98.37% when tested on standard benchmark datasets.

The use of artificial intelligence (AI) methods in the histopathology of ovarian cancer is examined in [14] evaluating the research's quality and emphasizing the constraints and biases present in existing studies.

Koch *et al.* [15] provided an analysis of computer-aided diagnostics in the preoperative diagnosis of ovarian cancer. The systematic review encompassed 31 studies. Computer-aided diagnosis (CAD) utilizing ultrasound, CT, and MRI exhibited encouraging outcomes. Sensitivities varied, with ultrasound ranging from 40.3% to 100%, CT from 84.6% to 100%, and MRI from 66.7% to 100%. Specificities also showed variability, with ultrasound ranging from 76.3% to 100%, CT from 69% to 100%, and MRI from 77.8% to 100%.

Singh *et al.* [16] introduced a methodology known as the Error-Guided Artificial Bee Colony (EABC) algorithm designed for the training of neural networks. A comparative analysis was conducted contrasting EABC with existing algorithms such as Artificial Bee Colony (ABC), Genetic Algorithm (GA), Backpropagation, and Particle Swarm Optimization (PSO) in terms of their performance based on mean square error metrics. The primary objective of these training algorithms is to address the challenges associated with slow training processes and the occurrence of over-fitting in neural networks.

Hiremath *et al.* [17] introduced a technique for identifying follicles through the utilization of geometric characteristics, showcasing its effectiveness. The experimental outcomes were compared with manual inferences carried out by healthcare professionals. Recognition of follicles was done using geometric features extracted from ultrasound images and were classified based on area, ratio, compactness, extent, and centroid of follicles.

T. Saba [18] conducted a comprehensive examination of machine-assisted methodologies for cancer detection across diverse cancer types and compared state-of-the-art techniques on benchmark datasets for different cancers. The author also highlighted the limitations of existing techniques in cancer detection and classification.

Raveendran *et al.* [19] focused on identifying boundaries of domains in spatial data analysis and proposed binary segmentation algorithms for clustering homogeneous sub-regions. Extensive literature on image recognition and boundary identification in statistical imaging was done.

Noha [20] evaluated contemporary approaches and contributions to brain tumor segmentation research and described several image segmentation techniques used by different researchers. Deep learning, texture features, and kernel sparse coding are all included in the study.

The use of machine learning (ML) in ultrasound imaging for detection, segmentation, and classification in non-destructive evaluation (NDE) and medical diagnostics was studied by Monica *et al.* [21]. Techniques for deep learning, elastography, K-means clustering, and unsupervised learning were studied.

Byale *et al.* [22] introduced a framework that employs machine learning techniques to enhance the accuracy of classification. The approach encompasses initial data processing focused on noise reduction through adaptive median filtering, segmentation via a Gaussian Mixture Model (GMM) to pinpoint the region of interest, feature extraction utilizing the Grey Level Co-occurrence Matrix GLCM to extract the characteristics of diverse tumor types, and classification through Neural Networks (NN) for the confirmation and classification of tumors as either benign or malignant. The proposed framework is evaluated against traditional machine learning algorithms like Adaboost (Adaptive Boosting) that assign the image to distinct classes (Normal, Benign, Malignant).

Yinhui Deng *et al*. [23] emphases on the automated identification of Polycystic Ovary Syndrome from ultrasound images. Their study introduces a method for PCOS detection utilizing an adaptive morphological filter and underscores the necessity for automated detection in light of the challenges associated with manual counting issues.

Patil *et al.* [24] performed a study that entailed applying different denoising filters to transvaginal ultrasound images showcasing ovarian tumors. They proceeded to assess the efficacy of different denoising techniques using performance measures such as mean square error (MSE), peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Universal image quality index (UQI). The results of the experiments indicated that the block-matching 3-D filter surpassed all other methodologies in terms of performance.

3. Research Gap

Ovarian cancer ranks third among all cancer types worldwide and is the fifth most often caused malignancy. Early ovarian cancer detection does not yield reliable outcomes. Transvaginal ultrasonography is a standard screening method that aids in tumor detection, however patients frequently experience adnexal lumps. It's difficult to categorize the masses as benign or malignant. A current development in medical imaging is the ability to identify tumors in their earliest stages. Ovarian malignancies are not, however, correctly diagnosed. Using ultrasound image is possible to make a preliminary diagnosis of ovarian cyst as benign or malignant without undergoing surgery.

- Current diagnostic methods often detect ovarian masses at later stages, limiting treatment options and reducing the chances of successful intervention. There is a need for innovative approaches and advanced imaging techniques to enable early and accurate detection of ovarian masses.
- Variations in imaging protocols, equipment, and operator expertise can lead to differences in image quality and interpretations.
- Ovarian mass detection systems should aim to provide real-time decision support to radiologists and clinicians during image interpretation. Integrating CAD systems into the clinical workflow to assist in image analysis and diagnosis can enhance efficiency and accuracy in ovarian mass detection.

4. Contribution

- To develop an algorithm for the detection of tumors using annotated ovarian masses for early cancer detection.
- To use feature extraction, segmentation and classification methods for classification of ovarian masses and improve recognition and accuracy rate.
- To show the usefulness of the suggested approach, a thorough comparative study is done with a variety of performance measures for both segmentation and classification.

The paper is organized in following sections. Section V describes the proposed method and its algorithm. Performance evaluation is provided in Section VI, result analysis is explained in Section VII, and an assessment of performance values is provided in Section VIII.

5. Proposed Methods

The primary objective of our research is to improve, segment, extract features, and classify ovarian masses into different categories like benign, malignant and PCOD. It consists of four main steps: pre-processing (denoising), segmentation, feature extraction, and classification. Firstly, the transvaginal ultrasound images of ovarian masses undergo pre-processing using the BM3D filter. Following pre-processing, separate segmentation algorithms, namely Binary segmentation and Watershed segmentation, are applied. From the segmented images, features are extracted utilizing GLCM, Tamura, Gabor, and Edge feature extraction methods. Finally, the images are categorized using both the Random Forest (RF) classifier and k-Nearest Neighbors (KNN) classifier. Figure 3 gives detail explanation of proposed work.

5.1 Block Matching-3D Filter (BM3D)

The primary objective of the pre-processing is to enhance the image's pixel values while also getting rid of any unwanted noise and background information. Frequency domain and spatial domain concepts are applied in the BM3D filtering algorithms to improve the ovary images

Block Matching 3D (BM3D) is a widely used image denoising algorithm that is effective in reducing noise and enhancing the quality of images. It was introduced as an extension of the Block Matching and 3D filtering (BM3D) algorithm, which was originally designed for video denoising. The BM3D algorithm has since been adapted for image denoising and has gained popularity due to its excellent performance.

The main principle behind the BM3D filter is to exploit the self-similarities present in natural images. In natural images, similar patterns and structures often repeat throughout the image, and these repetitions can be leveraged to better estimate and reduce noise. The BM3D algorithm works in two main steps: collaborative filtering and aggregation.

5.1.1. Collaborative Filtering

In the first step, the image is divided into small overlapping blocks, and similar blocks are grouped together based on their content. A similarity measure, such as the mean squared error (MSE) or the structural similarity index (SSIM), is used to find similar blocks. These similar blocks form a collaborative group, and the denoising is performed within this group. By considering the similarities between blocks, BM3D effectively captures the inherent structure of the image.

5.1.2. Aggregation

In the next step, the denoising process is carried out within each collaborative group. A 3D transform, like a 3D discrete wavelet transform (DWT) and a 3D discrete cosine transform (DCT) is applied to the grouped blocks to convert them into a transformed domain. In this domain, the noise is often represented by sparse coefficients, while the signal (non-noise) coefficients are dense.

5.2 Segmentation Techniques

Segmentation involves the separation of an image into distinct regions of interest, enabling the identification and isolation of specific structures or abnormalities, such as tumors.

5.2.1. Binary Segmentation

Using binary segmentation, an image can be divided into two distinct regions based on certain criteria. The primary objective of binary segmentation is to categorize each pixel in the image as belonging to either the foreground or the background**.** It involves selecting a threshold value that separates the foreground and background pixels based on a specific image feature, such as intensity, color, or texture. Pixels whose feature values exceed the threshold are classified as foreground, and pixels below the threshold are classified as background. Figure 4 shows segmented image using binary segmentation.

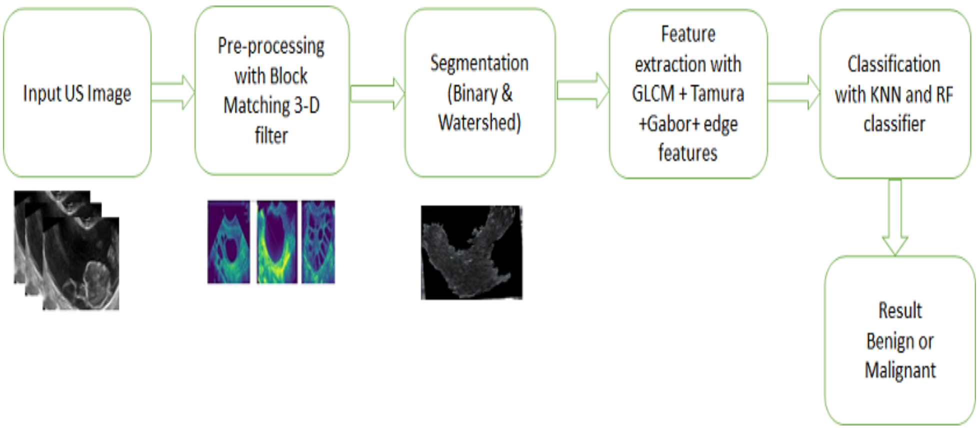

Figure 3. Proposed flow diagram

Figure 4. Binary Segmented Image

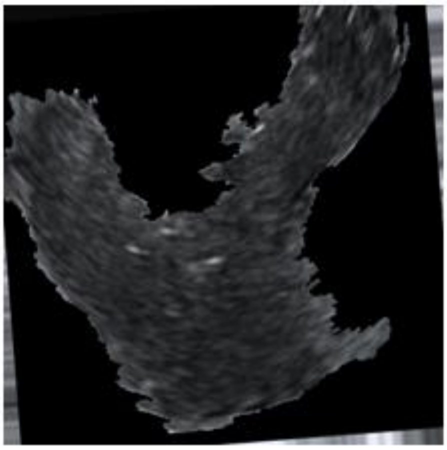

Figure 5. Watershed segmented image

5.2.2. Watershed Segmentation

Because standard binary segmentation and contour detection will not produce accurate results, complex images are segmented using the watershed technique. The watershed transform is based on the concept of treating an image as a topographic surface, where intensity values represent elevations. The process simulates a flooding scenario, where water is poured into the regional minima of the image and gradually rises to form watershed lines along ridges. These watershed lines separate different regions or objects in the image. Accurate background and foreground data collecting forms the basis of the watershed approach. After that, the precise boundaries are established by running markers through the watershed. Watershed segmentation is particularly useful for segmenting objects or regions with irregular boundaries or when multiple objects are close together. The basic principle of watershed segmentation can be summarized in the following steps:

i. Gradient Computation: The first step is to compute the gradient of the input image. The gradient represents the intensity variations or edges in the image and is crucial for identifying potential boundaries.

ii. Markers Initialization: Watershed segmentation requires the placement of markers on the image. These markers act as seeds for region growing and play a significant role in the segmentation process. Markers can be manually specified by users or generated automatically using techniques like morphological operations or thresholding.

iii. Region Growing: Starting from the markers, the watershed algorithm performs a region growing process to expand the segmented regions. The algorithm simulates the flooding of a terrain, where water rises from the markers until it reaches the boundaries or ridges between different regions.

iv. Watershed Lines: As the region growing process continues, the algorithm forms watershed lines along the boundaries where water from different regions meets. These watershed lines act as the final segmentation boundaries. Figure 5 shows segmented image using watershed segmentation.

5.3. Feature Extraction Techniques

Feature extraction is a pivotal step in medical image processing, enabling the conversion of complex image data into meaningful and actionable information. An important aspect of feature extraction is that it generates numerical features from raw data while keeping the original data set's information intact. These features act as the foundation for subsequent stages like disease diagnosis and tumor segmentation. We have used following methods for extracting features from raw ultrasound segmented images from ovarian masses.

5.3.1 GLCM Feature Extraction

Gray-Level Co-occurrence Matrix (GLCM) is a powerful texture analysis technique widely used in medical image processing. GLCM captures spatial relationships between pixel intensities in an image and provides valuable information about the texture properties of tissues. In the context of ovarian tumor detection, GLCM can reveal subtle textural differences that are indicative of tumor presence. A GLCM is computed from a grayscale image by analyzing the spatial relationship between pairs of intensity values in a predetermined direction. (e.g., horizontal, vertical, diagonal). Each pixel's occurrence of intensity pairs is tabulated in a matrix. Each element (i, j) represents the frequency of intensity pairs occurring at a particular distance and angle. From the GLCM, various statistical measures like contrast, energy, entropy and correlation are extracted as features that characterize the texture of the image region. These features are used to quantify the relationships between pixel intensities.

5.3.2 Gabor Features

The Gabor filter captures both frequency and orientation information in an image, making it ideal for highlighting specific texture patterns. Different textures can be captured using Gabor filters by analyzing spatial frequency content from various perspectives. The process of Gabor feature extraction involves convolving an image with a bank of Gabor filters. Each filter responds to a specific frequency and orientation combination, highlighting patterns that match the filter's properties. From the convolution results, various statistical measures or features can be extracted to characterize the texture patterns highlighted by the Gabor filters. Mean, variance, energy and entropy are some common Gabor features.

5.3.3 Edge Features

Edges represent abrupt changes in intensity and are often indicative of important structures and boundaries within an image. In ovarian tumor detection, edge features can help delineate tumor boundaries and provide valuable information about the shape and structure of the tumor. Edge detection techniques are used to identify and highlight regions of significant intensity variation in an image. We have used various edge detection methods like sobel operator, Prewitt operator, Canny edge detector etc.

5.3.4 Tamura Features

Tamura features focus on capturing various aspects of texture perception, such as roughness, coarseness, and contrast. Based on human perceptions of texture, Tamura proposed a set of features to describe texture.

After computing and combining all the features from ultrasound images of ovarian masses, these features are given as input for classification algorithms to differentiate between benign and malignant tumors.

5.4 Classification

The process of classification involves assigning labels or classifications to ovarian tumors based on features.

5.4.1 K-Nearest Neighbor Classifier

KNN is an instance-based classifier that identifies new data points by classifying them according to the k-nearest neighbor labels in the feature space. In KNN, the parameter 'k' indicates the number of neighbors that are taken into account when classifying a given data point. The choice of 'k' is critical, as a smaller 'k' can be sensitive to noise, while a larger 'k' can lead to over-smoothing. Cross-validation techniques are used to determine the optimal value of 'k' for the given dataset.

The following steps are followed when a new image is presented to the KNN classifier for classification-

1. Euclidean distance is calculated as the difference between the feature vector of the new image and the feature vector of all the images in your training set.
2. Select the 'k' nearest neighbors with the smallest distances.
3. Allocate the class label to the new image based on the majority class among the 'k' neighbors.

5.4.2. Random Forest Classifier

An ensemble learning technique named Random Forest classifier builds several decision trees and aggregates their predictions to increase accuracy, robustness, and generalization. The random forest is generated by combining N decision trees, and then predicting each tree produced in phase one. When a new image is presented to the Random Forest classifier for classification, each decision tree in the forest independently predicts the class label based on the features of the new image. The class that receives the most votes or has the highest average among all decision trees is assigned as the final prediction.

6. Performance Measures

6.1 Mean Square Error (MSE)

MSE quantifies the similarity between two images by calculating the average of the squared pixel intensity differences between corresponding pixels in the images. The Mean Squares Error must be determined as per Equation (1).

$$MSE = \sum_{M_i N_i} \frac{[I(m_1,n_1) - I(m_2,n_2)]^2}{M_1 \times N_1} \tag{1}$$

where M1 and N1 are the input image's M1 and N1 corresponding row and column counts.

6.2 Peak Signal to Noise Ratio (PSNR)

PSNR provides a quantitative assessment of how well the processed image preserves the quality of the original image. Equation (2) provides the PSNR value for the input image with the highest degree of variation.

$$PSNR = 10 \log_{10}\left(\frac{R1^2}{MSE}\right) \tag{2}$$

6.3 Root Mean Square Error

It provides a measure of the overall discrepancy between a reference (original) image and a processed or reconstructed image. RMSE is a variation of the Mean Square Error (MSE) and is particularly useful for evaluating the quality of image reconstruction, denoising, or restoration processes. The equation (3) yields the Root Mean Square Error.

$$RMSE = \sqrt{MSE} \tag{3}$$

6.4 Accuracy

Out of all the instances in the dataset, it calculates the percentage of correctly identified instances (or pixels in the case of image segmentation). In image processing, accuracy is often employed to assess the success of classifying pixels or regions within an image.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \tag{4}$$

where, respectively, True Positive, True Negative, False Positive, and False Negative are given as TP, TN, FP, FN.

6.5 Sensitivity (Recall)

It is the proportion of the actual malignant territory that the categorization system has successfully identified. It used to calculate the performance of classification and segmentation tasks. It measures the percentage of actual positive cases that are correctly identified by a classifier or segmentation algorithm. Equation (5) provides the recall or sensitivity measure of different cases of malignancies.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN+TP}} \times 100\% \tag{5}$$

6.6 Specificity

It is the proportion of the malignant region's true background that the categorization algorithm has successfully identified. It measures the percentage of actual negative instances that are correctly identified as negative by a classifier or segmentation algorithm. Equation (6) gives specificity as a performance measure for correctly identifying negative instances.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \times 100\% \tag{6}$$

6.7 Precision

It reveals what proportion of the detected area actually is the region. In this case, it is the percentage of instances correctly predicted as positive (true positives) among all instances in the database where the classifier or segmentation algorithm predicted them to be positive. Equation (7) gives precision as a measure of correct predictions done.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \text{X}100\% \tag{7}$$

7. Experiment and result analysis

7.1 Image Denoising Results

Table 1 gives comparative results of all image denoising techniques. Out of various denoising techniques applied, Block Matching 3D filter gives better performance.

Figure 6 shows the output images of various techniques after denoising. Figure 7 shows graphical performance analysis of all denoising techniques.

7.2 Image Segmentation, Feature Extraction and Classification Results

Table 2 gives performance analysis of binary and watershed segmentation techniques with KNN and RF classifiers on various types of ovarian images.
The outcome of our research shows that when employing binary segmentation and RF classifiers, total accuracy is higher than with conventional classifiers.

Figures 8 and 9 give comparative analysis of Binary and watershed segmentation with KNN and RF classifiers. The proposed study effort is compared to the currently utilized approaches with the classification accuracy in Table 3. Figure 10 shows performance analysis of overall accuracy.

Table 1. Comparative Analysis of Various Denoising Techniques

Sr.No.	Filtering Methods	PSNR	MSE	SSIM	RMSE
	Ideal Values	**INF**	**0**	**1**	**0**
1	Gaussian Filter	24.79	54.31	0.701	0.1837
2	Median filter	23.52	53.91	0.61	0.2127
3	Gabor Filter	4.35	93.36	0.07	1.844
4	bilateral Filter	32.24	39.69	0.872	0.0913
5	Non-local means filter	33.3	22.08	0.906	0.0673
6	**BM3D Filter**	**71.88**	**0**	**0.999**	**0.001**

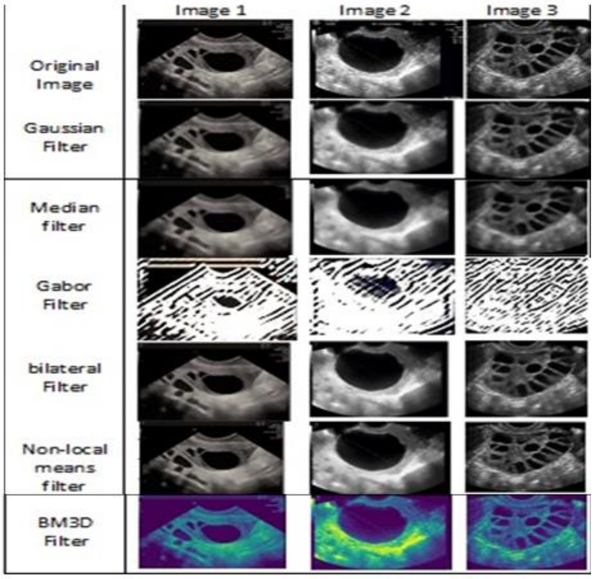

Figure.6. Output of filtering Techniques

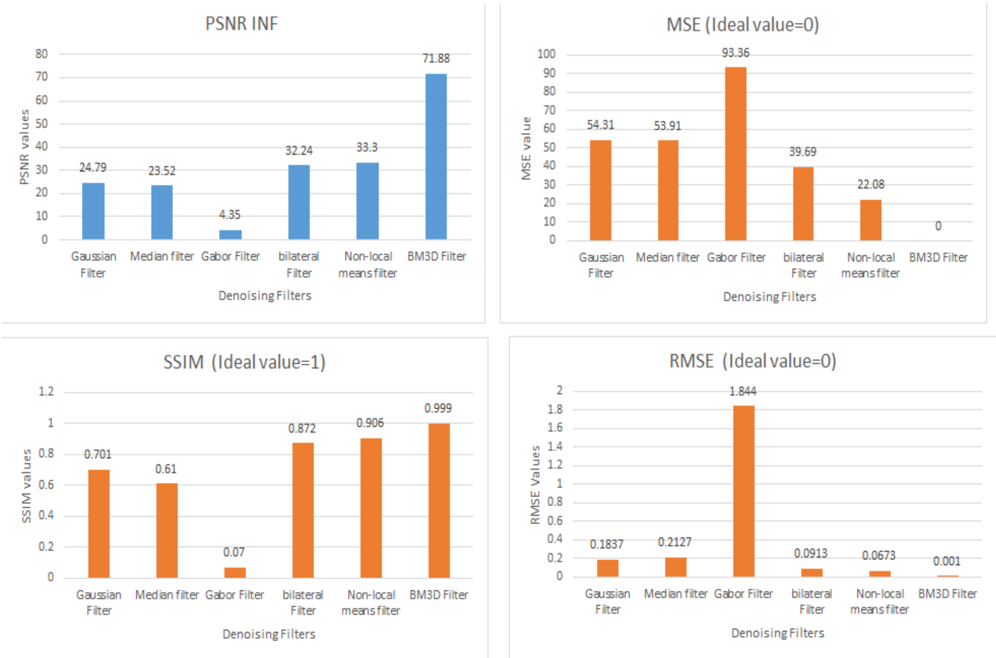

Figure 7. Performance analysis of all filters

Table 2. Analysis of segmentation and classification performance value

Sr. no	Type	Performance Measure	Binary Segmentation		Watershed Segmentation	
			KNN Classifier	RF Classifier	KNN Classifier	RF Classifier
1	Malignant	Precision	0.67	0.84	0.86	0.84
	Benign		0.88	0.79	0.85	0.78
	PCOS		0.72	0.94	0.68	0.93
2	Malignant	Recall	0.7	0.8	0.72	0.83
	Benign		0.63	0.91	0.73	0.85
	PCOS		0.1	0.9	0.1	0.9
3	Malignant	F1-score	0.68	0.82	0.79	0.83
	Benign		0.74	0.85	0.79	0.82
	PCOS		0.84	0.92	0.81	0.91
4	Overall Accuracy		0.75 (75%)	0.86 (86%)	0.79(79%)	0.85(85%)

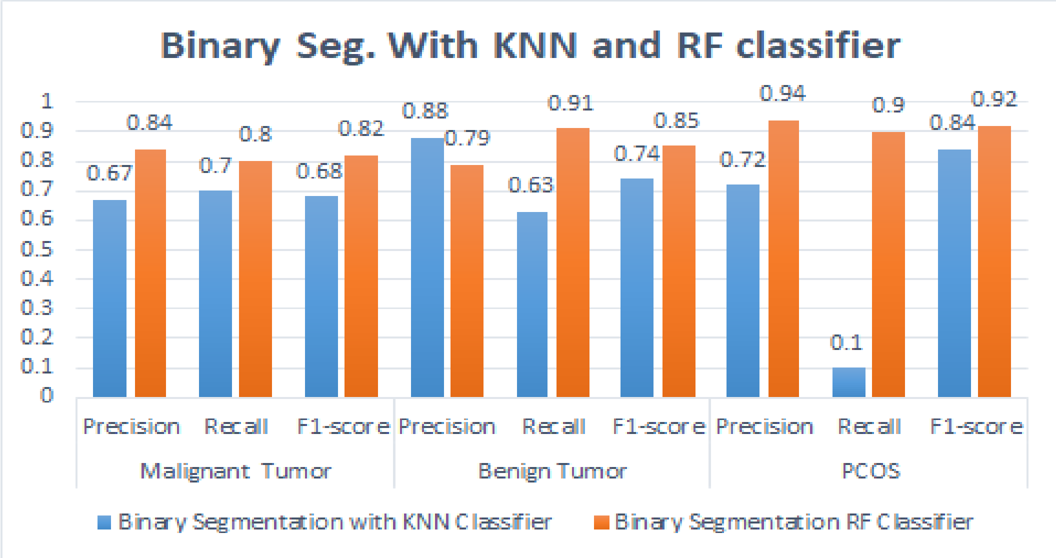

Figure 8. Comparative analysis of Binary segmentation with KNN and RF classifier

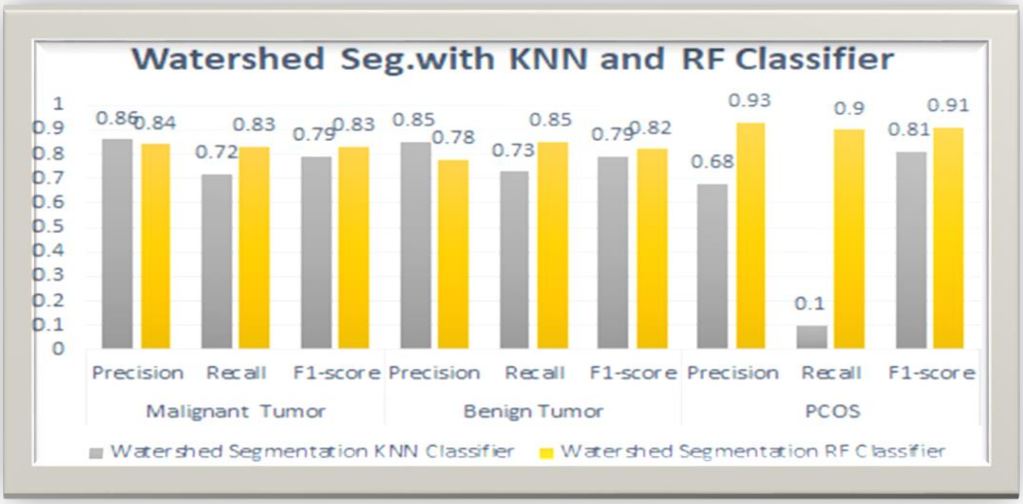

Figure 9. Comparative analysis of Watershed segmentation with KNN and RF classifier

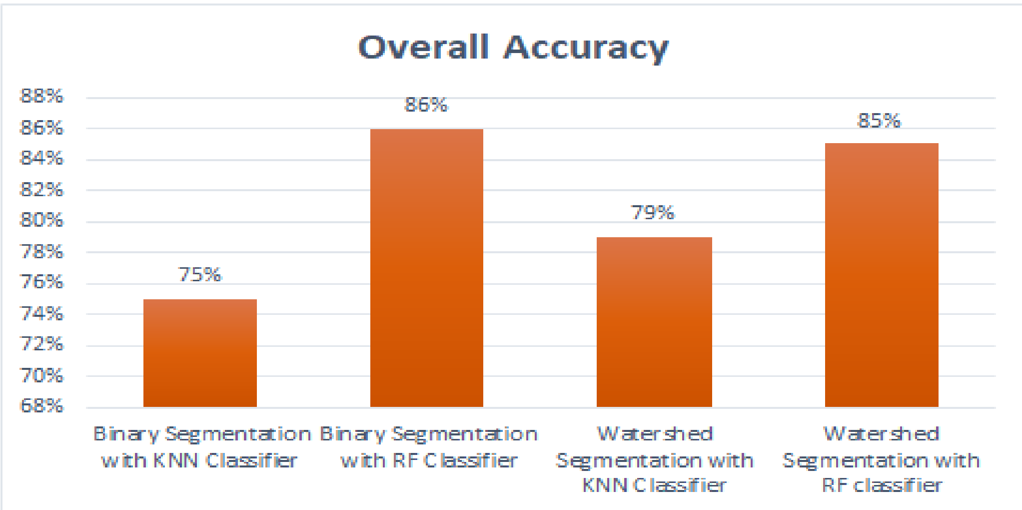

Figure 10. Performance analysis of accuracy

Table 3. Comparison of different methods and their results

Study	Method	Accuracy
Martínez- Más *et.al*.[3]	K-Nearest Neighbors (KNN), Linear Discriminant (LD), Support Vector Machine (SVM) and Extreme Learning Machine (ELM).	60% for KNN More than 85% for LD,SVM and ELM
Singh *et tal* [16]	Error Guided Artificial Bee Colony	83.3%
Hiremath and Tegnoor [17]	By combining horizontal and vertical scanline thresholding region of interest (ROI) Deng	82%
Proposed Method	**Binary segmentation and RF classifier**	86% MSE=0

To the best of our knowledge, the research effort that is being offered, which uses binary segmentation and the RF classifier approach is not being used and yields a better outcome with an accuracy of 86%. Table 3 compares the proposed research work to existing methodologies.

8. Conclusion

The application of computer-aided detection using binary segmentation and random forest technology holds significant promise in diagnosis of ovarian masses. The combination of these advanced techniques offers a robust framework for accurately identifying and delineating regions of interest within transvaginal ultrasound images of ovarian masses thereby assisting healthcare professionals in making informed decisions. When compared to the other two algorithms, Binary segmentation and RF classifier (86%) offer the best results, according to the performance analysis's final findings. To the best of our knowledge, we have used the combination of binary segmentation with Random Forest classifier for the first time with an accuracy of 86%. The ability of this CAD system to analyse ultrasound images and provide accurate classifications can aid healthcare professionals in making informed decisions and facilitating prompt patient management.

9. Future Goal

The results of our research show that the Binary and Watershed segmentation method can generate precise segmentation results that can be used to differentiate the texture of ultrasound images. Imaging modalities like MRI and CT scans can be combined with ultrasound images to provide more information about ovarian masses, potentially improving classification accuracy. Deep learning can also enhance the accuracy of ovarian mass classification.

References

[1] F.H. Juwono, W.K. Wong, H.T. Pek, S. Sivakumar, & D.D. Acula, Ovarian cancer detection using optimized machine learning models with adaptive differential evolution. Biomedical Signal Processing and Control, 77, (2022) 103785. https://doi.org/10.1016/j.bspc.2022.103785

[2] N. Taleb, S. Mehmood, M. Zubair, I. Naseer, B. Mago, M.U. Nasir, (2022) Ovary Cancer Diagnosing Empowered with Machine Learning. International Conference on Business Analytics for Technology and Security (ICBATS), IEEE, Dubai, https://doi.org/10.1109/ICBATS54253.2022.9759010

[3] J. Martínez-Más, A. Bueno-Crespo, S. Khazendar, M. Remezal-Solano, J.P. Martínez-Cendán, S. Jassim, H. Du, H.A. Assam, T. Bourne, D. Timmerman, Evaluation of machine learning methods with Fourier Transform features for classifying ovarian tumors based on ultrasound images. PLoS One, 14(7), (2019) e0219388. https://doi.org/10.1371/journal.pone.0219388

[4] O. Faust, U.R. Acharya, K.M. Meiburger, F. Molinari, J.E. Koh, C.H. Yeong, P. Kongmebhol, K. H. Ng, Comparative assessment of texture features for the identification of cancer in ultrasound images: a review. Biocybernetics and Biomedical Engineering, 38(2), (2018) 275-296. https://doi.org/10.1016/j.bbe.2018.01.001

[5] S. Rashid, R. Kaur, Automatic detection of ovarian cancer based on improved DWT transformation. International Journal of Engineering & Technology, 7(2.27), (2018) 104-108. https://doi.org/10.14419/ijet.v7i2.27.12575

[6] Q. Huang, F. Zhang, X. Li, Machine learning in ultrasound computer-aided diagnostic systems: a survey. BioMed research international, 2018, (2018). https://doi.org/10.1155/2018/5137904

[7] V. Kiruthika, S. Sathiya, M.M. Ramya, Automatic texture and intensity based ovarian classification. Journal of medical engineering & technology, 42(8), (2018) 604-616. https://doi.org/10.1080/03091902.2019.1588407

[8] Z. Zhang, Y. Han, Detection of ovarian tumors in obstetric ultrasound imaging using logistic regression classifier with an advanced machine learning approach. IEEE Access, 8, (2020) 44999-45008. https://doi.org/10.1109/ACCESS.2020.2977962

[9] S. Srivastava, P. Kumar, V. Chaudhry, & A. Singh, Detection of ovarian cyst in ultrasound images using fine-tuned VGG-16 deep learning network. SN Computer Science, 1, (2020) 1-8. https://doi.org/10.1007/s42979-020-0109-6

[10] A. Nabilah, R. Sigit, T. Harsono, A. Anwar, (2020). Classification of ovarian cysts on ultrasound images using watershed segmentation and contour analysis. In 2020 International Electronics Symposium (IES) IEEE, Indonesia. https://doi.org/10.1109/IES50839.2020.9231695

[11] L.J. Belaid, W. Mourou, Image segmentation: a watershed transformation algorithm. Image Analysis & Stereology, 28(2), (2009) 93-102. https://doi.org/10.5566/ias.v28

[12] S.A. Suha, & M.N. Islam, An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Scientific Reports, 12(1), (2022) 17123. https://doi.org/10.1038/s41598-022-21724-0

[13] T.N. Ravishankar, H.M. Jadhav, N.S. Kumar, & S. Ambala, N. Muthuvairavan Pillai, A deep learning approach for ovarian cysts detection and classification (OCD- FCNN) using fuzzy convolutional neural network. Measurement: Sensors, 27, (2023) 100797. https://doi.org/10.1016/j.measen.2023.100797

[14] J. Breen, K. Allen, K. Zucker, P. Adusumilli, A. Scarsbrook, G. Hall, N.M. Orsi, N. Ravikumar, Artificial intelligence in ovarian cancer histopathology: a systematic review. NPJ Precision Oncology, 7(1), (2023) 83. https://doi.org/10.1038/s41698-023-00432-6

[15] A.H. Koch, L.S. Jeelof, C.L.P. Muntinga, T.A. Gootzen, N. M.A. van de Kruis, J. Nederend, T. Boers, F. van der Sommen & J.M.J. Piek, Analysis of computer-aided diagnostics in the preoperative diagnosis of ovarian cancer: a systematic review. Insights into Imaging, 14(1), (2023) 34. https://doi.org/10.1186/s13244-022-01345-x

[16] A. Singh, D. Kumar, (2017) Novel ABC based training algorithm for ovarian cancer detection using neural network. In 2017 International Conference on Trends in Electronics and Informatics (ICEI), IEEE, India. https://doi.org/10.1109/ICOEI.2017.8300771

[17] P.S. Hiremath, Jyothi R. Tegnoor, (2009) Recognition of follicles in ultrasound images of ovaries using geometric features. International Conference on Biomedical and Pharmaceutical Engineering, IEEE, Singapore. https://doi.org/10.1109/ICBPE.2009.5384097

[18] T. Saba, Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenge. Journal of Infection and Public Health, 13(9), (2020) 1274-1289. https://doi.org/10.1016/j.jiph.2020.06.033

[19] N. Raveendran, G. Sofronov, Binary segmentation methods for identifying boundaries of spatial domains. Communication papers of the Federated Conference on Computer Science and Information Systems, 13, (2017) 95–102.

[20] Noha A. El-Hag, Literature Review of Segmentation and Classification of Medical Images. Egyptian Journal of Artificial Intelligence, 1(1), (2022). https://doi.org/10.21608/ejai.2022.142684.1005

[21] M. Monica, A. Iula. Recent advances in machine learning applied to ultrasound imaging. Electronics, 11(11), (2022) 1800. https://doi.org/10.3390/electronics11111800

[22] H. Byale, G.M. Lingaraju, S. Sivasubramanian, Automatic segmentation and classification of brain tumor using machine-learning techniques. International Journal of Applied Engineering Research, 13(14), (2018) 11686-11692.

[23] Y. Deng, Y. Wang, P. Chen, (2008) Automated detection of polycystic ovary syndrome from ultrasound images. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, Canada. https://doi.org/10.1109/IEMBS.2008.4650280

[24] S.D. Patil, & P.J. Deore, Machine Learning Approach for Comparative Analysis of De-Noising Techniques in Ultrasound Images of Ovarian Tumors. International Journal on Recent and Innovation Trends in Computing and Communication, 11, (2023) 230-236. https://doi.org/10.17762/ijritcc.v11i2s.6087

Acknowledgement

We would like to extend our sincerest thanks to Dr. Vivek Mahajan, Radiologist, Avani Diagnostic Centre, Shirpur for facilitating discussions and for performing manual segmentation of Ovarian Masses, PCOS and also classification in ultrasound images of Ovaries.

Authors Contribution Statement

Smital D. Patil - Conceptualization, Methodology, Experimental result finding, Implementation and writing - Original Draft. Pramod J. Deore- Database collection, Investigation, Experimental result finding, Writing - Review & Editing. Vaishali Bhagwat Patil- Formal Analysis, Visualization, paper formatting, Supervision. All the authors read and approved the final version of the manuscript.

Funding sources

The authors declare that no funds, grants, or other support were received during this work.

Conflict of Interest:

The authors report there are no competing interests to declare.

Availability of data:

The data that support the findings of this study are openly available in at https://osf.io/n9abq/.

Has this article screened for similarity?

Yes

About the License