PATHOLOGY DETECTION OF PCOS AND ITS SEVERITY GRADING USING MACHINE LEARNING ALGORITHM

Suruthi M[1], Thendralmani J V[2], Rajalakshmi R [3]

[1] *Student, Electrical & Electronics Engineering,Bannari Amman Institute of Technology, Tamilnadu,India*
[2] *Student, Electrical & Electronics Engineering,Bannari Amman Institute of Technology, Tamilnadu,India*
[3] *Student, Electrical & Electronics Engineering,Bannari Amman Institute of Technology, Tamilnadu,India*

ABSTRACT

Polycystic ovarian syndrome (PCOS) stands as a significant threat among gynaecological conditions due to its tendency to manifest with subtle symptoms, often leading to diagnosis at advanced stages. Distinguishing between various types of ovarian cysts poses a formidable challenge for medical professionals. Among imaging techniques, ultrasound (US) imaging emerges as the preferred choice owing to its convenience, non-invasiveness, and real-time capabilities. However, the current screening methods for ovarian cysts via imaging still suffer from limitations, contributing to the poor prognosis associated with ovarian cysts. In recent years, the integration of deep learning models with US images has shown promising results in enhancing diagnostic efficiency, reducing mortality rates, and minimizing diagnostic delays. This project introduces an innovative approach for diagnosing ovarian cysts using US images, employing a Deep Convolutional Neural Network (DCNN) model enhanced with a Generative Adversarial Network (GAN) to address overfitting issues by augmenting training samples. Through this augmentation process, a more robust dataset is created, enabling the DCNN model to effectively classify different types of ovarian cysts. The proposed system not only aids in accurate diagnosis but also serves as a valuable tool for physicians in medical decision-making. By analysing the outcomes generated by the fused DCNN model, medical professionals can gain valuable insights into the nature of ovarian cysts and make informed treatment decisions.

The results obtained from this study demonstrate the superior precision and accuracy of the proposed model in diagnosing ovarian cancer and other types of cysts. This advancement marks a significant stride towards improving the management and prognosis of ovarian cysts, potentially saving lives through early detection and intervention.

Keyword: *Deep Convolutional Neural Network (DCNN), a Generative Adversarial Network (GAN), ultrasound (US), imaging techniques.*

1. INTRODUCTION

Polycystic ovary syndrome (PCOS) is widely recognized as one of the prevalent endocrine disorders affecting women during their reproductive years. Estimates suggest its prevalence ranges from 8–13% among all women and 6–18% among younger females. PCOS often leads to infertility in approximately 70% of diagnosed cases, primarily due to the development of multiple cysts within the ovaries, disrupting the normal ovulation process.

Accurate diagnosis of ovarian disorders, including monitoring follicle development and detecting cysts and PCOS, is imperative for effective management of infertility. Diagnostic tools like abdominal or intravaginal scans are commonly utilized to provide detailed images of ovarian follicles and cysts. However, interpreting these images presents challenges for physicians as distinguishing between follicular and non-follicular areas, as well as cystic and non-cystic regions, can be complex. Additionally, identifying interfollicular regions can be difficult, complicating analysis further. The similarity in appearance between normal ovarian follicles and cysts adds to the complexity, necessitating reliance on medical professionals' expertise for accurate interpretation. Manual detection and analysis may introduce errors, underscoring the need for intelligent automated detection and classification algorithms for ovarian abnormalities.

Deep Learning, an evolving technology, holds promise in addressing these challenges across various domains, including healthcare. By leveraging deep learning, healthcare professionals and researchers can extract insights from extensive datasets, enhancing medical diagnostics and decision-making processes. In the case of PCOS, which lacks effective diagnosis and treatment options, deep learning techniques offer potential solutions by enabling precise analysis of ovarian abnormalities, ultimately improving patient care and management strategies.The diagnosis of polycystic ovary syndrome (PCOS) poses considerable challenges, but the emergence of deep learning technology offers promising solutions in various fields. Firstly, in medical diagnosis, intelligent automated detection and classification algorithms driven by deep learning can accurately pinpoint ovarian abnormalities, assisting healthcare providers in effectively monitoring and treating PCOS patients. Additionally, solutions for managing infertility can combine advanced imaging methods with deep learning algorithms to detect ovarian cysts promptly, facilitating timely interventions and personalized treatment strategies tailored to individual requirements.

However the integration of telemedicine and deep learning offers a groundbreaking approach to diagnosing and managing ovarian disorders. By enhancing accessibility, empowering early detection, and unlocking the potential of data-driven research, this synergy has the power to revolutionize healthcare for women around the world.

1.1 Objectives

The primary objective of this study is to advance the field of ultrasound image analysis through the utilization of Generative Adversarial Network (GAN) technology, with a specific focus on discriminating between benign and malignant ovarian cysts. By harnessing the power of GANs, we aim to develop a sophisticated computerized system capable of accurately differentiating between these two types of cysts, thereby enhancing diagnostic accuracy and clinical decision-making in the management of ovarian cysts. Additionally, we seek to develop a deep convolutional neural network (DCNN) model designed to automate the evaluation of ultrasound images related to ovarian cysts. This model aims to improve upon existing methods by providing a more efficient and precise means of diagnosing ovarian cysts. By leveraging deep learning algorithms and cutting-edge image analysis techniques, our goal is to facilitate a more accurate and timely diagnosis of ovarian cysts, ultimately leading to improved patient outcomes and enhanced clinical care in the realm of gynaecological health.

1.2 Problem Identification

The problem revolves around refining diagnostic precision and clinical decision- making concerning ovarian cyst management, with a specific emphasis on distinguishing between benign and malignant cysts. Through the integration of advanced technologies like Generative Adversarial Network (GAN) and deep convolutional neural network (DCNN), the objective is to develop computerized systems for ultrasound image analysis capable of accurately discerning between these cyst types. Leveraging the capabilities of GANs, the aim is to craft a sophisticated algorithm adept at detecting nuanced discrepancies in ultrasound images to precisely classify cysts. Furthermore, the creation of a DCNN model aims to automate ultrasound image evaluation, transcending the constraints of current methodologies and furnishing clinicians with a more efficient and accurate diagnostic instrument. Ultimately, the overarching aim is to heighten diagnostic accuracy and streamline the diagnostic pathway for ovarian cysts, thereby fostering improved patient outcomes and enriched healthcare provision within the realm of gynaecological health.

1.3.Outcome

Infertility stemming from disruptions in the maturation process of ovarian follicles, leading to the manifestation of polycystic ovaries (PCO), presents a significant challenge in women's reproductive health. Traditionally, the diagnosis of PCO has relied heavily on manual assessment by gynaecologists, necessitating meticulous counting and sizing of follicles within the ovaries. However, this approach is not only time-consuming but also demands a high degree of accuracy.

1.3.1 PERFORMANCE METRICS

The performance evaluation of our model encompasses the following measures:

- True Positive for PCOS (TP): Instances where the patient is diagnosed with PCOS and the model correctly classifies them as such.
- False Positive for PCOS (FP): Instances where the patient does not have PCOS, yet the model erroneously classifies them as having PCOS.
- True Negative for Non-PCOS (TN): Cases where the patient is identified as not having PCOS, and the model accurately classifies them as non- PCOS.

False Negative for Non-PCOS (FN): Scenarios where the patient actually has PCOS, but the model incorrectly labels them as not having PCOS.

1.3.2 ACCURACY

Accuracy serves as a crucial metric indicating the effectiveness of a model or algorithm in training and performance evaluation. In the context of this thesis, accuracy gauges the model's proficiency in detecting cyst types within the TV Ovarian Ultrasound Image Dataset

1.3.3 PRECISION

Precision represents the proportion of positively predicted cases that are truly positive. In the context of this thesis, precision quantifies the fraction of objects identified as cyst type that genuinely exist within TV ultrasound images.

1.3.4 RECALL

Recall, in essence, captures the ratio of actual positive cases correctly identified as positive by the model. In the context of this thesis, recall quantifies the fraction of follicles predicted as PCOS among all the PCOS instances.

1.3.4 TRAINING TIME

Training time is crucially assessed to measure the duration necessary for training selected machine learning algorithms on the dataset.

1.3.5 F1 SCORE

The F1 score serves as a comprehensive metric blending precision and recall, reflecting the model's accuracy in identifying ovarian cysts within TV ultrasound images. A high F1 score suggests minimal false positives and false negatives, indicating the model's proficiency.

1.3.6 LOSS FUNCTION

Additionally, the loss function plays a pivotal role in aligning the ground truth with the output of the segmentation network. It aids in optimizing network weights based on features extracted across various resolutions, rather than solely focusing on individual pixel-level details.

2. OBJECTIVES OF THE PROPOSED WORK

The current method of manually analysing ovarian follicles in ultrasound images has limitations. Firstly, subjectivity between doctors can lead to inconsistencies and missed diagnoses. Secondly, the manual process is time-consuming This project proposes an automated diagnostic system to address these shortcomings.

The system employs a combination of automatic thresholding and a neural network to precisely detect follicles and capture their shapes more accurately than existing methods. Following detection, the system analyzes the size and count of the follicles. This information is then fed into a Convolutional Neural Network (CNN) which classifies the type of cyst present, including differentiating between normal ovaries, complex cysts, and various PCOS classifications. This cyst classification is a novel aspect of the proposed system.

To further enhance the system's capabilities, a Generative Adversarial Network (PCO-GAN) is used to create high-resolution, editable ultrasound images. These synthetic images not only improve the training data for the system but also allow for the incorporation of sketch guidance, resulting in more realistic and structurally accurate images. Overall, this project presents a groundbreaking end- to-end framework that leverages machine learning and image synthesis techniques to achieve a more efficient and reliable automated diagnosis system for ovarian ultrasound analysis.

2.1.Software procedure

- Server Side : Python 3.7.4(64-bit) or (32-bit)
- Client Side : HTML, CSS, Bootstrap
- IDE : Flask 1.1.1
- Back end : MySQL 5.
- Server : Wampserver 2i
- OS : Windows 10 64 –bit or Ubuntu 18.04 LTS “Bionic Beaver”

2.1.1 . WampServer

WampServer is a Windows web development environment. It allows you to create web applications with Apache2, PHP and a MySQL database. Alongside, PhpMyAdmin allows you to manage easily your database. WAMPServer is a reliable web development software program that lets you create web apps with MYSQL database and PHP Apache2. With an intuitive interface, the application features numerous functionalities and makes it the preferred choice of developers from around the world. The software is free to use and doesn’t require a payment or subscription.

2.1.2 My SQL 5

It is fast, scalable, and easy to use database management system in comparison with Microsoft SQL Server and Oracle Database. It is commonly used in conjunction with PHP scripts for creating powerful and dynamic server-side or web-based enterprise applications. It is developed, marketed, and supported by MySQL AB, a Swedish company, and written in C programming language and C++ programming language. The official pronunciation of MySQL is not the My Sequel; it is My Ess Que Ell. However, you can pronounce it in your way. Many small and big companies use MySQL. MySQL supports many Operating Systems like Windows, Linux, MacOS, etc. with C, C++, and Java languages.

2.1.3 Flask 1.1.1

Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website. Flask is often referred to as a micro framework. It aims to keep the core of an application simple yet extensible. Flask does not have built-in abstraction layer for database handling, nor does it have formed a validation support. Instead, Flask supports the extensions to add such functionality to the application. Although Flask is rather young compared to most Python frameworks, it holds a great promise and has already gained popularity among Python web developers. Let's take a closer look into Flask, so-called "micro" framework for Python.

2.1.4 Bootstrap

Bootstrap is a free and open-source tool collection for creating responsive websites and web applications. It is the most popular HTML, CSS, and JavaScript framework for developing responsive, mobile-first websites. It solves many problems which we had once, one of which is the cross-browser compatibility issue. Nowadays, the websites are perfect for all the browsers (IE, Firefox, and Chrome) and for all sizes of screens (Desktop, Tablets, Phablets, and Phones). All thanks to Bootstrap developers -Mark Otto and Jacob Thornton of Twitter, though it was later declared to be an open-source project.

3. PROPOSED WORK MODULES

In this project, a Deep Learning-based Convolutional Neural Network (CNN) coupled with PCO-GAN (ML-CNN) is proposed for the detection, classification, and verification of ovarian cysts. The methodology involves employing a single modality classification algorithm to extract features from transvaginal ultrasound images. A typical PCO-GAN comprises two main components: a generator responsible for feature generation and a discriminator utilized for sorting input images based on "feature maps" representing various object features. These features could include corners, lines, circular arches, etc., which are relatively invariant to position distortions or shifting. The feature extractor may consist of multiple optional sub-sampling and convolutional layers.

3.1MODULE DESCRIPTION:

3.1.1 DATA PRE-PROCESSING

The input images fed into the DNN models needed to be a specific size: 224x224 pixels and have three color channels (RGB) for proper processing. Since the original images came in various sizes, a process called down-sampling and resizing was applied. To achieve this resizing, a technique known as nearest- neighbor interpolation was used. It's important to note that due to the non-uniform scaling of the original images, the resulting resized images might have slightly different physical resolutions. To ensure consistency across the data, an additional step was taken. The pixel values of each color channel (red, green, and blue) were standardized. This standardization involved subtracting the mean value (average) of each channel across the entire training dataset and then dividing each pixel value by the standard deviation. The same standardization process was applied to both the validation and test sets to maintain consistency throughout the training process.

3.1.2 SEGMENTATION

Imagine an image as a whole region, let's call it S. Segmentation is the process of dividing this entire region S into smaller, distinct sub-regions. We can represent these sub-regions as S1, S2, S3, and so on, all the way up to Sp.

Here's what makes a proper segmentation:

Completeness: Every single pixel in the original image (region S) must belong to one and only one of the sub-regions (S1, S2, S3, etc.). There can't be any pixels left out or unassigned.

Connectivity: Within each sub-region, all the points (pixels) should be connected to each other in some way. Imagine exploring the sub-region; you should be able to move from any point to any other point within that region without ever leaving its boundaries.
Disjointness: No overlap is allowed between the sub-regions. Each pixel can only be a part of a single sub-region, avoiding any ambiguity about which region it belongs to.

3.1.3 FEATURE EXTRACTION

During the initial processing stage, a range of techniques is applied to enhance the quality of ultrasound tumor images. These methods primarily focus on reducing noise, particularly speckle noise, by employing Adaptive Anisotropic Diffusion Filters. Contrast enhancement is then achieved through Contrast Limited Adaptive Histogram Equalization (CLAHE). Subsequently, segmentation of ovarian tumors is carried out utilizing algorithms such as GrabCut and the FL-SNN Model.

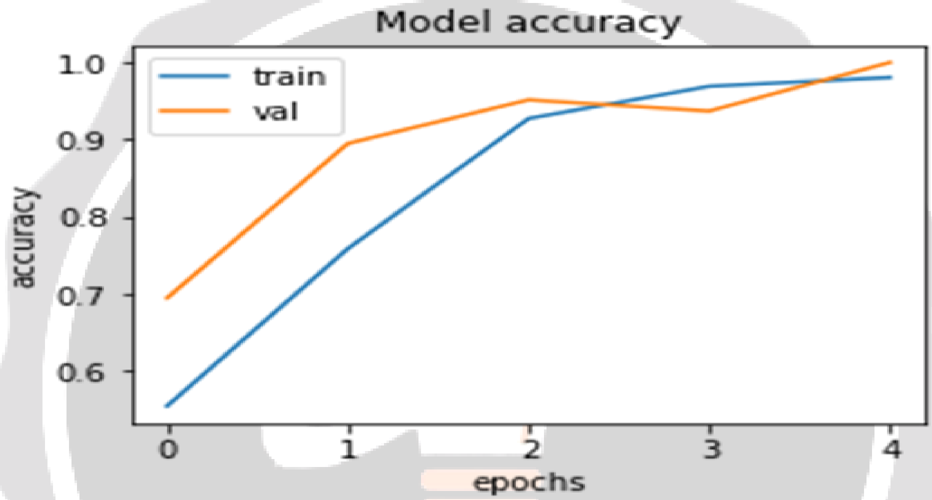

Fig 1-Accuracy

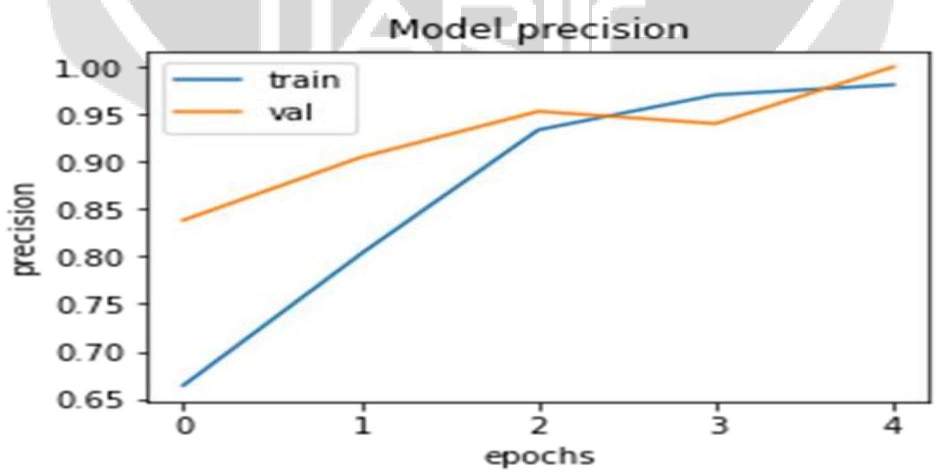

Fig 2-Precision

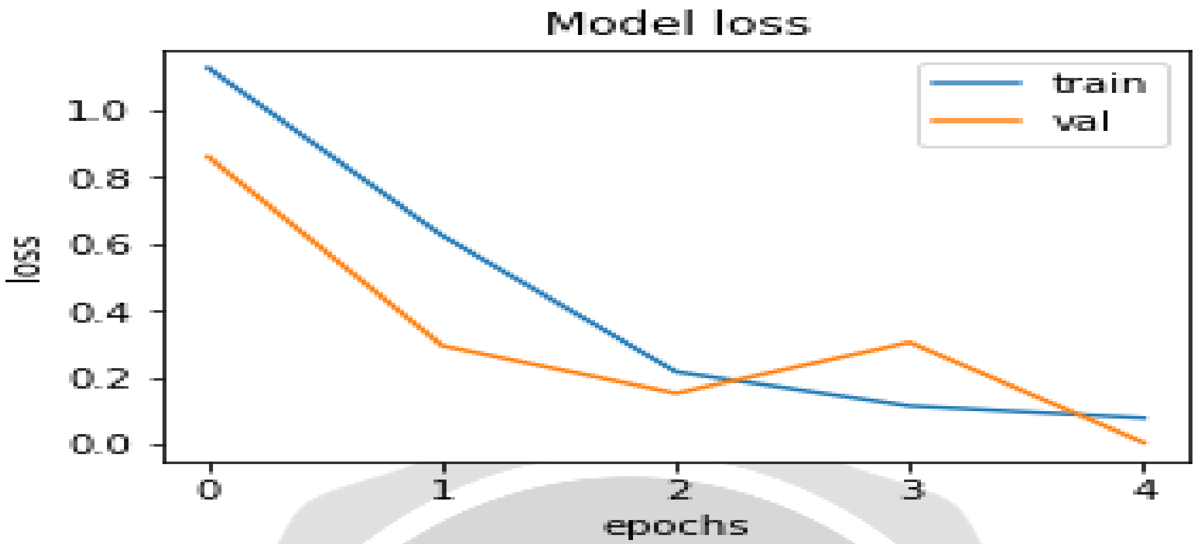

Fig 3- Loss

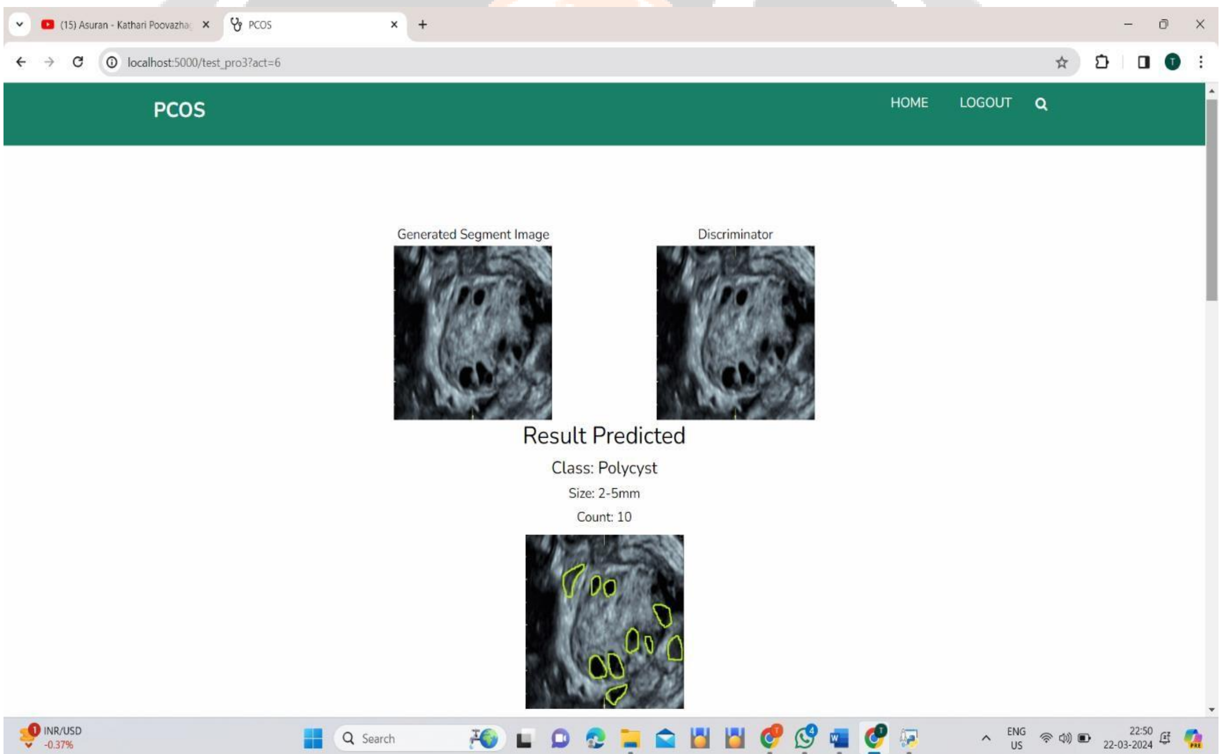

Fig 4-Output

Gray level co-occurence matrix

GLCM, a method of second-order statistical texture analysis, assesses the spatial relationships among pixels in an image by investigating the occurrence frequency of specific pixel combinations at various directions (Θ) and distances (d). Typically, images are quantized into 16 gray levels (0–15), producing four GLCMs (M) for each of the four directions (Θ = 0, 45, 90, and 135 degrees) with a distance of d = 1. From each GLCM, five features (Equations 13.30–13.34) are extracted, resulting in a total of 20 features per image. Before inputting these features into classifiers, normalization to a range between 0 and 1 is performed to ensure consistency across all classifiers.

The extracted features can be classified into three groups. The first group consists of first-order statistics, which encompass descriptors such as maximum intensity, minimum intensity, mean, median, 10th percentile, 90th percentile, standard deviation, variance of intensity value, energy, entropy, and other relevant metrics.

4. CONCLUSIONS

The conclusion of the project, the advent of deep neural network architectures, particularly Generative Adversarial Networks (GANs), has opened up new avenues for efficiently synthesizing realistic images. Leveraging this technology, our project introduces a deep learning framework for the automated detection of PCO, ovarian cysts, and functional cysts from transvaginal ultrasound (US) images of the ovaries. Central to this framework is the utilization of Convolutional Neural Networks (CNNs) for feature mapping, segmentation, and classification tasks.

Through the deployment of a CNN-based classifier, our approach not only identifies abnormalities but also distinguishes between different types of ovarian cysts, such as Dermoid cysts, Hemorrhagic cysts, and Endometrioma cysts. The integration of these algorithms into a unified DeepGAN methodology has yielded promising results, achieving an impressive accuracy rate of 98%.

Moreover, we have devised a novel approach to computing the loss function by incorporating PCO-GAN, further enhancing the diagnostic capabilities of our automated tool. By streamlining the diagnostic process and reducing both cost and time associated with manual assessment, our proposed method aims to expedite accurate treatment interventions for PCO patients. Ultimately, this automated diagnostic tool has the potential to significantly aid clinicians in the prognosis and management of PCO, thereby improving patient outcomes and reproductive health care delivery.

5. REFERENCES

[1]. K. C and J. M. Shivaram, "Comparative Analysis of Image Enhancement Techniques for Ultrasonic Ovarian Cyst Images", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 976-980, 2021.

[2]. S. Y. Liu, R. H. Zhu, Z. T. Wang, W. Tan, L. Zhang, Y. Q. Wang, et al., "Landscape of Immune Microenvironment in Epithelial Ovarian Cancer and Establishing Risk Model by Machine Learning", Journal of Oncology, 2021.

[3]. Z. Yue, C. Sun, F. Chen, Y. Zhang, W. Xu, S. Shabbir, et al., "Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis", Biomedical optics express, vol. 12, no. 5, pp. 2559-2574, 2021.

[4]. M. Lu, Z. Fan, B. Xu, L. Chen, X. Zheng, J. Li, et al., "Using machine learning to predict ovarian cancer", International Journal of Medical Informatics, vol. 141, pp. 104195, 2020

[5]. A. Laios, A. Gryparis, D. DeJong, R. Hutson, G. Theophilou and C. Leach, "Predicting complete cytoreduction for advanced ovarian cancer patients using nearest-neighbor models", Journal of Ovarian Research, vol. 13, no. 1, pp. 1-8, 2020.

[6]. A. Nabilah, R. Sigit, T. Harsono and A. Anwar, "Classification of Ovarian Cysts on Ultrasound Images Using Watershed Segmentation and Contour Analysis", 2020 International Electronics Symposium (IES), pp. 513-519, 2020.

[7]. C. Gopalakrishnan and M. Iyapparaja, "A Detailed Research on Detection of Polycystic Ovary Syndrome from Ultrasound Images of Ovaries", International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 2 S11, September 2019.

[8]. A. Denny, A. Raj, A. Ashok, C. M. Ram and R. George, "i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques", TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), pp. 673-678, 2019.

[9]. Ö. İnik, A. Ceyhan, E. Balcıoğlu and E. Ülker, "A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network", Computers in Biology and Medicine, vol. 112, 2019.

[10]. P. Soni and S. Vashisht, "Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques", 2018 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 816-820, 2018.