Auto-PCOS Classification Challenge

Palak Handa[a], Anushka Saini[b], Siddhant Dutta[c], Harsh Pathak[d], Nishi Choudhary[e], Nidhi Goel[f], Jasdeep Kaur Dhanao[f]

[a] *Dept. of ECE, DTU, Delhi*
[b] *Dept. of ECE-AI, IGDTUW, Delhi*
[c] *Dept. of CSE, SVKM's DJSCOE, Mumbai*
[d] *Dept. of ME, DTU, Delhi*
[e] *Dept. of Obstetrics and Gynecology, Lady Hardinge Medical College, Delhi*
[f] *Dept. of ECE, IGDTUW, Delhi*

1. Objective

The aim of the Auto-PCOS classification challenge is to provide an opportunity for the development, testing and evaluation of Artificial Intelligence (AI) models for automatic PCOS classification of healthy and un-healthy frames extracted from ultrasound videos. This challenge encompasses diverse training and test datasets, fostering the creation of vendor-agnostic, interpretable, and broadly applicable AI models.

2. Motivation

Polycystic ovary syndrome (PCOS) is one of the most common endocrine and metabolic disorders in premenopausal women. Heterogeneous by nature, PCOS is defined by a combination of signs and symptoms of androgen excess and ovarian dysfunction in the absence of other specific diagnoses. The aetiology of this syndrome remains largely unknown, but mounting evidence in the recent literature suggests that PCOS might be a complex multigenic disorder with strong epigenetic and environmental influences, including diet and lifestyle factors [1].

The World Health Organization (WHO) estimates that a staggering 116 million women globally grapple with PCOS. Statistics report that about 70% of the women suffering from PCOS remain undiagnosed which highlights the substantial prevalence and under-recognition of this condition [2]. In India, as per the Indian Fertility society, the prevalence of PCOS ranges from 3.7% to 22.5%.

This lifestyle-related ailment leads to a spectrum of metabolic and psychological challenges, including irregular menstrual cycles, hirsutism, sudden weight gain, type 2 diabetes, thyroid irregularities, and increased risk of depression and other psychiatric disorders, significantly affecting overall quality of life.

In the last 25 years, several attempts have been made by institution and societies like National Institutes of Health (NIH), European Society of Human Reproduction and Embryology (ESHRE) and American Society for Reproductive Medicine (ASRM) and Androgen Excess Society (AES) to standardize the diagnostic criteria for PCOS [3, 4]. They are based on various combinations of otherwise unexplained hyperandrogenism, anovulation, and the presence of polycystic ovaries observed through ultrasound imaging [5]. This observation is time-consuming, dependent on the sensitivity of the ultrasound equipment, the skill of the operator, the approach (vaginal v/s abdominal) and the weight of the patient [6].

Considering the underdiagnoses of PCOS, lack of experts in ultrasound imaging, and overall low patient-physician ratio across globe, there arises a need of modern diagnostic methods to fight PCOS. AI is predicted to have profound effects on the future of ultrasound imaging technology in the context of contrast enhancement, quality assessment, abnormality-based video summarization, annotations, measurements, and its artefact removal [7, 8]. Machine learning-based algorithms may help in automatic classification, detection and segmentation of polycystic ovaries observed in ultrasound imaging. These algorithms may prove to enhance diagnostic accuracy, reduce the manual steps, and overcome operator-independency.

3. PCOSGen dataset

The PCOSGen dataset is first of its kind, consists of different training and test datasets which have been collected from multiple internet resources like YouTube, ultrasoundcases.info, and Kaggle. PCOSGen-train and PCOSGen-test consists of 3200 and 1468 healthy and un-healthy instances respectively. Both training and testing datasets have been medically annotated with the help of experienced gynaecologist based in New Delhi, India. Figure 1 illustrates the flowchart of the data processing pipeline. Figure 2 and Figure

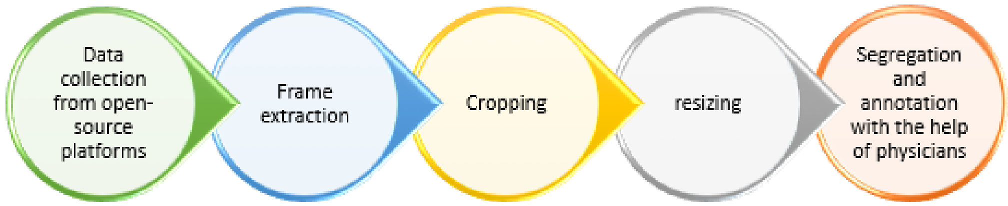

Figure 1: Data preparation pipeline.

Figure 2: Healthy PCOSGen training images.

Figure 3: Un-healthy PCOSGen training images.

3 showcase examples of healthy and unhealthy images from the PCOSGen training dataset respectively.

Release date of training dataset on zenodo: 25 December 2023

Release date of testing dataset on zenodo: 15 January 2024

Dataset links

NOTE: The following research and dataset development are solely intended for academic and research purposes. All data, information, and findings presented in this paper are for scholarly discussion, experimentation, and advancement within the scientific community.

References

[1] 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.

[2] A. Jabeen, V. Yamini, A. Rahman Amberina, M. Dinesh Eshwar, S. Vadakedath, G. S. Begum, V. Kandi, Polycystic ovarian syndrome: Prevalence, predisposing factors, and awareness among adolescent and young girls of south india, Cureus 14 (8) (2022) e27943.

[3] M. Asgharnia, F. Mirblook, M. Ahmad Soltani, The prevalence of polycystic ovary syndrome (PCOS) in high school students in rasht in 2009 according to NIH criteria, Int J Fertil Steril 4 (4) (2011) 156–159.

[4] B. O. Yildiz, G. Bozdag, Z. Yapici, I. Esinler, H. Yarali, Prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome under different diagnostic criteria, Human Reproduction 27 (10) (2012) 3067–3073. arXiv:https://academic.oup.com/humrep/article-pdf/27/10/3067/2238605/des232.pdf, doi:10.1093/humrep/des232.
URL https://doi.org/10.1093/humrep/des232

[5] S. Franks, Diagnosis of Polycystic Ovarian Syndrome: In Defense of the Rotterdam Criteria, The Journal of Clinical Endocrinology Metabolism 91 (3) (2006) 786–789. arXiv:https://academic.oup.com/jcem/article-pdf/91/3/786/10779615/jcem0786.pdf, doi:10.1210/jc.2005-2501.
URL https://doi.org/10.1210/jc.2005-2501

[6] C. Bansal, P. Handa, N. Goel, Comparing different models for polycystic ovary syndrome diagnosis: An empirical investigation on a large clinical dataset, in: 2023 IEEE Women in Technology Conference (WINTECHCON), 2023, pp. 1–6. doi:10.1109/WINTECHCON58518.2023.10277608.

[7] A. Makhdoomi, N. Jan, P. Palak, N. Goel, Machine learning techniques for medical images in pcos, in: 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST), 2022, pp. 1–6. doi:10.1109/AIST55798.2022.10064817.

[8] N. Jan, A. Makhdoomi, P. Handa, N. Goel, Machine learning approaches in medical image analysis of pcos, in: 2022 International Conference on

Machine Learning, Computer Systems and Security (MLCSS), 2022, pp. 48–52. doi:10.1109/MLCSS57186.2022.00017.