



YOLOv11 Model for PCOS Diagnosis: A Significant Contribution with Clinical Insights

Gianluca Mondillo

Department of Woman, Child and of General and Specialized Surgery, Università degli Studi della Campania “Luigi Vanvitelli”, Naples, Italy

I read with great interest the work of Zhao et al.¹ on the YOLOv11 model for the automatic recognition of ovarian ultrasound images in polycystic ovary syndrome (PCOS). The authors made a noteworthy contribution by conducting the first large-scale prospective study in women from East Asia, demonstrating excellent performance with an average training accuracy of 95.7% and validation accuracy of 97.6-97.8%, coupled with a remarkable processing speed nearly 50 times faster than that of clinicians.

The methodological rigor of the study deserves particular recognition. The adoption of the Rotterdam criteria as the diagnostic gold standard, the randomized stratification of patients across centers, and, most importantly, the detailed analysis of image structural independence using SSIM and MI indices are quality elements that set this work apart within the literature on artificial intelligence applied to PCOS. This aspect is especially relevant given that, according to a recent systematic review, only 10 of the 31 available studies (32%) on artificial intelligence (AI)/ML in PCOS diagnosis employed standardized diagnostic criteria (NIH, Rotterdam, or Revised International)², underscoring the robustness of the methodological framework established by the authors. However, some methodological aspects raise important concerns regarding the transition to clinical application. The exclusion rate of 46.7% of initially screened patients, though justified by the need for adequate image quality, creates a significant gap between experimental conditions and real-world clinical practice. Paradoxically, the very cases that are most challenging to diagnose ultrasonographically, precisely those where AI assistance would be most valuable, were excluded from the study. Comparable studies have addressed these limitations by employing advanced preprocessing techniques to manage suboptimal images³, suggesting that future developments could focus on strategies to broaden the applicability of the system.

From a technical perspective, several fundamental details essential for reproducibility are missing. The authors do not specify which variant of YOLOv11 was used among the five available, a detail far

from trivial given the different trade-offs between computational speed and diagnostic accuracy. Similarly, the absence of information on batch size, number of epochs, early stopping criteria, and hardware configuration restricts the possibility of replicating the reported results.

The geographic restriction of the study to East Asian populations, while understandable, raises concerns about the generalizability of the findings. The authors themselves note that 52% of Chinese PCOS patients present O + PCOM or H + PCOM subtypes, compared to only 30.2% in Western populations¹, indicating potential phenotypic differences that could affect model performance across diverse cohorts. In addition, the systematic exclusion of adolescent patients represents a notable limitation. Given that early diagnosis is critical for effective management of PCOS and its long-term complications, extending the model to younger populations could substantially enhance its clinical utility and the overall impact.

In light of the Rotterdam criteria, which integrate ultrasound, laboratory, and clinical data, a logical next step would be the development of a multimodal AI agent⁴, a true “Rotterdam agent.” Such a system could combine the YOLOv11 model with LLMs trained to process anthropometric, hormonal, clinical, and anamnestic information. This architecture would enable a shift from a monomodal recognition tool to a genuine decision-support tool, capable of delivering more comprehensive and balanced assessments, including for adolescent patients and those at menarche. Moreover, the adoption of a continuous and adaptive learning strategy would help sustain high performance while improving the system’s generalizability across diverse populations.

In conclusion, despite certain methodological limitations, the work of Zhao et al.¹ provides a robust foundation for the advancement of AI in diagnostic gynecology. The public availability of the code on GitHub represents a valuable contribution to the scientific community, fostering transparency and reproducibility.



Corresponding author: Gianluca Mondillo, Department of Woman, Child and of General and Specialized Surgery, Università degli Studi della Campania “Luigi Vanvitelli”, Naples, Italy

e-mail: gianluca.mondillo@gmail.com

Received: August 29, 2025 **Accepted:** September 25, 2025 **Available Online Date:** January 02, 2026 • **DOI:** 10.4274/balkanmedj.galenos.2025.2025-8-295

Available at www.balkanmedicaljournal.org

ORCID iDs of the authors: G.M. 0009-0000-7756-8145.

Cite this article as: Mondillo G. YOLOv11 Model for PCOS Diagnosis: A Significant Contribution with Clinical Insights. *Balkan Med J.*; 2026; 43(1):52-3.

Copyright@Author(s) - Available online at <http://balkanmedicaljournal.org/>

Most importantly, the potential integration of this system into real-time clinical workflows, particularly in settings with limited access to specialized expertise, highlights its promise as a transformative tool in the management of PCOS.

REFERENCES

1. Zhao B, Wen L, Huang Y, et al. A deep learning-based automatic recognition model for polycystic ovary ultrasound images. *Balkan Med J.* 2025;42:419-428. [\[CrossRef\]](#)
2. Barrera FJ, Brown EDL, Rojo A, et al. Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. *Front Endocrinol (Lausanne).* 2023;14:1106625. [\[CrossRef\]](#)
3. Sumathi M, Chitra P, Sheela S, Ishwarya C. Study and implementation of automated system for detection of PCOS from ultrasound scan images using artificial intelligence. *Imaging Sci J.* 2024;72:828-839. [\[CrossRef\]](#)
4. Wang W, Ma Z, Wang Z, et al. A survey of LLM-based agents in medicine: how far are we from Baymax? *arXiv.* 2025. [\[CrossRef\]](#)