A Machine Learning Approach to Early PCOS Detection:

Evaluating Extreme Gradient Boost Performance Against Other Models and Standard Diagnostic Pathways

Authors: Tanha Ahmed, WEP Clinical, London, UK, Dania Al Dulaimy, WEP Clinical, London, UK

MT5

INTRODUCTION

Polycystic Ovarian Syndrome (PCOS) is described as a heterogenous endocrine-metabolic condition and a syndrome of ovarian dysfunction and is the most common endocrine disorder faced by women of reproductive age. PCOS typically manifests features of hyperandrogenism and / or polycystic ovarian morphology (PCOM).[1] Though PCOS is often referred to as one condition, its heterogeneity means the best diagnostic method has been an area of debate. Certain criteria, such as the Rotterdam criteria for formal diagnosis, rely on a combination of clinical, biochemical and imaging findings, which may not be readily accessible or sensitive enough to detect early disease. Diagnostic challenges leave up to 70% of women with PCOS undiagnosed, resulting in a higher risk of long-term physical and emotional conditions.[2]

The 2023 International Guideline for the Assessment and Management of PCOS emphasizes the need for refined diagnostic criteria and a simplified diagnostic algorithm, alongside recognition for metabolic, cardiovascular, sleep, and psychological features.[3]

Within the last decade, machine learning (ML) has been explored as a method for identifying PCOS risk. Gradient-boosted tree methods, such as Extreme Gradient Boosting (XGBoost), an open-source, scalable ML library that implements decision-trees to solve prediction problems, have been consistently competitive in trials vs other ML models. Through converting symptom patterns into calibrated risk predictions, ML has potential in flagging high-risk PCOS patients.[4]

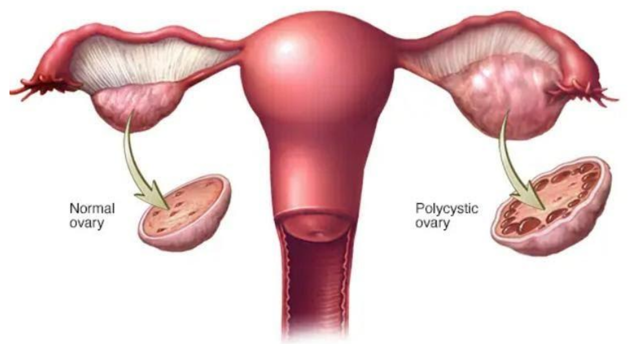

Figure 1: Comparative visual imaging of polycystic vs a normal ovarian morphology.[5]

OBJECTIVE

1. To systematically evaluate the performance of XGBoost against other machine learning models, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Logistic Regression (LR) and K-Nearest Neighbors (KNN), for the early detection and diagnosis of PCOS, using area under the curve (AUC), precision and F1 score
2. To consider the impact of excluding clinically relevant anthropometric variables, such as body mass index and waist circumference, on model performance and applicability to clinical practice
3. To assess the extent to which current approaches address calibration, threshold and real-world applicability across diverse populations

METHODOLOGY

Peer-reviewed literature up to December 2025 were systematically identified, screened and consolidated to capture studies evaluating the performance of XGBoost-based ML models for the early detection and diagnosis of PCOS. This poster prioritizes studies that (a) use real-world clinical data from hospital settings or clearly described cohorts, (b) compare ML models within-study, and / or (c) report external validation and calibration evaluation. Searches were conducted across major scientific databases using predefined keywords related to PCOS, ML, and gradient-boosting techniques. Eligible studies included those that investigated XGBoost models with standard diagnostic measures, such as the Rotterdam criteria, in clinical and retrospective datasets. Key clinical variables examines across studies include anti-Müllerian hormone (AMH), a biomarker assessing ovarian reserve and follicular dysfunction in PCOS, alongside ultrasound, biochemical and anthropometric features.

Studies were included if they reported quantitative performance metrics enabling objective comparison, including AUC, overall accuracy, precision and F1 score, a measure that balances precision and recall to provide a single indicator of a model's predictive reliability. This poster also examines the clinical interpretability of such models using Shapley Additive explanations (SHAP), a technique quantifying each individual feature's contribution to a model's prediction, helping identify which clinical variables are most influential in determining PCOS risk. Extracted data were synthesized to evaluate trends in model performance, robustness across populations and potential clinical utility of XGBoost in improving PCOS diagnosis.

RESULTS

3 studies meeting the criteria were identified for inclusion in the analyses.

A study evaluated ML models (ANN, SVM, LR, KNN and XGBoost), in which XGBoost demonstrated superior and most consistent performance in PCOS diagnosis. The highest performance was achieved by XGBoost's model using combined clinical, ultrasound (USG) and AMH features (AUC = 0.9947; accuracy = 0.9553), followed by clinical and USG data alone (AUC = 0.9852; accuracy = 0.9384).[4]

In imaging-based applications, a hybrid deep learning approach combining XGBoost with other models found that inclusion of XGBoost for PCOS detection significantly increased accuracy of PCOS diagnosis. Combining XGBoost with other models and techniques such as Ada boost, random forest, decision tree and a hybrid model yielded an accuracy of 97.2% from a dataset of 2,004 ovarian ultrasound images.[5] Additionally, combining XGBoost with VGGNet-19 achieved an accuracy of 99.6% in diagnosing PCOS. This demonstrates the versatility of XGBoost as part of multimodal or hybrid architectures of ML for classification of PCOS.[6]

Additionally, a large-scale study involving approximately 1,600 women in a Chinese hospital setting reported strong predictive performance for early PCOS detection using XGBoost (training cohort: AUC = 0.919, 95% CI: 0.896-0.942; validation cohort: AUC: 0.923, 95% CI: 0.893-0.953), compared to the predictive performance of seven other ML models, with the second ranked being the Random Forest Model (AUC = 0.918). XGBoost also demonstrated an F1 Score of 0.805 and precision of 0.8312.[7]

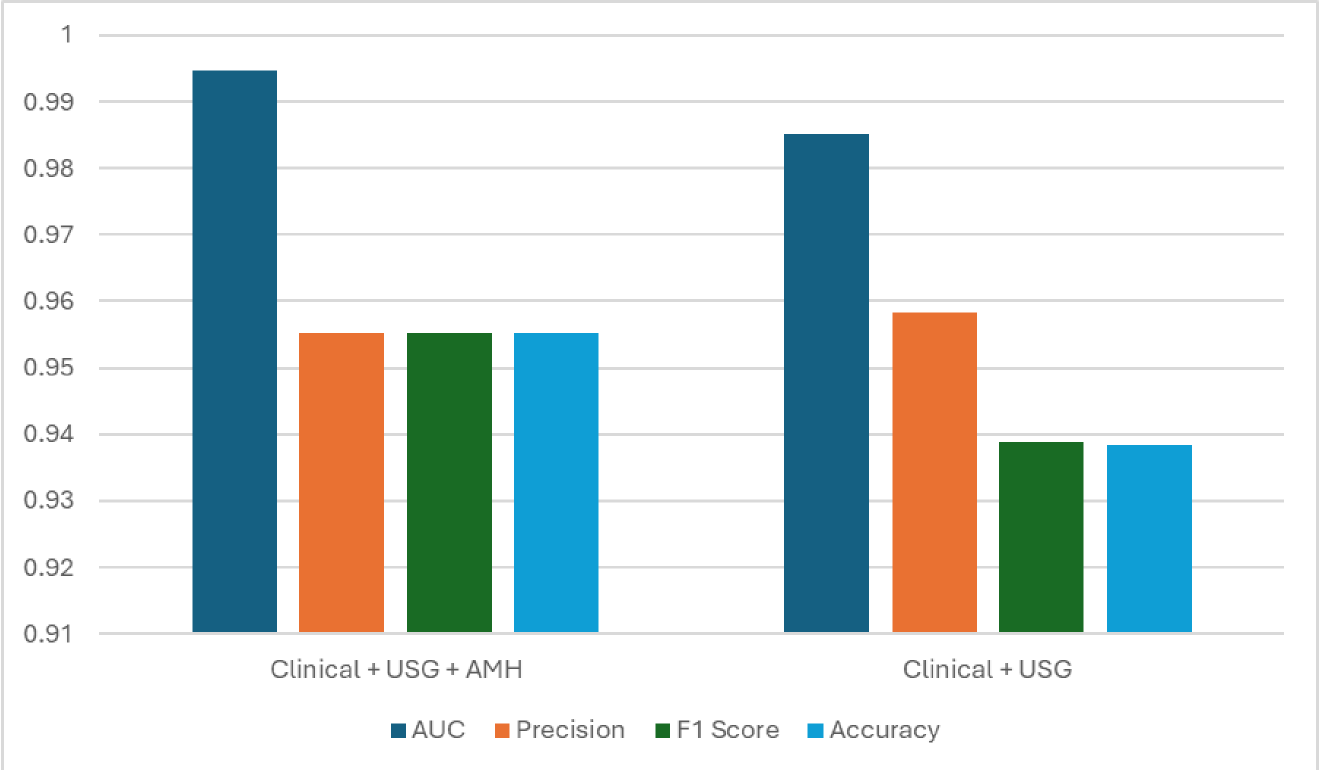

Figure 2: Final XGBoost model showed strong performance across feature sets: Clinical + USG + AMH (AUC = 0.9947, Precision/F1/Accuracy = 0.9553) and Clinical + USG (AUC = 0.9852, Precision = 0.9583, F1 = 0.9388, Accuracy = 0.9384). Key features included follicle count, weight gain, AMH, hair growth, menstrual irregularity, fast food intake, pimples, and hair loss. External validation (n = 320) achieved perfect performance (AUC/Precision/F1/Accuracy = 1.0).[4]

DISCUSSION & CONCLUSION

The evolution from traditional, observer-dependent methods to automated and non-invasive ML methods highlights a significant shift in PCOS diagnosis and this may address the critical need for earlier screening to prevent long-term complications such as type 2 diabetes, cardiovascular disease, infertility and reproductive cancers. XGBoost consistently demonstrates superior performance as a method for PCOS detection, especially when applied to multimodal datasets combining clinical, USG and imaging features. When combining XGBoost with models such as VGGNet-19, which alone has an accuracy of 96%,[8] it can also increase the accuracy in the detection of PCOS. This supports its suitability for capturing the complex, multifaceted nature of PCOS. It is important to note, that although XGBoost in a hybrid model demonstrated strong results, many of these studies did not capture the performance of each model as a standalone method for detecting PCOS diagnosis.

Near-perfect AUC values should be interpreted cautiously, as they may reflect overfitting or limited diversity in the datasets. Many studies prioritize discrimination (AUC) without addressing calibration and clinically meaningful trade-offs between sensitivity and specificity, which are crucial for real-world screening and early detection. Additionally, a recurring limitation across studies is the absence of important anthropometric variables, including body mass index and waist circumference.[7] These are routinely used in clinical practice and are closely linked to the metabolic and obesity-related complications of PCOS, so excluding them likely limits the clinical relevance and the predictive power of existing models. Therefore, taking this into consideration in future studies could meaningfully improve risk stratification and allow for more comprehensive screening tools. On the interpretability side, explainable AI methods like SHAP offer useful insights into which features drive predictions, but they can be difficult for clinicians to use in practice.[7] Moreover, these models often focus on ovarian morphology alone, which fails to capture the full spectrum of PCOS. Future work should prioritize large scale, multi-center prospective validation of XGBoost-based models in PCOS, alongside standardized feature sets aligned to the 2023 International Guidelines. Evaluating model fairness across diverse populations, robustly assessing calibration and integrating findings into clinical decisions are essential steps toward translating these promising results into better diagnostic outcomes.

In conclusion, while XGBoost based models show real potential detecting PCOS early on, translating them into clinical use will require stronger validation, and would also require including clinically relevant variables and interpretability tools that clinicians can work with to improve outcomes for the 70% of undiagnosed women with PCOS.

References

1. Azziz R, Carmina E, Chen Z, et al. Polycystic ovary syndrome. Nat Rev Dis Primers. 2016;2(1). doi:10.1038/nrdp.2016.57
2. World Health Organization. Polycystic ovary syndrome. World Health Organization. Published 2025. Accessed April 14, 2026. https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome
3. Teede HJ, Tay CT, Laven JJE, et al. Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. J Clin Endocrinol Metab. 2023;108(10):2447-2469. doi:10.1210/clinem/dgad463
4. Agirsoy M, Oehlschlaeger MA. A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features. Sci Rep. 2025;15(1). doi:10.1038/s41598-025-10453-9
5. Khan Sarguroh F, R S. An analytical review of AI-driven techniques for polycystic ovary syndrome prediction. Int J Sci Res (IJSR). 2025;14(12):1347-1352. doi:10.21275/sr251217131307
6. Pawar S, Dhanraj D, Dheple P, Doshi N. PCOS detection using hybrid CNN-XGBoost model—a multimodal data approach. Biomed Pharmacol J. 2025;18(3):2258-2269
7. Yao B, Yu X, Zhang Y, et al. Development and validation of an explainable machine learning and nomogram model for early detection and risk stratification of polycystic ovary syndrome: a multicenter study. Front Endocrinol. 2025;16. doi:10.3389/fendo.2025.1719631
8. Praneesh M, Nivetha N, Maidin SS, Ge W. Optimized deep learning method for enhanced medical diagnostics of polycystic ovary syndrome detection. Journal of Applied Data Sciences. 2024 Sep 17;5(3):1399-411.