

# Incremental Discrimination of Clinical, Biochemical, and Ultrasound Features for Polyendocrine Metabolic Ovarian Syndrome (Formerly PCOS): A Secondary Analysis of a Public Multicenter Dataset

Mariam A Alkhatib<sup>1</sup>, Ahed J Alkhatib<sup>2\*</sup> and Nawal A Alkhatib<sup>3</sup>

<sup>1</sup>Faculty of Nursing, Jordan University of Science and Technology, Jordan

<sup>2</sup>Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science and Technology, Jordan

<sup>3</sup>Faculty of Pharmacy, Jordan University of Science and Technology, Jordan

**\*Corresponding Author:** Ahed J Alkhatib, Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science and Technology, Jordan.

**Received:** September 17, 2026; **Published:** October 07, 2026

## Abstract

**Background:** PMOS is diagnosed with combinations of reproductive, androgenic and ovarian features; it was previously known as polycystic ovary syndrome (PCOS). The public datasets on PCOS have been used widely for machine-learning classification. However, not much attention has been paid to the extra information offered by the clinical, biochemical, and ultrasound domains or the impact of obvious data-quality errors.

**Methods:** We conducted a secondary cross-sectional analysis of a dataset available in the public domain claimed to consist of 541 women assessed in 10 hospitals in Kerala, India; 177 (32.7%) were classified as PCOS positive. Values that were highly implausible were retained as missing value using fairly conservative rules, but a sensitivity analysis was performed to exclude the complete nine records containing such values. Multivariable logistic regression was used to analyze clinical associations. Four ridge-logistic classification models were internally validated using repeated stratified 10-fold cross-validation (20 repeats): clinical (Model A), clinical plus biochemical (Model B), clinical plus biochemical plus ultrasound (Model C), and restricted clinical model not featuring prominent diagnostic-criterion features (Model D).

**Results:** In the adjusted clinical model ( $n = 538$ ), OR 4.05, 95% CI 2.35-7.00 hair growth, OR 3.85, 95% CI 2.24-6.59 darkening of skin, aOR 3.67, 95% CI 2.08-6.45 irregularity of cycles, and OR 3.31, 95% CI 1.79-6.13 weight gain had strongest independent associations with PCOS status. Model A's AUC was 0.881, 95% bootstrap CI was 0.848-0.912. Improvement in discrimination was not evident upon adding the biochemical variables (Model B AUC 0.879; 95% CI 0.845-0.909; delta AUC -0.002; 95% CI -0.008 to 0.004). By incorporating ultrasound variables, the Area Under Curve (AUC) improved to 0.949. Moreover, the most significant contributor to this gain was the isolated follicle count. The limited non-criteria clinical model still provided useful separation (AUC 0.843, 95% CI 0.805-0.877). When we omitted all nine records with grossly implausible values, results were nearly the same.

**Conclusion:** In this public dataset, readily available clinical features showed good discrimination, routine biochemical variables provided little incremental information, and ultrasound, in particular follicle count, greatly improved classification. As several influential variables coincide with the diagnostic criteria, the interpretation of these results is essentially, discrimination in a labelled dataset that already exists, not a prediction of a candidate or diagnostic validation which is independent.

**Keywords:** Polyendocrine Metabolic Ovarian Syndrome; Polycystic Ovary Syndrome; Ultrasound; Anti-Müllerian Hormone; Diagnostic Discrimination; Secondary Analysis

## Introduction

Polyendocrine metabolic ovarian syndrome (PMOS), which has previously been known as polycystic ovary syndrome (PCOS), is a common, heterogeneous endocrine and metabolic condition with reproductive, dermatologic, metabolic, and psychological manifestations. The occurrence of this disease varies significantly depending on the diagnostic criteria utilized and the population investigated. Systematic reviews indicate that the occurrence of the disease is substantially high-single-digit to low-double-digit in reproductive-age women [1-3]. Women with obesity and central adiposity are more common and can worsen reproductive and metabolic abnormalities [4,5] while long-term cardiometabolic risk deserves further attention across the life course [6,7]. In May 2026, an international multi-step consensus process adopted the term PMOS to portray better the multi-system nature of the condition and to move away from the erroneous implication that pathological ovarian cysts define the condition [8]. This analysis uses the historical nomenclature of PCOS with respect to outcome variable contained in the source dataset.

Diagnostic definitions have progressively shifted from the Rotterdam consensus toward current global evidence-based recommendations [9-11]. The latest guidance indicates the combined presence of ovulatory dysfunction, clinically or biochemically proven hyperandrogenism, and ultrasound demonstration of polycystic ovarian morphology after exclusion of other disorders, to constitute the diagnosis. As imaging resolution has improved, ultrasound thresholds have changed, and follicle number per ovary has been particularly useful morphological measure [12-14]. According to the 2023 international recommendations AMH estimation in adults is permitted as another method to define polycystic ovarian morphology within a proper diagnostic algorithm, as opposed to using AMH as a stand-alone diagnostic test [11]. Previous research that combined examined the meta-analyses analytically also demonstrated that AMH has the potential to diagnose [15]. This creates a significant methodological problem for classification studies that are driven by data; for instance, menstrual irregularity, hirsutism, follicle number and AMH can be model inputs but also closely related to how the diagnosis label has been assigned.

A large number of machine-learning studies using the public dataset of 541 women from Kerala, India have reported high classification performance using combinations of clinical symptoms and laboratory and ultrasound features of ovary [16,17]. Reanalyzing the same data has often focused on the accuracy of an algorithm and the ranking of features. Nonetheless, high apparent performance is difficult to interpret clinically when diagnostic-criterion variables are included among predictors, preprocessing is not transparent, there is no external validation, and performance is assessed via a single random train-test split. Studies show that well-specified logistic regression may be as good at making clinical prediction as complex machine learning methods [18]. Furthermore, important considerations of discrimination, calibration, validation and risk of bias need to be reported reliably [19-22].

## Aim of the Study

The current investigation, therefore, focused on information increment and not algorithms. We had the following aims (1) To compare the clinical, biochemical and ultrasound features according to PCOS status; (2) To identify the clinical features that are independently associated with the diagnosis recorded; (3) To quantify the incremental discrimination achieved by adding the biochemical and ultrasound domains to a clinical model; and (4) To test the robustness of results to obvious data-quality errors and to removal of prominent diagnostic-criterion features.

## Methods

### Study design and data source

This refers to the secondary cross-sectional analysis of a publicly available PCOS dataset from Kaggle. As stated by its repository and prior publications, the dataset contains clinical, biochemical, lifestyle and ultrasound information from 541 women evaluated in 10 hospitals in Kerala, India.

The dataset includes 44 columns in the version analyzed here, including binary outcome 'PCOS (Y/N)'. Since the present study did not gather primary data, statements concerning original recruitment dates, sampling procedures, diagnostic adjudication, laboratory assays, and primary ethical approval are limited to what is available from the public dataset and prior publications. All of these items should be checked against original source documentation prior to journal submission.

### **Outcome and predictor domains**

Binary PCOS status (1=positive; 0=negative) was the outcome. We organized the predictors into domains a priori. The clinical model consisted of age, recalculated body mass index (BMI), waist-to-hip ratio, irregularity of cycle, length of cycle, weight gain reported, hair growth, darkening of skin, hair loss, pimples, fast food consumption and regular exercise. The addition of BCH layer included follicle-stimulating hormone (FSH), luteinizing hormone (LH), thyroid-stimulating hormone (TSH), AMH, prolactin (PRL) and random blood sugar (RBS). The ultrasound layer counted follicle as total count was left ovary plus right ovary means size and endometrial thickness. The variable for FSH/LH was not included together with FSH and LH to avoid redundancy by deterministic means. The primary classification models did not incorporate blood group, record identifiers, beta-HCG measures, pregnancy status, and abortion history.

### **Data-quality assessment and cleaning**

Binary PCOS status (1=positive; 0=negative) was the outcome. We organized the predictors into domains a priori. The clinical model consisted of age, recalculated body mass index (BMI), waist-to-hip ratio, irregularity of cycle, length of cycle, weight gain reported, hair growth, darkening of skin, hair loss, pimples, fast food consumption and regular exercise. The addition of BCH layer included follicle-stimulating hormone (FSH), luteinizing hormone (LH), thyroid-stimulating hormone (TSH), AMH, prolactin (PRL) and random blood sugar (RBS). The ultrasound layer counted follicle as total count was left ovary plus right ovary means size and endometrial thickness. The variable for FSH/LH was not included together with FSH and LH to avoid redundancy by deterministic means. The primary classification models did not incorporate blood group, record identifiers, beta-HCG measures, pregnancy status, and abortion history.

### **Statistical analysis**

Due to the highly skewed nature of many hormonal variables, continuous variables were summarized as median and interquartile range (IQR) and compared by Mann-Whitney U test. To show that a categorical variable would work in a biochemical character or drug competition, we used the automated DNA copy number tag. Reported are crude odds ratios (ORs) and 95% confidence intervals (CIs) for all binary characteristics. Continuous characteristics were also compared with standardized mean differences (SMDs) to describe effect size.

In order to identify independent clinical associations, a multivariable logistic regression model was fitted including age (per 5 years), BMI (per 5 kg/m<sup>2</sup>), waist-to-hip ratio (per 0.1), cycle irregularity, cycle length, weight gain, hair growth, skin darkening, hair loss, pimples, fast food consumption and regular exercise. The association model used complete cases (n = 538), with adjusted odds ratios (aORs) and 95% CIs reported.

Four prespecified ridge-logistic models were evaluated for discrimination analysis. Model A analysed the clinical variables; Model B added biochemical variables; Model C added ultrasound variables; the restricted clinical model (Model D) excluded cycle irregularity, cycle length, and hair growth, thus limiting direct overlap with key diagnostic-criterion features. Predictors were standardized within the modeling pipeline, with median imputation of missing values within each training fold. Using closed-loop design, 10-fold repeated cross-validation was used. Average probability of predictions was calculated outside fold on repeats for each subject. The discrimination level was assessed by means of the area under the receiver operating characteristic curve (AUC), and 95% CIs were estimated by use of 3000 bootstrap resamples of the averaged out-of-fold predictions. A probability threshold of 0.50 classification metrics was used

to compute and report Brier score, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), calibration intercept, and calibration slope. Both discrimination and calibration were reported according to standard recommendations for evaluation of clinical prediction models [22,23] and ridge penalization was used to reduce coefficient instability in this moderate-sized dataset. Nested models were compared using bootstrap distributions to assess the differences in paired AUC.

Two analyses on sensitivity were carried out. To start with, nine records containing grossly implausible values were removed. Subsequently, the contribution of ultrasound was assessed by juxtaposing two models: one that incorporated follicle size, endometrial thickness, but not follicle count; and the other that included total follicle count, compared to the clinical model. We considered the number of candidate predictors and the absence of external validation cohort when interpreting the model performance, as the size and validation strategy affect optimism and transportability in prediction research [24]. In Python analysis was done with the libraries, SciPy, statsmodels and scikit-learn. The recommendations from STROBE for observational analyses as well as the TRIPOD/TRIPOD+AI and PROBAST principles for reporting and assessment of prediction-models were followed for reporting [19-21,25].

Ethical considerations and data availability

The study used a public dataset with no identifiers and did not contact any participants in the study. Prior to submission, confirm with the authors’ institution whether a formal exemption or institutional determination is required. You can find the source dataset on the Kaggle repository referenced in the References. The analytical code and a dictionary of cleaned variables should optimally be deposited with the final manuscript.

Results

Sample characteristics and data quality

Of the 541 subjects, 177 were diagnosed with PCOS (32.7%) and 364 were normal PCOS-negative (67.3%). There were not many missing values, but there were some isolated nonnumeric, miscoded, or grossly implausible values which required inspection. There was at least one grossly implausible numerical measurement in the records of nine participants. Instead of dropping these study participants from the main analysis, only the specific measurements associated with these participants were made missing; the complete-record exclusion sensitivity analysis yielded almost identical model performance.

Continuous and categorical differences

Participants with PCOS were younger than those without (median ages, 29 vs 32 years,  $p < 0.001$ ) and had a higher recalculated BMI (25.15 vs 23.62 kg/m<sup>2</sup>,  $p < 0.001$ ). the median AMH of the PCOS-positive was 5.90 as compared to 3.20 ng/mL in the PCOS-negative group and was statistically significant with  $p < 0.001$ . The greatest ongoing contrasts were seen for ovarian follicle count. The median total follicle count was 20 [IQR 16-25] in the PCOS-positive group and 8 [IQR 5-12] in the PCOS-negative group ( $p < 0.001$ ; SMD 1.76) (Table 1).

| Variable                               | Non-PCOS (N) | Non-PCOS median [IQR] | PCOS (N) | PCOS median [IQR]   | p value | SMD    |
|----------------------------------------|--------------|-----------------------|----------|---------------------|---------|--------|
| Age (years)                            | 364          | 32.00 [28.00-36.00]   | 177      | 29.00 [27.00-33.00] | <0.001  | -0.365 |
| BMI (kg/m <sup>2</sup> , recalculated) | 364          | 23.62 [21.37-26.09]   | 177      | 25.15 [23.01-28.26] | <0.001  | 0.423  |
| Waist-to-hip ratio                     | 364          | 0.89 [0.86-0.93]      | 177      | 0.90 [0.86-0.93]    | 0.652   | 0.026  |
| Cycle length (days)                    | 363          | 5.00 [5.00-6.00]      | 177      | 5.00 [3.00-5.00]    | <0.001  | -0.373 |
| FSH (mIU/mL)                           | 363          | 5.01 [3.40-6.55]      | 177      | 4.48 [3.23-5.76]    | 0.009   | -0.03  |
| LH (mIU/mL)                            | 364          | 2.30 [1.03-3.60]      | 176      | 2.21 [1.00-4.30]    | 0.410   | 0.169  |
| TSH (mIU/L)                            | 364          | 2.17 [1.48-3.59]      | 177      | 2.31 [1.57-3.51]    | 0.715   | -0.023 |

|                         |     |                      |     |                       |        |       |
|-------------------------|-----|----------------------|-----|-----------------------|--------|-------|
| AMH (ng/mL)             | 363 | 3.20 [1.90-5.33]     | 177 | 5.90 [2.60-10.20]     | <0.001 | 0.525 |
| PRL (ng/mL)             | 364 | 21.16 [14.73-29.59]  | 177 | 22.90 [14.13-30.33]   | 0.593  | 0.011 |
| RBS (mg/dL)             | 364 | 96.00 [92.00-107.25] | 177 | 100.00 [92.00-107.00] | 0.345  | 0.097 |
| Follicles, left         | 364 | 4.00 [2.00-6.00]     | 177 | 10.00 [7.00-12.00]    | <0.001 | 1.493 |
| Follicles, right        | 364 | 4.00 [2.00-7.00]     | 177 | 11.00 [8.00-13.00]    | <0.001 | 1.702 |
| Total follicle count    | 364 | 8.00 [5.00-12.00]    | 177 | 20.00 [16.00-25.00]   | <0.001 | 1.759 |
| Mean follicle size (mm) | 364 | 15.00 [13.00-17.50]  | 177 | 16.00 [14.00-17.50]   | 0.006  | 0.295 |
| Endometrium (mm)        | 364 | 8.30 [7.00-9.50]     | 177 | 8.90 [7.60-10.00]     | 0.005  | 0.234 |

Table 1: Continuous characteristics according to recorded PCOS status.

The group was significantly more likely to experience clinical manifestations. Skin darkening (crude OR 9.03, 95% CI 5.95-13.69), hair growth (OR 8.96, 95% CI 5.85-13.74), weight gain (OR 7.32, 95% CI 4.90-10.92), irregular cycles (OR 6.30, 95% CI 4.18-9.51), and fast food (OR 5.89, 95% CI 3.89-8.94) were all associated strongest with recorded PCOS-status in unadjusted analyses (Table 2).

| Characteristic        | Non-PCOS n/N (%) | PCOS n/N (%)    | Crude OR | 95% CI     | p value |
|-----------------------|------------------|-----------------|----------|------------|---------|
| Irregular cycle       | 56/364 (15.4%)   | 94/176 (53.4%)  | 6.3      | 4.18-9.51  | <0.001  |
| Weight gain           | 83/364 (22.8%)   | 121/177 (68.4%) | 7.32     | 4.90-10.92 | <0.001  |
| Hair growth           | 47/364 (12.9%)   | 101/177 (57.1%) | 8.96     | 5.85-13.74 | <0.001  |
| Skin darkening        | 56/364 (15.4%)   | 110/177 (62.1%) | 9.03     | 5.95-13.69 | <0.001  |
| Hair loss             | 143/364 (39.3%)  | 102/177 (57.6%) | 2.1      | 1.46-3.03  | <0.001  |
| Pimples               | 142/364 (39.0%)  | 123/177 (69.5%) | 3.56     | 2.43-5.22  | <0.001  |
| Fast-food consumption | 139/363 (38.3%)  | 139/177 (78.5%) | 5.89     | 3.89-8.94  | <0.001  |
| Regular exercise      | 83/364 (22.8%)   | 51/177 (28.8%)  | 1.37     | 0.91-2.06  | 0.129   |
| Ever pregnant         | 142/364 (39.0%)  | 64/177 (36.2%)  | 0.89     | 0.61-1.28  | 0.521   |

Table 2: Categorical clinical and lifestyle characteristics associated with recorded PCOS status.

Multivariable clinical associations

Based on multivariable clinical model (n = 538), among them hair growth had the strongest independent association with recorded PCOS status OR 4.05 (95% CI 2.35-7.00; p < 0.001), then skin darkening OR 3.85 (95% CI 2.24-6.59; p < 0.001) and irregular cycles OR 3.67 (95% CI 2.08-6.45; p < 0.001). Weight gain OR 3.31 (95% CI 1.79-6.13; p < 0.001) was less often but still strongly associated. The association of pimples and fast-food consumption with the outbreak of acne stayed the same for this study. With increase in age recorded status of PCOS decreased (OR per 5 years 0.69 95% CI 0.54-0.89; p = 0.004). After adjustment, BMI, waist-to-hip ratio, hair loss and regular exercise were not independent (Table 3 and figure 1).

| Variable              | Adjusted OR | 95% CI    | p value |
|-----------------------|-------------|-----------|---------|
| Age per 5y            | 0.69        | 0.54-0.89 | 0.004   |
| BMI per 5 kg/m²       | 0.87        | 0.60-1.26 | 0.461   |
| WHR per 0.1           | 1.04        | 0.58-1.86 | 0.901   |
| Irregular cycle       | 3.67        | 2.08-6.45 | <0.001  |
| Cycle length, per day | 0.85        | 0.71-1.01 | 0.067   |

|                  |      |           |        |
|------------------|------|-----------|--------|
| Weight gain      | 3.31 | 1.79-6.13 | <0.001 |
| Hair growth      | 4.05 | 2.35-7.00 | <0.001 |
| Skin darkening   | 3.85 | 2.24-6.59 | <0.001 |
| Hair loss        | 0.87 | 0.51-1.48 | 0.604  |
| Pimples          | 2.3  | 1.36-3.88 | 0.002  |
| Fast food        | 2.01 | 1.15-3.50 | 0.014  |
| Regular exercise | 1.45 | 0.82-2.55 | 0.204  |

**Table 3:** Multivariable logistic regression of clinical characteristics associated with recorded PCOS status (n = 538).

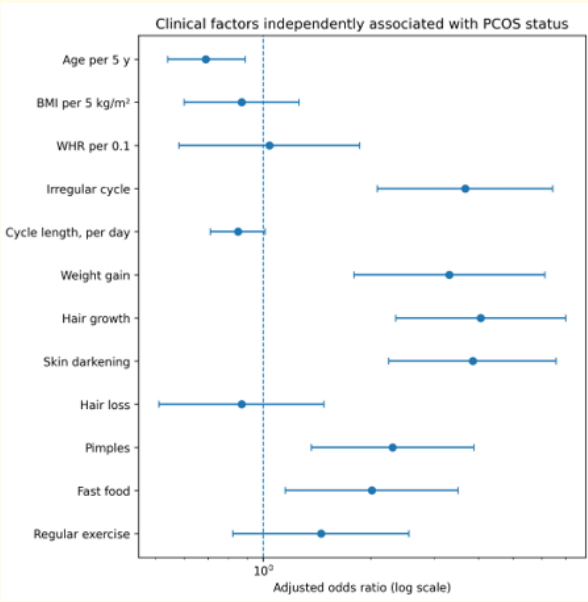


**Figure 1:** Adjusted odds ratios from the multivariable clinical logistic regression model.

Incremental model discrimination

Model A (clinical model) demonstrated strong discrimination (AUC 0.881; 95% bootstrap CI: 0.848, 0.912; Brier score 0.120; sensitivity 0.729; and specificity 0.909). Including biochemical variables did not increase discrimination: The AUC of model B was 0.879 (95% confidence interval 0.845-0.909) and the paired delta AUC was -0.002 (95% confidence interval -0.008 to 0.004) in relative to model A.

Including ultrasound variables greatly enhanced distinction. MODEL C AUC 0.949 SENSITIVITY 0808 SPECIFICITY 0945 in an additional analysis of predictive performance we assessed the predictive performance of clinical tests and prediction. Compared to the clinical model, the AUC improvement was 0.068 (95% CI 0.040-0.097). The calibration intercepts were close to zero and calibration slopes were near one of all models, though these estimates symbolize internal rather than external validation.

The restrictive non-criterion clinical model (Model D) showed useful discrimination even without cycle features and hair growth, evidenced by an AUC of 0.843 (95%CI 0.805-0.877).

By decomposing the ultrasound layer, we can see that inclusion of follicle size and endometrial thickness along with follicle count yielded an AUC of around 0.889 only which was just fair. The clinical model added total follicle count yielded area under curve of approximately 0.952. Thus, most of the ultrasound performance improvement was due to the counting of follicles (Table 4, Figure 2 and 3).

| Model                                 | AUC   | 95% CI      | Brier | Sensitivity | Specificity | Accuracy | PPV   | NPV   |
|---------------------------------------|-------|-------------|-------|-------------|-------------|----------|-------|-------|
| A Clinical                            | 0.881 | 0.848-0.912 | 0.12  | 0.729       | 0.909       | 0.85     | 0.796 | 0.873 |
| B Clinical + biochemical              | 0.879 | 0.845-0.909 | 0.122 | 0.729       | 0.909       | 0.85     | 0.796 | 0.873 |
| C Clinical + biochemical + ultrasound | 0.949 | 0.924-0.969 | 0.074 | 0.808       | 0.945       | 0.9      | 0.877 | 0.91  |
| D Restricted non-criterion clinical   | 0.843 | 0.805-0.877 | 0.144 | 0.644       | 0.874       | 0.799    | 0.713 | 0.835 |

Table 4: Internally validated performance of prespecified classification models.

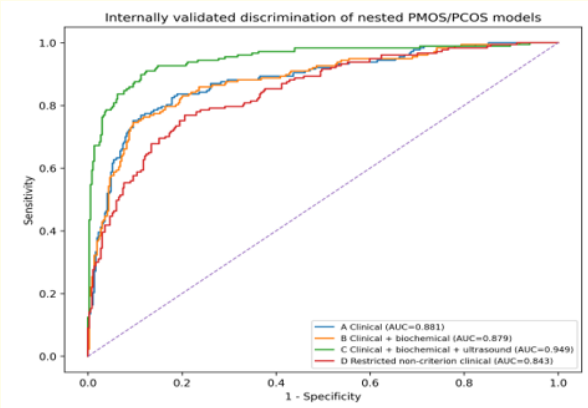


Figure 2: Receiver operating characteristic curves for internally validated Models A-D.

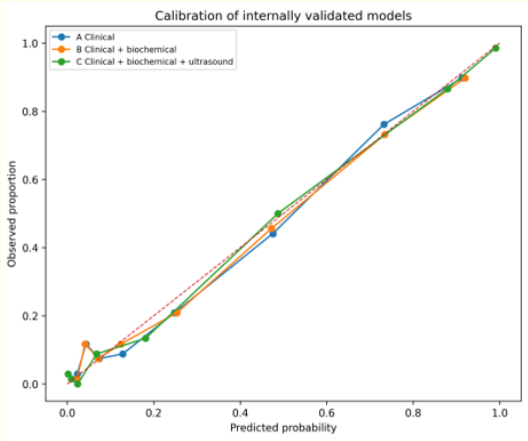


Figure 3: Calibration of the internally validated clinical, clinical-plus-biochemical, and full ultrasound models.

### Sensitivity analysis

After excluding all nine records showing grossly implausible values, the AUCs were 0.882 for model A, 0.879 for model B, 0.949 for model C, and 0.843 for model D. Namely, these values did not substantially change from the primary analysis, suggesting that key conclusions were not driven by the identified errors.

### Discussion

There were three main findings from this. To begin with, easily obtainable clinical information provided high discrimination between records labeled PCOS-positive and PCOS-negative. In addition, the clinical model performance was not enhanced by the addition of the selected hormonal and biochemical variables. Thirdly, variables related to ultrasound had a significant strengthening effect on performance although this was mostly due to follicle count which is quite related to well-established diagnostic constructs. The outcomes of our study suggest the favoring of a clinically interpretable incremental-information framework over another comparison of increasingly complex machine-learning algorithms. Also, this is in line with other evidence that greater algorithmic complexity does not guarantee better clinical prediction performance [18].

The robust adjusted links found in our study for increased body hair, irregular periods, weight gain, dark skin, and pimples correlate strongly with the diverse reproductive, androgens, and metabolic signs of PMOS as described in expert reviews and guidelines [10,11,26,27]. “Obesity is common in women with PCOS and can worsen both metabolic and reproductive manifestations of the disorder, although PCOS also affects women without obesity [4,5]”. While lifestyle management is an important part of care, especially regarding metabolic health, one should not assign a causal interpretation to an observational association with a single dietary variable, which in this instance is fast food consumption. Moran, *et al.* [28] and Cowan, *et al.* [29] are relevant in this regard. In the present data set, many different variables are symptoms and/or signs that may have contributed directly to the original diagnostic label. Thus, age is thought to be inversely associated with the label as well as an adjustment for fast food consumption. It is important to note that these are associations within the present sample and should not be viewed as etiologic effects.

It is appreciable that there is no increase after adding FSH, LH, TSH, AMH, PRL, RBS. In univariate tests, AMH varied markedly between two groups. However, after we controlled for clinical variables, it contributed little. It would be wrong to take this as indication that biochemical evaluation does not have clinical value. Laboratory tests are significant for excluding alternate endocrine disorders and conceptualizing hyperandrogenism and metabolic risk [10,11]. The association of AMH with follicular phenotype of PCOS was strong. Its apparent diagnostic utility was shown in meta-analysis. However, assay-specific thresholds, age effects and overlap with ovarian morphology lessen interpretation as single diagnostic [13,15]. This labeled dataset has no specific biochemical variables, where the ones selected adds cross-validated discrimination over and above information already present in the clinical picture.

It's important to interpret the ultrasound results carefully. The total follicle count had the largest continuous standardized difference between groups of all the features examined and was responsible for virtually all of the increase from an AUC of ~0.88 to ~0.95. The presence of polycystic ovarian morphology is included in Rotterdam diagnostic criteria for PCOS. With improvements in trans-vaginal ultrasound, the recommended follicle number thresholds have become greater than those that were used in the early criteria [9,12,13]. Ovarian morphology remains the central diagnostic domain and, under specific conditions, AMH is permitted as an alternative approach in adults, according to current international guidance [11]. As a result of including follicle count, high classification performance may reflect criterion incorporation rather than a discovery of a new independent diagnostic marker. The restricted clinical analysis and ultrasound decomposition were specifically introduced to highlight this issue and not to conceal it.

The previous machine-learning publications based on same dataset of public 541-participant also can be contextualized with these findings. Previous research has claimed a high level of performance based on Ensembles, Feature Selection, and Explainable AI [16,17].

Studies may show computation feasibility, but their diagnostic accuracy may be hard to convert into clinical usefulness if they include highly diagnostic variables, are not externally validated, or do not reveal their preprocessing decision. It is argued in model guidance on prediction that only discrimination is not enough, but also that calibration, internal validation, applicability and risk of bias are essential [19-23]. The current analysis thus highlights repeated out-of-fold validation, calibration, transparent cleaning of data, and incremental contributions in the domains.

The assessment of data quality revealed just a handful of entries that were very grossly implausible; yet, their presence is significant because public datasets are reused repeatedly all over and preprocessing assumption may get carried over to studies. The main outcomes were robust to excluding every affected record, which bolsters confidence that conclusion is not an artifact of those errors. At the same time, plausibility cleaning cannot recover original values, and lack of access to source records limits verification. This is especially true for a dataset that has gone through redistribution and repeated usage for model development [16,17].

There are quite a few limitations on the study. The dataset is of cross-sectional nature and it is available for public redistribution. Sampling information, study recruitment date information, diagnosis adjudication information, assay methods information, ultrasound protocol information, treatment status information and other original data cleaning information is incomplete. The population from which they are drawn may not be representative of the general population or patients from other geographic areas. Well-established geographic and criterion-related differences in PCOS prevalence exist [1-3]. The version of the dataset that was analyzed did not include total or free testosterone and other key androgen measures limiting further alignment with current diagnostic recommendations [11]. The outcome that is recorded may use some of the same predictors that were used in the models. The small sample size for prediction modeling; no independent external cohort available limits claims of transportability [21,24]. In the end, we cannot link cause and effect to self-reported lifestyle variables.

The study is methodologically strong despite these limitations. We pre-specified the predictor domains, did not redundantly combine derived ratios with their components, restricted missing-value imputation and scaling to the cross-validation pipelines, assessed performance with repeated stratified cross-validation rather than a single split, and reported calibration and Brier scores along with AUC. The choices are in line with existing reporting and evaluation recommendations for prediction-models [19,20,22,23]. The sensitivity analyses examined gross data errors as well as diagnostic-feature overlap, thus enhancing the interpretability of a dataset.

## Conclusion

This secondary analysis of a publicly available 541-participant PCOS dataset found that clinical features alone provided strong internal discrimination; selected biochemical measures offered no measurable incremental improvement; and ultrasound substantially improved classification through follicle count primarily. A limited clinical model retained useful discrimination after removing several strong diagnostic-criterion features. One should not interpret these findings as prospective risk prediction nor as independent validation of a diagnostic tool. They instead show how the apparent classification performance of data-driven PMOS/PCOS depends on which diagnostic domains are included. Furthermore, they stress the need for transparent data-quality assessment, explicit handling of criterion incorporation and external validation in well-characterized contemporary cohorts.

## Bibliography

1. Bozdag G., *et al.* "The prevalence and phenotypic features of polycystic ovary syndrome: A systematic review and meta-analysis". *Human Reproduction* 31.12 (2016): 2841-2855.
2. Lizneva D., *et al.* "Criteria, prevalence, and phenotypes of polycystic ovary syndrome". *Fertility and Sterility* 106.1 (2016): 6-15.

3. Salari N., *et al.* "Global prevalence of polycystic ovary syndrome in women worldwide: A comprehensive systematic review and meta-analysis". *Archives of Gynecology and Obstetrics* 310.3 (2024): 1303-1314.
4. Lim SS., *et al.* "Overweight, obesity and central obesity in women with polycystic ovary syndrome: A systematic review and meta-analysis". *Human Reproduction Update* 18.6 (2012): 618-637.
5. Lim SS., *et al.* "The effect of obesity on polycystic ovary syndrome: A systematic review and meta-analysis". *Obesity Reviews* 14.2 (2013): 95-109.
6. Kakoly NS., *et al.* "Ethnicity, obesity and the prevalence of impaired glucose tolerance and type 2 diabetes in PCOS: A systematic review and meta-regression". *Human Reproduction Update* 24.4 (2018): 455-467.
7. Wekker V., *et al.* "Long-term cardiometabolic disease risk in women with PCOS: A systematic review and meta-analysis". *Human Reproduction Update* 26.6 (2020): 942-960.
8. Teede HJ., *et al.* "Polyendocrine metabolic ovarian syndrome, the new name for polycystic ovary syndrome: A multistep global consensus process". *Lancet* 407.10545 (2026): 2329-2339.
9. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. "Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome". *Fertility and Sterility* 81.1 (2004): 19-25.
10. Legro RS., *et al.* "Diagnosis and treatment of polycystic ovary syndrome: An Endocrine Society clinical practice guideline". *Journal of Clinical Endocrinology and Metabolism* 98.12 (2013): 4565-4592.
11. Teede HJ., *et al.* "Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome". *Journal of Clinical Endocrinology and Metabolism* 108.10 (2023): 2447-2469.
12. Lujan ME., *et al.* "Updated ultrasound criteria for polycystic ovary syndrome: Reliable thresholds for elevated follicle population and ovarian volume". *Human Reproduction* 28.5 (2013): 1361-1368.
13. Dewailly D., *et al.* "Definition and significance of polycystic ovarian morphology: A task force report from the androgen excess and polycystic ovary syndrome society". *Human Reproduction Update* 20.3 (2014): 334-352.
14. Christ JP and Cedars MI. "Current guidelines for diagnosing PCOS". *Diagnostics* 13.6 (2023): 1113.
15. Iliodromiti S., *et al.* "Can anti-Mullerian hormone predict the diagnosis of polycystic ovary syndrome? A systematic review and meta-analysis of extracted data". *Journal of Clinical Endocrinology and Metabolism* 98.8 (2013): 3332-3340.
16. Zigarelli A., *et al.* "Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: Observational study". *JMIR Formative Research* 6.3 (2022): e29967.
17. Khanna VV., *et al.* "A distinctive explainable machine learning framework for detection of polycystic ovary syndrome". *Applied System Innovation* 6.2 (2023): 32.
18. Christodoulou E., *et al.* "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models". *Journal of Clinical Epidemiology* 110 (2019): 12-22.
19. Collins GS., *et al.* "Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement". *Annals of Internal Medicine* 162.1 (2015): 55-63.
20. Collins GS., *et al.* "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods". *British Medical Journal* 385 (2024): e078378.

21. Wolff RF, *et al.* "PROBAST: A tool to assess the risk of bias and applicability of prediction model studies". *Annals of Internal Medicine* 170.1 (2019): 51-58.
22. Van Calster B, *et al.* "Calibration: The Achilles heel of predictive analytics". *BMC Medicine* 17 (2019): 230.
23. Steyerberg EW, *et al.* "Assessing the performance of prediction models: A framework for traditional and novel measures". *Epidemiology* 21.1 (2010): 128-138.
24. Riley RD, *et al.* "Calculating the sample size required for developing a clinical prediction model". *British Medical Journal* 368 (2020): m441.
25. von Elm E, *et al.* "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies". *Lancet* 370.9596 (2007): 1453-1457.
26. Azziz R, *et al.* "Polycystic ovary syndrome". *Nature Reviews Disease Primers* 2 (2016): 16057.
27. Escobar-Morreale HF. "Polycystic ovary syndrome: Definition, aetiology, diagnosis and treatment". *Nature Reviews Endocrinology* 14.5 (2018): 270-284.
28. Moran LJ, *et al.* "Lifestyle changes in women with polycystic ovary syndrome". *Cochrane Database of Systematic Reviews* 7 (2011): CD007506.
29. Cowan S, *et al.* "Lifestyle management in polycystic ovary syndrome - beyond diet and physical activity". *BMC Endocrine Disorders* 23 (2023): 14.
30. Kottarathil P. "Polycystic ovary syndrome (PCOS) [dataset on the Internet]". Kaggle [date unknown].

**Volume 15 Issue 10 October 2026**

**©All rights reserved by Ahed J Alkhatib, *et al.***