© 2026 Nahar S, et al. This is an open access article distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (https://creativecommons.org/licenses/by-nc-sa/4.0/). ISSN 2518-6299 (print), 2520-5048 (online)

Review Article

Artificial Intelligence for Prediction and Early Detection of Polycystic Ovary Syndrome: A Review

Shamsun Nahar[1], Musammat Shamima Akter[2], Sabab Aosaf[3]

Abstract

Polycystic ovary syndrome (PCOS) is one of the commonest endocrine disorders in women of reproductive age, yet diagnosis is often delayed because symptoms are varied and routine care may fail to integrate menstrual history, androgen excess, laboratory findings, and ovarian imaging. Earlier recognition matters because PCOS affects fertility, metabolic health, pregnancy, psychological well-being, and quality of life. AI has therefore attracted growing interest in prediction and early detection. Relevant tools include machine-learning models for structured clinical data, deep-learning systems for ultrasound images, natural-language processing of free-text reports, and digital platforms that generate risk scores from symptoms or menstrual-tracking data. Published studies suggest promise in symptom-based triage, EHR case finding, ultrasound standardization, and emerging omics or mobile-health approaches. At the same time, literature has important weaknesses. Many studies are retrospective, single-center, or based on small, curated datasets. Diagnostic definitions are not always consistent, external validation is limited, and only a minority of studies explain how predictions would be used in real clinical practice. There are also age-specific concerns: adult criteria cannot simply be transferred to adolescents, because current guidance does not support the use of ovarian morphology or anti-mullerian hormone (AMH) as diagnostic criteria in adolescents. At present, AI should be seen as an aid to earlier case finding, triage, and standardized interpretation rather than as a replacement for clinical judgment. The most useful future systems will be the ones that are clinically understandable, fair across patient groups, validated in diverse settings, and tested in routine care pathways.

Keywords: Polycystic ovary syndrome, PCOS, Artificial intelligence, Machine learning, Early detection, Prediction, Ultrasound, Electronic health records.
Received: September 12, 2025; **Accepted:** October 29, 2025
https://doi.org/10.3329/emcj.v11i1.90025

Introduction

Polycystic ovary syndrome (PCOS) is a chronic and heterogeneous condition that affects reproductive, metabolic, dermatologic, and psychological health. In daily practice, patients may first come with irregular menstrual cycles, infertility, hirsutism, acne, weight gain, insulin resistance, mood symptoms, or an ultrasound report that mentions polycystic ovaries. This broad clinical range is one reason why PCOS is both common and often missed. Current public-health and epidemiologic sources suggest that PCOS affects about one in ten women of reproductive age, although the exact prevalence varies with the diagnostic criteria used[1,2]. Reviews and guideline documents also emphasize that many women remain undiagnosed for years[3-8].

The diagnostic problem in PCOS is not only whether a patient meets a definition at one point in time. The harder problem is how to recognize the syndrome early enough to improve care. Earlier recognition can support counseling about lifestyle, menstrual health, fertility, contraception, metabolic risk, and long-term follow-up. It can also reduce repeated visits in which individual symptoms are treated separately without bringing them together under one diagnosis[3-8]. Because PCOS is common, multifactorial, and diagnosed from information that already exists in history, examination, laboratory tests, and imaging, it is a natural target for AI-based support.

In this field, AI does not refer to one single method. It includes machine-learning models built from symptoms and laboratory tests, deep-learning models that read ultrasound images, natural-language processing systems that convert free-text reports into structured information, and digital tools that estimate risk from menstrual and symptom data[9-11]. These approaches are attractive because they may help at different stages of care. A non-invasive symptom-based model may help in community or primary-care triage. An EHR-based model may help a hospital identify women who have suggestive features but no diagnosis. An ultrasound system may help standardize morphology

[1]Professor, Department of Obstetrics & Gynecology, Central Medical College & Hospital, Cumilla, Bangladesh.
[2]Professor, Department of Obstetrics & Gynecology, Eastern Medical College & Hospital, Cumilla, Bangladesh.
[3]MSc Student, Department of Computer Science and Engineering, BUET, Dhaka, Bangladesh.

Address of Correspondence: Prof. Dr. Shamsun Nahar, Professor & Head, Department of Obstetrics & Gynecology, Central Medical College, Cumilla, Bangladesh. Email: drsn101@gmail.com

assessment after imaging has already been performed[11-14]. This review summarizes where AI may currently help in PCOS, which approaches seem most promising, which claims require caution, and what is still needed before safe routine use[8-10,15-18].

Why early diagnosis matters in PCOS

Early diagnosis matters because PCOS is not only a fertility disorder. It is associated with ovulatory dysfunction, infertility, obesity, insulin resistance, impaired glucose tolerance, type 2 diabetes, sleep problems, anxiety, depression, and reduced quality of life[5-8]. Some women seek help for cosmetic symptoms such as acne or excess hair growth, while others present later with infertility or metabolic disease. When diagnosis is delayed, opportunities for earlier counseling and prevention may be lost.

Recent guideline updates and major reviews have strengthened this point. The 2023 international guideline process, supported by broad evidence synthesis, emphasized that care for PCOS should include attention to metabolic, reproductive, and psychological health, not only menstrual regularity or fertility[7,8,19]. In practice, this means that earlier recognition may prompt screening for diabetes risk, weight-related counseling, assessment of mental health, and discussion of pregnancy planning before complications arise. Patient experience studies show how long the diagnostic journey can be. Women often report that symptoms are normalized, explained away as isolated problems, or addressed by different clinicians without a joined-up assessment. Diagnostic delay has been linked with dissatisfaction, repeated consultations, and a feeling that concerns were not taken seriously[3,4]. For clinicians, this is important because an AI tool that shortens time to recognition may improve the patient's journey even before it changes to hard clinical outcomes. Underdiagnosis also has a population-level dimension. When researchers compare expected community prevalence with coded diagnoses in healthcare records, routine systems usually capture far fewer cases than would be expected[20,21]. This gap means that many women with likely PCOS are invisible in ordinary datasets and ordinary care. It also means that AI models trained only on formal diagnoses may learn patterns of who gets recognized by the health system, not just patterns of disease itself[20-22]. For early detection, that distinction is crucial.

Diagnostic criteria and what they mean for AI

Any review of AI in PCOS must begin with the diagnostic criteria, because model performance is only meaningful if the outcome label is clinically sound. Current international guidance states that in adults the diagnosis is made after excluding other causes and identifying two of the following three features: ovulatory dysfunction, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology[8,19]. This matters for AI because no single laboratory value, symptom, or ultrasound image is equal to the full clinical diagnosis. A model may help estimate risk, but it still has to fit within a proper diagnostic pathway.

A major change in recent guidance is the role of AMH. In adults, AMH may be used as an alternative to ultrasound for defining polycystic ovarian morphology in an appropriate clinical context[8,19]. This is important for AI because it supports multimodal adult models that combine symptoms, biochemical markers, and ovarian morphology. At the same time, the message is not that AMH alone is enough. A strong AMH-based classifier may be useful in endocrinology or infertility practice, but it cannot replace the need to consider menstrual history, androgen excess, and exclusion of other disorders[8,19,23].

Adolescent diagnosis needs special caution. Current evidence-based recommendations state that adolescents should be diagnosed only when both irregular menstrual cycles and hyperandrogenism are present after excluding mimicking disorders. Ovarian morphology and AMH should not be used diagnostically in this age group[8,19,24]. This has important implications for AI. A model that performs well in adult ultrasound data may be inappropriate for adolescent patients, even if its accuracy looks impressive on paper. Adult and adolescent tools should therefore be treated as separate clinical products, not as simple variants of the same model. The criteria also remind us that PCOS is a diagnosis of inclusion plus exclusion. Thyroid disease, hyperprolactinemia, nonclassic congenital adrenal hyperplasia, androgen-secreting tumors, and other endocrine disorders may mimic parts of the syndrome[8,19,24]. Some AI studies ignore this issue and focus only on separating women already labeled as PCOS from those labeled as controls. Clinically, however, a useful early-detection system should also help avoid false reassurance and false diagnosis. This is where differential-diagnosis models, such as steroid metabolomics combined with machine learning, may become especially valuable[25].

Review scope and overall evidence landscape

This paper is a narrative review intended for practical clinical interpretation rather than exhaustive meta-analysis. It focuses on literature available up to 6 March 2026 and emphasizes studies relevant to structured clinical datasets, EHRs, ultrasound, omics, and digital screening tools, alongside current guideline and methodological papers[8,15-19,24].

Table-I: Diagnostic context for PCOS and the main implications for AI systems

Clinical setting	Current guidance	What it means for AI	Common mistake
Adult diagnosis	Use two of three domains after exclusion of mimics: ovulatory dysfunction, hyperandrogenism, and polycystic ovarian morphology.	Advanced models can integrate symptoms, laboratory findings, ultrasound results, and AMH-based indicators when appropriate.	However, relying on a single feature-particularly ovarian morphology does not constitute a complete diagnosis.
Adult use of AMH	AMH may be used as an alternative to ultrasound for defining polycystic ovarian morphology in adults within the wider diagnostic framework.	Biomarker-rich models may improve triage or classification when reliable assays are available.	Assuming AMH alone is enough for universal screening in all settings.
Adolescents	Diagnosis should rely on menstrual irregularity and hyperandrogenism; ovarian morphology and AMH should not be used diagnostically.	Adolescent models should focus on menstrual pattern, androgen-related features, exclusion of mimics, and follow-up.	Applying adult image-based or AMH-based tools to adolescents.
EHR and claims data	Routine records often lack enough variables to recreate formal criteria and may undercapture the true burden of disease.	Prediagnostic models must handle missing data, label uncertainty, and incomplete documentation.	Treating absence of a coded diagnosis as proof that PCOS is absent.
Community or app-based screening	These tools should be used as triage aids, not as stand-alone diagnostic systems.	Non-invasive risk scores may shorten the path to assessment and counseling.	Overstating consumer tools as replacements for clinical evaluation.

Literature has expanded quickly over the past few years. Systematic reviews show that investigators have tested a wide range of models, including logistic regression, decision trees, random forests, support vector machines, gradient boosting, convolutional neural networks, and hybrid multimodal systems[9-11]. Many papers report high performance, often using measures such as accuracy or area under the receiver operating characteristic curve (AUC). For clinicians, the important point is that these numbers show technical promise, but they do not by themselves prove clinical usefulness.

Barrera and colleagues reviewed 31 studies and found a wide spread of performance, with AUCs from 0.73 to 1.00, diagnostic accuracy from 89% to 100%, sensitivity from 41% to 100%, and specificity from 75% to 100%[9]. Ghaderzadeh and colleagues later expanded the field to 80 studies and again found strong reported performance, especially for ultrasound imaging and structured tabular data[10]. Wang and colleagues described the broader movement of AI in PCOS management from proof of concept toward more clinically integrated systems[11]. Taken together, these reviews show that the field is active and promising, but also uneven in quality.

How AI is being used for prediction and early detection

Clinical, symptom-based, and laboratory models

A large part of the PCOS-AI literature uses structured clinical data. These models are built from variables such as age, body mass index, menstrual pattern, acne, hirsutism, weight change, blood pressure, lipid profile, androgen levels, AMH, and ultrasound-derived follicle counts[9-12,26-31]. From a clinical point of view, this is a sensible starting point because these are the same types of information doctors already use in ordinary diagnostic reasoning. A key practical advantage of structured data models is their ability to operate in a stepwise manner. A first model can estimate risk using only non-invasive variables such as cycle history and symptoms. A second model can refine the estimate once laboratory results or ultrasound findings become available. This staged approach fits real practice, where many patients are first seen in primary care or through general gynecology rather than in a reproductive endocrinology clinic[12,29].

Zigarelli and colleagues provided an early and clinically useful example. Using data from 541 women from 10 hospitals in Kerala, India, they built machine-aided self-diagnostic models for PCOS[12]. Their non-invasive model achieved an accuracy of about 81% to 82.5%, while accuracy increased to about 87.5% to 90.1% when more invasive clinical variables were added. Important predictors included irregular menstrual cycles, acne, hirsutism, acanthosis nigricans, weight gain, and age[12]. The study is valuable because it shows that meaningful triage may be possible even before hormone testing or imaging is complete.

Other symptom-based systems point in the same direction. Tiwari and colleagues proposed a smart diagnostic system that combined machine learning

OPEN ACCESS

with symptoms and clinical data[26]. Elmannai and colleagues focused on optimized feature selection and explainable methods to improve PCOS detection[27]. Khanna and colleagues developed an explainable framework that highlighted the importance of combining interpretable modeling with strong performance[28]. These studies do not all answer the same clinical question, but together they show that useful prediction is possible from relatively familiar clinical variables.

More recent work has pushed this approach toward earlier detection in more formal clinical settings. Panjwani and colleagues reported an optimized model for early detection in symptomatic women and found that obesity and high cholesterol were influential variables, with an AUC of 0.93[29]. Yao and colleagues developed and externally validated an explainable machine-learning model and nomogram for early detection, reporting AUCs of about 0.92 in both internal and external validation[30]. Their work is especially important because it moved beyond one-center development and included a more clinically interpretable output.

Laboratory-rich models may be especially useful in secondary or tertiary care. Tong and colleagues built a machine-learning diagnostic model in which AMH alone performed strongly, while multivariable endocrine data improved discrimination further[31]. The clinical lesson is simple: symptom-based models may help identify whom to investigate, while laboratory-rich models may help strengthen or weaken suspicion once evaluation has started. However, these biomarker-heavy studies are often based on smaller samples than EHR or imaging studies, so performance estimates should be interpreted cautiously[30,31].

Overall, structured clinical and laboratory models are among the most immediately usable forms of AI in PCOS. They fit the way clinicians already think, they can be deployed in stages, and they can often be explained in terms of familiar clinical features. Their main weaknesses are limited generalizability, variable dataset quality, and the temptation to overstate classification accuracy as proof of real-world early detection[9,10,12,26-31].

Electronic health records, administrative data, and natural-language processing

Among all PCOS-AI applications, EHR-based models come closest to the real meaning of early detection. Instead of separating known cases from known controls at the same point in time, these models attempt to identify women who are likely to have PCOS before the diagnosis is formally entered into the record[13]. That shift is clinically important because it matches the health-system problem: women often have suggestive symptoms and laboratory patterns long before someone names the syndrome.

Zad and colleagues studied 30,601 women aged 18 to 45 years in a large urban health system and trained several machine-learning models to predict PCOS ahead of clinical diagnosis[13]. Average AUCs were roughly 0.80 to 0.85 across different outcome definitions, and obesity together with endocrine measurements were important features. This work is notable not because it produced the highest published number, but because it addressed a meaningful clinical task. A model like this could be used to search records for women who need more focused review. Natural-language processing offers another practical route. In many hospitals, ultrasound findings relevant to PCOS are buried in free-text reports. Cheng and Mahalingaiah showed that text-mining methods could identify polycystic ovary morphology from more than 39,000 pelvic ultrasound reports with high accuracy[32]. This does not diagnose PCOS by itself, but it turns narrative radiology text into a structured flag that can be combined with cycle history, laboratory findings, and coded symptoms.

These approaches also reveal a major weakness in routine data. Atiomo and colleagues demonstrated that research on PCOS using EHR data is often constrained by incomplete data capture and substantial variability in how diagnoses are documented[21]. Hatoum and colleagues found that health-system and insurer records capture far fewer cases than population-based estimates would suggest[20]. In simple terms, many women who likely have PCOS are not labeled as such in ordinary records, and many records do not contain all the variables required to recreate formal diagnostic criteria[20,21]. There is also a fairness concern. Silva and colleagues reported underdiagnosis patterns associated with individual and spatial social vulnerability[22]. If a model is trained on historical diagnoses without accounting for these biases, it may learn who was previously recognized by the system rather than who truly had the syndrome. For clinicians, this means that an EHR model should be seen as a prompt for review, not as a silent automated verdict[20-22].

Ultrasound and imaging AI

Ultrasound has long been central to PCOS assessment in adults, but it is also one of the most operator-dependent parts of the workup. Follicle counting takes time, image quality varies, and interpretation can differ between clinicians and between machines. These problems make ultrasound a natural target for AI[9-11,14,33]. In theory, an imaging system can improve speed, consistency, and reproducibility.

Table-II: Representative studies using structured data, EHRs, and natural-language processing for earlier PCOS recognition

Study and data source	Model	Main result	Clinical comment
Zigarelli et al.[12], 541 women from 10 hospitals, symptom-based and clinical tabular data	CatBoost	About 81% to 82.5% accuracy with non-invasive variables; about 87.5% to 90.1% after invasive clinical variables were added.	Good example of practical symptoms-first triage, but still retrospective and geographically specific.
Panjwani et al.[29], symptomatic dataset with 12 attributes	Weighted optimized ensemble	Accuracy 92.8%; AUC 0.93; obesity and high cholesterol were influential features.	Useful early-screening concept, but prospective testing is still limited.
Tong et al.[31], endocrine biomarker dataset	Machine-learning classifier	AMH alone reached AUC 0.925; multivariable endocrine modeling improved discrimination further.	Promising for specialist settings, but small biomarker studies may overfit.
Yao et al.[30], multicenter early-detection cohort	Explainable gradient-boosted model plus nomogram	AUC 0.919 internally and 0.923 externally; DHEAS, AMH, triglycerides, and age were important.	One of the stronger validation designs in the structured-data literature.
Zad et al.[13], 30,601 women in an urban EHR system	Logistic regression, support vector machine, random forest, gradient boosting	Prediagnostic average AUCs around 0.80 to 0.85 across outcome definitions.	Closer to true early detection because prediction preceded formal diagnosis.
Cheng and Mahalingaiah[32], 39,093 ultrasound reports from 25,535 women	Text classification	97.6% and 96.1% accuracy on a 1,000-report evaluation set for polycystic ovary morphology.	It is a useful way to structure radiology text, but it classifies reports rather than images.

Many imaging studies report excellent performance, especially with deep learning. However, clinicians should interpret these results cautiously, because accuracy is often inflated in small or curated datasets and may drop in routine practice with more variable scanning techniques, devices, and patient characteristics[9,10]. The most encouraging imaging evidence comes from stronger validation designs.

Zhao and colleagues developed a deep-learning model for PCOS ultrasound recognition and tested it in a prospective, multicenter setting. They reported AUCs above 0.95 in training and validation cohorts and showed that the model could assess a single ovary much faster than clinicians. This kind of study is more persuasive than small retrospective image-classification papers because it begins to address transportability across centers[14].

Imaging is also moving toward multimodal integration. Agirsoy and Oehlschlaeger integrated ultrasound features with clinical data instead of relying on imaging alone as a definitive answer[33]. This is a clinically sensible direction. In adults, ovarian morphology is one diagnostic domain, not the whole syndrome. A useful imaging system should therefore fit into a broader pathway that also considers menstrual history, androgen excess, and exclusion of mimicking disorders[8,19,33]. At present, ultrasound AI may be one of the most useful practical tools in specialty care, especially for standardizing interpretation and reducing operator dependence. But it must be used in the right patients and in the right age group. In adolescents, current guidelines do not support using ovarian morphology for diagnosis, so adult imaging models should not be transferred into adolescent care simply because they appear accurate[8,14,19,24].

Omics, biomarkers, and multimodal molecular models

A newer frontier in PCOS-AI research is the use of high-dimensional molecular data. These studies examine broad panels of lipids, steroids, lipoproteins, or other biomarkers and then apply machine learning to detect patterns that are hard to see by routine clinical inspection. The appeal of this approach is biological depth. It may help identify subtypes of disease or improve differential diagnosis in complex cases[25,34,35].

Chen and colleagues used lipid-omics methods to screen serum biomarkers in women with PCOS and reported a test-set AUC of 0.815, with specificity higher than sensitivity[34]. These results are not as striking as the best ultrasound or structured-data models, but the study is still important. It suggests that lipid signatures may contain useful information about disease biology and may add value when routine clinical data are not decisive.

Yan and colleagues went a step further by combining AMH with lipoprotein subfractions in women with high follicle number per ovary and found that performance improved when the molecular panel was added. This is clinically interesting because it shows how AI can combine established markers with newer molecular features rather than treating them as competitors. The downside is that such models depend on specialized assays, high analytical quality, and careful external validation[35].

Steroid metabolomics may be especially valuable in differential diagnosis. Bachelot and colleagues combined metabolomics with machine learning to distinguish nonclassic 21-hydroxylase deficiency from PCOS[25]. This is highly relevant to clinical endocrinology because early detection is useful only if it does not increase misdiagnosis of other androgen excess disorders. A model that helps separate PCOS from an important mimic may sometimes be more clinically valuable than one that simply reports a high overall accuracy in a case-control dataset.

In summary, omics-based AI is promising but not yet ready for broad screening. It may have greater short-term value in specialist practice, difficult diagnostic cases, and research on biologic subtypes of PCOS. For everyday early detection, it is still more complex, costly, and less validated than symptom-based, EHR-based, or ultrasound-based approaches[25,34,35].

Table-III: Representative imaging, omics, and digital-health studies relevant to early PCOS recognition

Study and method	Model	Main result	Main limitations or clinical issue
Agirsoy and Oehlschlaeger[33], ultrasound plus clinical features	Multimodal machine-learning classifier	Shows the move from image-only diagnosis toward combined morphology and clinical context.	Performance must still be interpreted in light of adult-only applicability of morphology-based criteria.
Zhao et al.[14], prospective multicenter ultrasound imaging study	Deep learning	AUC 0.953 in training, 0.973 internally, and 0.967 externally; F1 96.9% in validation cohorts; faster than clinician assessment.	Needs broader device-level and setting-level validation before routine use.
Chen et al.[34], lipid-omics serum biomarkers	Ensemble machine learning	Test-set AUC 0.815; accuracy 0.74; specificity 0.88; sensitivity 0.70.	Omics pipelines are costly and need strong analytical standardization.
Yan et al.[35], lipoprotein subfractions plus AMH	Feature-selected machine learning	AUC improved from 0.750 to 0.874 when seven lipoprotein subfractions were added to AMH.	External validation remains limited and criterion dependence is high.
Bachelot et al.[25], steroid metabolomics for differential diagnosis	Machine-learning classifier	Separated nonclassic 21-hydroxylase deficiency from PCOS in the external test cohort.	Very useful clinically, but targeted differential diagnosis is not the same as population screening.
Rodriguez et al.[36], menstrual-app risk scoring	Bayesian network-based digital triage	Correlation of 0.82 with physician-assigned risk scores in a virtual functionality assessment.	Early proof of concept only; real-world impact remains uncertain.

Mobile health, wearables, and digital self-screening tools

Digital tools may offer the earliest point of all because they operate before the patient enters a formal diagnostic pathway. Menstrual-tracking apps, symptom checkers, and other connected tools can capture cycle irregularity, acne, body-weight change, and other features that may raise suspicion of PCOS[36-38]. For many women, this may be the first time their symptoms are brought together in one place.

Rodriguez and colleagues evaluated a mobile-health approach that used a Bayesian network to estimate PCOS risk from menstrual and symptom information[36]. The study was small and based on a virtual functionality assessment, but it showed proof of concept for digital risk scoring. The message is not that an app can diagnose PCOS. Rather, an app may identify women who should seek structured clinical evaluation. Scoping reviews suggest that women with PCOS already use digital technologies widely for information seeking, self-monitoring, and symptom management[37,38]. This creates not only an opportunity for earlier education and triage but also carries risks. Consumer-facing tools can easily overpromise, especially when the underlying evidence is thin or when users are not clearly told that formal diagnosis requires clinical assessment[36-38].

OPEN ACCESS

Digital screening is best understood as the front end of a longer care pathway. It may shorten the time from symptoms to consultation, but it should not replace examination, appropriate laboratory testing, or exclusion of mimicking disorders[8,36-38].

What the current literature really shows

The published literature on AI in PCOS is encouraging, but it is easy to read it too optimistically. A paper may report high accuracy, high sensitivity, or a very strong AUC and still have limited clinical value. To judge these studies well, clinicians need to look beyond the headline number and ask a few basic questions: What exactly was the model predicting? Was the diagnosis defined clearly? Was the model tested outside the original dataset? Could the result change care?[9,10,15-18]

Diagnostic labels and outcome definitions

The first major issue is label quality. Not all studies use the same definition of PCOS, and not all papers describe clearly how cases and controls were assigned[9,10]. Some use Rotterdam criteria, some use local clinical labels, and some rely mainly on dataset annotations that do not fully explain the diagnostic process. If the outcome is vague or inconsistent, a model can appear accurate while actually learning a narrow institutional pattern[8-10].

This problem has become even more important since the 2023 guideline update. Adult models may include AMH in a clinically appropriate way, but adolescent models should not use AMH or ovarian morphology diagnostically[8,19,24]. A study that mixes these contexts without care may be difficult to apply in real practice. For doctors, the key point is simple: before trusting performance metrics, first ask whether the model matches the current diagnostic rules for the patient group in which it will be used[8,19,24].

Sampling, class balance, and generalizability

The second major issue is generalizability. Many PCOS-AI studies are retrospective and single-center, and some are based on public datasets that may not reflect routine clinical populations[9,10]. Case-control separation may be cleaner in research datasets than in everyday practice, where obesity, irregular cycles, infertility, thyroid disease, and incomplete records create diagnostic overlaps. This can make performance look better in the study than it will look after deployment.

Generalizability also depends on who is in the dataset. If the study population is narrow by age, ethnicity, referral pathway, or disease severity, the model may perform less well in other settings. Studies by Yao and Zhao are important partly because they included external or multicenter validation, which is still uncommon in the field[14,30]. Such work should become standard rather than exceptional.

Calibration, thresholds, and clinical usefulness

A third issue, often overlooked, is calibration. Discrimination tells us whether a model ranks patients reasonably well from lower to higher risk. Calibration tells us whether the predicted risk is numerically close to the real probability. In practice, a clinician may be more interested in knowing whether a 30% predicted risk truly means about 30% risk than in a small difference in AUC[15,16]. Many PCOS studies report discrimination but say little about calibration, predictive values, or decision thresholds.

This gap matters because an early detection tool is useful only when it leads to action. A hospital may decide that women above a certain risk score should have structured endocrine review. A primary-care tool may prompt laboratory testing or gynecology referral. Without clear thresholds and workflow links, performance numbers remain hard to use[15-18].

Explainability and clinician trust

Explainability is especially important in PCOS because the syndrome is clinically diverse and often debated at the margins. Doctors are more likely to trust an AI model if they can see that it is relying on plausible inputs such as menstrual irregularity, biochemical hyperandrogenism, obesity, AMH in adults, or clearly defined imaging features. Explainable models and feature-attribution methods can therefore improve both trust and error checking[27,28,30].

Explainability is also a safeguard. If a model is making predictions based on hidden artifacts, local coding habits, or age-inappropriate variables, clinicians need a chance to notice that before the tool is used widely. For this reason, understandable models may sometimes be preferable to slightly more accurate black-box systems, especially during early clinical adoption[15-18,27,28].

Fairness, underdiagnosis, and data provenance

Fairness is not a side issue in PCOS. Underdiagnosis is already shaped by healthcare access, documentation quality, and social vulnerability[20-22]. If training labels reflect these inequities, an AI model may reproduce them. For example, a model may learn that women without a formal diagnosis are low risk, even when the real problem is that they were never fully evaluated.

This is why data provenance matters. Researchers need to show where the labels came from, which patients were included, how missing data were handled, and whether performance changes across subgroups. Routine reporting of subgroup

performance, bias assessment, and applicability should become part of good PCOS-AI research rather than an optional extra[15,16,22].

Special issues in adolescents

Adolescent PCOS remains one of the clearest examples of why disease context matters more than technical performance alone. Menstrual irregularity is common in normal maturation, and overdiagnosis can cause unnecessary anxiety. Current adolescent recommendations therefore require a careful combination of persistent menstrual irregularity, hyperandrogenism, and exclusion of other causes, while specifically avoiding ovarian morphology and AMH as diagnostic criteria[24].

A model developed in adult infertility practice may therefore be clinically inappropriate in adolescents even if its numerical performance is strong. The future need is for adolescent-specific models that support follow-up, risk enrichment, and timely specialist review rather than premature labeling[24]. In this group, clinical judgment and longitudinal observation remain essential.

Table-IV: Common translational gaps in PCOS-AI research and what is needed before implementation

Main gap	Why it matters	What should be done	Useful guidance or source
Nonstandard or weak outcome labels	Performance cannot be interpreted well if the model is predicting a vague or outdated definition of PCOS.	Use current guideline-based labels, describe adjudication clearly, and separate adult from adolescent use cases.	Teede et al.[8]; Pena et al.[24]
Single-center or convenience samples	Models may capture local workflow or referral bias rather than transportable disease signals.	Use multicenter development, temporal validation, and true external validation.	Yao et al.[30]; Zhao et al.[14]
Poor calibration and unclear thresholds	A high AUC does not show when to act or how many false positives to expect.	Report calibration, predictive values at likely prevalence, and threshold-linked actions.	TRIPOD+AI[15]; DECIDE-AI[18]
Limited explainability	Clinicians need to know whether the model is relying on plausible and age-appropriate features.	Use interpretable models where possible and provide feature attribution or saliency analysis.	Elmannai et al.[27]; Yao et al.[30]
Inherited underdiagnosis bias in EHR labels	Historical diagnosis patterns may reflect inequity in access rather than true disease absence.	Audit subgroup performance, test for label bias, and compare against adjudicated or enriched cohorts.	Atiomo et al.[21]; Silva et al.[22]
Adult models applied to adolescents	Image-based or AMH-heavy tools may conflict with current adolescent recommendations.	Create adolescent-specific models focused on menstrual irregularity, hyperandrogenism, and follow-up.	Pena et al.[24]
Lack of prospective impact studies	Even accurate models may fail to improve time to diagnosis, referrals, or outcomes.	Run implementation studies measuring workflow efficiency, diagnostic delay, patient experience, and downstream care.	DECIDE-AI[18]; World Health Organization[1]

Clinical integration: where AI can help now

The most realistic role for AI in PCOS today is decision support rather than autonomous diagnosis. In primary care, symptom-based tools may identify women who need structured evaluation; in hospitals, EHR and natural-language processing systems may support case finding; and in specialty practice, ultrasound or molecular models may aid standardized assessment and difficult differential diagnosis[8,13,14,25,32,33,36-38].

This layered view is clinically useful because it matches the real patient pathway. A woman may first notice irregular cycles or acne and use a digital tool. She may then present to primary care, where a non-invasive risk score or checklist prompts targeted questioning. If laboratory tests or ultrasound are ordered, those results can be added to the assessment. If suspicion remains high, specialist review can confirm the diagnosis, exclude other causes, and start long-term management[8,13,19,36-38].

In this model, AI helps organize care at several points, but it does not replace clinical reasoning. A good tool should make the next step clearer. It should not tempt clinicians to skip careful history taking, physical examination, exclusion of mimicking disorders, or age-specific judgment. This is especially important in adolescents and in women with overlapping endocrine problems[8,19,24,25]. For health systems, the best near-term use case may be active case finding through records. Women with repeated menstrual complaints, infertility visits, androgen-related symptoms, obesity, insulin

OPEN ACCESS

resistance, or suggestive ultrasound reports can be flagged for review[13,32]. For imaging departments, the best near-term use case may be standardization of ovarian morphology assessment in adults[14,33]. For specialist endocrine practice, steroid metabolomics or other biomarker-rich models may eventually improve difficult differential diagnoses[25,34,35]. The clinical question is therefore no longer whether AI can separate labeled PCOS cases from controls in research datasets. The more important question is where the tool fits into everyday workflow, who should act on the result, and whether the result shortens time to diagnosis or improves the quality of follow-up. These are the outcomes that will decide whether PCOS-AI moves from promising papers to useful practice[13-18].

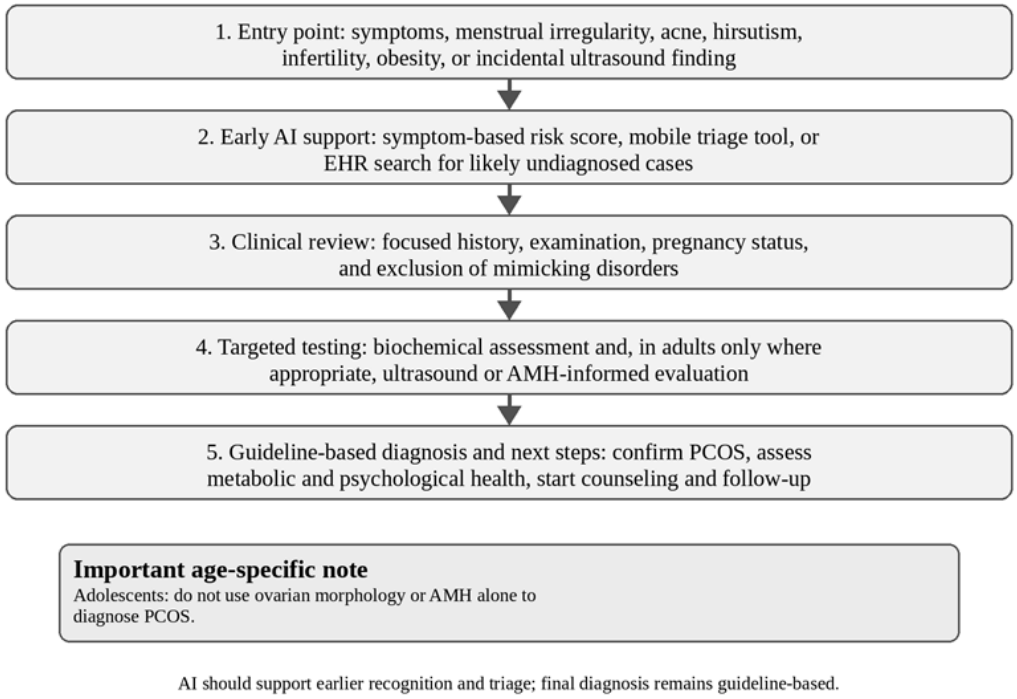

Figure-1: Practical AI-enabled pathway for earlier recognition of PCOS

Future directions

The next stage of PCOS-AI research should be more clinical and less purely technical. First, future studies should use clear, contemporary diagnostic definitions and separate adults from adolescent applications from the start[8,19,24]. This may sound obvious, but it is still not routine. Without clear labels, even a sophisticated model is hard to trust.

Second, multicenter validation must become the rule. Datasets should include a wider range of ages, body sizes, ethnic and social groups, referral pathways, and healthcare settings. This is necessary not only to improve transportability but also to reduce the risk that the model simply reflects one clinic's local habits[13,14,30]. External validation should not be treated as an optional bonus result.

Third, reporting quality needs to improve. AI studies in medicine now have established methodological frameworks. TRIPOD+AI addresses reporting of prediction models, PROBAST+AI focuses on bias and applicability, STARD-AI guides diagnostic accuracy studies, and DECIDE-AI supports early clinical evaluation of decision-support systems[15-18]. Using these frameworks would make PCOS studies easier to compare and easier to judge.

Fourth, future work should measure impact in practice. The key outcomes are not only AUC and accuracy. They include time to diagnosis, time to referral, appropriate metabolic screening, appropriate fertility counseling, patient understanding, clinician workload, and unintended harms such as over testing or false reassurance[15-18]. A model that performs well statistically but does not improve any of these outcomes may have limited value.

Finally, AI for PCOS should move toward models that are understandable to clinicians and acceptable to patients. In a syndrome as common and nuanced as PCOS, trust, context, and workflow fit may matter as much as algorithmic sophistication[15-18,27,28,30].

Conclusion

Artificial intelligence has real potential to improve the prediction and early detection of PCOS, but the present evidence should be read with balance. The strongest current use cases are symptom-led triage, EHR-enabled case finding, and adult ultrasound support. These approaches fit natural points in the

care pathway and have the best chance of shortening time to recognition[12-14,30,32,33,36].

At the same time, the field is not yet mature enough for AI to replace guideline-based diagnosis. Many studies still rely on retrospective datasets, weak or inconsistent labels, limited external validation, and incomplete reporting of calibration, fairness, and implementation strategy[9,10,15-18]. For clinicians, the safest view is that AI can help find women who deserve earlier evaluation, but the diagnosis of PCOS still depends on proper clinical assessment, exclusion of mimicking disorders, and age-appropriate interpretation[8,19,24,25].

The future of this field will depend less on achieving another impressive accuracy value and more on producing tools that are simple to understand, externally validated, fair across patient groups, and useful in real practice. If those conditions are met, AI may become an important support for earlier, better, and more consistent care for women with PCOS[13-18,30].

Key Messages for Clinicians

• PCOS is common and often diagnosed late, so earlier recognition matters for reproductive, metabolic, and psychological care.

• The strongest near-term AI use cases are symptom-based triage, EHR-enabled case finding, and adult ultrasound support.

• High accuracy in a research dataset does not by itself mean a model is ready for practice; external validation, calibration, and fairness are also needed.

• Adult models should not simply be transferred to adolescents without age-specific design and validation.

References

1. World Health Organization. Polycystic ovary syndrome. Geneva: World Health Organization; 2025. Available at: https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome [Accessed on August 12, 2025]
2. Neven ACH, Forslund M, Ranasinha S, Sethi P, Dhungana RR, Mousa A, et al. Prevalence of polycystic ovary syndrome: a global and regional systematic review and meta-analysis. Hum Reprod Update. 2026: dmaf030. doi: 10.1093/humupd/dmaf030.
3. Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed Diagnosis and a Lack of Information Associated with Dissatisfaction in Women with Polycystic Ovary Syndrome. J Clin Endocrinol Metab. 2017; 102 (2): 604-12. doi: 10.1210/jc.2016-2963.
4. Presswala B, De Souza LR. The diagnostic experience of polycystic ovary syndrome: A scoping review of patient perspectives. Patient Educ Couns. 2023; 113: 107771. doi: 10.1016/j.pec.2023.107771.
5. Joham AE, Norman RJ, Stener-Victorin E, Legro RS, Franks S, Moran LJ, et al. Polycystic ovary syndrome. Lancet Diabetes Endocrinol. 2022; 10 (9): 668-80. doi: 10.1016/S2213-8587(22)00163-2.
6. Stener-Victorin E, Teede H, Norman RJ, Legro R, Goodarzi MO, Dokras A, et al. Polycystic ovary syndrome. Nat Rev Dis Primers. 2024; 10 (1): 27. doi: 10.1038/s41572-024-00511-3.
7. Tay CT, Mousa A, Vyas A, Pattuwage L, Tehrani FR, Teede H. 2023 International Evidence-Based Polycystic Ovary Syndrome Guideline Update: Insights from a Systematic Review and Meta-Analysis on Elevated Clinical Cardiovascular Disease in Polycystic Ovary Syndrome. J Am Heart Assoc. 2024; 13 (16): e033572. doi: 10.1161/JAHA.123.033572.
8. Teede HJ, Tay CT, Laven J, Dokras A, Moran LJ, Piltonen TT, et al. International PCOS Network. Recommendations from the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. Fertil Steril. 2023; 120 (4): 767-93. doi: 10.1016/j.fertnstert.2023.07.025.
9. Barrera FJ, Brown EDL, Rojo A, Obeso J, Plata H, Lincango EP, 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. doi: 10.3389/fendo.2023.1106625.
10. Ghaderzadeh M, Garavand A, Salehnasab C. Artificial intelligence in polycystic ovary syndrome: a systematic review of diagnostic and predictive applications. BMC Med Inform Decis Mak. 2025; 25 (1): 427. doi: 10.1186/s12911-025-03255-6.
11. Wang J, Chen R, Long H, He J, Tang M, Su M, et al. Artificial intelligence in polycystic ovarian syndrome management: past, present, and future. Radiol Med. 2025; 130 (9): 1409-41. doi: 10.1007/s11547-025-02032-9.
12. Zigarelli A, Jia Z, Lee H. Machine-Aided Self-diagnostic Prediction Models for Polycystic Ovary Syndrome: Observational Study. JMIR Form Res. 2022; 6 (3): e29967. doi: 10.2196/29967.
13. Zad Z, Jiang VS, Wolf AT, Wang T, Cheng JJ, Paschalidis IC, et al. Predicting polycystic ovary syndrome with machine learning algorithms from electronic health records. Front Endocrinol (Lausanne). 2024; 15: 1298628. doi: 10.3389/fendo.2024.1298628.
14. Zhao B, Wen L, Huang Y, Fu Y, Zhou S, Liu J, et al. A Deep Learning-Based Automatic Recognition Model for Polycystic Ovary Ultrasound Images. Balkan Med J. 2025; 42 (5): 419-28. doi: 10.4274/balkanmedj.galenos.2025.2025-5-114.
15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024; 385: e078378. doi: 10.1136/bmj-2023-078378.
16. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025; 388: e082505. doi: 10.1136/bmj-2024-082505.

OPEN ACCESS

17. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, et al. STARD-AI Steering Committee. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025; 31 (10): 3283-9. doi: 10.1038/s41591-025-03953-8.
18. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. DECIDE-AI expert group. Reporting guidelines for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022; 28 (5): 924-33. doi: 10.1038/s41591-022-01772-9.
19. Teede HJ, Mousa A, Tay CT, Costello MF, Brennan L, Norman RJ, et al. Summary of the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome: an Australian perspective. Med J Aust. 2024; 221 (7): 389-95. doi: 10.5694/mja2.52432.
20. Hatoum S, Amiri M, Hopkins D, Buyalos RP, Bril F, Azziz R. Population-Based vs Health System and Insurer Records: Significant Underdiagnosis of PCOS. J Clin Endocrinol Metab. 2025; 110 (6): e1759-e1766. doi: 10.1210/clinem/dgaf037.
21. Atiomo W, Rizwan MNH, Bajwa MH, Furniturewala HJ, Hazari KS, Harab D, et al. Prevalence and Diagnosis of PCOS Using Electronic Health Records: A Scoping Review and a Database Analysis. Int J Environ Res Public Health. 2024; 21 (3): 354. doi: 10.3390/ijerph21030354.
22. Silva EL, Lane KJ, Cheng JJ, Popp Z, van Loenen BD, Coull B, et al. Polycystic Ovary Syndrome Underdiagnosis Patterns by Individual-level and Spatial Social Vulnerability Measures. J Clin Endocrinol Metab. 2025; 110 (6): 1657-66. doi: 10.1210/clinem/dgae705.
23. van der Ham K, Laven JSE, Tay CT, Mousa A, Teede H, Louwers YV. Anti-Mullerian hormone as a diagnostic biomarker for polycystic ovary syndrome and polycystic ovarian morphology: a systematic review and meta-analysis. Fertil Steril. 2024; 122 (4): 727-39. doi: 10.1016/j.fertnstert.2024.05.163.
24. Peña AS, Witchel SF, Boivin J, Burgert TS, Ee C, Hoeger KM, et al. International evidence-based recommendations for polycystic ovary syndrome in adolescents. BMC Med. 2025; 23 (1): 151. doi: 10.1186/s12916-025-03901-w.
25. Bachelot G, Bachelot A, Bonnier M, Salem JE, Farabos D, Trabado S, et al. Combining metabolomics and machine learning models as a tool to distinguish non-classic 21-hydroxylase deficiency from polycystic ovary syndrome without adrenocorticotropic hormone testing. Hum Reprod. 2023; 38 (2): 266-76. doi: 10.1093/humrep/deac254.
26. Tiwari S, Kane L, Koundal D, Jain A, Alhudhaif A, Polat K, et al. SPOSDS: A smart polycystic ovary syndrome diagnostic system using machine learning. Expert Syst Appl. 2022; 203: 117592. doi: 10.1016/j.eswa.2022.117592.
27. Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El-Sappagh S, et al. Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. Diagnostics (Basel). 2023; 13 (8): 1506. doi: 10.3390/diagnostics13081506.
28. Khanna VV, Chadaga K, Sampathila N, Prabhu S, Bhandage V, Hegde GK. A distinctive explainable machine learning framework for detection of polycystic ovary syndrome. Appl Syst Innov. 2023; 6 (2): 32. doi:10.3390/asi6020032.
29. Panjwani B, Yadav J, Mohan V, Agarwal N, Agarwal S. Optimized Machine Learning for the Early Detection of Polycystic Ovary Syndrome in Women. Sensors (Basel). 2025; 25 (4): 1166. doi: 10.3390/s25041166.
30. Yao B, Yu X, Zhang Y, Chen J, Zhu X, Zhang C, 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 (Lausanne). 2025; 16: 1719631. doi: 10.3389/fendo.2025.1719631.
31. Tong C, Wu Y, Zhuang Z, Yu Y. A diagnostic model for polycystic ovary syndrome based on machine learning. Sci Rep. 2025; 15 (1): 9821. doi: 10.1038/s41598-025-92630-4.
32. Cheng JJ, Mahalingaiah S. Data mining polycystic ovary morphology in electronic medical record ultrasound reports. Fertil Res Pract. 2019; 5: 13. doi: 10.1186/s40738-019-0067-7.
33. Agirsoy M, Oehlschlaeger MA. A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features. Sci Rep. 2025; 15 (1): 33638. doi: 10.1038/s41598-025-10453-9.
34. Chen JY, Chen WJ, Zhu ZY, Xu S, Huang LL, Tan WQ, et al. Screening of serum biomarkers in patients with PCOS through lipid omics and ensemble machine learning. PLoS One. 2025; 20 (1): e0313494. doi: 10.1371/journal.pone.0313494.
35. Yan X, Yang Z, Zhao H, Feng G, Li S, Li Y, et al. Unveiling lipoprotein subfractions signature in high-FNPO PCOS: implications for PCOM diagnosis and risk assessment using advanced machine learning models. BMC Med. 2025; 23 (1): 289. doi: 10.1186/s12916-025-04120-z.
36. Rodriguez EM, Thomas D, Druet A, Vlajic-Wheeler M, Lane KJ, Mahalingaiah S. Identifying Women at Risk for Polycystic Ovary Syndrome Using a Mobile Health App: Virtual Tool Functionality Assessment. JMIR Form Res. 2020; 4 (5): e15094. doi: 10.2196/15094.
37. Graca S, Alloh F, Lagojda L, Dallaway A, Kyrou I, Randeva HS, et al. Polycystic Ovary Syndrome and the Internet of Things: A Scoping Review. Healthcare (Basel). 2024; 12 (16): 1671. doi: 10.3390/healthcare12161671.
38. Wright PJ, Burts C, Harmon C, Corbett CF. Availability and Use of Digital Technology Among Women With Polycystic Ovary Syndrome: Scoping Review. JMIR Infodemiology. 2025; 5: e68469. doi: 10.2196/68469.

Citation of this article
Nahar S, Akter MS, Aosaf S. Artificial Intelligence for Prediction and Early Detection of Polycystic Ovary Syndrome: A Review. Eastern Med Coll J. 2026; 11 (1): 82-92.
doi https://doi.org/10.3329/emcj.v11i1.90025