Frontiers in Biomedical Signal Processing
http://dx.doi.org/10.62762/FBSP.2025.529389

RESEARCH ARTICLE

Advancing PCOS Diagnosis: Harnessing the Power of AI and Machine Learning for Enhanced Accuracy and Efficiency

Seyed-Ali Sadegh-Zadeh[1], Atena Behforouz[2], Masoumeh Chaparnia[3], Razieh Sadri[4], Amir M. Hajiyavand[5], Parisa khatibi Damavandi[6], Zahra Naderi[7], Elaheh Mobaser[8], Kaveh Kavianpour[9], Alireza Soleimani Mamalo[10,*], Hajar Abbasi[2] and Farnaz Moshiri[2]

[1] Department of Computing and Esports, School of Digital, Technology, Innovation and Business, Staffordshire University, Stoke-on-Trent ST4 2DE, United Kingdom
[2] Preventative Gynaecology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3] Iran University of Science and Technology (IUST), Tehran 16846-13114, Iran
[4] Faculty of Broadcast Engineering, Islamic Republic of Iran Broadcasting University (IRIBU), Tehran, Iran
[5] Department of Mechanical Engineering, School of Engineering, University of Birmingham, Birmingham, United Kingdom
[6] Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[7] Department of Obstetrics and Gynecology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[8] Tehran University of Medical Sciences, Tehran, Iran
[9] Department of Computer Science and Mathematics, Amirkabir University of Technology, Tehran 1591634311, Iran
[10] Student Research Committee, Urmia University of Medical Sciences, Urmia, Iran

Abstract

This research presents a diagnostic method for Polycystic Ovary Syndrome (PCOS), a common hormonal disorder among women of childbearing age. The study applied machine learning classifiers—Random Forest, CatBoost, and MLP—to the Kaggle PCOS dataset, enhanced by BorutaShap and SMOTE feature selection methods. The ensemble classifier achieved an F1 score of 93.54% and an accuracy of 95.15%. These results demonstrate AI's potential to improve PCOS diagnosis for broader clinical applications. Future research should integrate genetic and epigenetic factors into AI models, validated through clinical trials. Key contributions include: using an ensemble of machine learning classifiers, advanced feature selection methods, and achieving high F1 and accuracy scores, showing the model's clinical effectiveness.

Keywords: polycystic ovary syndrome (PCOS), hormonal disorder, reproductive age, infertility, early detection, classifiers, feature selection, artificial intelligence.

Submitted: 09 July 2025
Accepted: 26 July 2025
Published: 26 February 2026

Vol. 1, **No.** 1, 2026.
10.62762/FBSP.2025.529389

***Corresponding author:**
✉ Alireza Soleimani Mamalo
Soleymanialireza688@gmail.com

Citation
Sadegh-Zadeh, S. A., Behforouz, A., Chaparnia, M., Sadri, R., Hajiyavand, A. M., Damavandi, P. K., Naderi, Z., Mobaser, E., Kavianpour, K., Mamalo, A. S., Abbasi, H., & Moshiri, F. (2026). Advancing PCOS Diagnosis: Harnessing the Power of AI and Machine Learning for Enhanced Accuracy and Efficiency. *Frontiers in Biomedical Signal Processing*, 1(1), 49–70.

1 Introduction

Polycystic ovary syndrome (PCOS) is a common hormonal disorder in women of childbearing age [1]. The one of the most frequent cause of female infertility is PCOS (40%) [2]. Menstrual cycles in women with PCOS may be oligomenorrhea or irregular, and androgen levels may be elevated [4]. The ovaries grow numerous little follicles and consistently stop the ovulation. Excessive androgen, an irregular menstrual cycle, and polycystic ovaries are the main signs and symptoms of PCOS [5].

Patients with PCOS not only struggle with conception, but also with hair loss, imbalanced female hormones, high amounts of male hormones, hirsutism, obesity and even endometrial cancer [3]. In addition, PCOS can lead to additional severe metabolic conditions like high blood pressure, mental health issues, heart disease, type 2 diabetes, and endocrine problems. Therefore, the early prediction of PCOS mostly depends on physical exams (such as hirsutism or other hyper androgenic symptoms) as well as biochemical and sonographic features of ovaries, which are important factors and can be helpful in the treatment process [6–8].

PCOS may not have a definite treatment, but symptoms are frequently managed [9]. However, fewer than 50% of women receive an accurate diagnosis, meaning that many women with PCOS go untreated [10]. Because PCOS directly causes ovarian dysfunction with a greater risk of miscarriage, infertility, or even gynaecological cancer and mental anguish for the patients due to time and money waste, early detection and diagnosis of PCOS with minimal tests and imaging techniques is of the utmost importance and great significance [11, 12].

Machine learning, which is primarily a branch of artificial intelligence, has gained significant attention in the digital sphere as a vital element of modernization approaches [13, 14]. Understanding data structure and incorporating it into models that people can understand and use is the general goal of machine learning [15, 16].

Extraction of logical conclusions to diagnose diseases has become a difficult undertaking as a result of the introduction of numerous technologies in biomedicine and healthcare leading to the accumulation of enormous amounts of data. In order to help clinicians and speed up the process of diagnosis and prediction, machine learning techniques can learn interactions and patterns among data to make sensible decisions about unseen data in various fields of study, particularly the diagnosis of diseases. Additionally, because recent advances in machine learning techniques have produced positive outcomes for disease detection that aid in early treatment and a decrease in the death rate, scientists have been compelled to use various machine learning approaches to anticipate diseases [17–19].

Although some research [20–22] have concentrated on the identification of follicles and categorization of PCOS in females using ultrasound imaging, in general, the first stage of PCOS prognosis is patient screening based on clinical tests. Therefore, in this study, various machine learning algorithms were applied to the Kaggle PCOS clinical dataset in order to identify PCOS. Moreover, after using various feature selection techniques, the accuracy rate of each classifier was assessed.

The remainder of the paper is structured as follows: In the second section, related works and studies using the Kaggle PCOS dataset to diagnose PCOS using various machine learning algorithms and feature selection methods are presented. The third section includes material and methods which are a brief explanation of the dataset as well as the used classification algorithms and feature selection methods. The fourth section also included an explanation of the results. The fifth section discussed the experimental findings and the conclusions drawn from results.

2 Related Work

In the rapidly advancing field of machine learning, the interplay between features, feature selection, and classifiers forms a cornerstone for building effective and efficient predictive models [23]. At its essence, machine learning involves teaching algorithms to learn patterns from data in order to make accurate predictions or decisions. Features, also known as attributes or variables, play a pivotal role in the characteristics of the data that the algorithms utilize to discern these patterns [24]. However, not all features are created equal; some might be noisy, irrelevant, or redundant, thereby potentially hindering the performance of the model. This introduces the concept of feature selection, a critical process that involves identifying and retaining only the most informative features, consequently enhancing model performance by reducing complexity and improving generalization [25].

The relationship between features, feature selection, and classifiers becomes especially significant as it directly impacts the overall success of machine learning tasks. Classifiers, such as decision trees, support vector machines, and neural networks, utilize the selected features to make predictions or categorize data points into distinct classes [26]. The chosen classifier's effectiveness is profoundly influenced by the quality and relevance of the features it operates on. Thus, a meticulous selection of features is imperative to enable classifiers to uncover meaningful relationships within the data and produce reliable outcomes [27].

Traditional and ensemble classifiers were used to analyse the Kaggle PCOS dataset in the work of Danaei et al. [28]. Investigated was the effectiveness of a number of classifiers, including Ensemble Random Forest, Extra Tree, Adaptive Boosting (AdaBoost), and Multi-Layer Perceptron (MLP). The dataset with all features and reduced feature subsets produced by the filter, embedding, and wrapper feature selection methods were used for this. The experimental results showed that feature selection strategies improved the performance of all classifiers in a positive way.

In their study of PCOS patients, Meena et al. [29] suggested Neural Fuzzy Rough Subset Evaluating (NFRSE) as a feature selection technique and used Information Gain Subset Evolution (IGSE) to offer selected features of their dataset for ID3 and J48 decision tree classifiers. According to the findings, NFRSE-ID3 obtained a lower error rate than the others. They evaluated the performance of their proposed hybrid feature selection and classification method (a combination of Neural Fuzzy Rough Set (NFRS) and ANNs) and various feature selection methods (PCA, Gain Ration, Information Gain and Correlation based Feature Selection (CFS)) and classifiers (SVM, ANNs, Decision Tree and NB) in the classification of their PCOS dataset.

The PCOS dataset from Obafemi Awolowo University was classified by Balogun et al. [30] using C4.5 Decision Tree, NB, and MLP. The results showed that C4.5 Decision Tree and MLP outperformed NB by an accuracy rate of 74.359. When comparing the performance of three classifiers, the NB, Decision Tree, and ANNs, in the classification of PCOS, Vikas et al. [31] found that the accuracy rate of the NB classifier was greater (97.65%) than that of the other two.

For the classification of two datasets collected from Raman spectra of the Shengjing Hospital of China Medical University, Zhang et al. [32] utilised an XGB, KNN, RF and stacking classifier model with KNN, RF and XGB in its first layer and XGB in its second layer. The dataset also contained two sets of information from plasma and follicular fluid samples. According to experimental findings, the stacking classifier model could outperform other classifiers and produced a higher accuracy rate (89.32%) using follicular fluid than utilising plasma samples. To categorise the Kaggle PCOS dataset, Nandipati et al. [33] compared several feature selection and classification techniques using Python and Rapid Minder tools. Their findings demonstrated that RF (using RapidMiner) could obtain an accuracy rate of 93.12% without feature selection, and KNN and SVM could achieve an accuracy rate of 90.83 by employing ten selected features.

Inan et al. [34] selected features from the Kaggle PCOS dataset using statistical techniques like Chi-Square and Analysis of Variance (ANOVA) to offer inputs for several classifiers like SVM, KNN, RF, Ensemble Adaptive Boosting (AdaBoost), XGBoost, NB, and MLP. According to experimental findings, XGBoost was able to outperform the other classifiers in terms of classification accuracy rate, achieving a score of 95.83. In order to diagnose PCOS using the Kaggle PCOS dataset with nine features, Munjal et al. [35] employed three machine learning techniques: Extra Trees, RF and Decision Tree, and Genetic Algorithms (GA) as a feature selection approach. With an accuracy rate of 88%, the results demonstrated that Extra Trees could outperform the other two classifiers.

A PCOS dataset was acquired from an infertility clinic and research centre, and Pushkarini et al. [36] used RF, Linear Regression, and K-Nearest Network (KNN) algorithms to classify the dataset. Experimental findings showed that, when compared to the other two approaches, RF could produce the highest R2. In order to classify the PCOS Kaggle dataset, Neto et al. [37] examined the effectiveness of various classifiers (i.e., SVM, MLP, RF, LR, and Gaussian NB). By obtaining an accuracy rate of 95% and a precision of 96%, RF might outperform other classifiers. Using the Kaggle PCOS dataset, Bharati et al. [38] implemented a filter-based univariate feature selection strategy to choose the most crucial features. Additionally, they classified PCOS using a set of ten variables utilising gradient boosting, RF, Logistic Regression (LR), and Hybrid Random Forest and Logistic Regression (RFLR). The outcomes showed that RFLR could outperform other classifiers with classification accuracy of 91.01% and recall value of 90%.

Hassan et al. [39] classified the Kaggle PCOS dataset using a variety of machine learning techniques, including RF, SVM, CART, NB, and LR. Additionally, testing findings showed that RF may outperform other classifiers by obtaining an accuracy rate of 96%. To choose the most useful features for their PCOS dataset, Prapty et al. [40] used a Decision Tree. They also used RF, SVM, NB, and KNN to categorise the dataset. Results showed that RF may perform better than other classifiers at a rate of 93.5% accuracy. Tanwani et al. [41] classified the Kaggle PCOS dataset using LR and KNN after using filter methods to find correlations between dataset features. The results showed that LR had a higher accuracy rate of 92% compared to KNN. In conclusion, the preceding research in the field of classifying clinical PCOS datasets has laid a foundation for our current work. The dataset sourced from hospitals and scanning centers in Thodupuzha served as a valuable resource for our study. Notably, the hybrid classification approach employed by Thomas et al. [42], utilizing a combination of Artificial Neural Networks (ANNs) and Naive Bayes (NB), demonstrated promise by achieving an accuracy of 95% in categorizing the dataset. Despite this progress, certain aspects remain unexplored and opportunities for improvement persist. Motivated by the existing gaps in the literature, our proposed method seeks to build upon these foundations and push the boundaries further. Our key contribution lies in achieving an impressive accuracy of 95.15%, underscoring the efficacy of our novel approach in addressing the challenges posed by PCOS dataset classification.

3 Materials and Methods

3.1 Dataset Description

The PCOS dataset for this investigation was obtained from the Kaggle Dataset Repository by 10 different Indian hospitals. The class feature determines whether a person is diagnosed with PCOS in 177 instances or without PCOS in 364 instances. The dataset contains 43 features that are based on physical and medical evaluations of 541 women.

3.2 Methodology

Figure 1 presents a block diagram of the method we propose. This method encompasses several key stages: Initially, the data is cleaned and outliers are removed. To balance the dataset, oversampling techniques are employed. Next, we apply a range of feature selection techniques, including the feature extractor PCA. The data is then partitioned into training and test datasets. Various classifiers are trained using the training data. These models are then evaluated using the test data. Based on these evaluations, the best classifiers are chosen. Finally, these top-performing models are integrated using the stacking ensemble method to yield the final result. The subsequent sections will delve into each aspect of this proposed method in detail.

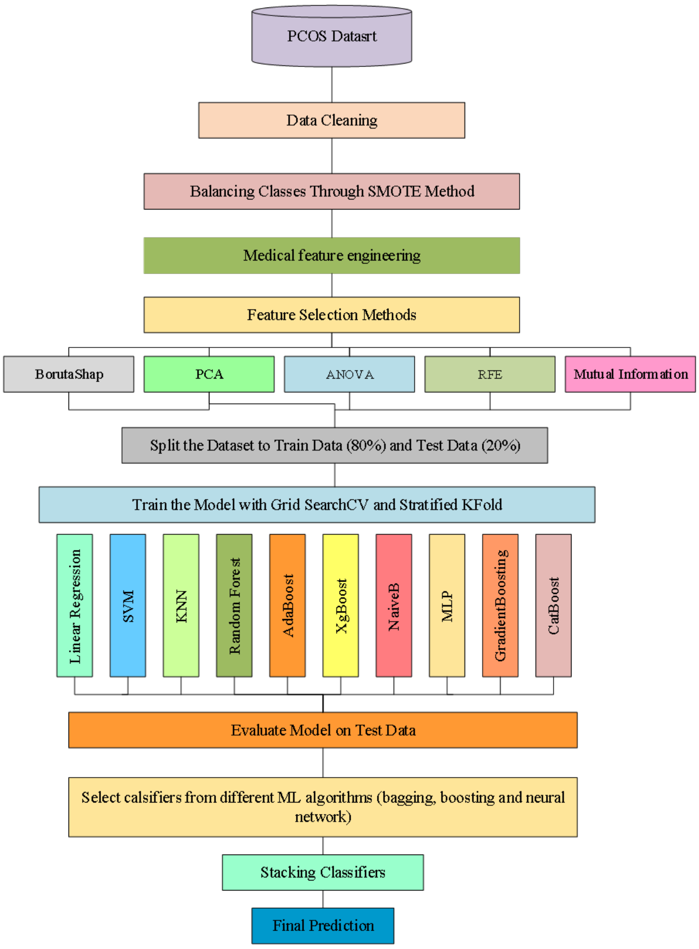

Figure 1. Block diagram of the proposed method.

Figure 1 outlines a structured workflow that begins with the PCOS Dataset and proceeds through various stages of data preprocessing, including Data Cleaning and Balancing Classes via the SMOTE method. It then involves Medical Feature Engineering and applies multiple Feature Selection Methods such as BorutaShap, PCA, ANOVA, RFE, and Mutual Information to refine the dataset. The next phase splits the dataset into training (80%) and testing (20%) sets, followed by model training using Grid SearchCV for hyperparameter tuning and Stratified K-Fold for validation. Various classifiers, including Linear SVM, RNN, Random Forest, XGBoost, AdaBoost, MLP, Gradient Boosting, and LightGBM, are trained. Subsequently, the model's performance is evaluated on the test data. The final stages involve selecting classifiers from different machine learning algorithms and a Stacking Classifier approach for the final prediction. This diagram effectively visualizes the comprehensive machine learning pipeline and the advanced ensemble techniques employed, culminating in the model that reportedly achieved high accuracy and F1 score, suggesting the effectiveness of this AI-assisted approach in diagnosing PCOS.

3.3 Data cleaning and Removing outliers

The Inter quartile range (IQR) approach [43–45] is then used to remove outliers in each characteristic once null values have been removed. We identify the interval in which a feature in the dataset has the maximum likelihood of being rich and connected, and we remove features outside of this region. We apply the data clustering algorithm Density-based spatial clustering of applications with noise (DBSCAN) [46, 47] to eliminate the remaining outlier data after removing the irrelevant data in all features using IQR. The available data should be divided into two groups using this algorithm: original data and noise. This technique results in the identification of 19 data as noise. After removing the outliers, we get at 512 data from the initial dataset's 541 data (including 345 data with negative class and 167 data with positive class).

3.4 Medical Feature Engineering

In accordance with the recommendations from the doctors' panel, our approach involves preliminary processing of the clean data. This pre-processing phase is organized into three distinct categories of features. The first group, Group1, remains unchanged. The second set, Group2, undergoes encoding. Meanwhile, the features in Group3, which include attributes like Weight, Height, FSH, LH, HIP, and the Waist/hip ratio, are omitted from subsequent analysis. As detailed in Table 1, we specifically apply encoding techniques to the features in Group2.

3.5 Data Sampling with SMOTE

Figure 2 clearly shows that the number of negative class samples is nearly triple the number of positive class samples, indicating an imbalance in our data. To address this, we used the Synthetic Minority Oversampling Technique (SMOTE) to increase the quantity of samples in the minority class. SMOTE draws a new sample at a location along a line that connects examples in the feature space after choosing examples that are near the feature space [59].

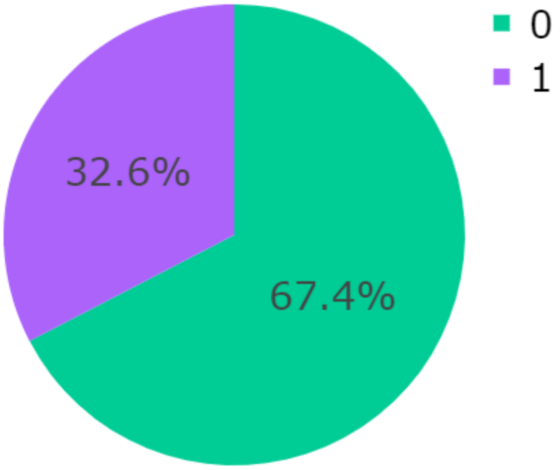

Figure 2. Pie chart of data distribution.

The SMOTE algorithm can be described as follows: First, determine the difference between a sample and its nearest neighbor. Next, this difference is divided by a randomly chosen number between 0 and 1. Then, to generate a new synthetic data point in the feature space, this calculated difference is added to the original sample. Finally, this process is repeated with the next nearest neighbor until a user-specified number of synthetic examples are created.

3.6 Feature selection methods

Here are a number of possible feature selection techniques, which can be grouped into one of three categories. Each technique has pros and cons of its own. The following is a description of the categories from [60, 61]: Filtering Methods- Independent of the selected predictor, use filtering techniques to choose subsets of variables as a preliminary step. Wrapper Methods - Use the target learning system as a "black box" to rank subsets of variables according. Embedded Methods - perform variable selection during the training process and are typically unique to a particular learning machine.

Feature selection aims to streamline the number of input variables, which not only cuts down on the computational costs involved in model development but also enhances the performance of the model. This process targets the removal of predictors that are redundant or carry minimal information, thereby refining the model's efficiency. Feature

Table 1. Medical feature engineering.

FEATURE NAME	CATEGORIES	REFERENCE
BMI	Underweight $< 18.5kg/m^2$, Normal weight: $18.5 - 24.9kg/m^2$, Over weight: $25 - 29.9kg/m^2$, Obesity: $\geq 30kg/m^2$	[48, 49, 67]
CYCLE LENGTH	$\leq$ 8day normal, $>$ 8 day menorrhagia	[50, 51]
BETA HCG	negative $< 5mIU/mL$, Positive$\geq 5mIU/mL$	[52]
LH/FSH RATIO	$>$ 2 high probability for pcos	[53, 54]
WAIST	$\geq$ 31.5 inch: suggestive of metabolic syndrome	[55]
TSH	$<$ 0.1 hyperthyroidism, $0.1 - 5$ normal range, >5: hypothyroidism	[56]
PRL	$>$ 20 hyper prolactinemia, $\leq$ 20 normal	[57]
VITAMIN D	$>$ 20 sufficient	[58]

selection with supervision, unsupervised learning, and semi-supervised learning are a few classification techniques [46, 62, 63]. These strategies differ primarily in how they use the target variable in feature selection. Unlike correlation-based methods, unsupervised methods do not consider the target variable when selecting features. The selection of features for wrappers, base filters, or embeddings, which are typically unsupervised, is evaluated based on their performance with the retained set of features and data. This is a key aspect of various segmentation methods. Wrapper-based approaches function by comparing different models, each retaining and selecting specific features, and ultimately choosing the most effective model. These methods are applicable to any type of variable, though they can be resource-intensive. An example of a wrapper approach is Recursive Feature Elimination (RFE), which involves the systematic removal of features. On the other hand, filtering techniques evaluate the relevance of predictors independently of predictive models. They then incorporate only those predictors that fulfill certain criteria into the model. Some models also have built-in feature selection mechanisms, ensuring that only predictors contributing to improved accuracy are included in the model. In these situations, the model can decide which data representation is preferable. These techniques include penalised regression and the Lasso decision tree [64, 65].

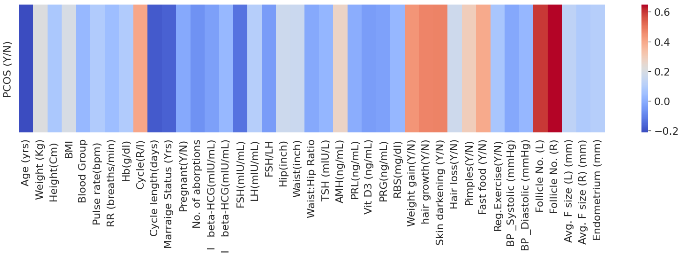

Figure 3. Pearson correlation heatmap PCOS(Y/N) with 41 other features (original data).

Figure 3 illustrates the linear correlation between the diagnosis of PCOS (Yes/No) and a range of 41 other features, likely clinical parameters and laboratory results relevant to the condition. In the heatmap, varying shades of blue and red represent the strength and direction of the correlation: blue indicates a positive correlation, while red signifies a negative correlation with the intensity of the color correlating to the strength of this relationship. The figure shows that certain features have a stronger positive correlation with PCOS, such as LH/FSH ratio and testosterone levels, which is consistent with the known pathophysiology of PCOS where elevated levels of these hormones are common. Conversely, features like SHBG (sex hormone-binding globulin) appear negatively correlated, aligning with the clinical understanding that women with PCOS often have lower SHBG levels.

3.7 Pearson's correlation coefficient

A value between -1 and 1 known as a Pearson correlation describes the degree to which two variables are linearly connected [66]. This statistic is used to assess how much the input feature X and the output feature Y are linearly correlated. Its value can be anything between +1 and -1, with 1 denoting a total positive correlation and -1 denoting a total negative correlation. As a result, a value of 0 indicates that there is no linear relationship [60]. Figures 3 and 4 display the heatmap of the Pearson's correlation between the PCOS(Y/N) target feature and the other 41 features.

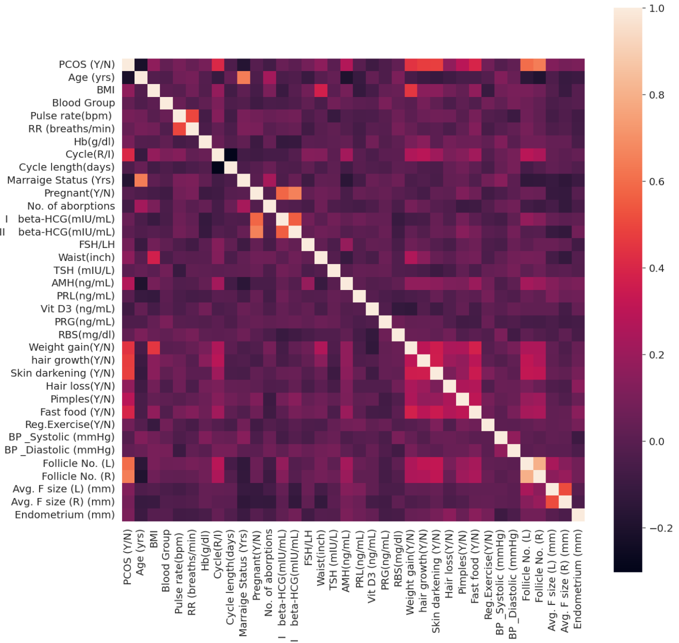

Figure 4. Correlation heat map of the pre-processed data features obtained from the filter based on Pearson correlation.

Figure 4 focuses on various factors associated with Polycystic Ovary Syndrome (PCOS). The variables include both clinical measurements such as hormone levels (LH, FSH, TSH, AMH, PRL, Vit D3, PRG, beta-HCG), physiological data (BMI, blood pressure, pulse rate), and patient history (marriage status, number of abortions, pregnancy status). The diagonal line of white squares from the top left to the bottom right represents the correlation of each variable with itself, which is always 1. Surrounding this line, we can identify clusters of variables that show some degree of correlation with each other, indicated by lighter shades. For example, there may be a correlation between hormone levels and certain symptoms or patient history factors, which could be significant in diagnosing or understanding PCOS.

3.8 BorutaShap Feature Selection

A wrapper feature selection technique called BorutaShap combines the Boruta feature selection algorithm and the Shapley values. The Boruta approach is an elegant way to wrap the Random Forest model. The method is a development of the concept outlined in the Party On paper[64], which establishes the relative relevance of genuine features to random probes in order to assess the significance of each feature [64]. Although Boruta is a reliable approach for feature selection, it heavily depends on the estimation of feature importance, which may be inaccurate or inadequate for the data. The team now includes SHAP (SHapley Additive exPlanations) [68–70]. The Boruta SHAP Feature Selection Algorithm is created by employing SHAP Values as the feature selection strategy in Boruta. This method combines the robustness of the Boruta algorithm to ensure that only significant variables remain in the set with the powerful feature explanations that are present in the SHAP method.

The algorithm operates through the following steps:

1. It starts by creating duplicates of all features in the data set, naming them 'shadow' followed by the original feature name. These shadow features are then shuffled to eliminate any existing correlations with the response variable.
2. The classifier is run using both the original data and these randomized shadow features. Subsequently, the features are ranked based on their importance, typically measured through permutation importance in the original algorithm.
3. The highest relevance score among the shadow features is used as a benchmark. Any feature exceeding this score is marked as a hit.
4. For each feature that hasn't been marked yet, a two-sided T-test for equality is conducted.
5. Features are categorized based on their importance relative to the threshold: those significantly lower are deemed 'unimportant' and eliminated, while those significantly higher are considered 'important.'
6. This process is repeated, excluding the shadow features, until each feature is assigned an importance value or the algorithm completes its pre-set number of iterations.

The BorutaShap feature selection method pinpointed essential factors including Age, Cycle Regularity, Marital Duration, AMH levels, Weight Gain, Hair Growth, Skin Darkening, Pimple Occurrence, Fast Food Intake, and the Number and Size of Follicles in both ovaries. Figures 5 to 8 effectively illustrate these crucial attributes. Figure 5 elucidates two aspects: first, a regression analysis illustrating BMI trends over time in individuals with and without PCOS; second, a depiction of menstrual cycle lengths. Figure 6 presents two graphs: one depicting follicle distribution in ovaries, and another exploring the correlation between age and menstrual irregularity. Figure 7

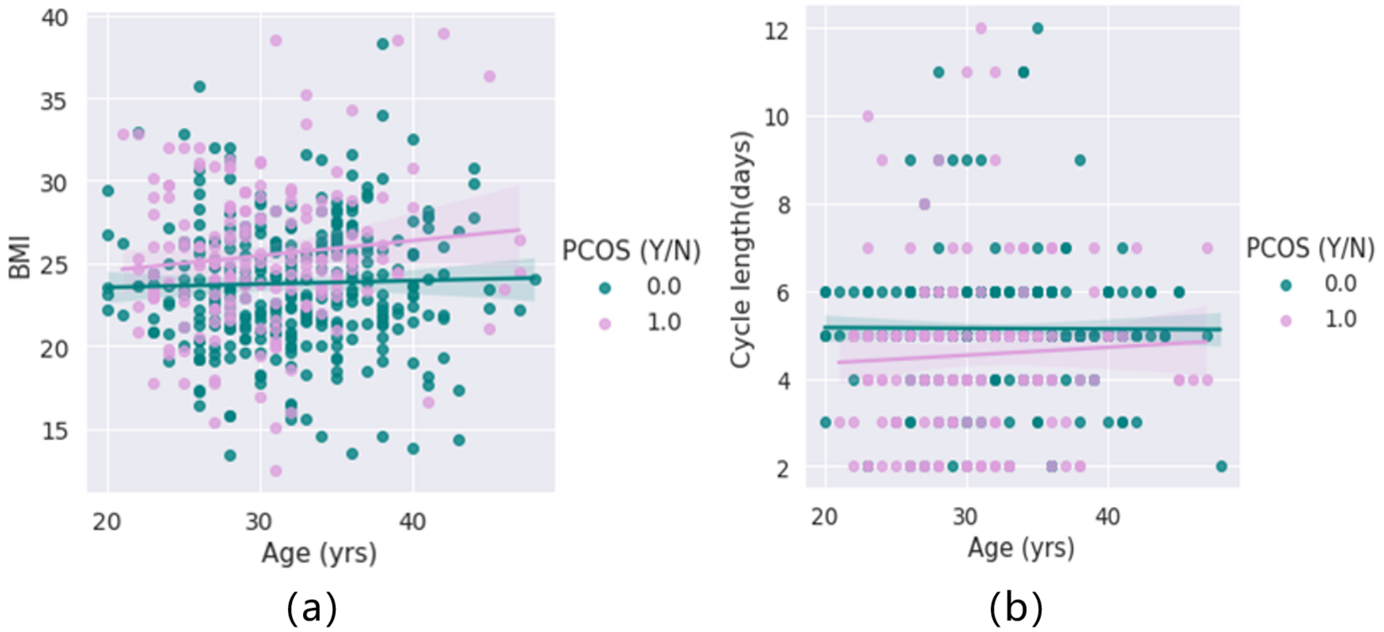

Figure 5. (a) Visualizes the linear relationship as determined through regression for a pattern of weight gain (BMI) over years in PCOS and Normal, (b) Visualizes the linear relationship as determined through regression for the length of the menstrual phase in P.

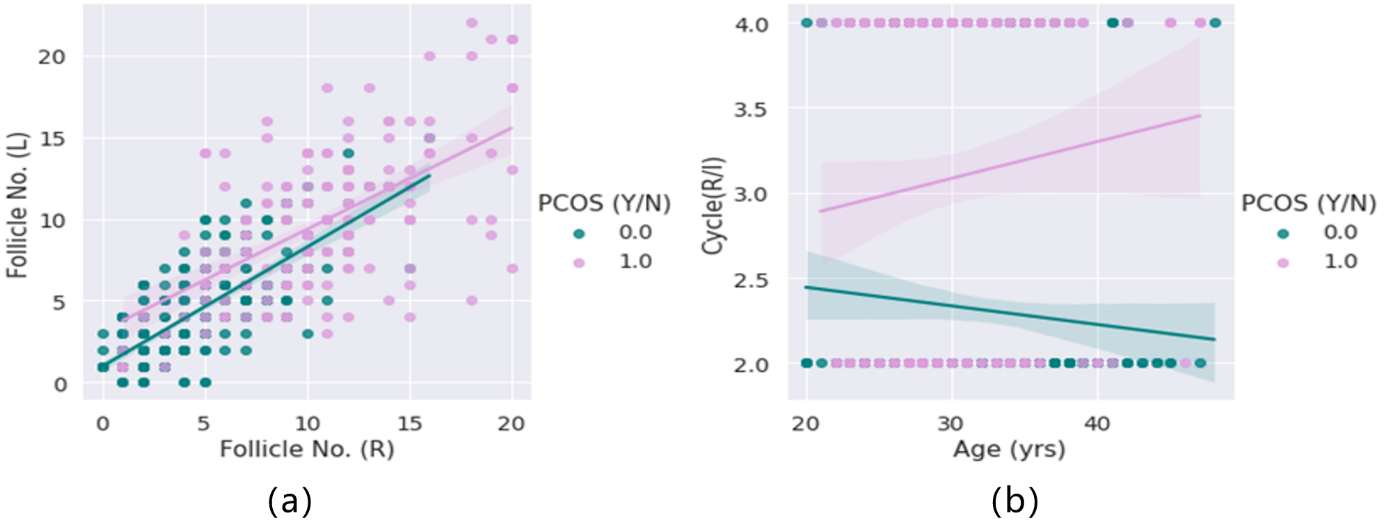

Figure 6. (a) Visualizes the linear relationship as determined through regression for distribution of follicles in both ovaries, (b) Visualizes the linear relationship as determined through regression for cycle IR concerning age.

displays histograms correlating various factors like follicle count, age, BMI, cycle length, and endometrial thickness with PCOS presence. In Figure 8, bar charts represent the link between PCOS and several categorical features such as Blood Type, Pregnancy Status, Weight Gain, Hair Growth, Skin Darkening, Hair Loss, Pimples, and Fast Food Habits. Finally, Figure 9 specifically highlights the distribution of different blood groups in relation to PCOS.

Figure 5 presents two scatter plots exploring the relationship between age, Body Mass Index (BMI), and menstrual cycle length in the context of Polycystic Ovary Syndrome (PCOS). Panel (a) shows the correlation between age and BMI, with the presence of PCOS indicated by two colors, suggesting that BMI may increase with age regardless of the PCOS status but with a possible difference between those with and without PCOS.

Figure 5(a) shows the correlation between age and BMI, with the presence of PCOS indicated by two colors, suggesting that BMI may increase with age regardless of the PCOS status but with a possible difference between those with and without PCOS. Panel (b) illustrates the relationship between age and menstrual cycle length, again differentiated by PCOS status. Both panels include a trend line, which likely represents a regression line, offering a statistical model of the relationship. The distribution of the data points and the slope of the trend lines can provide insights into the nature of these relationships. This visual analysis could be used to harness AI and machine learning to improve the diagnosis and understanding of PCOS by identifying patterns that may not be easily discernible through traditional statistical methods.

Figure 6(a) shows a linear relationship between the number of follicles in the left (L) and right (R) ovaries, with points representing individual observations. The presence of PCOS is denoted by two colours— blue for 'No' (0.0) and pink for 'Yes' (1.0). The overlaid line and shaded area suggest a regression line with its confidence interval, indicating a positive correlation between follicle numbers in both ovaries. Figure 6(b) presents a regression analysis of cycle irregularity (Cycle IR) across different ages with PCOS status differentiated by colour. The lines indicate a clear trend of cycle irregularity with age, and how this trend differs between individuals with and without PCOS. The shaded areas around the lines represent the uncertainty or variability in the regression prediction. These visualizations serve as a tool for assessing the association between the number of ovarian follicles, age, cycle irregularity, and the occurrence of PCOS, showcasing the potential of AI and machine learning in enhancing diagnostic accuracy and efficiency in medical research.

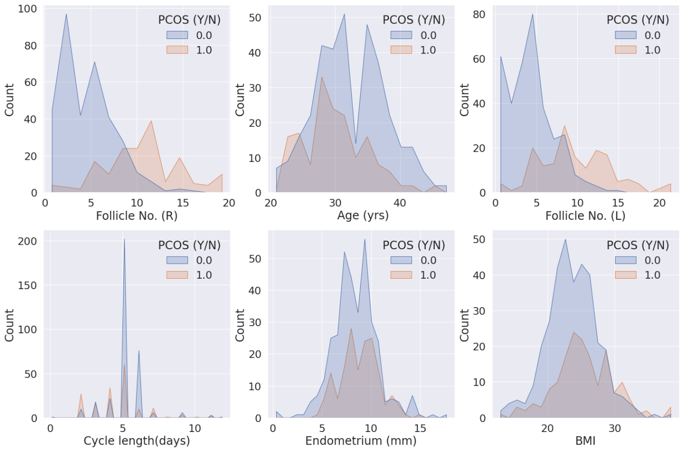

Figure 7. Histogram plots to display the joint distribution of Follicle No. (L), Follicle No. (R), Age (yrs), BMI, Cycle length(days), Endometrium (mm) based on target variable PCOS (Yes/No).

Figure 7 is a collection of histogram plots. These plots are used to display the joint distribution of various clinical and biological measures, with separate histograms for individuals with and without PCOS, indicated by blue and orange colors, respectively. The variables compared are the number of follicles on the right (R) and left (L) ovaries, age in years, Body Mass Index (BMI), and cycle length in days, as well as endometrium thickness in millimeters. This visual analysis aims to identify patterns and correlations that can be fed into AI algorithms to improve the diagnostic accuracy for PCOS. The histograms suggest potential differences in these variables between individuals with and without PCOS, which could be crucial in developing predictive models. The clear separation in some of the distributions (like the number of follicles) indicates strong potential predictors, while others show more overlap, suggesting a more complex relationship with the condition.

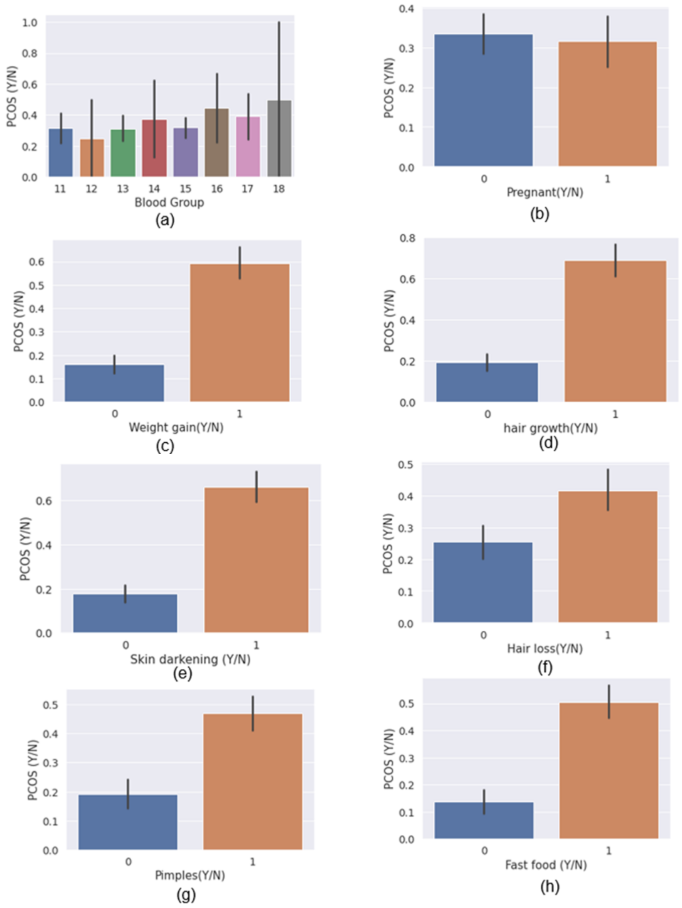

Figure 8. Bar plots to display the categorical features (a) Blood Group, (b) Pregnant(Y/N) (c) Weight Gain(Y/N) (d) Hair Growth(Y/N), (e) Skin Darkening(Y/N), (f) Hair loss(Y/N), (g)Pimples(Y/N), (h) Fast food(Y/N) and target variable PCOS (Yes/No).

Plots in Figure 8 display the categorical features: (a) Blood Group, (b) Pregnancy status, (c) Weight Gain, (d) Hair Growth, (e) Skin Darkening, (f) Hair Loss, (g) Pimples, and (h) Fast Food consumption, and their relationship to the target variable PCOS

(Yes/No). Each bar represents the proportion of the PCOS condition corresponding to the presence (1) or absence (0) of these features, presumably calculated using machine learning or AI algorithms to evaluate their significance in diagnosing PCOS. The error bars on top of the bars indicate the variability or uncertainty in the measurements. The use of different colors for 'Yes' and 'No' categories enhances the visual distinction between the groups. This kind of visualization is typically used to summarize the output of a classification algorithm, helping to identify which features may be more strongly associated with the presence of PCOS, thus providing insights into potential diagnostic criteria or risk factors.

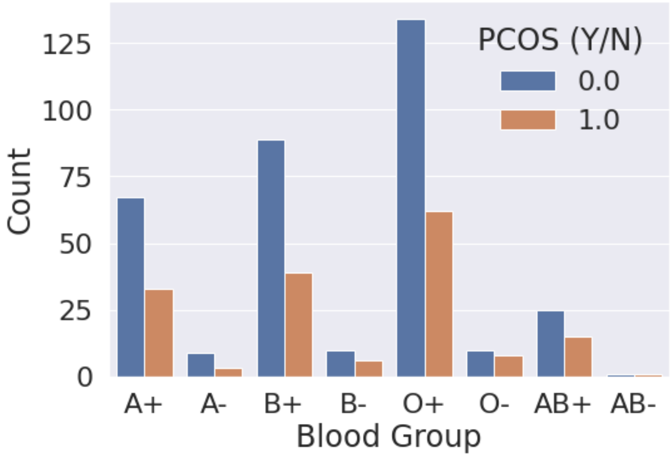

Figure 9. Blood groups plot based on PCOS.

Figure 9 represents a bar chart representing the count of individuals with and without Polycystic Ovary Syndrome (PCOS) across different blood groups. The x-axis categorizes the blood types (A+, A-, B+, B-, O+, O-, AB+, AB-), and the y-axis shows the count of individuals. Two sets of bars are depicted for each blood group: one for individuals without PCOS (labeled "0.0" and shown in blue) and one for individuals with PCOS (labeled "1.0" and shown in orange). At a glance, blood type O+ has the highest count of individuals without PCOS, while the count of individuals with PCOS seems relatively lower across all blood types, indicating a possible correlation between blood type and the presence of PCOS, which could be of interest in the medical research community.

3.9 Recursive feature elimination (RFE) method

One of the base wrapper approaches, the Recursive feature elimination (RFE) method aims to enhance the generalisation performance by deleting the least significant features. The RFE approach is appropriate for small sample situations since it makes use of the support vector machine's inherent generalisation ability. Despite its strong performance, RFE frequently omits "weak" features that, when paired with other features, may significantly boost performance [68, 71].

3.10 Analysis of Variance (ANOVA)

Analyzing variance, or ANOVA, is one of the filter-based techniques. When there is a parametric dependent variable and one or more independent variables, an ANOVA is a statistical test that is used to find differences between group means [72, 73].

3.11 Mutual information

A measure of statistical independence known as mutual information has the main properties of being able to analyse any kind of relationship, including nonlinear interactions between random variables, and of remaining invariant under various feature space mappings [74].

3.12 Principal Component Analysis Feature Extractor

Feature extraction and dimensionality reduction are the main applications of the unsupervised linear transformation method known as principal component analysis (PCA). It projects the data into a new subspace with the same number of dimensions as the original [74–76]. We undertake feature extraction using PCA, with the aim of generating a new, more condensed set of features that nonetheless effectively captures the majority of the relevant data.

3.13 Machine learning classifiers

3.13.1 Logistic regression

To determine the likelihood of a target variable, supervised classification techniques such as logistic regression are used. It is known as "logistic regression" since the basic methodology is relatively similar to linear regression. The Logit function, which is used in this categorization method, is referred to as "logistic." There are just two groups because the aim or dependent variable is dichotomous.

3.13.2 Support Vector Classifier

Support Vector Machine algorithms for regression, classification, and outlier detection are supervised machine learning techniques. In SVM, the data is essentially represented as points in an n-dimensional space, where n is the quantity of features. The search method looks for a hyperplane that can divide the plotted points into the necessary or predetermined number of classes [77].

3.13.3 K-Nearest Neighbor (K-NN)

The supervised learning algorithm K-Nearest Neighbor, sometimes referred to as the lazy learner algorithm, is used for both classification and regression. Instead of learning the dataset right away, it initially stores it before taking action on it when it comes time to classify[78].

3.13.4 Random Forest

A supervised machine learning technique called Random Forest is employed for both classification and regression. To obtain a more precise and consistent forecast, it constructs numerous decision trees and blends them [79].

3.13.5 AdaBoost Classifier

The Boosting method utilised as an Ensemble Method in Machine Learning is called the AdaBoost algorithm, or Adaptive Boosting. When used to solve classification problems, these models attain accuracy that is just a little bit above random chance. One-level decision trees are the most effective and are therefore used the most commonly with AdaBoost. It is a type of meta-estimator that begins by adapting a predictor to the provided dataset, then adapts subsequent copies of the predictor to nearly the same dataset while updating the weights of instances that were incorrectly classified in order to ensure that the emphasis of subsequent classifiers is on difficult complex scenarios [80].

3.13.6 XGBoost Classifier

Extreme Gradient Boosting, or XGBoost, is a decision-tree-based ensemble machine learning technique that is used to construct a scalable and accurate gradient boosting system. This strategy is effective for many classification and regression predictive modelling problems. To train models quickly and with fine-grained control, the stochastic gradient boosting method is used with several hyperparameters [81].

3.13.7 Gaussian Naive Bayes

Naive Bayes are algorithms that compute the probability and determine which class the given data would belong to using the Bayes theorem. For our hypothesis, we have employed Gaussian Naive Bayes. It supports continuous data and adheres to the Gaussian Normal Distribution, which means there won't be any covariance between the features. Each predictor is presumed to be conditionally independent of the others by Naive Bayes. Each predictor's distribution is independent of the others for a certain target value. In reality, even when broken, this condition of independence results in a quick, computationally feasible approach as opposed to an intractable one and does not dramatically reduce the model's predictive accuracy [82].

3.13.8 MLP Classifier

Multi-layer Perceptron classifier, or MLPClassifier, is connected to a neural network by the name itself. MLPClassifier uses an underlying Neural Network to carry out the classification task, unlike other classification methods like Support Vectors or Naive Bayes Classifier. A feedforward artificial neural network model called the multilayer perceptron (MLP) converts input data sets into a collection of useful outputs. An MLP is made up of several layers, each of which is completely connected to the one before it. With the exception of the nodes in the input layer, the nodes of the layers are neurons with nonlinear activation functions. There may be one or more nonlinear [83].

3.13.9 Gradient Boosting Classifier

A collection of machine learning techniques called gradient boosting classifiers combine a number of ineffective learning models to create a potent predicting model. Gradient boosting often makes use of decision trees. A method for training several models sequentially is gradient boosting. It is a boosting variation that works with any differentiable loss function. It can be used to address issues with classification and regression [84].

3.13.10 CatBoost Classifier

Classifying Boosting The open-source boosting package CatBoost or is used for regression and classification. It works with a variety of data types, including audio, text, and image data as well as historical data. This approach uses various types of statistics on combinations of categorical features and combinations of categorical and numerical features to convert categorical values into numbers [85].

3.13.11 Stacking the Classifiers

Stacking classifiers is a sophisticated ensemble machine learning technique that aims to surpass the performance of individual base estimators. This method, often referred to as simply as "Stacking," employs a strategy of combining multiple estimators to mitigate their individual biases, potentially leading to enhanced overall performance. However, this advantage comes at the cost of increased computational time.

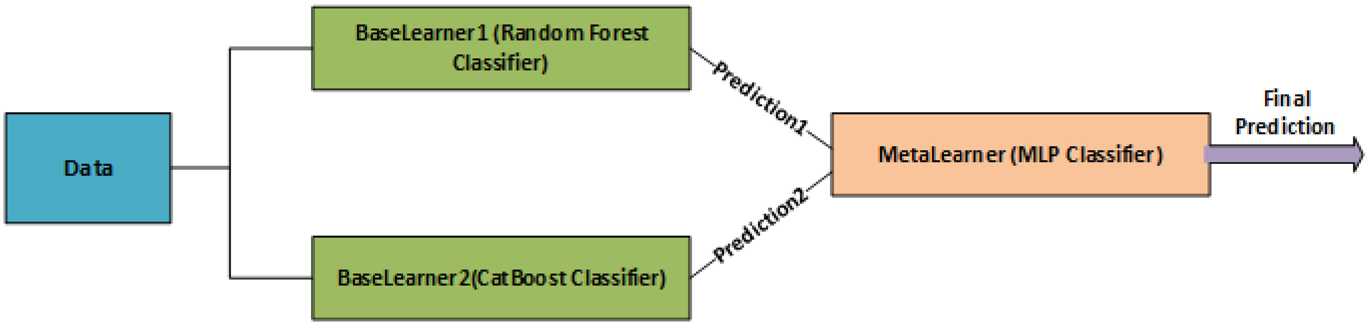

Figure 10. The block diagram of stacking best classifiers.

The core principle of stacking involves the use of generalization methods, such as bagging and boosting, to amalgamate predictions from various machine learning models applied to the same dataset. This technique addresses the dilemma of choosing the most reliable machine learning model among several competent ones, each excelling in distinct aspects of problem-solving. Stacking resolves this by introducing another machine learning model, which learns to appropriately trust or apply each model in the ensemble based on specific circumstances.

Distinct from bagging, stacking employs diverse models, all trained on the same dataset. Unlike boosting, it focuses on learning the best method to integrate predictions from these models, utilizing a single model for this purpose. The architecture of a stacking model typically comprises two layers: the base models (Level-0 Models) and a meta-model (Level-1 Model). The base models are trained on the dataset, and their predictions are compiled. The meta-model, trained on extrapolations from non-sample data provided by the base models, learns to optimally combine these predictions. Recent studies and experiments, such as those by Inan et al. [34], Bahad et al. [86], and Ibrahim et al. [87], have shown the effectiveness of certain features and methods in stacking classifiers. For instance, BorutaShap Features combined with Smote have demonstrated superior performance, guiding the preparation of input data for stacking classifiers.

In practice, stacking classifiers can integrate various machine learning algorithms, including bagging and boosting methods, and neural networks. An example of this is the stacking of classifiers like Random Forest, CatBoost, and MLP. In this setup, the Random Forest and CatBoost serve as base learners, with their predictions fed into the MLP meta-learner for final decision-making, as illustrated in Figure 10. It illustrates the architecture of a stacking ensemble method for classification tasks. This diagram represents a two-layer stacking model where the base layer consists of two different algorithms: a Random Forest Classifier and a CatBoost Classifier, both of which are used to independently predict outcomes from the input data. These predictions, referred to as Prediction1 and Prediction2, are then used as input features for the Meta Learner in the second layer, which in this case is an MLP (Multi-layer Perceptron) Classifier. The final prediction is then made by this meta-learner, combining the strengths of the base learners to ideally produce a more accurate model, as often used in complex classification tasks.

Table 2 presents a comparison of the average accuracy of various classifiers used for detecting PCOS, with and without the application of SMOTE, a method used to balance class distribution in datasets. The methods compared include the use of all features, BorutaShap features, PCA features, ANOVA features, RFE features, and Mutual features. The highest accuracy is achieved with BorutaShap Features with SMOTE at 91.26%, closely followed by Mutual Features with SMOTE at 91.257%. The lowest accuracies are observed with PCA features without SMOTE at 88.641% and PCA features with SMOTE at 89.32%. This suggests that feature selection techniques combined with SMOTE generally result in higher diagnostic accuracy for PCOS.

4 Results

For our training process, we allocated 80% of the clean dataset for training purposes and reserved the remaining 20% for testing. To enhance the effectiveness of our classifiers, we employed a grid search technique (as described by Lundberg et al. [88]). Additionally, we implemented stratified K-Fold cross-validation (referenced from Jolliffe [89]) to train our models, aiming to improve their generalizability and minimize the risk of overfitting. GridSearchCV is a pivotal method used in adjusting model hyperparameters to identify the optimal values for improved model

Table 2. Average accuracy of all classifiers in different methods for Polycystic Ovary Syndrome detection.

Method	Average accuracy of all models
All features w/o Smote	90.289
All features w Smote	90.484
BorutaShap Features w/o Smote	90.872
BorutaShap Features w Smote	91.26
PCA features w/o Smote	88.641
PCA features w Smote	89.32
Anova features without Smote	90.775
Anova features with Smote	90.678
RFE features without Smote	89.611
RFE features with Smote	89.805
Mutual Features without Smote	90.872
Mutual Features with Smote	91.257

performance. It's important to recognize that the hyperparameters of a model significantly influence its efficacy, and determining the best values for these hyperparameters is not straightforward. Since manually exploring every possible hyperparameter value is both time-consuming and resource-intensive, we leveraged GridSearchCV to streamline this process.

A common challenge in machine learning is overfitting, where a model is excessively tailored to its training dataset, leading to increased error rates when applied to test and holdout datasets. Cross-validation offers a solution to this problem. In this approach, we randomly segment our dataset into groups, using one group for testing and training the model on the remaining groups. This procedure is repeated for each group designated as the test group, and the average of the models is used to formulate the final model. Among the various forms of cross-validation, K-fold cross-validation is widely recognized, where 'k' represents the number of folds in the dataset. We utilized stratified K-fold cross-validation with k set to 10. This method is similar to simple K-fold cross-validation but differs by employing stratified sampling instead of random sampling. To evaluate the machine learning classifiers for Polycystic Ovary Syndrome (PCOS) classification, we applied the aforementioned methodology to various classifiers, including Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), AdaBoost (ADA), XGBoost (XG), Naive Bayes (NAIVEB), Multi-Layer Perceptron (MLP), Gradient Boosting (GB), and CatBoost (CB). We compared the performance of these models based on accuracy and F1 score, as detailed in Tables 3, 4, 5, 6.

Table 3 is a comparison of various machine learning classifiers based on their accuracy in detecting PCOS using different feature selection methods and data preprocessing techniques. LR and SVM consistently show high accuracy across all methods, maintaining accuracy above 91%. KNN has the most variability, with accuracy ranging from 82.52% to 86.41%. RF and MLP perform well, especially with BorutaShap feature selection and PCA with Smote, indicating robustness in handling imbalanced data. AdaBoost and XGBoost show lower performance compared to others but improve with feature selection techniques. Naive Bayes (NAIVEB) remains above 86% accuracy in all cases. GBOOST and CATBOOST show good performance, particularly with feature selection methods. The use of Smote for balancing the dataset seems to slightly decrease the accuracy for some models, while feature selection methods like BorutaShap and PCA generally maintain or improve classifier performance.

Table 4 presents the performance of various machine learning classifiers for detecting PCOS using three feature selection methods—ANOVA, RFE (Recursive Feature Elimination), and Mutual Features—both with and without the application of SMOTE. It's evident from the table that the XGBOOST and NAIVEB classifiers achieved the highest accuracy with ANOVA features and SMOTE, scoring 94.17% each. In contrast, the lowest accuracy is observed with KNN using RFE features without SMOTE at 84.47%. The utilization of SMOTE generally enhances the performance across most models, which suggests that handling class imbalance is beneficial for the accuracy of PCOS detection.

Table 5 presents a comparison of various machine learning classifiers based on their F1 scores, which is a measure of a test's accuracy, for the detection of PCOS. classifiers are evaluated across different feature selection methods: using all features, all features with and without SMOTE, BorutaShap feature selection with and without SMOTE, and PCA with a variance threshold of 0.95, with and without SMOTE. Performance varies across models and feature selection methods, with MLP achieving the highest F1 score of 91.18 using PCA(0.95) features with SMOTE, indicating a robust performance in this specific setup for PCOS detection.

Table 3. Performance of various classifiers in terms of accuracy on holdout test data using all features, BorutaShap and PCA features for Polycystic Ovary Syndrome detection.

Model	All features without Smote	All features with Smote	BorutaShap Features without Smote	BorutaShap Features with Smote	PCA(0.95) features without Smote	PCA(0.95) features with Smote
LR	92.23	91.26	92.23	89.32	91.26	92.23
SVM	92.23	92.23	93.2	93.2	93.2	93.20
KNN	82.52	86.41	85.44	89.32	84.47	86.41
RF	91.26	90.29	92.23	92.23	86.41	87.38
ADABOOST	89.32	90.29	92.23	93.2	86.41	86.41
XGBOOST	89.32	89.32	85.44	86.41	89.32	90.29
NAIVEB	92.23	89.32	91.26	91.26	86.41	86.41
MLP	91.26	92.23	93.2	92.23	93.2	94.17
GBOOST	92.23	91.26	92.23	93.2	87.38	89.32
CATBOOST	90.29	92.23	91.26	92.23	88.35	87.38

Table 4. Performance of various classifiers in terms of accuracy on holdout test data using Anova, RFE and Mutual features for Polycystic Ovary Syndrome detection.

Model	Anova features without Smote	Anova features with Smote	RFE features without Smote	RFE features with Smote	Mutual Features without Smote	Mutual Features with Smote
LR	91.26	91.26	90.29	93.2	92.23	91.26
SVM	89.32	93.2	92.23	92.23	91.26	94.17
KNN	84.47	85.44	84.47	85.44	85.44	88.35
RF	91.26	93.2	88.35	85.44	91.26	93.2
ADABOOST	89.32	90.29	89.32	91.26	89.32	92.23
XGBOOST	94.17	92.23	89.32	87.38	93.2	89.32
NAIVEB	94.17	90.29	91.26	91.26	91.26	92.23
MLP	89.32	89.32	90.29	90.29	92.23	89.32
GBOOST	93.2	91.26	89.32	92.23	91.26	92.23
CATBOOST	91.26	90.29	91.26	89.32	91.26	91.26

Table 6 presents a comparison of various classifiers used in the detection of PCOS in terms of the F1 score, which is a measure of a test's accuracy. The F1 score considers both the precision and the recall of the test to compute the score. Each model's performance is evaluated under six feature selection scenarios: ANOVA features without SMOTE, ANOVA features with SMOTE, RFE features without SMOTE, RFE features with SMOTE, Mutual Features without SMOTE, and Mutual Features with SMOTE. The SVM and XGBoost models show significant improvement when Mutual Features with SMOTE are used, achieving the highest F1 scores of 91.18 across all models and scenarios. It's notable that the use of SMOTE generally leads to an improvement in the F1 scores, indicating its effectiveness in addressing class imbalance in the dataset. The results indicate that feature selection and balancing methods significantly impact the performance of classifiers in detecting PCOS.

Table 7 presents the results of a stacking ensemble method applied to holdout data for the classification of PCOS. Two scenarios are compared: one with feature engineering and one without. In both scenarios, all metrics are notably high, suggesting effective classification performance. With feature engineering, precision, F1-score, and recall are slightly higher in both negative and positive class results than without feature engineering, indicating that the additional processing of data features improves the model's predictive capabilities. Accuracy and sensitivity

Table 5. Performance of various classifiers in terms of F1 score on holdout test data using all features, BorutaShap and PCA features for Polycystic Ovary Syndrome detection.

Model	All features without Smote	All features with Smote	BorutaShap Features without Smote	BorutaShap Features with Smote	PCA(0.95) features without Smote	PCA(0.95) features with Smote
LR	87.1	84.06	87.1	84.06	85.71	87.88
SVM	87.5	84.38	89.23	89.86	88.89	89.55
KNN	68.97	78.12	73.08	81.97	74.19	79.41
RF	84.75	88.52	87.5	87.5	77.42	80.0
ADABOOST	81.97	88.24	87.1	89.55	76.67	79.41
XGBOOST	84.51	84.93	77.61	80.0	83.08	85.29
NAIVEB	87.38	81.16	86.96	86.96	78.12	78.79
MLP	85.71	84.06	88.89	88.24	88.52	91.18
GBOOST	87.5	85.29	87.5	88.89	79.37	82.54
CATBOOST	83.33	87.50	85.71	87.88	81.25	79.37

Table 6. Performance of various classifiers in terms of F1 score on holdout test data using Anova, RFE and Mutual features for Polycystic Ovary Syndrome detection.

Model	Anova features without Smote	Anova features with Smote	RFE features without Smote	RFE features with Smote	Mutual Features without Smote	Mutual Features with Smote
LR	86.15	86.96	84.28	88.89	87.1	86.15
SVM	82.54	89.55	87.5	87.5	86.15	91.18
KNN	71.43	76.19	72.41	76.19	73.68	80.65
RF	85.71	88.89	81.25	76.19	84.75	88.52
ADABOOST	82.54	84.38	81.97	86.15	82.54	88.57
XGBOOST	90.62	88.24	84.06	81.16	89.55	84.51
NAIVEB	91.18	85.29	86.15	86.57	86.96	88.24
MLP	82.54	83.33	84.38	83.08	87.1	84.51
GBOOST	88.89	86.15	83.08	87.88	86.15	88.24
CATBOOST	85.25	84.85	86.5	82.54	85.71	86.96

Table 7. Results for Stacking Ensemble on Holdout Data for Polycystic Ovary Syndrome classification.

	All Results					Negative Class Results			Positive Class Results		
Metric	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Stacking with feature engineering	93.94	93.54	95.15	91.18	97.1	93.94	91.18	93.54	96.71	97.10	96.40
Stacking without feature engineering	93.75	90.91	94.17	88.24	97.1	93.75	88.24	90.91	94.37	97.10	96.71

are also marginally better with feature engineering, while specificity remains the same in both scenarios, suggesting that the model's ability to identify true negatives is not influenced by feature engineering.

Overall, the table suggests that feature engineering enhances the performance of the ensemble method for PCOS classification.

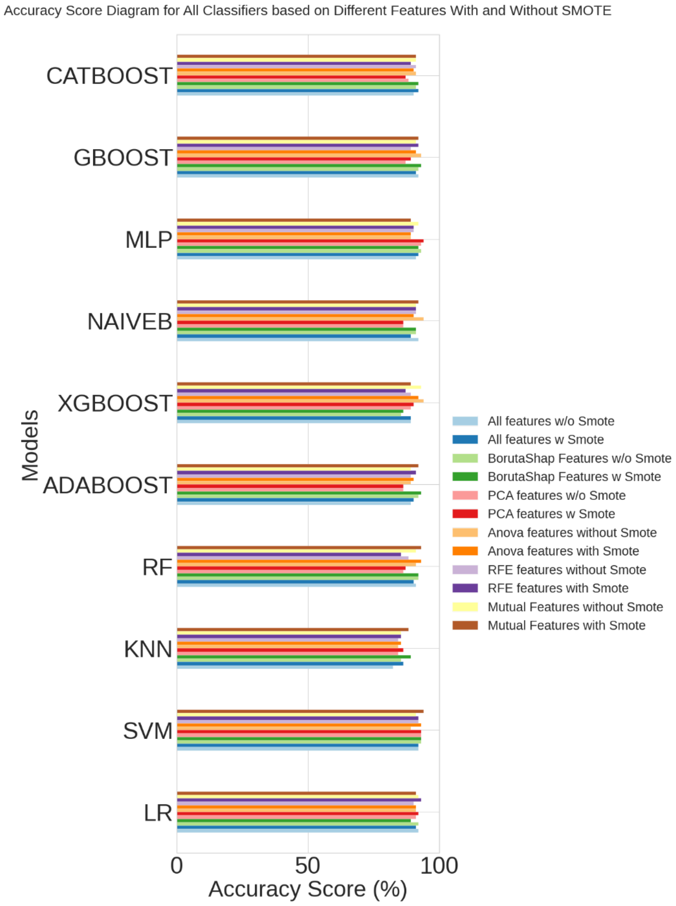

Figure 11. Accuracy score diagram for all classifiers based on different features with and without SMOTE.

Figure 11 is a bar chart comparing the performance of various machine learning classifiers on PCOS. Each classifier (e.g., CATBOOST, GBOOST, MLP, NAIVEB, XGBOOST, ADABOOST, RF, KNN, SVM, LR) is evaluated based on different feature selection methods with and without the application of SMOTE. The classifiers are benchmarked by their accuracy scores, represented as percentages. The color coding represents different combinations of features and whether SMOTE was applied. For instance, "All features w/o SMote" could indicate the use of all features without oversampling, while "BorutaShap Features w/ SMote" might refer to a subset of features selected by the BorutaShap algorithm with oversampling applied. This chart provides a visual representation of which combinations of features and oversampling techniques result in higher predictive accuracy for each classifier, which is critical for developing an effective AI-assisted diagnostic platform for PCOS.

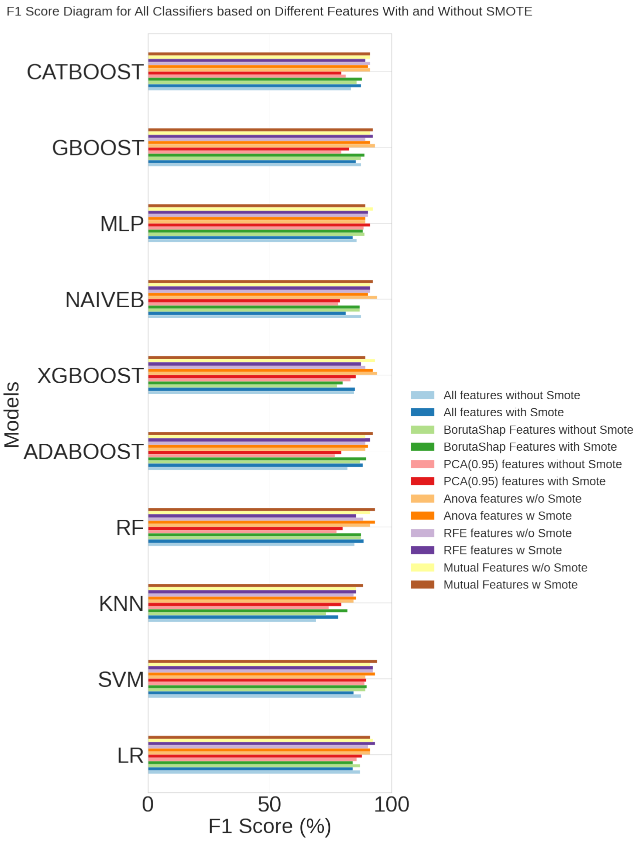

Figure 12. F1 score diagram for all classifiers based on different features with and without SMOTE.

Figure 12 presents a comparative analysis of various machine learning classifiers based on their F1 scores, a harmonic mean of precision and recall, considering different feature selection methods with and without the application of SMOTE. The feature selection methods used include all features, BorutaShap, PCA (with 95% variance), ANOVA, RFE, and Mutual Features. The diagram shows that the performance of each classifier varies with different feature selection methods, and the use of SMOTE generally seems to improve the F1 scores across most classifiers, indicating its effectiveness in handling class imbalance, which is a common issue in medical diagnoses like PCOS. The trends indicate that feature selection and data sampling techniques significantly impact the model's predictive performance.

5 Discussion and Conclusion

PCOS is a complex endocrine disorder that significantly impacts women in their reproductive years. It is characterized by diverse metabolic, reproductive, and psychological implications, PCOS presents significant diagnostic challenges. The integration of Artificial Intelligence (AI) in its diagnosis heralds a new era in gynaecological healthcare. This groundbreaking approach promises enhanced accuracy, speed, and personalized treatment [90–92]. Our research delves into this innovative domain, presenting an AI-assisted diagnostic platform tailor-made for PCOS. This platform represents a significant stride in gynaecological, potentially revolutionizing the diagnostic process by offering a more nuanced and efficient methodology. Such advancements are crucial given the varied and often overlapping symptoms of PCOS, which include menstrual irregularities, hyperandrogenism, and polycystic ovaries. The traditional diagnostic methods, reliant on clinical assessments and imaging, often fall short in addressing the complexity of PCOS, leading to potential misdiagnoses and treatment delays.

A timely and precise diagnosis of PCOS is critical for effective management and intervention. This condition transcends mere gynaecological issues, involving a spectrum of health problems such as metabolic syndrome, cardiovascular risks, and psychological distress [93]. The heterogeneous nature of PCOS, with its spectrum of symptoms and potential for numerous comorbidities, necessitates a more sophisticated diagnostic approach [94]. Our AI-assisted diagnostic platform leverages advanced machine learning classifiers, such as Random Forest, CatBoost, and MLP, combined with sophisticated feature selection methods like BorutaShap and SMOTE. These technologies enhance the diagnostic process, addressing the challenges posed by the variability of symptoms and the need for multiple tests. The ensemble of classifiers, fine-tuned through feature engineering, achieves high accuracy and F1 scores, highlighting the platform's reliability in differentiating between PCOS and non-PCOS cases. This precision is crucial for gynaecologists, allowing for timely and accurate diagnoses, which is essential for effective PCOS management and reducing the risk of long-term complications.

The field of gynaecological is experiencing a paradigm shift with AI's introduction in diagnosing conditions like PCOS. Traditional diagnostic methods, while foundational, have limitations in sensitivity and specificity. Our AI-assisted diagnostic platform transcends these conventional methods by utilizing a diverse array of machine-learning classifiers and feature selection techniques. The platform combines clinical and demographic data, providing a comprehensive view of a patient's health status. This approach addresses the critical challenges in PCOS diagnosis, such as the variability in clinical manifestations and the syndrome's heterogeneity. The model's adaptability to diverse clinical scenarios, indicated by its predictive accuracy and efficiency, is particularly beneficial for a condition as complex as PCOS. Moreover, the use of machine learning algorithms offers insights into underlying data patterns and correlations, potentially paving the way for more personalized treatment strategies.

Incorporating the Rotterdam criteria into our AI model is a significant step forward. These criteria include oligo-anovulation, clinical and/or biochemical signs of hyperandrogenism, and polycystic ovarian morphology. By aligning with these criteria, our model reflects the current clinical diagnostic standards, ensuring its applicability in modern gynecological practice. The AI model's sophisticated machine learning techniques, such as the ensemble approach combining Random Forest, CatBoost, and MLP classifiers, demonstrate an exceptional capacity to handle complex datasets. This capability is crucial in diagnosing PCOS, where symptomatic overlaps with other conditions can lead to uncertainties. The model's predictive capability, refined through feature selection methods and class imbalance techniques like SMOTE, captures the nuanced variations in PCOS presentations, making it a robust and reliable tool in gynecological healthcare.

Our research explores various machine learning classifiers, including Random Forest, CatBoost, and MLP, assessing their efficacy in diagnosing PCOS. The stacking ensemble of these classifiers exhibits superior performance, evidenced by high F1 and accuracy scores. This finding underscores the potential of machine learning in enhancing diagnostic precision in gynaecological, offering a more accurate, efficient, and personalized diagnostic process. The use of these advanced classifiers, each with its unique strengths, allows for a more holistic approach to diagnosing PCOS, addressing the diverse manifestations of this syndrome.

In medical diagnoses, class imbalances are a common issue that can skew results. Our use of the SMOTE addresses this challenge effectively. By balancing the dataset, SMOTE improves the performance of machine learning models, enhancing the overall quality of our AI-assisted diagnostic tools. This technique is particularly useful in conditions like PCOS, where the variety of symptoms and their prevalence can vary significantly among different patient populations. The application of SMOTE ensures that our model remains effective and accurate across a diverse range of clinical scenarios.

Feature selection is a critical aspect of our AI model's performance. We employed various methods, including BorutaShap and PCA, to determine the most impactful features for PCOS diagnosis. This approach is vital in a condition like PCOS, where symptoms and markers are diverse and multifactorial. The right selection of features significantly impacts the model's predictive accuracy, allowing for a more precise and individualized diagnostic process. This aspect of AI modelling is crucial for handling the complexity and variability inherent in PCOS symptoms and presentations.

The implementation of advanced algorithms like XGBoost and SVM, especially in combination with feature selection techniques, has shown remarkable improvements in predictive accuracy. These findings highlight the potential of these algorithms in refining the diagnostic process of AI tools used for PCOS. Advanced algorithms offer deeper insights and more nuanced analysis of data, which is crucial for accurately diagnosing a complex and multifaceted condition like PCOS.

Feature engineering plays a crucial role in the performance of our AI model. This process involves refining and selecting relevant features, leading to improved precision, F1-score, and recall. The importance of feature engineering lies in its ability to enhance the model's accuracy and reliability, making it a vital component in the development of efficient AI-assisted diagnostic platforms. The careful selection and optimization of features ensure that the model is finely tuned to the specific characteristics of PCOS, enhancing its effectiveness in real-world clinical settings.

Our AI-assisted diagnostic platform for PCOS represents a significant advancement in gynaecological health. It harnesses the power of AI and machine learning to provide healthcare professionals with an effective tool for diagnosing PCOS. This innovative approach has the potential to transform gynaecological care, offering immense benefits to women suffering from this complex syndrome. The platform's accuracy, efficiency, and personalized diagnostic capabilities make it a promising tool in the healthcare sector. As we continue to refine and validate this platform, its integration into clinical practice could revolutionize the diagnosis and management of PCOS, improving the quality of life for millions of women worldwide.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study was conducted in accordance with the ethical principles and national norms and standards for conducting medical research in Iran, as approved by the Research Ethics Committee of Urmia University of Medical Sciences (Approval ID: IR.UMSU.REC.1403.234, Approval Date: 2024-10-30). Written informed consent was obtained from all participants. The researchers ensured compliance with all professional and legal requirements, maintaining the confidentiality and anonymity of participant data.

References

[1] Riestenberg, C., Jagasia, A., Markovic, D., Buyalos, R. P., & Azziz, R. (2022). Health care-related economic burden of polycystic ovary syndrome in the United States: pregnancy-related and long-term health consequences. *The Journal of Clinical Endocrinology & Metabolism, 107*(2), 575-585. [Crossref]

[2] Notaro, A. L. G., & Neto, F. T. L. (2022). The use of metformin in women with polycystic ovary syndrome: an updated review. *Journal of assisted reproduction and genetics, 39*(3), 573-579. [Crossref]

[3] Dinsdale, N., Nepomnaschy, P., & Crespi, B. (2021). The evolutionary biology of endometriosis. *Evolution, medicine, and public health, 9*(1), 174-191. [Crossref]

[4] Harris, H. R., Titus, L. J., Cramer, D. W., & Terry, K. L. (2017). Long and irregular menstrual cycles, polycystic ovary syndrome, and ovarian cancer risk in a population-based case-control study. *International journal of cancer, 140*(2), 285-291. [Crossref]

[5] Zheng, S., Chen, Y., Ma, M., & Li, M. (2022). Mechanism of quercetin on the improvement of ovulation disorder and regulation of ovarian

CNP/NPR2 in PCOS model rats. *Journal of the Formosan Medical Association, 121*(6), 1081-1092. [Crossref]

[6] Schmid, J., Kirchengast, S., Vytiska-Binstorfer, E., & Huber, J. (2004). Infertility caused by PCOS—health-related quality of life among Austrian and Moslem immigrant women in Austria. *Human reproduction, 19*(10), 2251-2257. [Crossref]

[7] Homburg, R. (2003). The management of infertility associated with polycystic ovary syndrome. *Reproductive Biology and Endocrinology, 1*(1), 109. [Crossref]

[8] Ahmed, A. A., Moselhy, S. S., Kumosani, T. A., Huwait, E. A., Al-Ghamdi, M. A., Al-Madani, K. A., ... & Kumosani, A. T. (2020). Ultrasonographic and biochemical assessments as early prediction of polycystic ovarian syndrome in obese women. *African Health Sciences, 20*(2), 676-681. [Crossref]

[9] Kodama, S., & Torrealday, S. (2021). Pharmacologic treatments for PCOS patients. *Clinical obstetrics and gynecology, 64*(1), 55-64. [Crossref]

[10] Fernandez, R. C., Moore, V. M., Rumbold, A. R., Whitrow, M. J., Avery, J. C., & Davies, M. J. (2021). Diagnosis delayed: health profile differences between women with undiagnosed polycystic ovary syndrome and those with a clinical diagnosis by age 35 years. *Human Reproduction, 36*(8), 2275-2284. [Crossref]

[11] ESHRE, T. R., & ASRM-Sponsored PCOS Consensus Workshop Group. (2004). Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertility and sterility, 81*(1), 19-25. [Crossref]

[12] Gesteiro, E., Megía, A., Guadalupe-Grau, A., Fernandez-Veledo, S., Vendrell, J., & González-Gross, M. (2021). Early identification of metabolic syndrome risk: A review of reviews and proposal for defining pre-metabolic syndrome status. *Nutrition, Metabolism and Cardiovascular Diseases, 31*(9), 2557-2574. [Crossref]

[13] Wang, J., Chen, R., Long, H., He, J., Tang, M., Su, M., ... & Tang, L. (2025). Artificial intelligence in polycystic ovarian syndrome management: past, present, and future. *La radiologia medica*, 1-33. [Crossref]

[14] Sadi, R. P. R., Raju, Y., Saravanan, M. G., & Chandra, G. R. (2022). *AI APPLICATIONS. GCS PUBLISHERS.*

[15] Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature reviews Molecular cell biology, 23*(1), 40-55. [Crossref]

[16] Zigarelli, A., Jia, Z., & Lee, H. (2022). Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: observational study. *JMIR Formative Research, 6*(3), e29967. [Crossref]

[17] Allan, S., Olaiya, R., & Burhan, R. (2022). Reviewing the use and quality of machine learning in developing clinical prediction models for cardiovascular disease. *Postgraduate Medical Journal, 98*(1161), 551-558. [Crossref]

[18] Mahesswari, G. U., & Maheswari, P. U. (2024). SmartScanPCOS: a feature-driven approach to cutting-edge prediction of polycystic ovary syndrome using machine learning and explainable artificial intelligence. *Heliyon, 11*, e39205. [Crossref]

[19] Kumar, A., Sharma, D., Tyagi, A., Kapoor, A., & Mishra, A. (2024, April). Machine Learning Framework for Polycystic Ovary Syndrome. In *International Conference on Information and Communication Technology for Intelligent Systems* (pp. 367-379). Singapore: Springer Nature Singapore. [Crossref]

[20] Rachana, B., Priyanka, T., Sahana, K. N., Supritha, T. R., Parameshachari, B. D., & Sunitha, R. (2021). Detection of polycystic ovarian syndrome using follicle recognition technique. *Global Transitions Proceedings, 2*(2), 304-308. [Crossref]

[21] Gopalakrishnan, C., & Iyapparaja, M. (2021). Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning. *International Journal of System Assurance Engineering and Management*, 1-8. [Crossref]

[22] Jeevitha, S., & Priya, N. (2022, July). Optimized segmentation technique for detecting PCOS in ultrasound images. In *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 1* (pp. 759-771). Singapore: Springer Nature Singapore. [Crossref]

[23] Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., & Livingood, W. (2021). A review of machine learning in building load prediction. *Applied Energy, 285*, 116452. [Crossref]

[24] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering, 17*(4), 491-502. [Crossref]

[25] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics, 85*, 189-203. [Crossref]

[26] Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & De Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes, 4*(1), 299. [Crossref]

[27] Jaw, E., & Wang, X. (2021). Feature selection and ensemble-based intrusion detection system: an efficient and comprehensive approach. *Symmetry, 13*(10), 1764. [Crossref]

[28] Danaei Mehr, H., & Polat, H. (2022). Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques. *Health and*

Technology, 12(1), 137-150. [Crossref]

[29] Meena, K., Manimekalai, M., & Rethinavalli, S. (2015). Correlation of artificial neural network classification and NFRS attribute filtering algorithm for PCOS data. *Int. J. Res. Eng. Technol, 4*(3), 519-524. [CrossRef]

[30] Balogun, J. A., Egejuru, N. C., & Idowu, P. A. (2018). Comparative analysis of predictive models for the likelihood of infertility in women using supervised machine learning techniques. *Computer Reviews Journal, 2*(1), 313-330.

[31] Vikas, B., Anuhya, B. S., Chilla, M., & Sarangi, S. (2018). A critical study of Polycystic Ovarian Syndrome (PCOS) classification techniques. *Int J Comput Eng Manag, 4*, 1-7.

[32] Zhang, X., Liang, B., Zhang, J., Hao, X., Xu, X., Chang, H. M., ... & Tan, J. (2021). Raman spectroscopy of follicular fluid and plasma with machine-learning algorithms for polycystic ovary syndrome screening. *Molecular and cellular endocrinology, 523*, 111139. [Crossref]

[33] Nandipati, S. C., & Ying, C. X. (2020). Polycystic Ovarian Syndrome (PCOS) classification and feature selection by machine learning techniques. *Applied Mathematics and Computational Intelligence (AMCI), 9*, 65-74.

[34] Inan, M. S. K., Ulfath, R. E., Alam, F. I., Bappee, F. K., & Hasan, R. (2021, January). Improved sampling and feature selection to support extreme gradient boosting for PCOS diagnosis. In *2021 IEEE 11th annual computing and communication workshop and conference (CCWC)* (pp. 1046-1050). IEEE. [Crossref]

[35] Munjal, A., Khandia, R., & Gautam, B. (2020). A machine learning approach for selection of Polycystic Ovarian Syndrome (PCOS) attributes and comparing different classifier performance with the help of WEKA and PyCaret. *Int. J. Sci. Res, 9*, 1-5.

[36] Pushkarini, H., & Anusuya, M. A. (2020). A prediction model for evaluating the risk of developing PCOS. *Int Res J Eng Technol (IRJET), 7*(9), 1150-6.

[37] Neto, C., Silva, M., Fernandes, M., Ferreira, D., & Machado, J. (2021, February). Prediction models for Polycystic Ovary Syndrome using data mining. In *International Conference on Advances in Digital Science* (pp. 210-221). Cham: Springer International Publishing. [Crossref]

[38] Bharati, S., Podder, P., & Mondal, M. R. H. (2020, June). Diagnosis of polycystic ovary syndrome using machine learning algorithms. In *2020 IEEE region 10 symposium (TENSYMP)* (pp. 1486-1489). IEEE. [Crossref]

[39] Hassan, M. M., & Mirza, T. (2020). Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome. *Int. J. Comput. Appl, 975*(8887). [CrossRef]

[40] Prapty, A. S., & Shitu, T. T. (2020, December). An efficient decision tree establishment and performance analysis with different machine learning approaches on polycystic ovary syndrome. In *2020 23rd International conference on computer and information technology (ICCIT)* (pp. 1-5). IEEE. [Crossref]

[41] Tanwani, N. (2020). Detecting PCOS using machine learning. *Int J Modern Trends Eng Sci (IJMTES), 7*(1), 1-20.

[42] Thomas, N., & Kavitha, A. (2020). Prediction of polycystic ovarian syndrome with clinical dataset using a novel hybrid data mining classification technique. *Int J Adv Res Eng Technol (IJARET), 11*(11), 1872-1881.

[43] Dekking, F. M. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

[44] Kiconco, S., Earnest, A., Enticott, J., Hart, R., Mori, T. A., Hickey, M., ... & Joham, A. E. (2023). Normative cut-offs for polycystic ovary syndrome diagnostic features in adolescents using cluster analysis. *European journal of endocrinology, 188*(6), 494-502. [Crossref]

[45] Dapas, M., Lin, F. T., Nadkarni, G. N., Sisk, R., Legro, R. S., Urbanek, M., ... & Dunaif, A. (2020). Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. *PLoS medicine, 17*(6), e1003132. [Crossref]

[46] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).

[47] Karadayı Ataş, P. (2024). Exploring the molecular interaction of PCOS and endometrial carcinoma through novel hyperparameter-optimized ensemble clustering approaches. *Mathematics, 12*(2), 295. [Crossref]

[48] Expert Panel on the Identification, Treatment of Overweight, Obesity in Adults (US), National Heart, Lung, Blood Institute, ... & Kidney Diseases (US). (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report* (No. 98). National Institutes of Health, National Heart, Lung, and Blood Institute.

[49] Consultation, W. H. O. (2000). Obesity: preventing and managing the global epidemic. *World Health Organ Tech Rep Ser, 894*, 1-253.

[50] Fraser, I. S., Critchley, H. O., Munro, M. G., & Broder, M. (2007). A process designed to lead to international agreement on terminologies and definitions used to describe abnormalities of menstrual bleeding. *Fertility and sterility, 87*(3), 466-476. [Crossref]

[51] Fraser, I. S., Critchley, H. O. D., Munro, M. G., & Broder, M. (2007). Can we achieve international agreement on terminologies and definitions used to describe abnormalities of menstrual bleeding?. *Human reproduction, 22*(3), 635-643. [Crossref]

[52] Grenache, D. G. (2020). Current practices when reporting quantitative human chorionic gonadotropin test results. *The Journal of Applied Laboratory Medicine, 5*(5), 850-857. [Crossref]

[53] Ae, T. (1997). Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab, 82*, 2248-2256.

[54] Arroyo, A., Laughlin, G. A., Morales, A. J., & Yen, S. S. C. (1997). Inappropriate gonadotropin secretion in polycystic ovary syndrome: influence of adiposity. *The Journal of Clinical Endocrinology & Metabolism, 82*(11), 3728-3733. [Crossref]

[55] Alberti, K. G. M., Zimmet, P., & Shaw, J. (2005). The metabolic syndrome—a new worldwide definition. *The Lancet, 366*(9491), 1059-1062. [CrossRef]

[56] Cappola, A. R. (2019). The thyrotropin reference range should be changed in older patients. *Jama, 322*(20), 1961-1962. [Crossref]

[57] Schlechte, J., Sherman, B., Halmi, N., vanGilder, J., Chapler, F., Dolan, K., ... & Harris, C. (1980). Prolactin-secreting pituitary tumors in amenorrheic women: a comprehensive study. *Endocrine reviews, 1*(3), 295-308. [Crossref]

[58] Giustina, A., Adler, R. A., Binkley, N., Bouillon, R., Ebeling, P. R., Lazaretti-Castro, M., ... & Bilezikian, J. P. (2019). Controversies in vitamin D: summary statement from an international conference. *The Journal of Clinical Endocrinology & Metabolism, 104*(2), 234-240. [Crossref]

[59] Chen, S., & He, H. (2013). Nonstationary stream data learning with imbalanced class distribution. *Imbalanced learning: Foundations, algorithms, and applications*, 151-186.

[60] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182. [Crossref]

[61] Duch, W. (2006). Filter methods. In *Feature extraction: foundations and applications* (pp. 89-117). Berlin, Heidelberg: Springer Berlin Heidelberg. [Crossref]

[62] Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2015). Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM transactions on computational biology and bioinformatics, 13*(5), 971-989. [Crossref]

[63] Li, G., Yu, Z., Yang, K., Lin, M., & Chen, C. P. (2024). Exploring feature selection with limited labels: A comprehensive survey of semi-supervised and unsupervised approaches. *IEEE Transactions on Knowledge and Data Engineering, 36*(11), 6124-6144. [Crossref]

[64] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer. [Crossref]

[65] Liu, B., Xia, H., Zhang, J., Hu, Y., & Wang, Q. (2025). Building Personalized Risk Prediction Models fBuilding Personalized Risk Prediction Models for Polycystic Ovary Syndrome Using Machine Learning Techniques: A Retrospectivor Polycystic Ovary Syndrome Using Machine Learning Techniques: A Retrospective Study. *Current Science, 5*(2), 1942-1955.

[66] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia, 126*(5), 1763-1768. [Crossref]

[67] Lafortuna, C. L., Agosti, F., Marazzi, N., & Sartorio, A. (2004). The relationship between body composition and muscle power output in men and women with obesity. *Journal of endocrinological investigation, 27*(9), 854-861. [Crossref]

[68] Chen, X. W., & Jeong, J. C. (2007, December). Enhanced recursive feature elimination. In *Sixth international conference on machine learning and applications* (*ICMLA 2007*) (pp. 429-435). IEEE. [Crossref]

[69] Ejiyi, C. J., Qin, Z., Ukwuoma, C. C., Nneji, G. U., Monday, H. N., Ejiyi, M. B., ... & Bamisile, O. O. (2025). Comparative performance analysis of Boruta, SHAP, and Borutashap for disease diagnosis: a study with multiple machine learning algorithms. *Network: Computation in Neural Systems, 36*(3), 507-544. [Crossref]

[70] Samara, M. N., & Harry, K. D. (2025). Integrating Boruta, LASSO, and SHAP for Clinically Interpretable Glioma Classification Using Machine Learning. *BioMedInformatics, 5*(3), 34. [Crossref]

[71] Bulut, O., Tan, B., Mazzullo, E., & Syed, A. (2025). Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. *Information, 16*(6), 476. [Crossref]

[72] Sawyer, S. F. (2009). Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative Therapy, 17*(2), 27E-38E. [Crossref]

[73] Bouchene, M. M., & Fatima, M. (2025). Normalized mean difference (NMD): a novel filter-based feature selection method. *International Journal of Machine Learning and Cybernetics*, 1-19. [Crossref]

[74] Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural computing and applications, 24*(1), 175-186. [Crossref]

[75] Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998, June). Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data* (pp. 94-105). [Crossref]

[76] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics, 2*(4), 433-459. [Crossref]

[77] Agrawal, T. (2020). Hyperparameter optimization

using scikit-learn. In *Hyperparameter optimization in machine learning: make your machine learning and deep learning models more efficient* (pp. 31-51). Berkeley, CA: Apress. [Crossref]

[78] Kramer, O. (2016). *Machine learning for evolution strategies* (Vol. 20). Switzerland: Springer. [Crossref]

[79] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of clinical epidemiology, 122*, 56-69. [Crossref]

[80] Denny, A., Raj, A., Ashok, A., Ram, C. M., & George, R. (2019, October). i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 673-678). IEEE. [Crossref]

[81] Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augmented Human Research, 5*(1), 12. [Crossref]

[82] Tiwari, S., Kane, L., Koundal, D., Jain, A., Alhudhaif, A., Polat, K., ... & Althubiti, S. A. (2022). SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning. Expert Systems with Applications, 203, 117592. [Crossref]

[83] Rehman Javed, A., Jalil, Z., Atif Moqurrab, S., Abbas, S., & Liu, X. (2022). Ensemble adaboost classifier for accurate and fast detection of botnet attacks in connected vehicles. *Transactions on Emerging Telecommunications Technologies, 33*(10), e4088. [Crossref]

[84] Yu, D., Liu, Z., Su, C., Han, Y., Duan, X., Zhang, R., ... & Xu, S. (2020). Copy number variation in plasma as a tool for lung cancer prediction using Extreme Gradient Boosting (XGBoost) classifier. *Thoracic cancer, 11*(1), 95-102. [Crossref]

[85] Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning, 29*(2), 103-130. [Crossref]

[86] Bahad, P., & Saxena, P. (2019, December). Study of adaboost and gradient boosting algorithms for predictive analytics. In *International Conference on Intelligent Computing and Smart Communication 2019: Proceedings of ICSC 2019* (pp. 235-244). Singapore: Springer Singapore. [Crossref]

[87] Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost classifier with other machine learning methods. *International Journal of Advanced Computer Science and Applications, 11*(11), 738-748. [CrossRef]

[88] Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888.*

[89] Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094-1096). Springer, Berlin, Heidelberg. [Crossref]

[90] Johnstone, E. B., Rosen, M. P., Neril, R., Trevithick, D., Sternfeld, B., Murphy, R., ... & Cedars, M. I. (2010). The polycystic ovary post-Rotterdam: a common, age-dependent finding in ovulatory women without metabolic significance. *The Journal of Clinical Endocrinology & Metabolism, 95*(11), 4965-4972. [Crossref]

[91] Ehrmann, D. A. (2005). Polycystic ovary syndrome. *New England Journal of Medicine, 352*(12), 1223-1236. [Crossref]

[92] Dumesic, D. A., Oberfield, S. E., Stener-Victorin, E., Marshall, J. C., Laven, J. S., & Legro, R. S. (2015). Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocrine reviews, 36*(5), 487-525. [Crossref]

[93] Hoeger, K. M., Dokras, A., & Piltonen, T. (2021). Update on PCOS: consequences, challenges, and guiding treatment. *The Journal of Clinical Endocrinology & Metabolism, 106*(3), e1071-e1083. [Crossref]

[94] Gibson-Helm, M. E., Lucas, I. M., Boyle, J. A., & Teede, H. J. (2014). Women's experiences of polycystic ovary syndrome diagnosis. *Family practice, 31*(5), 545-549. [Crossref]