Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2024;13(3):568-74

Global interpretation of machine learning models in predicting polycystic ovary syndrome with the explainable artificial intelligence method SHAP

Ipek Balikci Cicek, Zeynep Kucukakcali

İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye

Received 05 December 2023; Accepted 02 February 2024
Available online 01 August 2024 with doi: 10.5455/medscience.2023.12.228

Abstract

Polycystic ovary syndrome (PCOS) is a complex condition characterized by high male hormone levels, irregular menstrual cycles, lack of ovulation, and sometimes small ovarian cysts. Often underdiagnosed, PCOS leads to significant health issues, making timely and efficient identification crucial. Recently, machine learning (ML) has shown promise in medical diagnoses, but the perceived "black box" nature of ML models necessitates explanations of key parameters influencing predictions. This study aims to provide global explanations using SHapley Additive exPlanations (SHAP) to ensure the efficiency, effectiveness, and reliability of the ML model. An open-access dataset with 300 PCOS patients was utilized to predict whether a patient's luteinizing hormone (LH) to follicle-stimulating hormone (FSH) ratio is up to 1 or more than 1. The study employed ML classifiers including AdaBoost, XGBoost, CatBoost, and Bagging methods, with Bagging performing the best. The modeling process used a 5-fold cross-validation approach, splitting the dataset into 80% training and 20% testing sets. The model's performance was evaluated using accuracy (ACC), balanced accuracy (b-ACC), specificity (SP), sensitivity (SE), negative predictive value (npv), positive predictive value (ppv), and F1-score. The Bagging method yielded the following performance metrics: ACC (99.0%), b-ACC (99.0%), SE (98.8%), SP (99.1%), ppv (97.7%), npv (99.5%), and F1-score (98.3%). SHAP analysis identified the top predictors for distinguishing between LH: FSH ratio categories as TTng/dL, BMI, AMH, age, family history, and menstrual cycle regulation. This study demonstrates that incorporating SHAP explanations enhances the interpretability and reliability of ML models in diagnosing PCOS.

Keywords: Polycystic ovary syndrome, explainable artificial intelligence, machine learning, SHapley Additive exPlanations

Introduction

Polycystic ovarian syndrome (PCOS) is a multifaceted endocrine illness characterized by hormonal dysregulation, the formation of cysts in the ovaries, and the failure to release eggs, all of which have a substantial influence on a woman's well-being. [1,2]. Oligomenorrhea and amenorrhea-like interruptions in the menstrual cycle can arise from the disturbance of reproductive hormones, such as LH, FSH, total testosterone (TT), and estrogen. PCOS is defined by excessive levels of androgens, irregular menstruation cycles, and the existence of ovarian cysts of different sizes, but there may be considerable variance across individuals. This intricate condition mostly impacts teenagers who are at an elevated risk of acquiring several comorbidities, such as cardiovascular disease, obesity, type II diabetes, endometrial dysplasia, infertility, and mental health disorders [3,4]. The recently recognized Rotterdam criteria, endorsed by an international evidence-based PCOS guideline, are employed for the diagnosis of PCOS [5,6].

According to specific studies, the brain has a dual role as both a regulator and an impacted organ in PCOS (the brain phenotype). The brain, equipped with many receptors and neurons together with their neurotransmitters, generates a higher frequency of gonadotropin pulses. Consequently, individuals with PCOS produce a greater amount of LH compared to FSH [6]. An alteration in the LH:FSH ratio results in an elevation in ovarian androgen synthesis. Several research has shown the most effective threshold values for the LH:FSH ratio in patients with PCOS. Nevertheless, because of the diverse sensitivity

CITATION
Balikci Cicek I, Kucukakcali Z. Global interpretation of machine learning models in predicting polycystic ovary syndrome with the explainable artificial intelligence method SHAP. Med Science. 2024;13(3):568-74.

OPEN ACCESS

Corresponding Author: Ipek Balikci Cicek, İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye
Email: ipek.balikci@inonu.edu.tr

and specificity, the precise threshold value remains uncertain [7-9]. Hsu et al. suggested that an LH:FSH ratio exceeding one provided the most favorable balance between sensitivity and specificity in diagnosing PCOS [7]. On the other hand, Papaleo et al. suggested that an LH:FSH ratio of 2 is a characteristic indicator of abnormal gonadotropin production in individuals with PCOS [10]. Moreover, hyperandrogenemia diminishes the inhibitory effect of estrogen on the hypothalamus, leading to an elevation in the frequency of LH pulses. Consequently, a harmful cycle is initiated, leading to various clinical manifestations of PCOS, such as hyperandrogenism [11].

The criteria-based diagnosis of PCOS is complicated by the variability in the clinical evaluation of hyperandrogenism and the identification of monthly irregularity. Moreover, the variation in PCOS standard norms worsens these challenges [12]. Estimations indicate that a significant proportion of women with PCOS receive their diagnosis more than two years after the onset of the condition. Nevertheless, it is probable that this assessment is lower than the actual number [13].

Artificial intelligence (AI) refers to computer-based systems that imitate human intellect. Machine learning (ML) is a specific branch of AI that emphasizes the ability to learn from past experiences and utilize this knowledge for making future decisions [14]. The significant progress made in AI and machine learning in the past ten years has great potential for significantly enhancing our ability to detect and manage PCOS. This is partially attributed to the ability of AI to analyze large quantities of diverse data, making it an ideal tool for assisting in the identification of complex diseases such as PCOS [15]. Given the opaque nature of ML algorithms, it is difficult to provide justifications for the specific patient characteristics that influence a particular prediction. The limited interpretability of ML techniques has hindered their widespread use in medical decision support. One of the main obstacles to using ML in the medical industry is the lack of intuitive understanding of ML models. [16,17]. Explainable AI (XAI) is characterized by its capacity to provide system transparency and interpretability. XAI has played a significant role in addressing several problems, including bias, unfairness, security, and causality, in recent times [18]. By incorporating XAI into AI-driven systems used for diagnostic, prognostic, and therapeutic purposes, it is possible to enhance the system's responsibility and foster trust in its medical conclusions. XAI facilitates the process of verifying the effectiveness of machine learning models in a healthcare setting [19].

XAI techniques improve model predictions' interpretability, understanding, and reliability. Thefore, the focus of this research is on the application of SHapley Additive exPlanations (SHAP) for PCOS screening. SHAP utilizes a game theory paradigm and provides quantifiable numbers for explanation accuracy and consistency. SHAP ranks features depending on their relevance [20]. Through global explanations, the current study attempts to provide model explanations to ensure effectiveness, efficiency, and trust in the created model.

Material and Methods

Dataset

The PCOS open-access dataset was taken from https://figshare.com/articles/dataset/updated_PCOS_row_data/19089371. The dataset had collected in Sudan, between October 2020 and September 2021. Using a convenience sampling strategy, the study sample had drawn from various infertility clinics. Women who were diagnosed with PCOS based on the Rotterdam criteria were considered eligible if they met at least two of the following criteria: oligomenorrhea/anovulation (menstrual cycles delayed by more than 35 days or fewer than 8 spontaneous bleeding episodes per year), hyperandrogenism (clinical hirsutism based on a modified Ferriman-Gallwey score of 8 or biochemical hirsutism), and polycystic ovaries observed on ultrasonography (presence of 12 or more follicles measuring 2 to 9 mm in diameter in each ovary, and/or an ovarian volume greater than 10 mL3). Women with pre-existing systemic illnesses such as diabetes mellitus and cardiovascular disease, who were already using medication such as glucocorticoids, ovulation induction medications, oral contraceptives, estrogenic or anti-androgenic medicines before the study, were not included in the research [21,22].

Table 1 shows the 6 independent variables and 1 dependent variable in the PCOS dataset, as well as their explanations.

Table 1. Variable descriptions from the dataset

Names of variables	Variable type
Age	Numerical
Family history (yes/no)	Categorical
Regulation of menstrual cycle (regular/irregular)	Categorical
BMI	Numerical
TotalTestesterone ng/dL	Numerical
Anti Mullein's Hormone	Numerical
Regulation of menstrual cycle	Numerical
LH:FSH ratio group (up to 1/more than 1)	Target variable/ categorical

Biostatistics Analysis

The qualitative variables included in the study were summarized by number (percentage). Whether the quantitative data showed normal distribution was evaluated with the Kolmogorov Smirnov test. Since quantitative data did not show a normal distribution, they were summarized with median (minimum-maximum). In statistical analyses, Mann-Whitney U test and Pearson chi-square test were used where appropriate. In the statistical analysis applied, $p<0.05$ value was considered statistically significant. All analyses were performed using IBM SPSS Statistics 26.0 for Windows (New York; USA).

Model Phase

There are methods used in ML to anticipate target values depending on independent variables values. The analyses and computations were conducted using Python Version 3.6.5, utilizing the sklearn, and SHAP packages. The methods that will be used to estimate the LH:FSH ratio group (up to 1 / more than 1) within the scope of this study are as follows:

Adaptive Boosting (AdaBoost)

AdaBoost algorithm is an ensemble learning method proposed by Freund and Schapire that can increase the accuracy of weakly learning classifiers by changing the distribution of sample weights. By training more than one classifier for the same training set, weak classifiers are integrated with each other. The AdaBoost algorithm basically changes the distribution of data samples [23]. The new data set with modified weights is trained again to obtain a new weak classifier. In the first repetition, the weights of all samples are the same. With each iteration, the weight of misclassified examples increases; The weight of the classified samples is reduced and all weights are normalized. In the last step, by voting according to the accuracy values of the prediction classes, the good ones are selected from the weak classifiers and they are integrated with each other to create a better classifier. [24].

Categorical Boosting (CatBoost) Method

CatBoost was developed by Yandex in 2017 [25]. As in boosting algorithms, the general principles of combining weak learners to As in boosting algorithms, it works by combining weak learners to obtain a strong learner, then calculating the errors after creating the first tree, using these errors to create the second tree, and continuing this process by bringing together many trees, resulting in a strong learner. It performs better, especially in processing categorical variables. It has emerged as a faster and higher prediction model that performs better in processing categorical variables and provides categorical variable support [26]. Unlike other boosting algorithms that convert categorical variables into numerical values, this model directly processes categorical variables without any loss of information. It is capable of performing operations like sorting and discretization on categorical variables. Executing these actions enhances the velocity and precision of the model for a dataset with a substantial amount of categorical variables [27].

Extreme Gradient Boosting (XGBoost)

XGBoost is a model that was originally proposed by Tianqi Chen and Carlos Guestrin in 2011 and has since undergone continuous modifications and enhancements in following scientific research [25]. XGBoost is an approach for classifying data based on decision trees. It has gained significant popularity and has been demonstrated to be a very effective technique. XGBoost is an extremely scalable and comprehensive method for boosting trees, used in machine learning for both classification and regression tasks. The framework is built on Boosting Trees, a technique for learning. Classic boosting tree models exclusively utilize the information from the first derivative. Performing distributed training for the nth tree is challenging due to the utilization of the residual from the previous n-1 trees. XGBoost utilizes a second-order Taylor expansion on the loss function and can automatically utilize the CPU's multithreading for parallel processing. In addition, XGBoost utilizes several techniques to prevent overfitting [28].

Bagging Method

The bagging method is an ensemble method devised by Leo Breiman in 1994. It is capable of doing classification and regression tasks. The purpose of this is to enhance the stability and precision of ML algorithms employed in classification and regression tasks. The system functions by amalgamating categorizations derived from randomly generated training sets to give a conclusive forecast. These techniques are frequently employed as a method to reduce variability by including randomness in their production process and subsequently constructing an ensemble from it [29]. Bagging methods have become popular due to their simplicity of implementation and improved accuracy. Bagging may be described as a "smoothing operation" that offers a benefit in enhancing the prediction performance of regression or classification trees. The fundamental concept behind this ensemble technique is that several "weak learners" might collaborate to form a "strong learner." Bagging generates a substantial quantity of decision trees. Individually, each decision tree is considered a "weak learner" in this scenario. However, when all of the decision trees are joined, they form a "strong learner." When classifying a new instance, the method is iterated for every tree in the ensemble. Each tree symbolizes a "vote" for a certain class. The class that receives the highest number of votes is declared as the final prediction for the class of the new instance [30].

The dataset is partitioned into two subsets, with 80% of the data allocated for training purposes and the remaining 20% reserved for testing. In this study, the 5-fold cross-validation technique, which is a type of resampling approach, was employed to guarantee the validity of the model. The n-fold cross-validation approach involves the initial separation of the dataset into n distinct subsets. Subsequently, the model is applied to each of these subsets.

- The second phase involves partitioning the dataset into n parts, with one part reserved for testing and the remaining n-1 parts utilized for training.
- In the last stage, the cross-validation strategy is assessed by calculating the average of the values obtained from the models.

The performance of the modeling was evaluated using metrics such as b-ACC, ACC, SP, SE, npv, ppv, and F1-score.

In the final stage, the study generated variable significance values in order to ascertain the risk variables influencing the LH:FSH ratio. These values offered insights into the extent to which the input factors accounted for the variation in the output variable.

SHapley Additive exPlanations (SHAP)

Given that one of the major disadvantages of ML algorithms is their lack of interpretability, XAI can give clear explanations for how an ML model affects model performance [31]. SHAP analysis is a cutting-edge XAI algorithm. The SHAP facilitates the understanding of model findings by providing a standardized approach. The SHAP value quantifies the influence of each component on the model's predictions, both at the individual level and over the entire dataset. The additive aspect of SHAP guarantees that the final output is obtained by summing all relevant measurements and baseline values. The model's output is derived from game theory through the process of linearly combining the input variables [32]. The "Shapley value" quantifies the individual contribution of each attribute. SHAP's model prediction approach may be used to understand even the most complex models [33].

Results

The dataset used in this research consists of a total of 300 patients, 87 (29.0%) with LH:FSH ratio up to 1 and 213 (71.0%) with LH:FSH ratio more than 1. The mean age of all patients is 29.08±5.882 years and the median age is 29 (17-75).

Biostatistical Analysis Results

When patients with LH: FSH ratio up to 1 and more than 1 were compared in terms of age, BMI, TT ng/dL, and AMH input variables; there was a statistically significant difference between the two groups in the TT ng/dL variable (p<0.05). However, no statistical significance was found in other variables. Additionally, when looking at whether there is a significant relationship between the group with LH:FSH ratio up to 1 and more than 1 and the family history, regulation of menstrual cycle input variables; a statistically significant relationship was found between the family history variable and the two groups (p<0.05). The results of the analyzes are given in Table 2.

Table 2. The analysis of the output variable in relation to the input variables

Variables	LH:FSH ratio group		p
	Up to1	More than 1	
	Median (min-max)	Median (min-max)	
Age	29 (17-43)	29 (17-45)	0.538*
BMI	29 (18.36-46.25)	28.3 (19.11-41.93)	0.102*
TT ng/dL	86 (10-439)	270 (12-580)	**<0.001***
AMH	5.9 (3.1-19.5)	6.2 (4-20.1)	0.345
	n (%)	**n (%)**	
Family history	28 (32.18)	95 (44.60)	**0.047****
	59 (67.82)	118 (55.40)	
Regulation of menstrual cycle	39 (44.83)	104 (48.83)	0.529**
	48 (55.17)	109 (51.17)	

Data are given as count (percentage) or median (minimum–maximum); *:Mann Whitney U test, **:Pearson chi-square test; min: minimum; max: maximum

The performance metrics of machine learning prediction models (AdaBoost, XGBoost, CatBoost, and Bagging) utilized for predicting the LH:FSH ratio up to 1 / more than 1 employing test datasets are presented in Table 3.

Table 3. The evaluation criteria for assessing the efficacy of the machine learning models

Performance metrics value (%) (95% CI)	ML methods			
	AdaBoost	Bagging	CatBoost	XGBoost
ACC	82.3 (78-86.6)	**99 (97.9-100)**	79 (74.4-83.6)	93.3 (90.5-96.2)
b-ACC	79 (74.4-83.6)	**99 (97.8-100)**	76.1 (71.3-80.9)	90.2 (86.8-93.6)
SE	72.4 (60.9-82)	**98.8 (93.7-100)**	71.4 (57.8-82.7)	97.7 (94.6-99.2)
SP	85.7 (80.4-90)	**99.1 (96.7-99.9)**	80.7 (75.2-85.5)	82.8 (73.2-90)
ppv	63.2 (52.2-73.3)	**97.7 (91.9-99.7)**	46 (35.2-57)	93.3 (89.1-96.2)
npv	90.1 (85.3-93.8)	**99.5 (97.4-100)**	92.5 (88.1-95.6)	93.5 (85.5-97.9)
F1-score	67.5 (62.2-72.8)	**98.3 (96.8-99.7)**	55.9 (50.3-61.6)	95.4 (93-97.8)

The values highlighted in bold indicate the ML method that achieved the highest level of performance

ACC: accuracy, b-ACC: balanced accuracy, SP: specificity, SE: sensitivity, npv: negative predictive value, ppv: positive predictive value

Modelling Results

The findings of the modeling analysis are presented in this section.

The performance metrics of machine learning prediction models (AdaBoost, XGBoost, CatBoost and Bagging) utilized for predicting the LH: FSH ratio up to 1/more than 1 employing test dataset are presented in Table 3.

Table 4 presents the variable importance values derived from the Bagging method, which yielded the greatest accuracy outcome during the modeling process. The graph presented in Figure 1 illustrates the variable importance values obtained from the Bagging method.

Table 4. The variable importance values obtained from the Bagging method's findings

Variables	Importance value
TT ng/dL	100
BMI	74.354
AMH	58.367
Age	32.307
Family history/yes	10.011
Regulation of menstrual cycle/irregular	9.025

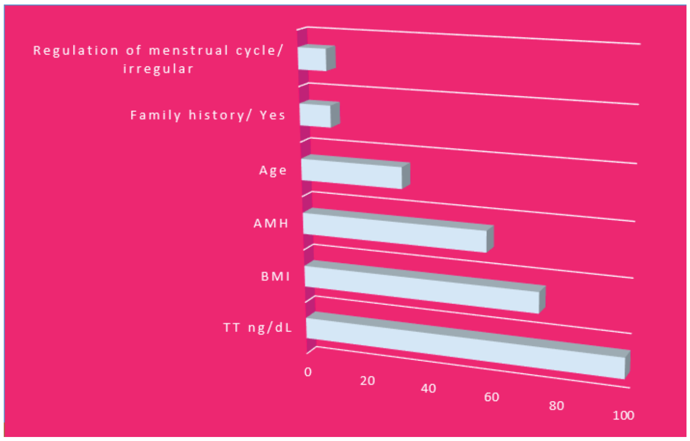

Figure 1. Variable Importance plot for Bagging method

SHAP Value Analysis (Feature Importance)

Bagging method emerged as one of the top-performing machine learning methods, achieving an accuracy of 99%. The utilization of SHAP values helped the explanation of the model predictions in question. Figure 2A and Figure 2B present the SHAP importance significance graph, which illustrates the impact of each variable on the target variable. Feature importance is a measure of the extent to which input features are useful in predicting a target variable. In order to achieve this objective, the methodology of feature importance assigns a numerical value to the input characteristics, hence offering valuable insights into the prediction model, particularly in the context of classification models. The prioritization of feature importance significantly enhances the efficacy and efficiency of the predictive modeling endeavor. The SHAP summary graphic was utilized in this investigation. The utilization of the SHAP summary graphic (Figure 2A and Figure 2B) offers dual benefits, namely feature ranking and the assessment of each feature's impact. The feature ranking in decreasing order is determined by the position in the y-axis In terms of significance, the following factors are ranked from highest to lowest relevance. The impact of each characteristic is assessed through the utilization of x-axis SHAP values. Positive SHAP values indicate a positive association with the target variable, while negative SHAP values suggest a negative association. Furthermore, the feature value shown by the color red signifies a greater value for the feature, while the color blue signifies a lower value for the feature. The presence of jagged overlapping spots provides a coherent representation of the distribution.

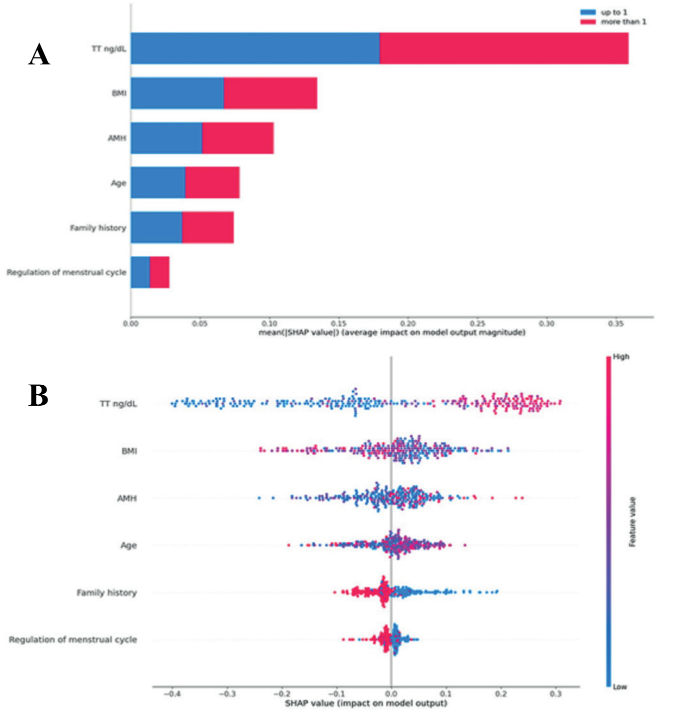

Figure 2. A. SHAP Feature Importance Plot, **B.** SHAP Feature Importance Plot

Discussion

PCOS is a widely recognized metabolic and reproductive endocrine disorder that impacts around 18% of women in their reproductive years. PCOS is characterized by the presence of irregular menstrual periods, which can lead to various complications. These complications include the formation of cysts (follicles) in one or both ovaries, difficulties in conceiving, disturbances in androgen hormone levels, elevated body mass index (BMI), and abnormalities in the levels of luteinizing hormone (LH), follicle-stimulating hormone (FSH), dehydroepiandrosterone sulfate (DHEAS), fasting insulin, and fasting blood sugar [34].

The ratio between LH and FSH is a hormonal marker that holds diagnostic and evaluative value in the context of PCOS. PCOS is a prevalent endocrine illness that mostly affects adults in their reproductive years. It is distinguished by a variety of symptoms, such as irregular menstruation periods, the presence of ovarian cysts, and heightened levels of androgens (male hormones) [35].

An imbalance in the ratio of LH to FSH is frequently observed in multiple cases of PCOS. Typically, in the context of the menstrual cycle, FSH levels exhibit a small elevation compared to LH levels, resulting in a ratio that approximates 1:1. Nevertheless, in the case of PCOS, it is common to observe an increased ratio, characterized by elevated levels of LH in comparison to FSH levels. The presence of a hormonal imbalance has the potential to play a role in the manifestation of specific symptoms associated with PCOS, including anovulation (the absence of regular ovulation) and heightened androgen production [36,37].

There has been a lack of consistency in the diagnostic procedures employed for women with PCOS. Indeed, there have been reports indicating that women expressed dissatisfaction, frustration, and confusion regarding their diagnostic journey, enduring an average waiting period of two years before obtaining an accurate diagnosis [38]. According to the findings of a research examining patient satisfaction with the diagnosis of PCOS, a mere 35.2% of participants reported being content with their overall experience[39]. This level of satisfaction was determined based on factors such as the duration of time it took to receive a diagnosis and the number of healthcare professionals consulted before obtaining a definitive diagnosis. Indeed, a considerable number of patients find themselves compelled to allocate a substantial amount of time to undertake several diagnostic tests and consult multiple professionals in order to receive proper attention and evaluation of their symptoms. Consequently, many individuals have resorted to self-diagnosis and self-management strategies for PCOS [39].

Thus, the goal of this study is to develop a high-performing diagnostic model utilizing ML in order to decrease and prevent human error in PCOS diagnosis. Therefore, the aim of the study is to present a model that can predict the LH:FSH ratio, which is one of the criteria used to determine the diagnosis of PCOS. The integration of ML into the diagnosis of PCOS has the potential to improve the quality of life for women affected by this condition. Timely diagnosis facilitated by ML can lead to early treatment of symptoms, address infertility concerns, and promote the adoption of a suitable lifestyle to minimize metabolic complications. In addition, SHAP, one of the explainable artificial intelligence methods, was used to explain and interpret the output and operation of the model. This will help the researcher to understand how the model works and make predictions, which will lead to more accurate results and help personalized treatment procedures.

The performance results obtained from ML models in this study are summarized in Table 3. Before applying the XAI method, the performances obtained from these ML models were analyzed to see which ML model gave the best performance. Accordingly, Bagging method gave the best performance among AdaBoost, Bagging, CatBoost, XGBoost models. As a result of the prediction model with Bagging method, the values of ACC, b-ACC, SE, SP, ppv, npv and F1-score performance metrics were 99%, 99%, 98.8%, 99.1%, 97.7%, 99.5%, and 98.3% respectively. Additionally, when the variable importance values produced by Bagging modeling were examined, it was determined that the most important features related to the LH:FSH ratio were TTng/dL, BMI, AMH, age, family history/yes, regulation of menstrual cycle/irregular, respectively.

The detailed outputs of the SHAP graphs acquired in the present investigation are provided below. According to the SHAP feature importance plot (Figure 2A/Figure 2B), the TTng/dL, BMI, AMH, age, family history, regulation of menstrual cycle variables are the most efficient in predicting LH:FSH (up to 1) / LH:FSH (more than 1) (target variable).

There are some limitations in this study. Using an open data set and collecting the data set from a specific region creates limitations in terms of the generalizability of the study.

When the results obtained from the study were examined, it was seen that the variable importance results obtained with the applied ML model Bagging method were similar to the results obtained with SHAP, although they were calculated with a model-independent method. The biggest limitation in the use of ML models is the black box feature, that is, the lack of understanding how it works and the low reliability of the model. In this study, to ensure model confidence, efficiency and effectiveness, the XAI method, SHAP, was also applied to the data and similar results were obtained. Integrating XAI with ML-based systems for diagnostic, prognostic and treatment purposes can help ensure accountability and increase confidence in medical decisions made by the system.

Conclusion

In conclusion, underscore the critical role of advanced ML techniques in enhancing the diagnostic accuracy of PCOS, paving the way for more personalized and effective treatments. Moreover, the integration of SHAP into the modeling process emphasizes the importance of model interpretability, which is crucial for clinical adoption and trust.

Conflict of Interests
The authors declare that there is no conflict of interest in the study.

Financial Disclosure
The authors declare that they have received no financial support for the study.

Ethical Approval
Ethics committee approval is not required in this study.

References

1. Escobar Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. Nat Rev Endocrinol. 2018;14:270-84.
2. Franks S. Polycystic ovary syndrome. NEJM. 1995;333:853-61.
3. El Hayek S, Bitar L, Hamdar LH, et al. Poly cystic ovarian syndrome: an updated overview. Front Physiol. 2016;7:124.

4. Goodarzi MO, Dumesic DA, Chazenbalk G, et al. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. Nat Rev Endocrinol. 2011;7:219-31.

5. Teede HJ, Misso ML, Costello MF, et al.; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Hum Reprod. 2018;33:1602-18.

6. Fauser B, Tarlatzis B, Chang J, et al. The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod. 2004;19:41-7.

7. Hsu MI, Liou TH, Liang SJ, et al. Inappropriate gonadotropin secretion in polycystic ovary syndrome. Fertil Steril. 2009;91:1168-74.

8. Cho LW, Jayagopal V, Kilpatrick ES, et al. The LH/FSH ratio has little use in diagnosing polycystic ovarian syndrome. Ann Clin Biochem. 2006;43:217-9.

9. Minakami H, Abe N, Oka N, et al. Prolactin release in polycystic ovarian syndrome. Endocrinol Jpn. 1988;35:303-10.

10. Papaleo E, Doldi N, De Santis L, et al. Cabergoline influences ovarian stimulation in hyperprolactinaemic patients with polycystic ovary syndrome. Hum Reprod. 2001;16:2263-6.

11. Ashraf S, Nabi M, Rashid F, et al. Hyperandrogenism in polycystic ovarian syndrome and role of CYP gene variants: a review. Egypt J Med Hum Genet. 2019;20:25.

12. Legro RS, Arslanian SA, Ehrmann DA, et al. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2013;98(12):4565-92.

13. Gibson Helm M, Teede H, Dunaif A, et al. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J Clin Endocrinol Metab. 2017;102:604-12.

14. Choi RY, Coyner AS, Kalpathy Cramer J, et al. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol. 2020;9:14.

15. Barrera FJ, Brown ED, Rojo A, et al. Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. Front Endocrinol. 2023;14:1106625.

16. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2:749-60.

17. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318:517-8.

18. Hagras H. Toward human-understandable, explainable AI. Computer. 2018;51:28-36.

19. Islam MR, Ahmed MU, Barua S, et al. A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Appl Sci. 2022;12:1353.

20. Zhang Y, Song K, Sun Y, et al. "Why should you trust my explanation?" Understanding uncertainty in LIME explanations. arXiv. 2019 Jun 4. doi: 10.48550/arXiv.1904.12991. [Epub ahead of print].

21. Teede H, Misso M, Costello M, et al. International evidence-based guideline for the assessment and management of polycystic ovary syndrome 2018. NHMRC. 2018: 1-198.

22. The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril. 2004;81:19-25.

23. Ying C, Qi-Guang M, Jia-Chen L, et al. Advance and prospects of AdaBoost algorithm. Acta Automatica Sinica. 2013;39:745-58.

24. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. JCSS. 1997;55:119-39.

25. Nalluri M, Pentela M, Eluri NR. A scalable tree boosting system: XG boost. Int J Res Stud Sci Eng Technol. 2020;7:36-51.

26. Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst. 2018;31:3342.

27. Dhananjay B, Sivaraman J. Analysis and classification of heart rate using CatBoost feature ranking model. Biomed Signal Process Control. 2021;68:102610.

28. Li S, Qin J, He M, et al. Fast evaluation of aircraft icing severity using machine learning based on XGBoost. Aerospace. 2020;7:36.

29. Breiman L. Bagging predictors. Machine Learn. 1996;24:123-40.

30. Ferreira AJ, Figueiredo MA. Boosting algorithms: a review of methods, theory, and applications. In: Zhang, C, Ma Y. eds. Ensemble machine learning. Springer, New York, NY.

31. Fatahi R, Nasiri H, Dadfar E, et al. Modeling of energy consumption factors for an industrial cement vertical roller mill by SHAP-XGBoost: a "conscious lab" approach. Sci Rep. 2022;12:7543.

32. Ekanayake I, Meddage D, Rathnayake U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). Case Stud Constr Mater. 2022;16:e01059.

33. Feng DC, Wang WJ, Mangalathu S, et al. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. J Struct Eng. 2021;147:04021173.

34. Dutta P, Paul S, Majumder M. An efficient SMOTE based machine learning classification for prediction & detection of PCOS. Published online November 8, 2021. doi: 10.21203/rs.3.rs-1043852/v1

35. Beydoun HA, Beydoun MA, Wiggins N, Stadtmauer L. Relationship of obesity-related disturbances with LH/FSH ratio among post-menopausal women in the United States. Maturitas. 2012;71:55-61.

36. Laven JS. Follicle stimulating hormone receptor (FSHR) polymorphisms and polycystic ovary syndrome (PCOS). Front Endocrinol. 2019;10:23.

37. Wu L, Zheng S, Zhao G, et al. Association of the LH/FSH ratio with electrocardiogram abnormalities in polycystic ovarian syndrome. Springer Nature. Published online October 11, 2022. doi: 10.21203/rs.3.rs-2141531/v1

38. Copp T, Cvejic E, McCaffery K, et al. Impact of a diagnosis of polycystic ovary syndrome on diet, physical activity and contraceptive use in young women: findings from the Australian Longitudinal Study of Women's Health. Hum Reprod. 2020;35:394-403.

39. Madhavan R. Prevalence of PCOS diagnoses among women with menstrual irregularity in a diverse, multiethnic cohort. MS thesis. Boston University, Boston, 2018.