IJASST
International Journal of
Applied Sciences and Smart Technologies
p-ISSN 2655-8564, e-ISSN 2685-943, pp. 261-278, 2026

Volume 08
Issue 01

Analysis of Key Features in PCOS Diagnosis Using Random Forest and XGBoost with SMOTE and SHAP

Aulia Firdatunnisa[1], Eka Wahyu Hidayat[1*], Siti Yuliyanti[1]

[1]*Department of Informatics, Faculty of Engineering,*
Siliwangi University, Indonesia
[*]*Corresponding Author: ekawahyu@unsil.ac.id*

(Received 18-07-2025; Revised 02-09-2025; Accepted 9-04-2026)

Abstract
Polycystic Ovary Syndrome (PCOS) is a hormonal disorder in women of reproductive age characterized by irregular cycles, hyperandrogenism, and polycystic ovarian morphology. Diagnosis is challenging because symptoms overlap with other endocrine disorders. This study proposes an interpretable machine learning approach for PCOS diagnosis using Random Forest and XGBoost. The Synthetic Minority Oversampling Technique (SMOTE) was applied to handle class imbalance, while Shapley Additive Explanations (SHAP) enhanced model interpretability. The dataset included 541 samples with 45 clinical and hormonal features, processed through preprocessing and hyperparameter tuning with GridSearchCV. XGBoost with SMOTE and GridSearchCV achieved the best performance, with 93% accuracy, 92% precision, 89% recall, and 90% F1-score. Random Forest obtained comparable results with 93% accuracy, 94% precision, 87% recall, and 90% F1-score. SHAP analysis highlighted key features such as follicle count, Anti-Müllerian Hormone (AMH), skin darkening, weight gain, and irregular cycles. Global SHAP interpretation identified the most influential predictors, while local SHAP provided patient-specific explanations that improved transparency. The consistency of SHAP results with the Rotterdam criteria supports the model's clinical validity and strengthens trust in AI-assisted tools. Overall, combining SMOTE, GridSearchCV, and SHAP not only improved predictive performance but also ensured transparent outcomes, indicating potential use for early PCOS screening.

Keywords: Machine Learning, PCOS, SHAP, SMOTE, XGBoost

1 Introduction

Polycystic Ovary Syndrome is a hormonal disorder that affects women childbearing age [1]. This hormonal disorder is characterized by symptoms of irregular menstrual cycles, abnormal hair growth, excess androgen hormones and ovarian cysts [2]. PCOS symptoms like irregular cycles, acne, and hirsutism often overlap with other hormonal

disorders such as hypothyroidism, hyperprolactinemia, and adrenal disorders, making diagnosis challenging. For instance, hypothyroidism causes irregular periods, weight gain, and fatigue, while congenital adrenal hyperplasia leads to hirsutism and elevated androgens [3]. Hyperprolactinemia, characterized by high prolactin levels, can also trigger menstrual disorders and infertility [4]. The similarity of these symptoms often leads to misdiagnosis if not accompanied by hormonal and ultrasound examinations.

Based on the Rotterdam Criteria (2003), the diagnosis of PCOS is established if the patient meets at least two of the three criteria: oligo/anovulation, hyperandrogenism, and polycystic ovarian morphology (PCOM) [5]. The prevalence of PCOS in Indonesia is increasing significantly, from 4-6% (1990) to 20% by 2023. If left untreated, PCOS can lead to insulin resistance, type 2 diabetes, and cardiovascular disease [6], [7]. Based on this, an accurate and interpretative diagnosis system is required.

Random Forest offers strong interpretability and stability with limited features, while XGBoost provides high accuracy and efficiency on large, complex data, effectively handling overfitting and high dimensionality [8] [9]. Both models have shown excellent performance in diagnosing diseases, including PCOS. However, class imbalance remains a key challenge in clinical datasets. SMOTE helps address this and has been shown to improve accuracy Random Forest achieved 93.26% [10] and XGBoost 96%[11]. For diagnostic clarity, Explainable AI methods like SHAP help reveal the most influential features, such as insulin, BMI, and LH/FSH ratio [12], [13], [14] .

There are not many studies that combine SMOTE and SHAP integrated in Random Forest and XGBoost respectively. This study aims to develop and evaluate the PCOS model. So that the model not only achieves high accuracy, but the model is also able to explain the diagnosis results.

2 Material and Methods

Research Data

This study uses the PCOS dataset from Kaggle uploaded by Prasoon Kottarathil, containing 541 female patients from a hospital in Kerala, India [15]. The patients' age range is 20–45 years old, with the majority in the 25–35 age group. Women of South

Asian ethnicity, including Indian, have a higher risk of PCOS (3.7–22.5%) [16]. This disorder can also be experienced by other ethnicities, especially those with unhealthy lifestyles. This demographic information is added so that readers can assess the extent to which the model can be generalized.

Data Preprocessing

Data preprocessing was performed by removing irrelevant columns and rows with missing values. During the preprocessing stage, irrelevant columns are removed because they do not contribute to the prediction and have the potential to cause noise, while missing values are removed to maintain input consistency [17]. This step is important because even though Random Forest is quite resistant to noise, irrelevant data still reduces stability, while XGBoost, which is more sensitive to missing values, actually improves performance when such data is removed [18], [19]. Missing values accounted for only about 0.18%, which did not have a significant effect and were removed to maintain data integrity [20]. In addition, imputation risks introducing bias in sensitive medical data. For future research, the use of more sophisticated imputation techniques such as model-based or regression-based imputation is recommended.

Exploratory Data Analysis

The Exploratory Data Analysis stage in this research is carried out to find out the relationship patterns between data. In addition, to find out the initial insight of the data before interpretation is carried out.

Model Implementation

This study uses Random Forest and XGBoost models, with SMOTE applied to prevent data leakage and GridSearchCV for optimal parameter tuning. Three schemes are tested: (1) baseline, (2) with SMOTE, and (3) with SMOTE + GridSearch. The complete methodology is data collection, preprocessing, modeling, evaluation, and SHAP-based interpretation shown in Fig. 1.

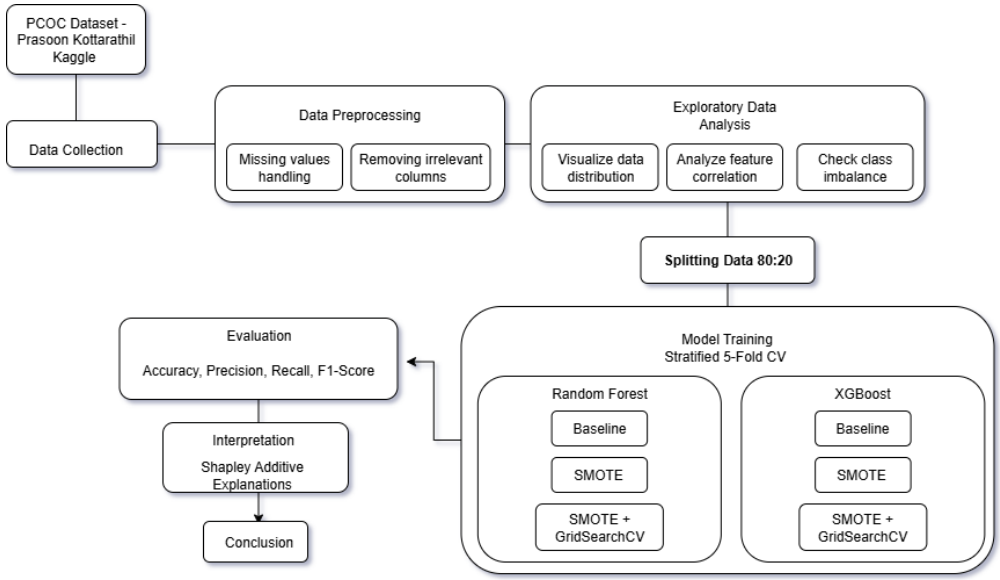

Figure 1. Research flow

Interpretability with SHAP

The SHAP approach was used to interpret the trained model, identifying key features contributing to the diagnosis. Interpretation was conducted at both global (entire dataset) and local (individual patient) levels.

3 Results and Discussions

Exploratory Data Analysis

The Exploratory Data Analysis stage was conducted to understand the distribution of data, the relationship between features, and see the differences in characteristics between PCOS and non-PCOS patients. The dataset used in this study consists of 541 patient data samples with 45 features containing physical, hormonal indicators, and clinical symptoms.

Based on EDA, it is found that the class distribution on the target label shows data imbalance. There are more non-PCOS patients than PCOS patients. The unbalanced class distribution is shown in Fig. 2.

Based on descriptive analysis, it was found that PCOS patients were generally between 25-30 years old and had higher BMI values compared to non-PCOS patients. The correlation heatmap illustrates the relationship between features and the PCOS label.

Features like follicle count (right and left), skin darkening, hair growth, and weight gain show strong correlations. Meanwhile, hormonal features such as AMH and LH/FSH ratio have lower linear correlations but may influence outcomes through non-linear relationships not captured by Pearson correlation. A visualization of the correlation heatmap is shown in Fig. 3.

Model Evaluation

Model evaluation was conducted with 3 schemes, including: (1) baseline, (2) model with SMOTE, (3) model with SMOTE and GridSearchCV as tuning parameter. The algorithms used are Random Forest and XGBoost. Table 1 presents the model evaluation results.

Based on the evaluation results in Table 1, applying SMOTE significantly improved model accuracy. GridSeacrhCV on XGBoost provided better metric balance for minority classes compared to Random Forest. In disease diagnosis like PCOS, recall is crucial as it reflects the model's ability to identify actual patients and reduce false negatives. SMOTE increased recall in both models, but XGBoost with SMOTE and GridSearchCV achieve stable recall (0.80), high precision (0.91), and F1-Score of 0.85 indicated a balanced and accurate detection of positive cases.

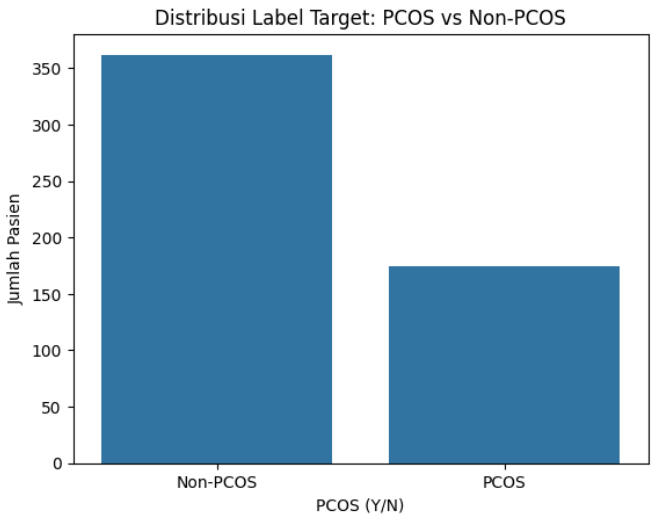

Figure 2. Label distribution

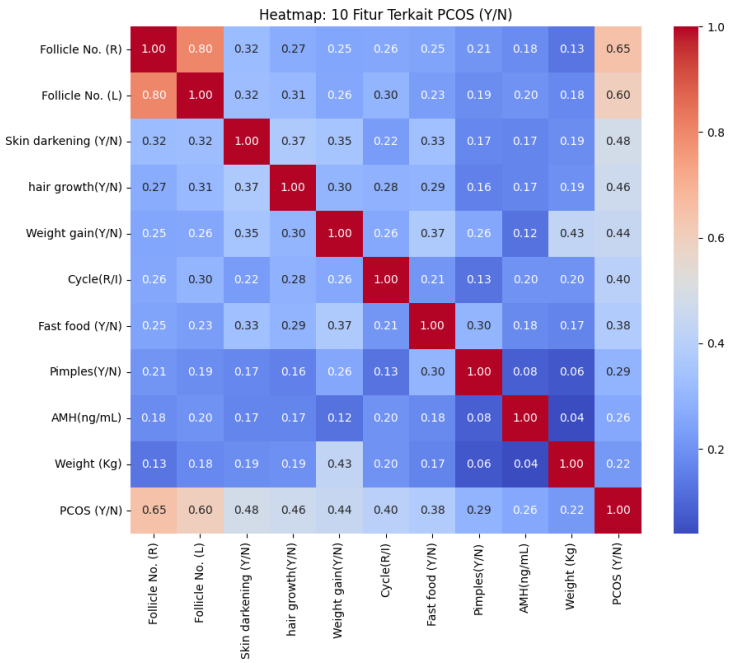

Figure 3. Heatmap correlation

Table 1. Evaluation results of Random Forest and XGBoost

Model	Schema	Accuracy
Random Forest	Baseline	0.89
Random Forest	SMOTE	0.95
Random Forest	SMOTE + GridSeacrhCV	0.93
XGBoost	Baseline	0.89
XGBoost	SMOTE	0.91
XGBoost	SMOTE + GridSearchCV	0.93

Model Interpretation with SHAP

The interpretation stage is carried out on both models with the aim of finding out which features are most influential on the model diagnosis results. Interpretation is done in two stages between global interpretation and local interpretation.

Random Forest Model

Global interpretation is done to determine the most influential features in general or overall, in the model. Follicle No. (R) and Follicle No. (L) are two main features in the

diagnosis, in line with the Rotterdam Criteria which mention PCOM as the main indicator of PCOS [5]. The third feature is the skin darkening, a common symptom in PCOS patients with insulin resistance [21] Fig. 4 shows the 10 most contributing features in the Random Forest model.

Based on the interpretation results presented in Fig. 4. Hirsutism or excessive hair growth occupies the fourth position. In a clinical context, this is a clinical manifestation of hyperandrogenism which is included in one of the Rotterdam Criteria [22]. Weight gain and Fast food (Y/N) are interrelated features that contribute significantly to the diagnosis. Both reflect an unhealthy lifestyle that triggers obesity and metabolic disorders, including insulin resistance [23][24]. AMH (ng/mL) also supports the diagnosis ah high AMH levels indicated an increased number of follicles, in accordance with the PCOM criteria in Rotterdam Criteria [25].Irregular menstrual cycle and cycle length features showed significant influence on diagnosis, in line with oligo/anovulation in Rotterdam Criteria [26]. Pimples (Y/N) reflecst acne symptons due to androgen excess, also part of the criteria [27].

Local interpretation is also performed on the Random Forest model, similar to the global interpretation, this interpretation is done to find out which features are most influential on the diagnosis results, but more specifically for one patient diagnosed with PCOS. Local interpretation on PCOS patients is presented in Fig. 5.

Based on Fig. 5, the patient was predicted to have PCOS with a probability of 92.3% (f(x) = 0.923), an increase from the model average prediction of 35.8% (F[f(x)] = 0.358).

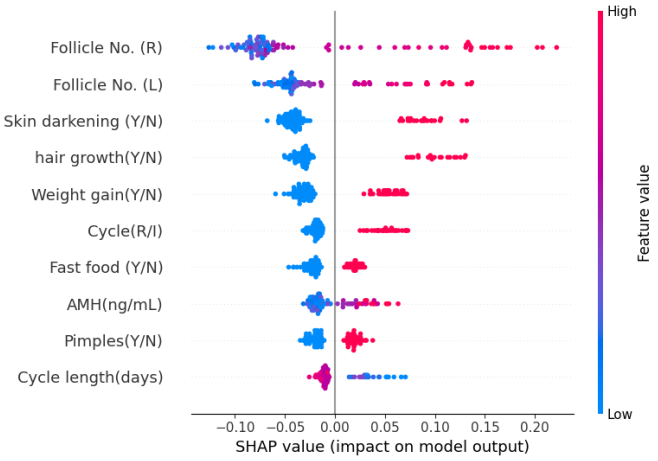

Figure 4. Global Random Forest interpretation results with SHAP

The Follicle No. (R) feature contributed the most (+0.15), indicating a high number of follicles. Other features that contributed positively included Hair Growth (+0.08), Skin Darkening (+0.08), Weight Gain (+0.05), and Follicle No. (L) (+0.05), all supporting the diagnosis of PCOS according to Rotterdam criteria [22][28].

Irregular menstrual cycle (Cycle (R/I) = 4, +0.05) and cycle duration (+0.03) reflected oligo/anovulation. The features Pimples (+0.04) and Fast-Food consumption (+small value) also strengthened the prediction of PCOS although their contribution was lower. These features cumulatively increased the predictive value towards a positive diagnosis.

XGBoost Model

The model not only evaluated in terms of performance, this study emphasized the importance of model interpretability in medical diagnosis, given that PCOS symptoms are similar to other hormonal diseases. Interpretation is necessary to reduce the risk of misdiagnosis and ensure the prediction results can be understood logically an clinically. The results of SHAP interpretation on the XGBoost model are presented in Fig. 6.

Based on Fig. 6, Follicle No. (R) is the most influential feature in the interpretation of the XGBoost model, with consistently high SHAP values increasing the prediction of PCOS, in accordance with the PCOM criteria in the Rotterdam Criteria [28].

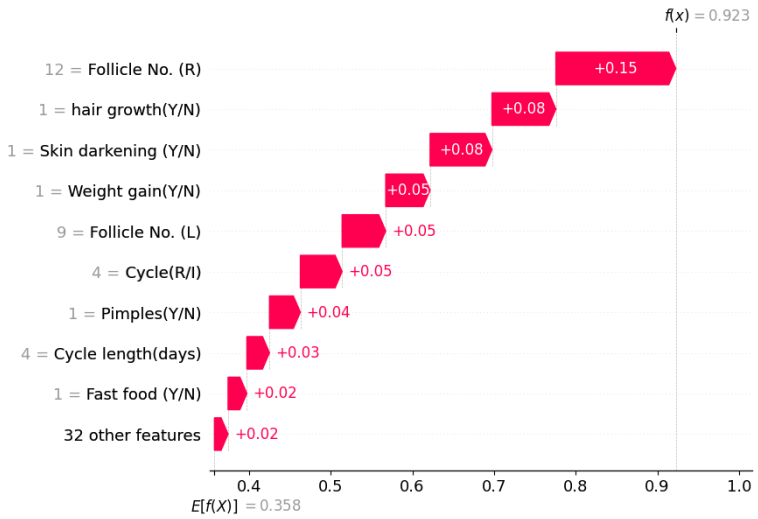

Figure 5. Local interpretation results of Random Forest model with SHAP

Hair growth (Y/N) feature followed as a major contributor, reflecting symptoms of hyperandrogenism [22].

Skin Darkening, Weight Gain, and AMH features also made significant contributions. Skin Darkening is related to insulin resistance and obesity [21]. Weight gain (Y/N) reflects obesity which exacerbates metabolic disorders [6], and AMH indicates a high follicle count [29], [30]. Other features such as Pimples (acne), Cycle (R/I) (menstrual cycle disorder), and Waist:Hip Ratio (central obesity) also contribute, all associated with PCOS symptoms [6][31]. PRL (ng/mL) is also considered although not the main criterion, as abnormal prolactin levels can cause menstrual cycle disorders and need to be differentiated from other hormonal disorders [32]. Follicle No. (L) also contributes, although smaller than the right side, it still supports the identification of PCOM [5].

Local interpretation of the XGBoost model using SHAP was performed to see the effect of each feature on the diagnosis result for one patient. The result shown in Fig. 7. Based on the local interpretation results shown in Fig. 7, the final prediction value of the model reaches ($f(x) = 8.63$), well above the model average ($E[f(x)] = -0.641$). Follicle No. (R) was the main contributor (+3.07), reflecting the number of excess follicles that fit the PCOM criteria [28]. The Skin Darkening (+0.98), Weight Gain (+0.85), Cycle (R/I) (+0.70), and AMH (+0.65) features also contributed greatly to the positive classification, as they are associated with insulin resistance, obesity, menstrual cycle disorders, and high follicle count [21].

The Pimples feature (+0.65) reflected symptoms of hyperandrogenism, while Hair Growth (-0.51) did not contribute significantly as it was 0. High Waist:Hip Ratio (+0.51) indicating central obesity, and PRL (+0.39), although within normal limits, was still considered relevant as it could affect the menstrual cycle. Overall, the model judged the features to be clinically relevant and strengthened the positive prediction of PCOS.

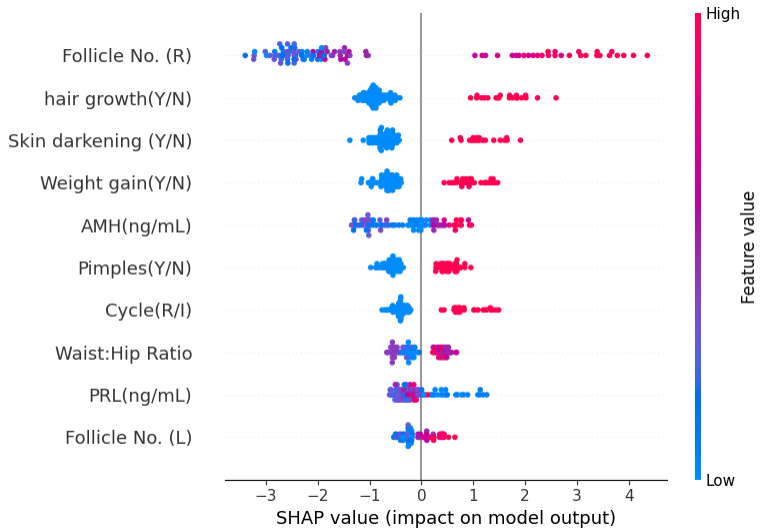

Figure 6. Global interpretation with SHAP on the XGBoost model

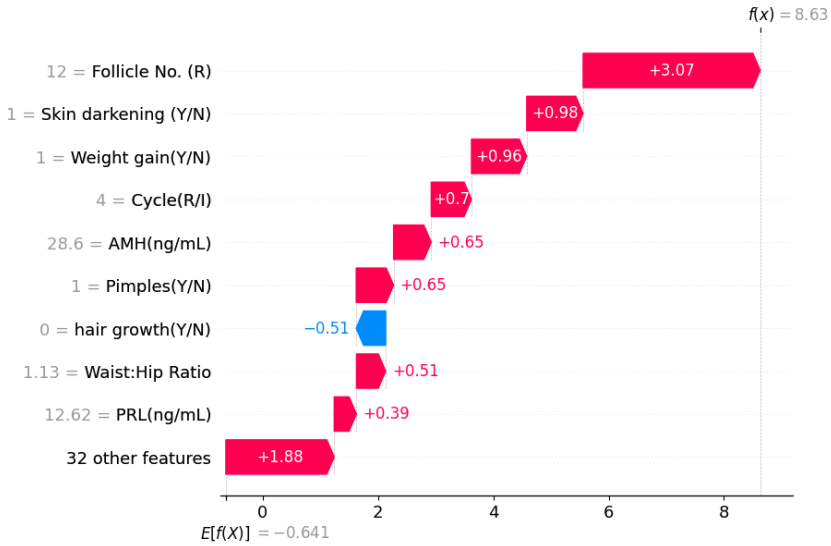

Figure 7. Local interpretation on XGBoost model with SHAP

Comparison of SHAP and Feature Importance

Feature importance and SHAP are model interpretation methods that differ in approach. Feature importance calculates feature contributions globally based on impurity or gain, without considering interactions or direction of influence. In contrast, SHAP uses game theory to calculate the contribution of individual features, both globally and locally,

and shows the direction and magnitude of influence, making it more transparent and in line with the principles of XAI. Fig. 8 shows a Venn diagram comparing features identified by SHAP and feature importance in both models.

Based on Fig. 8. Follicle No. (R) consistently appears as the most influential feature in both SHAP and feature importance, alongside clinical features like Hair Growth, Skin Darkening, Weight Gain, and Cycle (R/I). However, SHAP uniquely identifies features such as Pimples and Cycle Length, while feature importance highlights Blood Group and BP Systolic. This difference reflects SHAP's advantage in capturing local and non-linear interactions, offering more comprehensive and clinically relevant interpretations, especially in PCOS diagnosis.

Comparison of Models with Other Studies

The results of this study were compared with several previous studies that also examined PCOS diagnosis using machine learning. This was done to provide a more comprehensive context. Table 2 shows the results of this study and comparisons with other studies.

Table 2 shows that previous studies produced higher accuracy. However, most of them only emphasized predictive performance without considering clinical interpretability. Meanwhile, this study focuses on the balance between model performance and the transparency of model prediction results with the application of SHAP, so that it can provide added value in the context of medical decision systems.

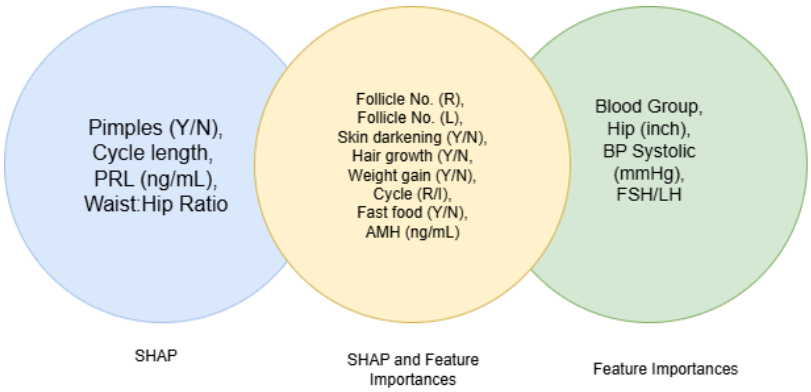

Figure 8. Comparison of SHAP and feature importance

Table 2. Evaluation results of Random Forest and XGBoost

Study	Method	Accuracy
[33]	Random Forest	0.91
[12]	XGBoost	0.90
Our Research	Random Forest + SMOTE + GridSeacrhCV	0.93
Our Research	XGBoost+ SMOTE + GridSeacrhCV	0.93

4 Conclusions

This study shows that the integration of Random Forest and XGBoost with SMOTE and SHAP interpretation produces an accurate and explainable PCOS diagnosis model. The best model was obtained from XGBoost with SMOTE and GridSearchCV, which provided balanced accuracy and recall. SHAP analysis highlighted important features such as follicle count, AMH, menstrual cycle, skin darkening, and weight gain. The interpretation is aligned with the Rotterdam criteria, which are widely recognized by the medical community, thus providing a strong basis for interpretation. However, further validation by medical experts is still needed to improve interpretability in clinical practice

This study is based on a collection of secondary clinical data obtained from hospitals, not from direct clinical examinations in laboratories or medical imaging. Therefore, the validity of the results is still limited to the available data patterns. Further research is recommended using real clinical data and integration with medical examinations such as ultrasound or laboratory tests. It should be emphasized that the main purpose of this study is as an initial screening tool or decision support system, not a substitute for medical diagnosis, so that it can help health workers identify cases that require further examination.

The limitation of this study lies in its predictive, rather than causal, model can only predict without definitively explaining the cause-and-effect relationship. If future studies use large datasets with high missing values, the application of modern imputation techniques such as regression-based imputation is recommended.

Acknowledgements

The author would like to thank the Department of Informatics, Faculty of Engineering, Universitas Siliwangi, for the support provided in completing this research.

References

[1] I. A. Begum, A. S. M. S. Hosen, D. Ghimire, and M. J. Park, "Association of Polycystic Ovary Syndrome with Clinical, Physical, and Reproductive Factors: A Data-Driven Analysis," *Diagnostics*, vol. 15, no. 6, Mar. 2025, doi: 10.3390/diagnostics15060711.

[2] C. Johnson, G. Garipoğlu, Y. Jeanes, G. Frontino, and A. Costabile, "The Role of Diet, Glycaemic Index and Glucose Control in Polycystic Ovary Syndrome (PCOS) Management and Mechanisms of Progression," Dec. 01, 2025, *Springer*. doi: 10.1007/s13668-024-00601-4.

[3] M. Yesiladali, M. G. K. Yazici, E. Attar, and F. Kelestimur, "Differentiating Polycystic Ovary Syndrome from Adrenal Disorders," Sep. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics12092045.

[4] K. van der Ham *et al.*, "The prevalence of thyroid dysfunction and hyperprolactinemia in women with PCOS," *Front Endocrinol (Lausanne)*, vol. 14, 2023, doi: 10.3389/fendo.2023.1245106.

[5] J. J. Kim, K. R. Hwang, D. Lee, S. Kim, and Y. M. Choi, "Adolescents diagnosed with polycystic ovary syndrome under the Rotterdam criteria but not meeting the diagnosis under the updated guideline," *Human Reproduction*, vol. 39, no. 5, pp. 1072–1077, May 2024, doi: 10.1093/humrep/deae042.

[6] X. Zeng, Y. jie Xie, Y. ting Liu, S. lian Long, and Z. cheng Mo, "Polycystic ovarian syndrome: Correlation between hyperandrogenism, insulin resistance and obesity," Mar. 01, 2020, *Elsevier B.V.* doi: 10.1016/j.cca.2019.11.003.

[7] Muna Dinur Rafifah, RB. Soeherman Herdiningrat, and Rizki Perdana, "Hubungan Obesitas dengan Polycystic Ovarian Syndrome (PCOS) di RSUD Al-Ihsan Provinsi Jawa Barat," *Bandung Conference Series: Medical Science*, vol. 5, no. 1, pp. 135–140, Jan. 2025, doi: 10.29313/bcsms.v5i1.16251.

[8] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic classification and biomarker identification of alzheimer's disease with random forest algorithm," *Brain Sci*, vol. 11, no. 4, Apr. 2021, doi: 10.3390/brainsci11040453.

[9] R. Chen *et al.*, "A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm," *BMC Med Inform Decis Mak*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12911-023-02140-4.

[10] P. Chauhan, P. Patil, N. Rane, P. Raundale, and H. Kanakia, "Comparative Analysis of Machine Learning Algorithms for Prediction of PCOS," Institute of Electrical and Electronics Engineers (IEEE), Aug. 2021, pp. 1–7. doi: 10.1109/iccict50803.2021.9510128.

[11] M. S. Khan Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee, and R. Hasan, "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 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 1046–1050. doi: 10.1109/CCWC51732.2021.9375994.

[12] B. J. Chelliah, S. K. Gahra, and A. Senthilselvi, "Enhancing PCOS Prediction Using Machine Learning and Explainable AI," in *2024 International Conference*

on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), IEEE, Nov. 2024, pp. 1–5. doi: 10.1109/IC-SIT63503.2024.10862800.

[13] B. V. Madhavi, A. Firdouse, and A. Kodipalli, "Detection of PCOS Leveraging Machine Learning and Interpretation Using Explainable AI," in *2024 1st International Conference for Women in Computing, InCoWoCo 2024 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/InCoWoCo64194.2024.10863596.

[14] P. Gethsia, J. Anitha, and S. Juliet, "Interpreting PCOS Risk Factors Using Explainable AI and Ensemble Modeling Techniques," in *Proceedings of the 4th International Conference on Ubiquitous Computing and Intelligent Information Systems, ICUIS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 33–38. doi: 10.1109/ICUIS64676.2024.10867148.

[15] P. Kottarathil, "Polycystic Ovary Syndrome (PCOS) Dataset," Kaggle Dataset.

[16] M. A. Ganie *et al.*, "Prevalence, Phenotypes, and Comorbidities of Polycystic Ovary Syndrome Among Indian Women," *JAMA Netw Open*, vol. 7, no. 10, p. e2440583, Oct. 2024, doi: 10.1001/jamanetworkopen.2024.40583.

[17] T. Shadbahr *et al.*, "The impact of imputation quality on machine learning classifiers for datasets with missing values," *Communications Medicine*, vol. 3, no. 1, Dec. 2023, doi: 10.1038/s43856-023-00356-z.

[18] Y. Akhiat, Y. Manzali, M. Chahhou, and A. Zinedine, "A New Noisy Random Forest Based Method for Feature Selection," *Cybernetics and Information Technologies*, vol. 21, no. 2, pp. 10–28, Jun. 2021, doi: 10.2478/cait-2021-0016.

[19] N. Shi, Y. Li, L. Wen, and Y. Zhang, "Rapid prediction of landslide dam stability considering the missing data using XGBoost algorithm," *Landslides*, vol. 19, no. 12, pp. 2951–2963, Dec. 2022, doi: 10.1007/s10346-022-01947-y.

[20] A. Arora *et al.*, "The value of standards for health datasets in artificial intelligence-based applications," *Nat Med*, vol. 29, no. 11, pp. 2929–2938, Nov. 2023, doi: 10.1038/s41591-023-02608-w.

[21] B. Khan and R. Basu, "Acanthosis nigricans in adolescents with polycystic ovary syndrome," *Int J Reprod Contracept Obstet Gynecol*, vol. 11, no. 3, p. 765, Feb. 2022, doi: 10.18203/2320-1770.ijrcog20220553.

[22] P. M. Spritzer, L. B. Marchesan, B. R. Santos, and T. M. Fighera, "Hirsutism, Normal Androgens and Diagnosis of PCOS," Aug. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics12081922.

[23] R. Heidari, F. Eftekhari, and M. Koushkie Jahromi, "Association of physical activity and nutrition-related knowledge, attitudes and practices with obesity indices in college-aged females with polycystic ovary syndrome in Shiraz, Iran: A cross-sectional study," *BMJ Open* , vol. 15, no. 5, May 2025, doi: 10.1136/bmjopen-2025-099260.

[24] E. J. Houston and N. M. Templeman, "Reappraising the relationship between hyperinsulinemia and insulin resistance in PCOS," May 01, 2025, *BioScientifica Ltd.* doi: 10.1530/JOE-24-0269.

[25] K. van der Ham, J. S. E. Laven, C. T. Tay, A. Mousa, H. Teede, and Y. V. Louwers, "Anti-müllerian hormone as a diagnostic biomarker for polycystic ovary syndrome and polycystic ovarian morphology: a systematic review and meta-analysis," Oct. 01, 2024, *Elsevier Inc.* doi: 10.1016/j.fertnstert.2024.05.163.

[26] S. Syamsiah, C. Adelin, and R. Kundaryanti, "THE RELATIONSHIP BETWEEN OBESITY, MENSTRUAL CYCLE, AND POLYCYSTIC OVARY SYNDROME CASES," *Jurnal Kebidanan Malahayati)*, vol. 10, no. 9, pp. 929–934, 2024, doi: 10.33024.

[27] S. A. Kanbour and A. S. Dobs, "Hyperandrogenism in Women with Polycystic Ovarian Syndrome: Pathophysiology and Controversies," *Androgens*, vol. 3, no. 1, pp. 22–30, Mar. 2022, doi: 10.1089/andro.2021.0020.

[28] R. Yang *et al.*, "Enhancing repeatability of follicle counting with deep learning reconstruction high-resolution MRI in PCOS patients," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-024-84812-3.

[29] M. I. Cedars, "Evaluation of Female Fertility - AMH and Ovarian Reserve Testing," Jun. 01, 2022, *Endocrine Society*. doi: 10.1210/clinem/dgac039.

[30] L. Melanie and E. Moolhuijsen, *Polycystic Ovary Syndrome and anti-Müllerian Hormone Integrating Genetics in Etiology and Diagnosis*. 2025. [Online]. Available: www.ridderprint.nl

[31] B. Keyif and A. Yavuzcan, "Visceral and Dysfunctional Adiposity Indices as Predictors of Insulin Resistance and Metabolic Syndrome in Women with Polycystic Ovary Syndrome: A Cross-Sectional Study," *Medicina (Lithuania)*, vol. 61, no. 3, Mar. 2025, doi: 10.3390/medicina61030424.

[32] Z. Davoudi, F. Araghi, M. Vahedi, N. Mokhtari, and M. Gheisari, "Prolactin level in polycystic ovary syndrome (Pcos): An approach to the diagnosis and management," *Acta Biomedica*, vol. 92, no. 5, Nov. 2021, doi: 10.23750/abm.v92i5.9866.

[33] S. Majumdar, S. K. Biswas, and B. Purkayastha, "Effectiveness of Class Imbalance Treatment for Diagnosis of PCOS Using Machine Learning Models: A

Comparative Study," in *10th International Conference on Advanced Computing and Communication Systems, ICACCS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1346–1351. doi: 10.1109/ICACCS60874.2024.10717284.