International Journal of Medical Toxicology & Forensic Medicine
2026, Volume 16, E53481

Research Paper

Clinical Feature-Based Polycystic Ovary Syndrome Detection Using an Interpretable Kolmogorov-Arnold Network and Hybrid Feature Selection

Veena Rani*, Chandrasekar Venkatachalam, Karthick Raghunath Kuppusamy Murugesan

Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed - to - be University), Kanakapura Campus, Bangalore, Karnataka, India.

Citation Rani V, Venkatachalam C, Raghunath KKM. Clinical Feature-Based Polycystic Ovary Syndrome Detection Using an Interpretable Kolmogorov-Arnold Network and Hybrid Feature Selection. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E53481.

https://doi.org/10.22037/ijmtfm.v16.53481

Article info:

Received: 16 September, 2026

Accepted: 25 September, 2026

Published: 11 November, 2026

Keywords:

Polycystic Ovary Syndrome, Clinical Data, Kolmogorov-Arnold Network, mRMR-Boruta, AdaPlus Optimizer, Explainable Artificial Intelligence

ABSTRACT

Background: Polycystic Ovary Syndrome (PCOS) is a prevalent hormonal disorder affecting women and can lead to menstrual irregularities, hormonal imbalance, infertility, and other health complications. The availability of clinical and physical parameters has encouraged the application of Artificial Intelligence (AI) and deep learning (DL) techniques for automated PCOS detection. However, existing approaches still face challenges in selecting the most relevant clinical features and learning complex nonlinear relationships between patient variables while maintaining the interpretability of predictions. To address these issues, this paper introduces an Interpretable Kolmogorov-Arnold Network-based framework for Polycystic Ovary Syndrome detection (IKAN-PCOS).

Methods: The proposed framework applies a hybrid mRMR-Boruta feature selection approach to identify relevant clinical attributes while reducing redundant information. A Tabular Kolmogorov-Arnold Network is then employed to learn nonlinear relationships among the selected clinical features and classify PCOS and non-PCOS cases. During model training, the AdaPlus optimizer is used to improve convergence and learning performance. Furthermore, SHAP and LIME-based explainable artificial intelligence techniques are incorporated to provide global and local explanations of the model predictions by identifying the contribution of individual clinical features. The explanations help identify the clinical attributes that most influence PCOS and No PCOS classification and provide a clearer understanding of the model's decision process.

Results: The proposed IKAN-PCOS framework achieved 96.93% accuracy in PCOS and No PCOS classification, demonstrating its effectiveness in accurately detecting PCOS from clinical features while providing interpretable insights into classification decisions.

Conclusion: The proposed IKAN-PCOS model supports accurate and interpretable PCOS classification by integrating effective feature selection, nonlinear learning, and XAI techniques.

*** Corresponding Author:**
Veena Rani, PhD
Department of Computer Science and Engineering, JAIN (Deemed to be University), Kanakapura Campus, Bangalore, Karnataka, India.
E-mail: *v4veena.rani@gmail.com*

Copyright © 2026 The Author(s);
This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0: https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Polycystic Ovary Syndrome (PCOS) is a common disorder characterized by excess androgen, polycystic ovaries, and ovulatory dysfunction. It is considered a systemic metabolic condition rather than simply an unhealthy or diseased condition [1]. The core characteristics of PCOS include hormonal imbalances, menstrual changes, metabolic effects, skin and hair symptoms, excess androgens, and polycystic ovarian morphology [2]. Early identification and diagnosis of PCOS in clinical settings can enable efficient planning and treatment. Recently, artificial intelligence (AI) and machine learning (ML) have advanced significantly in medicine by supporting diagnosis through efficient decision-making systems [3]. ML-based methods are well-suited to managing vast quantities of nonlinear information and integrating PCOS-related data from different modalities; various deep learning (DL)-based techniques have recently been used to obtain high accuracy in the diagnosis of PCOS in clinical settings [4]. To address these limitations, Explainable AI (XAI)-based models are designed to improve interpretability and provide meaningful insights into the decision-making process [5].

Recent studies have explored AI, ML, DL, and XAI-based approaches for accurate and interpretable PCOS detection. Athapaththu and Malkanthi [6] developed an innovative framework using XAI-based approaches for transparent, explainable, and timely PCOS diagnosis. Vijay et al. [7] developed an AI-based framework for early detection of PCOS using clinical records. Babu and Laxmanan [8] introduced an innovative approach by employing XAI-based models for early detection and diagnosis of PCOS. Malik et al. [9] developed an innovative XAI-based framework, which further integrates ML-based models for early PCOS detection. Baalebail et al. [10] designed a novel framework using DL-based models for precise PCOS detection and diagnosis. Jaganathan and Natesan [11] introduced an innovative approach for PCOS prediction while securely maintaining patients' clinical records and data. Patel et al. [12] developed a novel automated mechanism using medical records and ultrasound to detect and classify PCOS. Shah et al. [13] designed an innovative AI-based approach using features for early detection and diagnosis of PCOS.

Motivated by the shortcomings discussed above, an automated framework is designed to efficiently select appropriate clinical features, learn complex nonlinear relationships among patient features, and provide intelligible explanations for the generated predictions. Moreover, because of the complex associations among clinical attributes, a learning model is needed to capture nonlinear relationships. An appropriate explainability mechanism is needed to reveal the impact of individual clinical features. Hence, we present an Interpretable Kolmogorov-Arnold Network-based framework for Polycystic Ovary Syndrome detection (IKAN-PCOS). The significant contributions of this work are listed below:

- A hybrid minimum Redundancy Maximum Relevance (mRMR)-Boruta feature selection approach is developed to identify relevant clinical attributes while reducing redundant information in the PCOS dataset.

- A Tabular Kolmogorov-Arnold Network (TabKAN) is employed to model complex nonlinear relationships among the selected clinical features and perform PCOS and No PCOS classification.

- The AdaPlus optimizer is integrated with TabKAN to provide an adaptive parameter optimization mechanism during model training and improve the learning process.

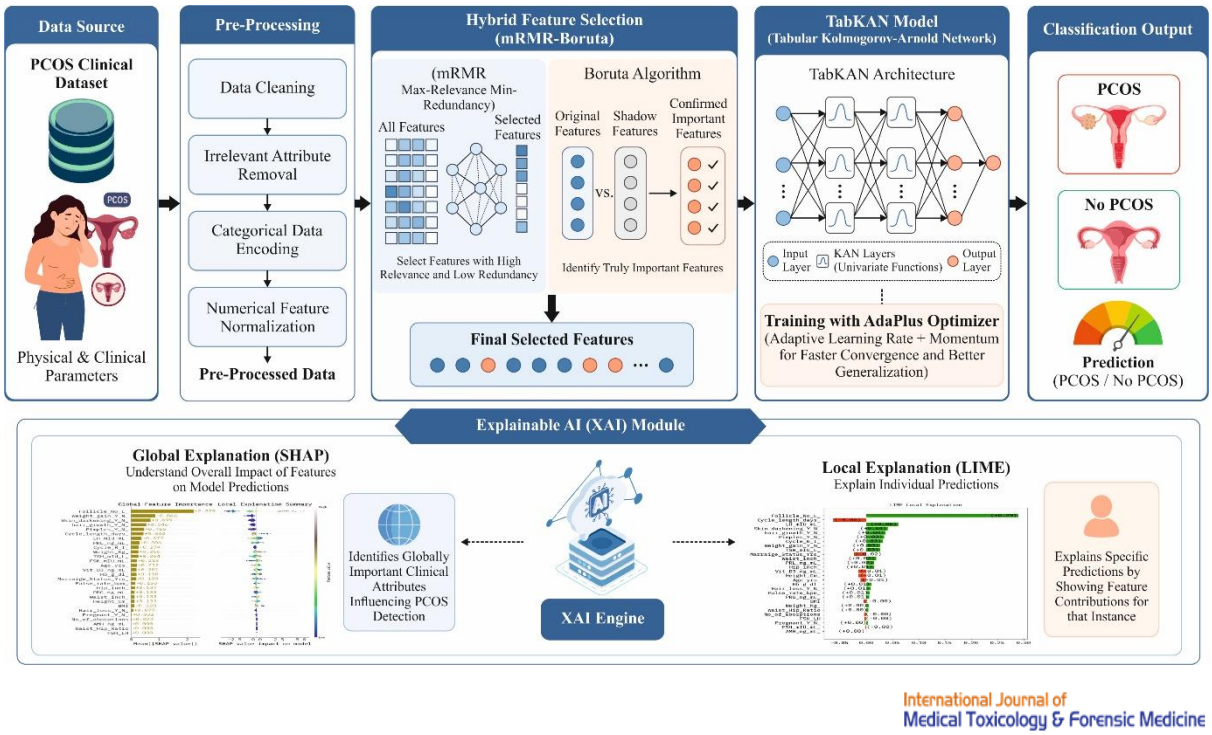
- SHAP and LIME-based explainable artificial intelligence (XAI) techniques are integrated to provide global and local explanations of the model predictions by identifying the contribution of individual clinical features to PCOS classification.

Materials and Methods

The proposed IKAN-PCOS model is designed to detect PCOS using clinical features, as illustrated in Figure 1. The framework includes data preprocessing, mRMR-Boruta feature selection, TabKAN classification, AdaPlus optimization, and SHAP-LIME-based interpretation. First, the clinical data are cleaned, missing values are handled, categorical features are encoded, and the data are normalized. The processed data are then given to mRMR-Boruta to select the most relevant features and remove redundant information. TabKAN uses these selected features for PCOS classification, while AdaPlus optimizes the model parameters. Finally, SHAP and LIME are used to explain the features influencing the model predictions.

Dataset Description and Acquisition

The clinical dataset employed in this study is the PCOS dataset obtained from Kaggle [14]. The PCOS



International Journal of Medical Toxicology & Forensic Medicine

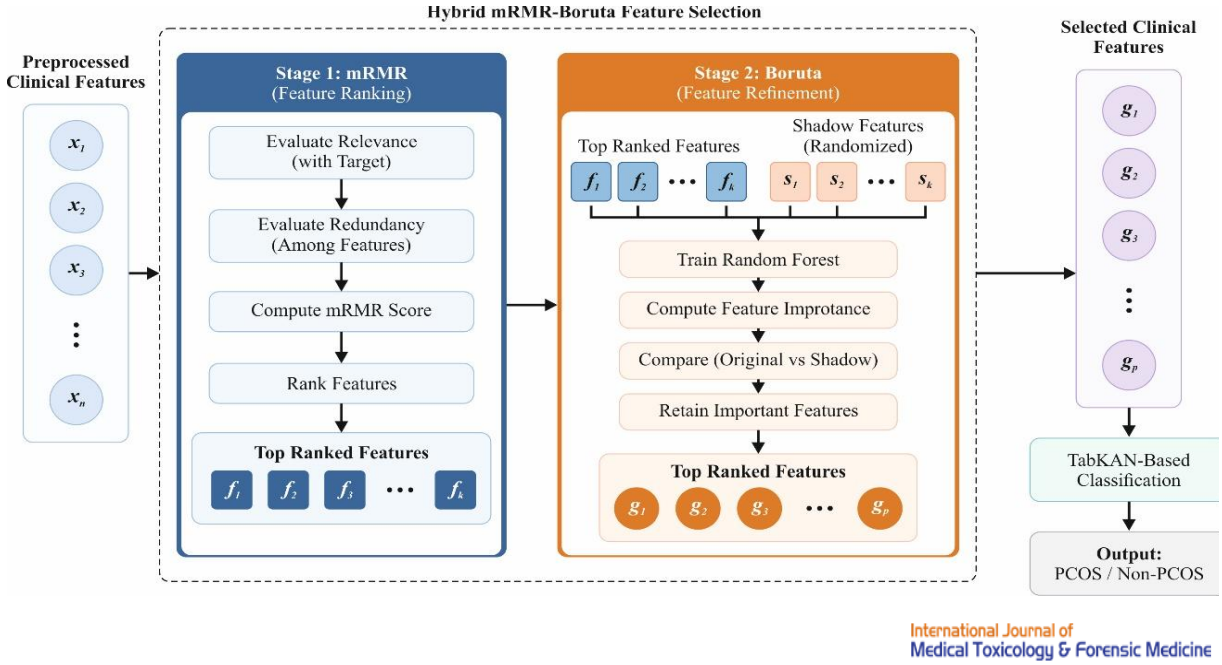
Figure 1. Proposed model for PCOS detection.

dataset contains 541 records and 44 features collected from 10 hospitals in Kerala, India. It includes 364 No PCOS and 177 PCOS cases, with demographic, physical, reproductive, hormonal, and clinical attributes. The features include numerical, categorical, and binary data, while the target variable indicates PCOS or No PCOS.

Hybrid mRMR-Boruta Feature Selection

Clinical data may contain irrelevant and redundant features that can affect PCOS classification. To select useful features and reduce redundancy, we use a hybrid mRMR-Boruta approach. Figure 2 shows the hybrid mRMR-Boruta feature selection process.

The first stage involves ranking features according to the mRMR criterion, which aims to identify features with maximum dependency on the output [15]. Given a candidate feature F_j and the target Y, maximum dependency and minimum redundancy between features are considered. The maximum relevance is



International Journal of Medical Toxicology & Forensic Medicine

Figure 2. Hybrid mRMR-Boruta feature selection process.

generally defined with respect to Mutual Information (MI), while minimum redundancy is defined based on the MI among the selected features.

$$I(X_j;Y) = \sum p\,(x_j,y) \log\left(\frac{p(x_j,y)}{p(x_j)p(y)}\right) \quad (2)$$

$$R(X_j,S) = \frac{1}{|S|}\sum_{X_k \in S} I\,(X_j;X_k) \quad (3)$$

$$\Phi(X_j) = I(X_j;Y) - R(X_j,S) \quad (4)$$

In this stage, the features selected in the first stage using the mRMR criterion are provided to the Boruta selection technique to refine the retained features. Boruta is a wrapper algorithm that selects features using a tree-based model and compares their importance through a randomization process.

$$X_s = \text{Boruta}(\text{mRMR}(X)) \quad (5)$$

The resulting compact feature representation is then fed to the TabKAN model for classification.

TabKAN-Based Classification

A TabKAN is applied to the selected clinical attributes for classifying PCOS. A TabKAN is chosen because clinical data typically contain nonlinear dependencies and interactions among different patient characteristics [16]. Unlike general tabular neural networks that usually adopt fixed activation functions, KAN models use learnable functions to describe transformations between variables.

$$y = f(X_s) \quad (6)$$

In the PCOS context, this characteristic becomes essential because each clinical feature might not contribute independently. Figure 3 represents the TabKAN-based PCOS classification process. The output provides a classification result for PCOS or non-PCOS based on a threshold value of $P(PCOS \mid X_s)$:

$$\hat{y} = \begin{cases} 1, & P(PCOS \mid X_s) \geq \tau \\ 0, & P(PCOS \mid X_s) < \tau \end{cases} \quad (7)$$

AdaPlus-Based Model Optimization

The performance of a neural network depends strongly on how effectively its parameters are updated during training. Therefore, the AdaPlus optimizer is employed for optimizing the trainable parameters of the TabKAN model [17]. In each training loop, the network parameters are updated using the gradient of the classification error to reduce prediction error.

$$\theta^{(t+1)} = \theta^{(t)} - \alpha g^{(t)} \quad (8)$$

The optimizer is applied only during model training, while the final trained TabKAN model is subsequently used for prediction and interpretation.

SHAP and LIME-Based Explainable AI

SHAP and LIME are used to explain the model decisions [18]. SHAP provides both global

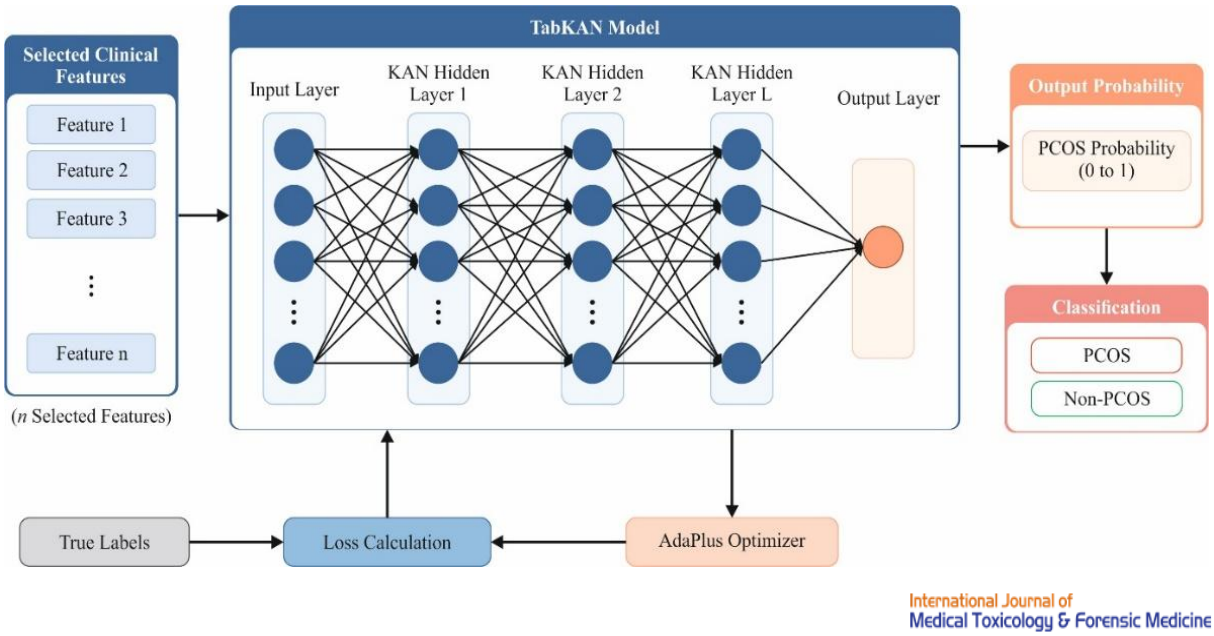


Figure 3. TabKAN classification process.

explanations of the selected clinical features in terms of their overall importance and local explanations for individual predictions. At the global level, the impact of the selected clinical attributes on the model outcomes is represented by averaging the absolute values of their SHAP contributions.

$$f(x) = E[f(x)] + \sum_{j=1}^{d_s} \phi_j \quad (9)$$

LIME is additionally used to provide a local interpretation of individual predictions. It mimics the behavior of the complex black-box model using an interpretable model, which is expressed as,

$$\xi(x) = \arg\min_{g \in G} [L(f, g, \pi_x) + \Omega(g)] \quad (10)$$

.

Results

The parameter settings are given as follows: learning rate: 0.01, dropout: 0.5, batch size: 5, and activation: ReLU. For experimental validation, 70% of the data is used for training and 30% for testing. Table 1 displays the PCOS classification results of IKAN-PCOS on an 80% training and 20% testing split. The proposed model achieves an average accuracy of 96.53% and an MCC of 92.16% in the training set,

Table 1. PCOS classification of IKAN-PCOS model using 80:20 split.

Splits	Accuracy	Precision	Recall	F1-Score	MCC
Training Set (80%)	96.53	95.83	96.33	96.07	92.16
Testing Set (20%)	95.41	94.56	95.17	94.85	89.72

International Journal of
Medical Toxicology & Forensic Medicine

whereas the corresponding testing results are 95.41% and 89.72%, respectively.

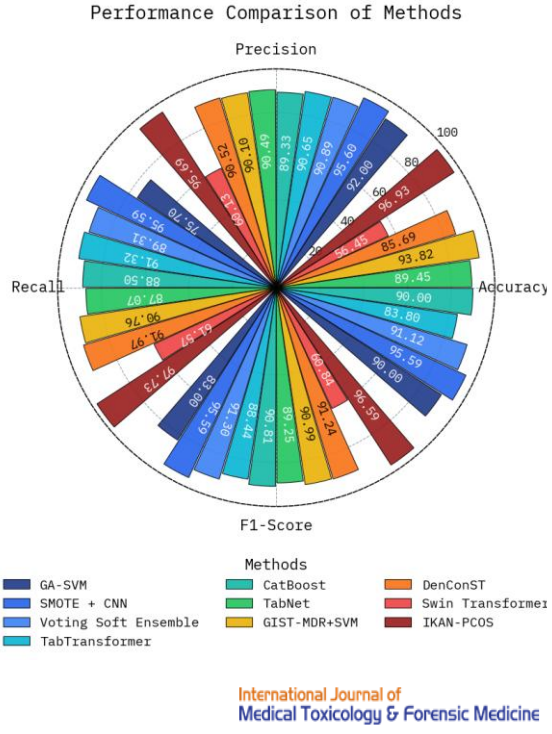
Table 2 describes the PCOS classification results of the IKAN-PCOS framework on 70% training and 30% testing splits. The proposed method attains an average accuracy of 97.88% with an MCC of 95.19% during training, while the testing results are consistently 96.93% and 93.40%, showing the robustness of the proposed framework for accurate PCOS classification.

Table 2. PCOS classification of the IKAN-PCOS model using a 70:30 split.

Splits	Accuracy	Precision	Recall	F1-Score	MCC
Training Set (70%)	97.88	98.00	97.19	97.58	95.19
Testing Set (30%)	96.93	95.69	97.73	96.59	93.40

Discussion

Figure 4 compares the IKAN-PCOS methodology with existing approaches [19,20]. The results show that the Swin Transformer [20] model performs poorly, with an accuracy of 56.45%, whereas the TabTransformer [20], DenConST [20], and TabNet [20] methodologies achieve slightly better performance. Moreover, CatBoost [20], GA-SVM [19], Voting Soft Ensemble [19], GIST-MDR+SVM [20], and SMOTE + CNN [19] achieve strong performance. Furthermore, the results show that the proposed IKAN-PCOS model achieves promising performance, with an



International Journal of
Medical Toxicology & Forensic Medicine

Figure 4. Comparative analysis of the IKAN-PCOS approach with existing methods.

accuracy of 96.93%, precision of 95.69%, recall of 97.73%, and an F1-score of 96.59%, respectively.

Figure 5 shows the global and local feature importance of the proposed model. The global explanation indicates each feature's overall importance, while the local explanation shows how the features affect an individual prediction.

Figure 6 presents the LIME explanation for a specific prediction. Green and red bars indicate positive and negative feature contributions, respectively. Follicle_No_L has the highest positive contribution (+0.29), while Cycle_length_days_ has the highest

negative contribution (−0.06). Features closer to zero have less effect on the prediction.

Table 3. Ablation study analysis of proposed model.

Methods	Accuracy	Precision	Recall	F1-Score
TabKAN	89.84	88.76	87.92	88.34
mRMR + TabKAN	92.16	91.24	90.48	90.82
mRMR + Boruta + TabKAN	94.37	93.52	94.16	93.83
IKAN-PCOS (mRMR + Boruta + TabKAN + AdaPlus)	96.93	95.69	97.73	96.59

International Journal of Medical Toxicology & Forensic Medicine

Table 3 shows the ablation study results of the proposed model. The results indicate that including mRMR, Boruta, TabKAN, and AdaPlus optimization improves performance and classification results.

Table 4. Cross-validation analysis of proposed model.

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	96.72	95.31	97.42	96.35
Fold 2	97.14	95.82	97.91	96.83
Fold 3	96.58	95.47	97.36	96.38
Fold 4	97.31	96.02	98.14	97.07
Fold 5	96.89	95.61	97.68	96.63
Fold 6	97.08	95.74	97.85	96.79
Fold 7	96.41	95.18	97.21	96.18
Fold 8	97.26	96.11	98.03	97.06
Fold 9	96.76	95.56	97.59	96.57
Fold 10	96.95	95.78	97.71	96.67
SD	±0.295	±0.294	±0.301	±0.296
Mean	96.91	95.66	97.69	96.65

International Journal of Medical Toxicology & Forensic Medicine

The cross-validation results of the proposed model are tabulated in Table 4. The experimental outcomes underlined the consistent performance of the proposed model under varying folds. The results showed that the proposed model achieved improved results across all folds.

Conclusion

This paper proposed the IKAN-PCOS framework for detecting PCOS from clinical features. This study integrated mRMR-Boruta feature selection to retain the most informative clinical features and reduce redundancy, a TabKAN classifier to learn intricate nonlinear relationships among the selected clinical attributes, the AdaPlus optimizer for parameter optimization, and SHAP and LIME methods for model explanation and transparency. The AdaPlus optimization procedure guided the parameter optimization of the TabKAN network. SHAP and LIME explanations provided insights into both the model's global decision-making patterns and the local decision patterns for individual patients. The experimental evaluation achieved an accuracy of 96.93% for the IKAN-PCOS framework in distinguishing between PCOS and non-PCOS instances in the clinical dataset. Future work should validate the framework using larger, multi-center, diverse, and longitudinal clinical datasets. Additional data sources, such as medical imaging and genomic information, can also be explored to improve the framework and support more comprehensive PCOS assessment.

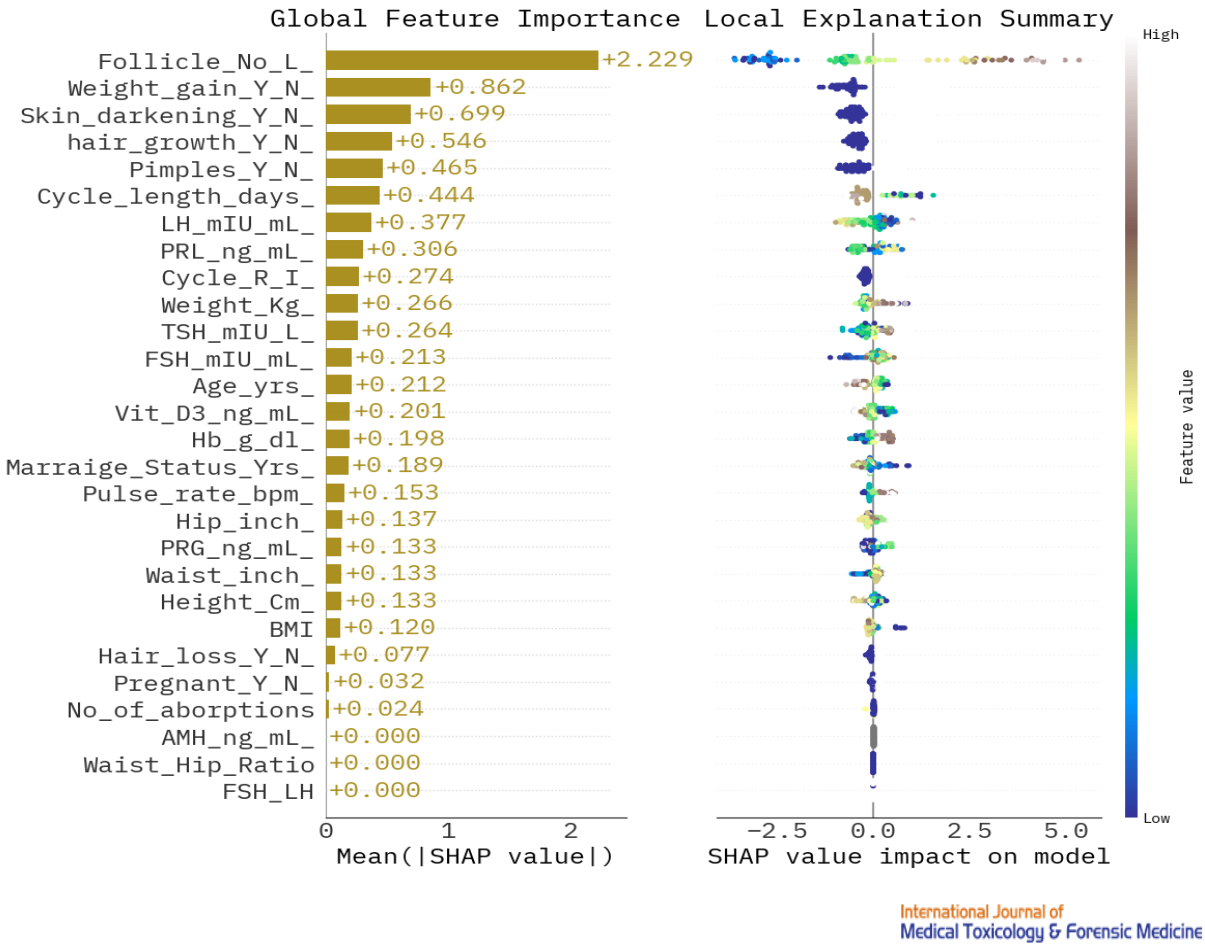


Figure 5. Global feature importance and local SHAP-based model explanations.

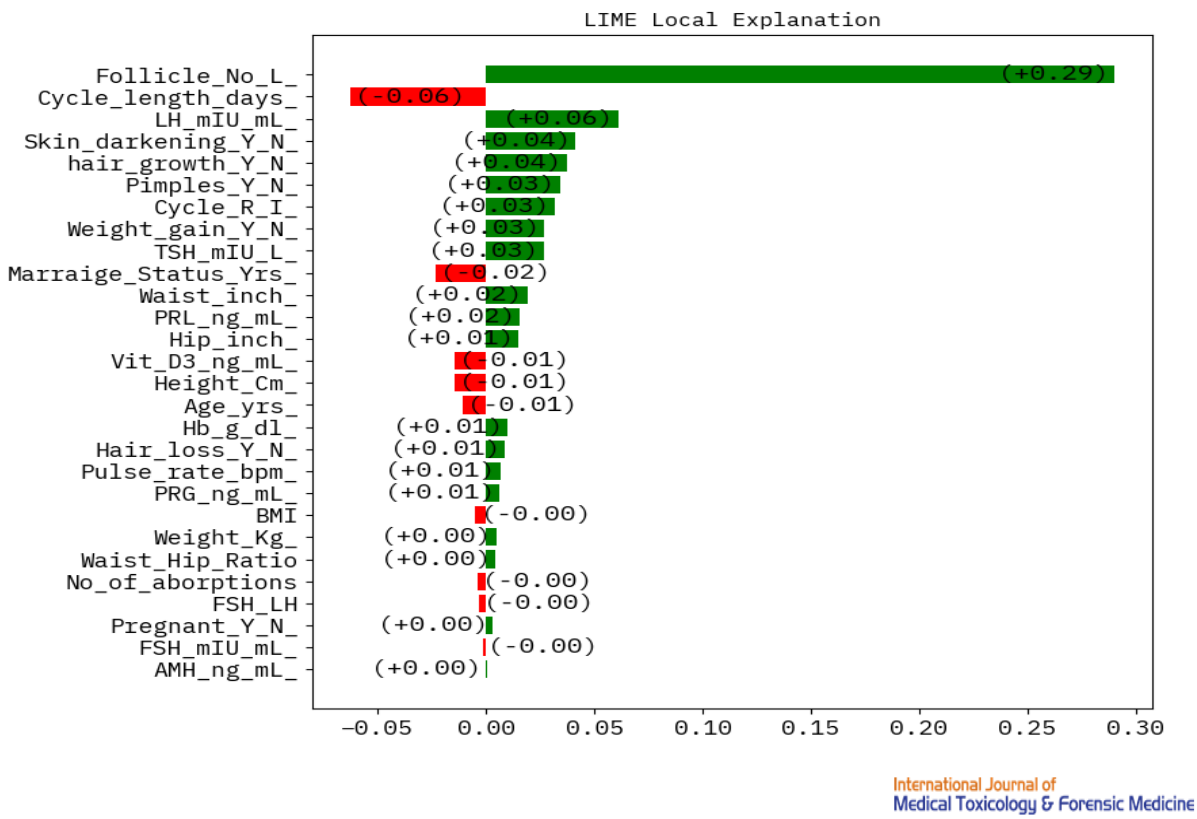


Figure 6. LIME-based Local Explanation.

Ethical Approval

Not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

[1] Ramteke M, Raut S. Prediction of polycystic ovary syndrome using machine learning with SFS and Boruta feature selection: an explainable AI approach. Systems Biology in Reproductive Medicine. 2025;71(1):439-60. [DOI: 10.1080/19396368.2025.2560839]

[2] Efty MRH, Rohman MN, Uddin KMM, Rahman MM, Uddin MA. A Two-Stage Hybrid Feature Selection and Ensemble Learning Framework with Explainable AI for Accurate PCOS Prediction. Health Science Reports. 2026;9(9):e73100. [DOI: 10.1002/hsr2.73100]

[3] Sah S, Myneni M, Sulaiman RB. A Fused Interpretable Architecture for PCOS Detection: Leveraging Deep learning and Ensemble learning with Explainable AI. Procedia Computer Science. 2026;282:369-78. [DOI: 10.1016/j.procs.2026.05.084]

[4] Nayem SI, Efat AH, Surhyth AI. X-decision tree: An explainable AI-based architecture for PCOS prediction with hyperparameter tuning and cross-validation. In: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN); 2025 July. IEEE; p. 1-6. [DOI: 10.1109/QPAIN66474.2025.11171903]

[5] Malode A, Injamuri V, Pawar V, Atkane S. Explainable AI-Based PCOS Detection Using SMOTE-Balanced Clinical Markers. International Journal of Computer Information Systems and Industrial Management Applications. 2026;18(15s):282-96. [DOI: 10.70917/ijcisim-2026-4403]

[6] Athapaththu SHSK, Malkanthi AMC. Explainable AI-Driven Early Screening of Polycystic Ovary Syndrome. In: 2026 6th International Conference on Advanced Research in Computing (ICARC); 2026 February. IEEE; p. 1-6. [DOI: 10.1109/ICARC68737.2026.11454090]

[7] Vijay M, Nandhini N, Mallikarjun I, Kumar GJ, Mounika K, Lavanya LL. Explainable AI-based Predictive Model for Early Diagnosis of Polycystic Ovary Syndrome (PCOS). In: 2026 Second International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS); 2026 January. IEEE; p. 1-7. [DOI: 10.1109/ICISCoIS62701.2026.11447799]

[8] Babu P, Laxmanan DL. Explainable AI-Driven LeNet-Deep Maxout Network-Based Polycystic Ovary Syndrome Detection Using Ultrasound Image. Journal of Bionic Engineering. 2026:1-30. [DOI: 10.1007/s42235-026-00916-8]

[9] Malik N, Srivastava I, Sachdeva R. Machine Learning for Early Diagnosis of PCOS Using Explainable AI. In: International Conference on Innovative Computing and Communication. Cham: Springer Nature Switzerland; 2026 February. p. 284-95. [DOI: 10.1007/978-3-032-28310-8_20]

[10] Baalebail B, Vrishank HJ, Rao ABR, Shetty AJ, Kodipalli A, Thomas M. Deep Learning-Driven Multi-Modal Diagnosis of Polycystic Ovary Syndrome (PCOS) with Explainable AI. In: 2026 Contemporary Computing Innovations Conference (CCIC); 2026 February. IEEE; p. 1-6. [DOI: 10.1109/CCIC68129.2026.11485971]

[11] Jaganathan G, Natesan S. Blockchain and explainable-AI integrated system for Polycystic Ovary Syndrome (PCOS) detection. PeerJ Computer Science. 2025;11:e2702. [DOI: 10.7717/peerj-cs.2702]

[12] Patel K, Rai S, Nischal J, Pandey D, Shrivastava RK, Hazra D. Polycystic Ovary Syndrome (PCOS) Detection Using Deep Learning and Explainable AI. In: 2025 2nd International Conference on Advanced Computing and Emerging Technologies (ACET); 2025 November. IEEE; p. 1-6. [DOI: 10.1109/ACET67282.2025.11430323]

[13] Shah IB, Yadav PK. Automated PCOS

Detection Using Fine-Grained Deep Feature Extraction and Explainable AI: A Transformer-Based Ensemble Approach. International Journal of Image, Graphics and Signal Processing. 2026;18(1):88. [DOI: 10.5815/ijigsp.2026.01.06]

[14] Polycystic Ovary Syndrome (PCOS) Dataset. Kaggle. [Link]

[15] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;27(8):1226-38. [DOI: 10.1109/TPAMI.2005.159]

[16] Farzanehpour M, Soleimanjahi H, Hassan Z, Amanzadeh A, Ghaemi A, Fazeli M. HSP70 modified response against HPV based tumor. European Review for Medical & Pharmacological Sciences. 2013;17(2):228. [Link]

[17] Guan L. AdaPlus: Integrating Nesterov momentum and precise stepsize adjustment on AdamW basis. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2024 April. IEEE; p. 5210-4. [DOI: 10.1109/ICASSP48485.2024.10447337]

[18] Salih AM, Raisi-Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, Menegaz G. A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems. 2025;7(1):2400304. [DOI: 10.1002/aisy.20]

[19] Ahmad R, Maghrabi LA, Khaja IA, Maghrabi LA, Ahmad M. SMOTE-based automated PCOS prediction using lightweight deep learning models. Diagnostics. 2024;14(19):2225. [DOI: 10.3390/diagnostics14192225]

[20] Rani V, Venkatachalam C, M KRK. Explainable Artificial Intelligence with Deep Representation Learning Model for Polycystic Ovary Syndrome Detection Using Clinical Features. Journal of Innovative Image Processing. 2026;8(3):916-37. [DOI: 10.36548/jiip.2026.3.009]

.