RESEARCH ARTICLE

Automated PCOS Disease Detection Using Clinical and Diagnostic Features

Sana Rubab[1], **Musarrat Shaheen**[1], **Zohra Tabassum**[1], **Samreen Liaqat**[1], **Memoona Saleem**[1] **and Aamir Ali**[2,*]

[1] Department of Nursing, Quaid-e-Azam College of Nursing and Allied Health Sciences, Sahiwal 57000, Pakistan
[2] Department of Computer Science, Quaid-e-Azam College of Engineering and Technology, Sahiwal 57000, Pakistan

Abstract

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting reproductive-age women, leading to infertility, hormonal imbalance, insulin resistance, and cardiovascular complications. Early diagnosis remains challenging due to heterogeneous manifestations, overlapping symptoms, and lack of automated screening tools. To address these issues, this study presents a comprehensive comparative framework for PCOS prediction using machine learning and deep learning on a public dataset of 541 patient records. The framework incorporates missing value imputation, feature standardization, SMOTE class balancing, and correlation-based feature selection. Five machine learning algorithms (Decision Tree, KNN, SVM, Random Forest, XGBoost) and two deep learning architectures (LSTM, CNN-ResNet) were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC, with confusion matrices employed for detailed classification analysis. KNN achieved the highest accuracy (92.66%), recall (91.67%), and F1-score (89.19%), while XGBoost delivered the best discriminative capability (ROC-AUC = 0.9540). Among deep learning models, LSTM consistently outperformed CNN-ResNet, demonstrating superior ability to capture complex clinical feature relationships. SHAP-based explainability identified ovarian follicle counts, menstrual irregularities, hair growth, weight gain, skin darkening, and anti-Müllerian hormone (AMH) levels as the most influential predictors. These findings indicate that explainable machine learning models, particularly KNN and XGBoost, provide accurate and interpretable decision support for early PCOS screening, enabling timely intervention and offering a promising foundation for intelligent healthcare decision-support systems.

Keywords: polycystic ovary syndrome, PCOS, machine learning, deep learning, k-nearest neighbors, XGBoost, SHAP, clinical decision support.

1 Introduction

Polycystic Ovary Syndrome (PCOS) is one of the most prevalent endocrine disorders affecting women of reproductive age and is characterized by a

Submitted: 19 June 2026
Revised: 10 July 2026
Accepted: 18 July 2026
Published: 09 August 2026

Vol. 2, **No.** 3, 2026.
10.62762/BISH.2026.255887

***Corresponding author:**
✉ Aamir Ali
amirali4436823@gmail.com

Citation
Rubab, S., Shaheen, M., Tabassum, Z., Liaqat, S., Saleem, M., & Ali, A. (2026). Automated PCOS Disease Detection Using Clinical and Diagnostic Features. *Biomedical Informatics and Smart Healthcare*, 2(3), 108–128.

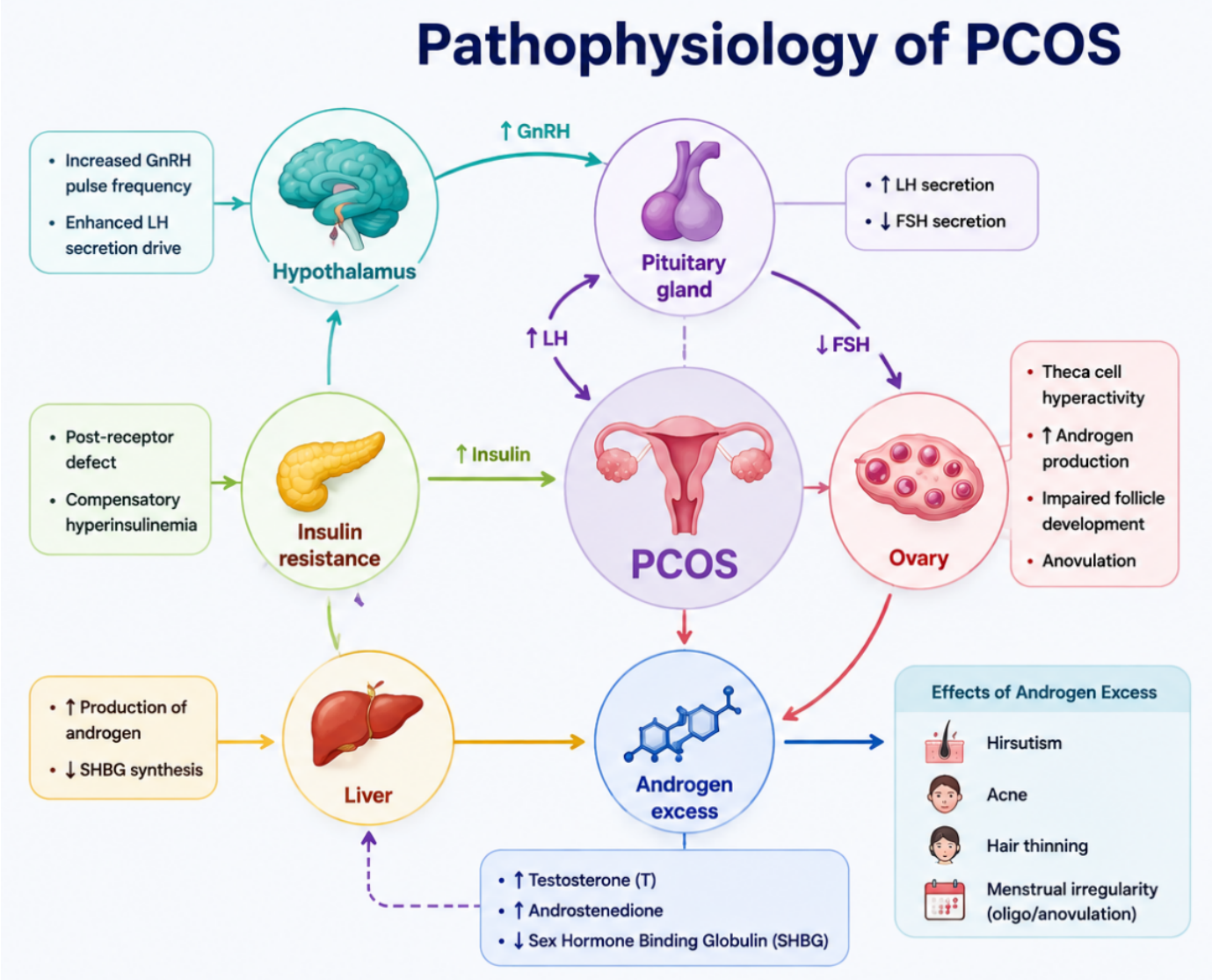

Figure 1. Overview of the pathophysiological mechanisms, clinical manifestations, and long term health consequences associated with PCOS.

combination of reproductive, metabolic, and hormonal abnormalities [1, 2]. Common manifestations include menstrual irregularities, hyperandrogenism, infertility, obesity, insulin resistance, and polycystic ovarian morphology. Beyond its reproductive consequences, PCOS has been associated with an increased risk of type 2 diabetes, cardiovascular complications, and psychological disorders, making early identification and management clinically important. Figure 1 illustrates the major hormonal, metabolic, and reproductive mechanisms associated with the development and progression of PCOS. The condition is characterized by endocrine dysfunction, insulin resistance, and androgen excess, which contribute to a wide range of clinical manifestations and long term health complications [3].

The diagnosis of PCOS remains challenging because its clinical presentation varies considerably among patients and often involves multiple interacting factors. Diagnostic assessment typically requires the evaluation of clinical symptoms, hormonal profiles, metabolic indicators, and ultrasound findings. This complexity can delay diagnosis and treatment, particularly in resource constrained settings where access to specialized expertise may be limited [2, 4]. As a result, there is increasing interest in the development of data-driven approaches that can support clinicians in the screening and assessment of PCOS.

Recent advances in artificial intelligence have significantly expanded the use of data-driven decision-support systems across healthcare applications, including disease screening, risk stratification, prognosis prediction, and personalized treatment planning. In women's healthcare, machine

learning has increasingly been applied to assist clinicians in the early identification of complex disorders characterized by heterogeneous clinical manifestations, thereby supporting timely intervention and improving diagnostic consistency [5]. Explainable artificial intelligence has further enhanced the clinical acceptance of predictive models by providing transparent explanations for model decisions, enabling healthcare professionals to better understand and trust automated recommendations. These developments highlight the growing importance of reliable and interpretable artificial intelligence systems for supporting PCOS diagnosis and management [6].

Building upon these advances in clinical decision-support systems, machine learning techniques have demonstrated considerable potential for analyzing complex healthcare datasets and identifying diagnostic patterns that may not be readily apparent through conventional statistical methods. Algorithms such as Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) have been successfully applied to a variety of medical prediction tasks [7, 8]. More recently, deep learning approaches, including lightweight CNN and LSTM architectures incorporating data balancing strategies, have also been explored because of their ability to model complex nonlinear relationships among clinical variables [9]. However, the relative effectiveness of traditional machine learning and deep learning models for PCOS prediction remains an active area of investigation [4, 10].

Although numerous studies have applied machine learning techniques for PCOS prediction, most investigations have focused on a limited number of algorithms or have evaluated either machine learning or deep learning approaches in isolation [11]. In addition, many studies emphasize predictive performance without examining the relative contribution of clinical features to model decisions. Consequently, there remains a need for a comprehensive comparison of conventional machine learning and deep learning models using a consistent preprocessing framework and standardized evaluation metrics, while also incorporating interpretable analyses to identify clinically relevant predictors associated with PCOS.

Unlike previous studies, which typically evaluate a limited subset of machine learning models, focus exclusively on either conventional machine learning or deep learning approaches, or adopt inconsistent preprocessing and evaluation strategies, this study establishes a unified benchmarking framework that enables fair and reproducible comparison across seven predictive models. All models are trained using the same preprocessing pipeline, hyperparameter optimization strategy, and evaluation metrics, thereby eliminating methodological inconsistencies that hinder objective comparison in earlier studies. Furthermore, this work extends existing research by incorporating SHAP based explainability together with model specific feature importance analysis to improve clinical interpretability. Rather than proposing another prediction algorithm, the primary contribution lies in identifying the most reliable and interpretable modelling strategy for structured clinical PCOS data under standardized experimental conditions, providing evidence that can better support future clinical decision support systems.

The major scientific contributions of this study include:

1. The development of a standardized benchmarking framework for fair comparison between machine learning and deep learning models,
2. Comprehensive evaluation using multiple performance metrics under identical preprocessing conditions,
3. Incorporation of SHAP-based explainability for clinically interpretable prediction,
4. Identification of clinically relevant predictors that may assist future PCOS screening systems.

The remainder of this paper is organized as follows. Section 2 reviews recent studies related to PCOS prediction using machine learning and deep learning techniques. Section 3 describes the dataset, preprocessing procedures, model development process, and evaluation methodology. Section 4 presents the experimental results and comparative performance analysis. Section 5 discusses the limitations of the study and outlines directions for future research. Finally, Section 6 presents the conclusions of the study.

2 Related Work

Recent advances in machine learning and deep learning have facilitated the development of automated approaches for PCOS prediction and diagnosis [12]. Various studies have employed traditional machine learning algorithms, including Decision Tree, K-Nearest Neighbors, Support

Vector Machine, Random Forest, and XGBoost, to analyze clinical, hormonal, and ultrasound-related features associated with PCOS [4, 8]. These approaches have generally demonstrated promising predictive performance, demonstrating the potential of data-driven methods to support clinical decision-making and early disease identification.

Deep learning techniques have also been explored for PCOS prediction. Convolutional Neural Networks have primarily been applied to ultrasound image analysis, while Long Short Term Memory networks and hybrid architectures have been utilized to model complex relationships within structured clinical datasets [17–19]. Recent studies have further demonstrated that integrating conventional machine learning algorithms with deep learning models such as CNNs and Vision Transformers can improve diagnostic accuracy and support clinical risk stratification for PCOS and related endocrine disorders [13]. Although many studies reported encouraging results, differences in datasets, preprocessing procedures, feature selection strategies, and evaluation protocols make direct comparison among models difficult. Furthermore, model interpretability and comprehensive comparisons between machine learning and deep learning approaches remain insufficiently investigated [4, 8].

Researchers have increasingly emphasized the importance of integrating both machine learning and deep learning techniques within a common evaluation framework for PCOS prediction [8]. Multimodal approaches that combine image-based and text-based clinical data have also been explored to leverage complementary diagnostic information [14]. Such comparative investigations enable a more objective assessment of model effectiveness, computational requirements, and generalization capabilities across different data characteristics [15]. Non-invasive diagnostic approaches combining ultrasound imaging with structured clinical and biochemical features have also been shown to improve prediction accuracy and support broader clinical applicability [16]. In addition, the growing demand for trustworthy artificial intelligence in healthcare has highlighted the need for explainable prediction models that can provide clinicians with transparent insights into the factors influencing diagnostic outcomes. Consequently, there is a need for studies that not only compare diverse predictive approaches but also examine their interpretability and practical applicability in real clinical environments.

Recent survey articles have highlighted significant progress in applying artificial intelligence for PCOS diagnosis while simultaneously identifying several unresolved research challenges. Ahmed et al. [8] reviewed machine learning techniques for PCOS detection. They concluded that existing studies often evaluate only a limited number of algorithms using heterogeneous datasets and inconsistent validation strategies, making objective performance comparison difficult. Similarly, Govindharajan et al. [4] emphasized that although deep learning methods have demonstrated promising predictive capability, the absence of standardized preprocessing pipelines, limited explainability, and inadequate clinical validation continue to restrict their practical adoption. Dhanka et al. [7] further reported that trustworthy AI remains an important challenge because many predictive models function as black boxes without providing clinically interpretable explanations. Furthermore, recent reviews have identified insufficient benchmarking between conventional machine learning and deep learning methods under identical experimental conditions, together with the limited use of explainable artificial intelligence techniques and the lack of external validation using diverse patient populations. These findings collectively demonstrate that existing literature still lacks a unified and reproducible evaluation framework capable of simultaneously comparing multiple predictive models while maintaining transparency and clinical interpretability. The proposed study directly addresses these challenges through a standardized preprocessing pipeline, comprehensive benchmarking of seven machine learning and deep learning models, hyperparameter optimization, and SHAP-based explainability for clinically meaningful feature interpretation.

These recent survey papers indicate that current PCOS prediction research is characterized by four major challenges: (i) inconsistent preprocessing and validation protocols that hinder objective comparison among predictive models, (ii) limited benchmarking between conventional machine learning and deep learning approaches under identical experimental settings, (iii) insufficient integration of explainable artificial intelligence for clinically interpretable predictions, and (iv) limited external validation using diverse patient populations. These research gaps highlight the need for a standardized and reproducible evaluation

Table 1. Summary of previous studies on PCOS prediction and their limitations.

Study	ML Models	DL Model	Standardized Evaluation Framework	Explainability	Hyperparameter Optimization	Major Limitation
[20]	✓	×	×	×	×	Limited algorithm comparison
[21]	✓	×	×	×	×	No deep learning evaluation
[22]	✓	×	×	×	×	Single model investigation
[23]	✓	×	×	Limited	×	Limited interpretability and real world validation
[24]	×	✓	×	×	×	High computational cost and dataset dependency
[3]	✓	✓	Partial	×	×	Evaluated using a single benchmark dataset
[25]	×	✓	×	×	×	No explainability analysis; limited to wrapper-based feature selection without cross-dataset validation
[26]	✓	✓	Partial	×	Partial	Lack of comprehensive cross dataset benchmarking
This Study	✓	✓	✓	✓(SHAP + Feature Importance)	✓	Single dataset evaluation; external validation remains future work

framework capable of objectively comparing multiple predictive models while maintaining clinical transparency. The present study was designed to address the first three challenges through unified benchmarking, standardized preprocessing, systematic hyperparameter optimization, and SHAP-based explainability.

Table 1 summarizes representative studies on PCOS prediction and highlights the limitations that motivated the present work.

The findings summarized in the recent survey papers and the representative studies presented in Table 1 demonstrate that, despite substantial advances in artificial intelligence for PCOS prediction, methodological inconsistencies continue to limit objective comparison across studies. These observations further support the need for a unified benchmarking framework that simultaneously evaluates conventional machine learning and deep learning models using identical preprocessing, optimization, and evaluation protocols while incorporating explainable artificial intelligence to improve clinical interpretability.

3 Methodology

3.1 Research Framework

This study employed a data driven framework for the prediction of PCOS using clinical patient data. The methodological workflow consisted of data acquisition, preprocessing, model development, model evaluation, and comparative performance analysis. The primary objective was to determine the effectiveness of both machine learning and deep learning approaches in identifying patients with PCOS based on demographic, clinical, hormonal, lifestyle, and ultrasound related attributes. Model performance was assessed using several classification metrics, including accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve. Confusion matrices and ROC curves were further employed to examine classification behaviour and discriminative capability. Finally, feature importance analysis was conducted to identify the clinical variables that contributed most substantially to PCOS prediction.

The present study followed a structured workflow for the prediction of PCOS using clinical patient data. The process began with dataset acquisition and preprocessing, followed by feature preparation, model development, performance evaluation, and comparative analysis. The complete methodological workflow adopted in this study is illustrated in Figure 2.

3.2 Dataset Description

The study utilized a publicly available clinical dataset comprising 541 patient records and 45 attributes [27]. Each record represented an individual patient and contained demographic, physiological, reproductive, hormonal, lifestyle, and ultrasound related information. The dataset was designed for binary classification, where the target variable was PCOS (Y/N). A value of 1 indicated the presence of

Table 2. Categories of features in the PCOS clinical dataset.

Category	Features Included
Demographic Features	Age (years), Weight (Kg), Height (cm), BMI, Blood Group
Vital Signs & Basic Health Indicators	Pulse Rate, Respiratory Rate, Hemoglobin
Menstrual & Reproductive Health	Cycle Type, Cycle Length, Marriage Duration, Pregnancy Status, Number of Abortions
Hormonal & Biochemical Parameters	β-hCG (I & II), FSH, LH, FSH/LH Ratio, TSH, AMH, PRL, Vitamin D3, PRG, RBS
Physical & Lifestyle Indicators	Weight Gain, Hair Growth, Skin Darkening, Hair Loss, Pimples, Fast Food Consumption, Regular Exercise
Body Measurements	Waist, Hip, Waist-to-Hip Ratio
Blood Pressure Measurements	Systolic BP, Diastolic BP
Ultrasound Imaging Features	Follicle Number (Left & Right), Average Follicle Size (Left & Right), Endometrium Thickness

PCOS, whereas a value of 0 indicated the absence of the condition [27].

The dataset comprised 541 patient records, including 364 non-PCOS cases (67.3%) and 177 PCOS cases (32.7%), indicating moderate class imbalance. During preprocessing, one missing value in *Marriage Status (Years)*, one missing value in *Fast Food (Y/N)*, and an *Unnamed: 44* column containing 539 missing values were identified; missing values were imputed using median imputation, while the highly incomplete *Unnamed: 44* attribute and identifier variables were removed prior to model development. Table 2 summarizes the principal categories of clinical variables included in the PCOS dataset [27]; the full cleaned feature set comprises 42 variables retained for model training following removal of identifier columns and the highly incomplete *Unnamed: 44* attribute.

The dataset encompasses demographic, reproductive, hormonal, lifestyle, anthropometric, and ultrasound-derived variables commonly associated with PCOS diagnosis and management. These attributes provide a comprehensive representation of patient characteristics and clinical indicators relevant to PCOS prediction.

3.3 Data Preprocessing

Data preprocessing was performed to improve data quality and ensure consistency across all predictive models. Non-numeric placeholder values were converted to missing values, applicable variables were transformed to numeric format, and missing values were imputed using median imputation. Identifier variables, including "Sl. No." and "Patient File No.", along with the highly incomplete Unnamed:44 column, were removed.

The original dataset contained 45 attributes. After data cleaning, 42 predictive variables remained. Predictor and target variables were separated, and the dataset was divided into training and testing subsets using an 80:20 stratified split with a fixed random seed of 42. Feature standardization was performed using StandardScaler, and SMOTE was applied exclusively to the training data to address class imbalance. Correlation analysis was conducted to identify variables associated with PCOS. Twenty features exhibited an absolute correlation coefficient $(|r|) \geq 0.10$ and were identified as the most relevant predictors. However, all 42 cleaned features were retained for subsequent model development and evaluation. For deep learning models, each patient record was reshaped into a one-dimensional sequence with input dimensions of $(42, 1)$. Figure 3 illustrates the preprocessing and model development workflow adopted in this study.

3.4 Model Optimization

To improve predictive performance and reduce the risk of overfitting, model specific hyperparameter optimization was conducted for each machine learning and deep learning model. Prior to model training, the dataset was standardized using StandardScaler and divided into training and testing subsets using an 80:20 stratified split. Class imbalance was addressed through the Synthetic Minority Oversampling

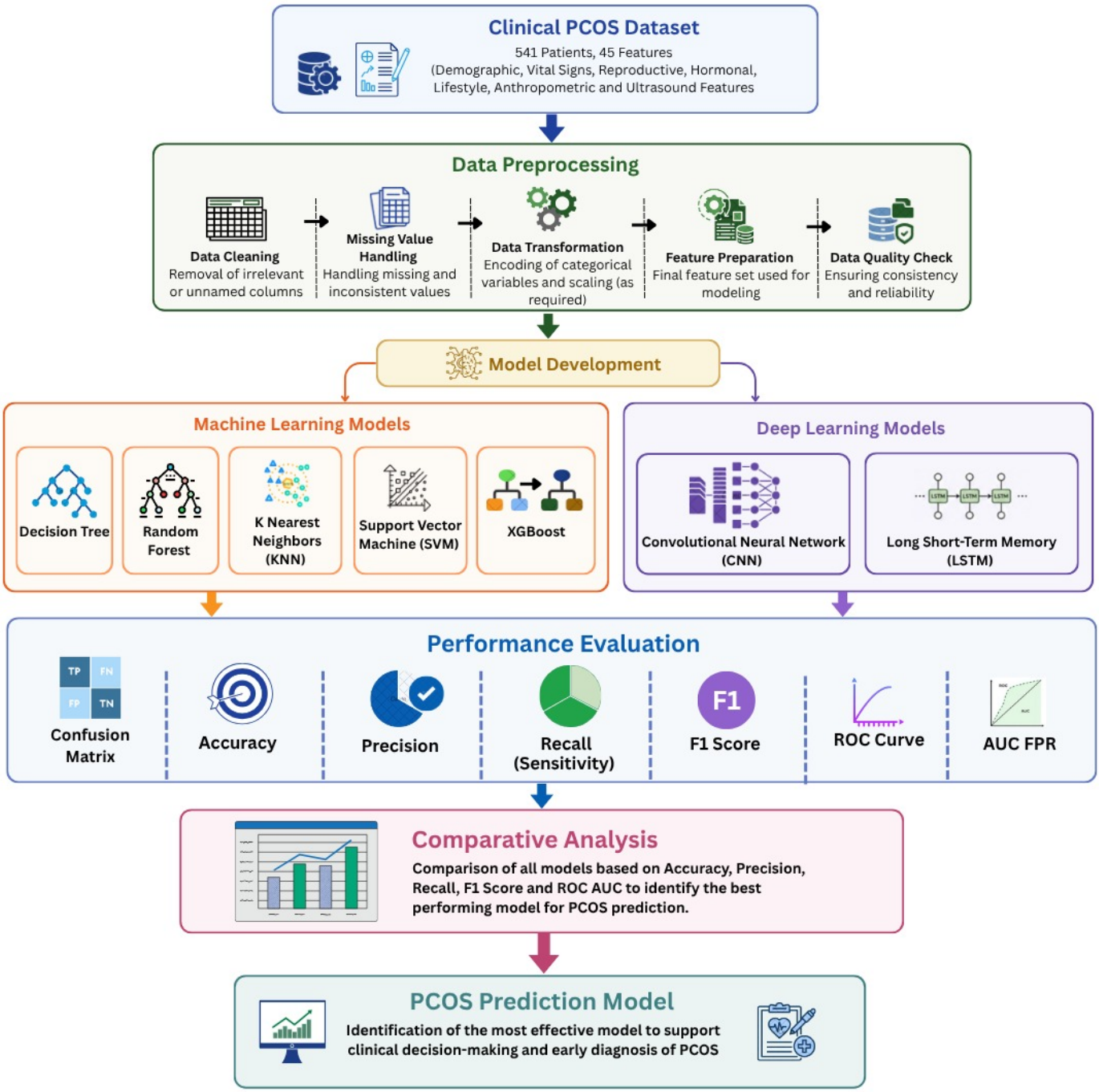

Figure 2. Methodological framework for PCOS prediction using machine learning and deep learning models.

Technique (SMOTE), which was applied exclusively to the training data to avoid information leakage.

Hyperparameter optimization was conducted prior to final model training. For the machine learning models, GridSearchCV with five-fold cross-validation was employed to identify optimal parameter configurations. The parameter set achieving the highest cross-validation score was selected for final model development. For the deep learning models, manual hyperparameter tuning was performed by evaluating multiple architectural and training configurations, and the model achieving the highest validation accuracy was selected. Table 3 summarizes the optimal configurations, best validation performance, and optimization times for all evaluated models.

Hyperparameter optimization was performed using five-fold cross-validation. A total of 28 parameter

Table 3. Optimal hyperparameter configurations selected after model tuning. Values marked with * for deep learning models represent validation set accuracy obtained during manual hyperparameter tuning rather than cross-validation scores.

Model	Best Hyperparameters	Best CV Score	Optimization Time (s)
Decision Tree	criterion=entropy, max_depth=3, min_samples_split=2, min_samples_leaf=1	0.9158	10.61
KNN	n_neighbors=15, weights=distance, p=1	0.9734	3.1
SVM	kernel=RBF, C=10, gamma=0.1	0.981	122.26
RF	n_estimators=50, max_depth=None, max_features=log2, bootstrap=False, min_samples_split=10, min_samples_leaf=1	0.9846	1577.5
XGBoost	n_estimators=100, max_depth=5, learning_rate=0.1, subsample=0.8, colsample_bytree=0.8, gamma=0	0.9806	613.91
LSTM	units=32, dropout=0.4, learning_rate=0.002, batch_size=64, epochs=80	0.8991*	95.79
CNN-ResNet	filters=128, residual blocks=3, dropout=0.2, learning_rate=0.0005, batch_size=16, epochs=150	0.8349*	191.9

combinations were evaluated for KNN, 810 for RF, and 729 for XGBoost. The parameter configuration achieving the highest cross-validation score was selected for final model training. To enhance convergence and training stability, both deep learning architectures incorporated the Adam optimization algorithm. Early stopping with restoration of the best weights was applied to prevent overfitting, while adaptive learning rate reduction was utilized when performance improvements plateaued during training.

3.5 Machine Learning Models

Five machine learning algorithms, namely Decision Tree (DT) [8], KNN [15], SVM [12], RF [8] and XGBoost [10] were evaluated for PCOS classification. All models were trained using the same preprocessed dataset and assessed using identical evaluation metrics to ensure a fair comparison.

Feature importance analysis was conducted to improve model interpretability. Intrinsic feature importance measures were used for DT, RF, and XGBoost, whereas permutation importance was applied to KNN. Additionally, SHAP analysis was performed for XGBoost to provide a comprehensive explanation of feature contributions. Feature importance analysis was not conducted for SVM, CNN-ResNet, and LSTM because these models require specialized explainability techniques beyond the scope of this study.

3.6 Deep Learning Models

Deep learning models were evaluated to investigate whether complex nonlinear feature interactions could improve PCOS classification performance. Two architectures, CNN-ResNet [28] and LSTM [18, 19], were implemented and trained using the same preprocessed dataset. A one-dimensional CNN-ResNet architecture was employed to learn hierarchical feature representations from the clinical dataset. The model consisted of convolutional, pooling, residual, and fully connected layers, with architectural details. The LSTM model was implemented to capture complex relationships among clinical variables. Although the dataset contained tabular rather than temporal data, each patient record was reshaped into a one-dimensional sequence, enabling the network to model potential dependencies among features. Prior studies have demonstrated the effectiveness of hybrid and ensemble deep learning approaches in improving PCOS classification performance on structured clinical datasets [9, 28], further motivating the inclusion of sequence-based architectures in the present benchmarking framework. To improve reproducibility and provide architectural transparency, the configurations of the deep learning models used in this study are summarized in Table 4.

All experiments were conducted in Python using Scikit-learn, TensorFlow/Keras, and XGBoost within the Google Colab environment, with fixed random seeds applied to ensure reproducibility.

Table 4. Deep Learning Architecture Configuration.

Component	CNN-ResNet	LSTM
Input Shape	(42, 1)	(42, 1)
Architecture	1D CNN with Residual Blocks	LSTM (32 units)
Initial Convolution Layer	Conv1D, 64 filters, kernel size = 7	—
Residual Blocks	2	—
Convolution Layers per Residual Block	2	—
Kernel Size	3	—
Activation	ReLU	Tanh
Pooling Layer	MaxPooling1D (pool size = 3)	—
Global Pooling	GlobalAveragePooling1D	—
Dense Layer	64, 32 neurons	32 neurons
Dropout	0.2	0.4
Optimizer	Adam	Adam
Loss Function	Categorical Crossentropy	Categorical Crossentropy

3.7 Performance Evaluation Metrics

The performance of all machine learning and deep learning models was assessed using multiple evaluation metrics to provide a comprehensive analysis of classification effectiveness. Since PCOS detection represents a binary classification problem, several complementary measures were employed to evaluate predictive performance from different perspectives.

A confusion matrix was generated for each model to provide detailed insight into classification outcomes. The confusion matrix summarizes predictions into four categories: 1) True Positives (TP): PCOS cases correctly classified as PCOS, 2) True Negatives (TN): Non PCOS cases correctly classified as non PCOS, 3) False Positives (FP): Non PCOS cases incorrectly classified as PCOS. 4) False Negatives (FN): PCOS cases incorrectly classified as non PCOS. The confusion matrix enables detailed examination of model strengths and weaknesses beyond aggregate performance measures. Based on these four outcomes, several evaluation metrics were computed to assess the predictive performance of the proposed models, as summarized in Table 5.

Feature importance analysis was conducted for Decision Tree, RF, KNN, and XGBoost models. SHAP (SHapley Additive exPlanations) [30] based interpretation was additionally performed for the XGBoost model to provide a more comprehensive explanation of feature contributions to PCOS classification.

4 Results and Discussion

All models were trained and evaluated using a standardized preprocessing pipeline comprising median-based missing value imputation, feature standardization, stratified train-test splitting, and SMOTE-based class balancing. This consistent framework enabled objective comparison of Decision Tree, KNN, RF, XGBoost, LSTM, and CNN-ResNet models using the same test dataset and evaluation metrics for PCOS prediction.

The Decision Tree (DT) classifier was utilized as a supervised learning algorithm to predict the presence of PCOS while maintaining model interpretability.

Figure 4 presents the feature importance rankings generated by the Decision Tree model. Right ovarian follicle count was the most influential predictor, followed by weight gain, hair growth, menstrual cycle regularity, and left ovarian follicle count. Demographic variables such as age, blood group, height, weight, and BMI contributed comparatively less to classification. These findings indicate that ovarian morphology, metabolic characteristics, and menstrual irregularities were the primary factors driving PCOS prediction.

Figure 5 presents the confusion matrix and ROC curve of the Decision Tree classifier. The model correctly classified 70 non-PCOS cases and 29 PCOS cases, while misclassifying 3 non-PCOS cases as PCOS (false positives) and failing to identify

Table 5. Performance Evaluation Metrics Used in PCOS Classification.

Metric	Formula	Description
Accuracy	Accuracy= $\frac{TP+TN}{TP+TN+FP+FN}$	Measures the proportion of correctly classified instances among all samples.
Precision	Precision= $\frac{TP}{TP+FP}$	Measures the proportion of predicted PCOS cases that are correctly classified.
Recall (Sensitivity)	Recall= $\frac{TP}{TP+FN}$	Measures the proportion of actual PCOS cases correctly identified by the model.
F1 Score	F1 $= 2 * \frac{Precision \cdot Recall}{Precision+Recall}$	Harmonic mean of precision and recall, providing a balanced assessment of classification performance.
ROC-AUC	Area under the Receiver Operating Characteristic (ROC) curve	Quantifies the model's ability to distinguish between PCOS and non-PCOS cases across all classification thresholds.

7 PCOS cases (false negatives), resulting in 99 correct predictions out of 109 test samples. The relatively low number of misclassifications indicates good discriminative performance, although the false negative cases remain clinically important because delayed detection may postpone diagnosis and treatment. The ROC curve further demonstrates the model's classification capability, achieving an AUC of 0.93, which indicates a strong ability to distinguish between PCOS and non-PCOS patients across different decision thresholds.

The predictive performance of the Decision Tree classifier was evaluated using multiple classification metrics, including Accuracy, Precision, Recall, F1 Score, and ROC–AUC, as summarized in Table 6.

Table 6. Performance Evaluation of the Decision Tree Model.

Metric	Value
Accuracy	0.9083
Precision	0.9063
Recall (Sensitivity)	0.8056
F1 Score	0.8533
ROC–AUC	0.9313

The Decision Tree model achieved an accuracy of 90.83% and an ROC–AUC of 0.9313, demonstrating strong predictive and discriminative performance. The confusion matrix yielded a precision of 90.63% and a recall of 80.56%, indicating effective identification of PCOS cases, although some positive cases remained undetected. Combined with its interpretable feature rankings and short optimization time of 10.61 seconds, these results suggest that the model is a computationally efficient and clinically relevant approach for PCOS classification.

The KNN classifier operates by identifying the nearest neighboring samples in the feature space and determining the class label through a majority voting mechanism.

Figure 6 presents the top ten features ranked by permutation importance in the KNN model. Right ovarian follicle count was the most influential predictor, followed by hair growth and left ovarian follicle count. Skin darkening, fast food consumption, menstrual cycle regularity, pimples, weight gain, AMH concentration, and age were also identified as important contributors, highlighting the combined influence of ovarian, hormonal, metabolic, and lifestyle factors in PCOS classification. The identified predictors are consistent with established clinical manifestations of PCOS and demonstrate the ability of the KNN model to capture relevant disease characteristics.

Figure 7 presents the confusion matrix and ROC curve of the KNN classifier. The model correctly classified 68 non-PCOS cases and 33 PCOS cases, with only 5 false positives and 3 false negatives, resulting in 101 correct predictions out of 109 test samples and an accuracy of 92.66%. The low number of false negatives indicates effective identification of PCOS cases, while the limited false positives demonstrate good specificity. The ROC curve further confirms the model's strong discriminative performance, achieving an AUC of 0.94 and maintaining a favorable balance between sensitivity and specificity across different

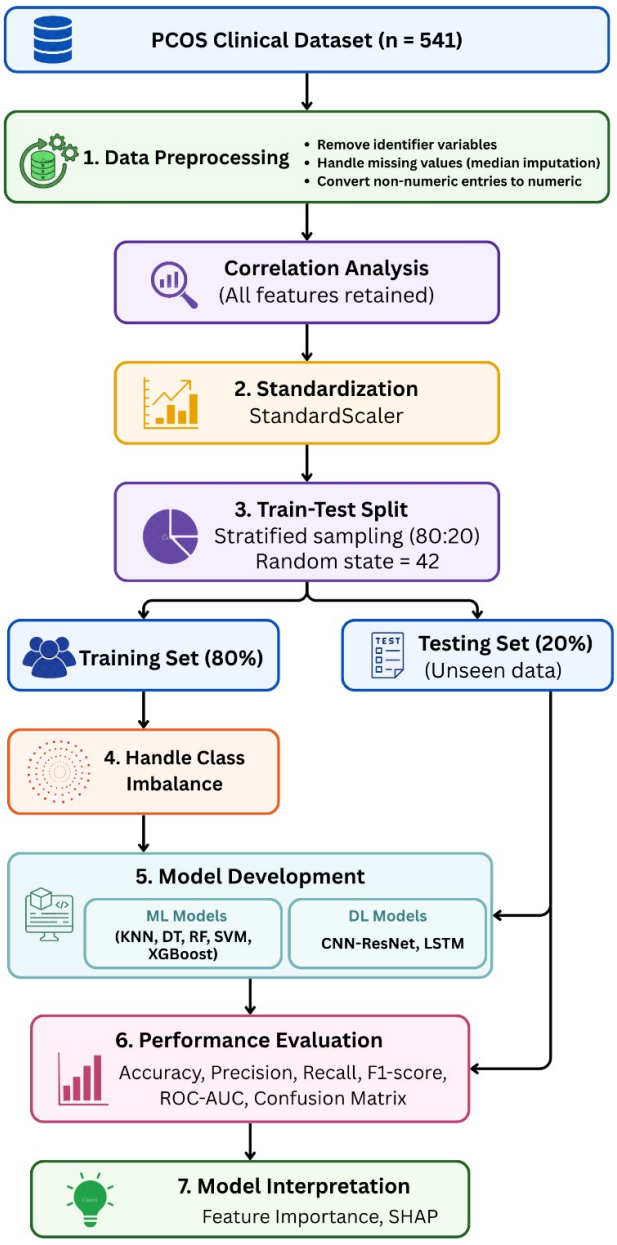

Figure 3. Overall workflow adopted for PCOS prediction.

classification thresholds. These results highlight the effectiveness of KNN for PCOS classification.

Table 7. Performance Evaluation Metrics of the KNN Model for PCOS Classification.

Metric	Value
Accuracy	0.9266
Precision	0.8684
Recall	0.9167
F1 Score	0.8919
ROC–AUC	0.9382

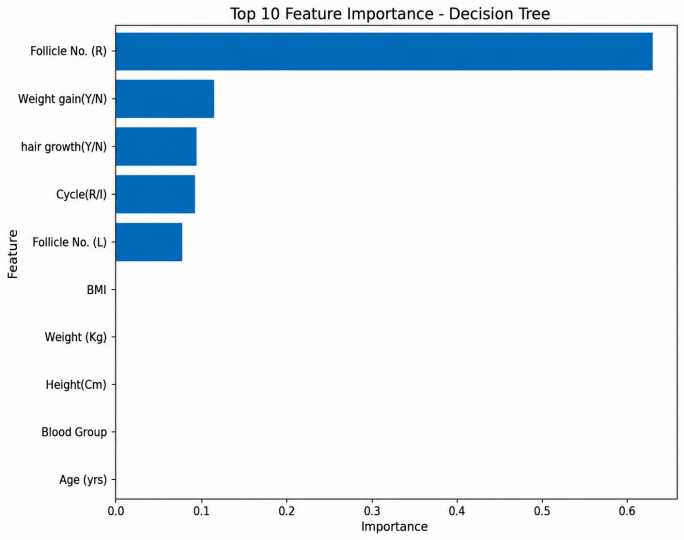

Figure 4. Top 10 Feature Importance Rankings Obtained from the Decision Tree Model.

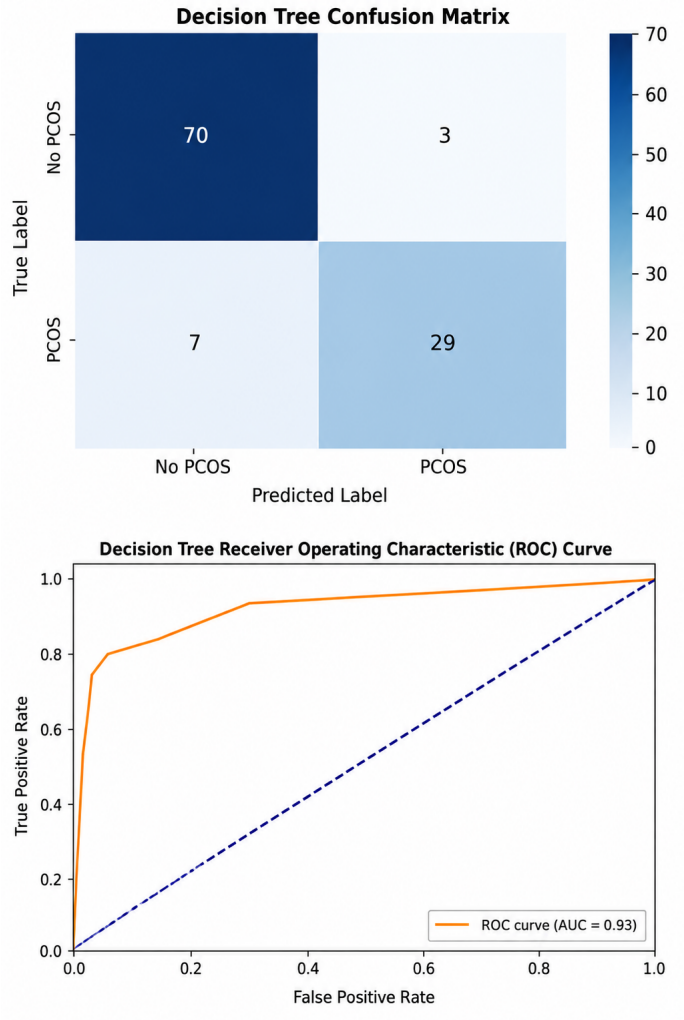

Figure 5. Confusion Matrix and ROC Curve of Decision Tree Model for PCOS Classification.

Table 7 summarizes the performance of the KNN model on the test dataset. The classifier achieved an

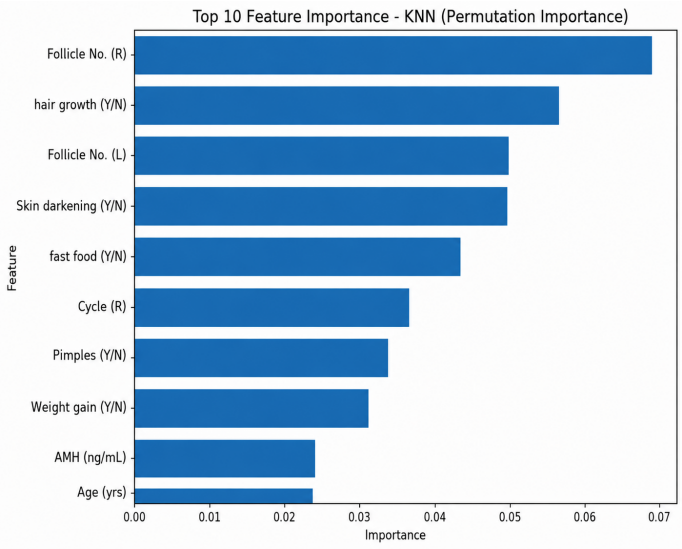

Figure 6. Top 10 Feature Importance Rankings Obtained Using Permutation Importance in the KNN Model.

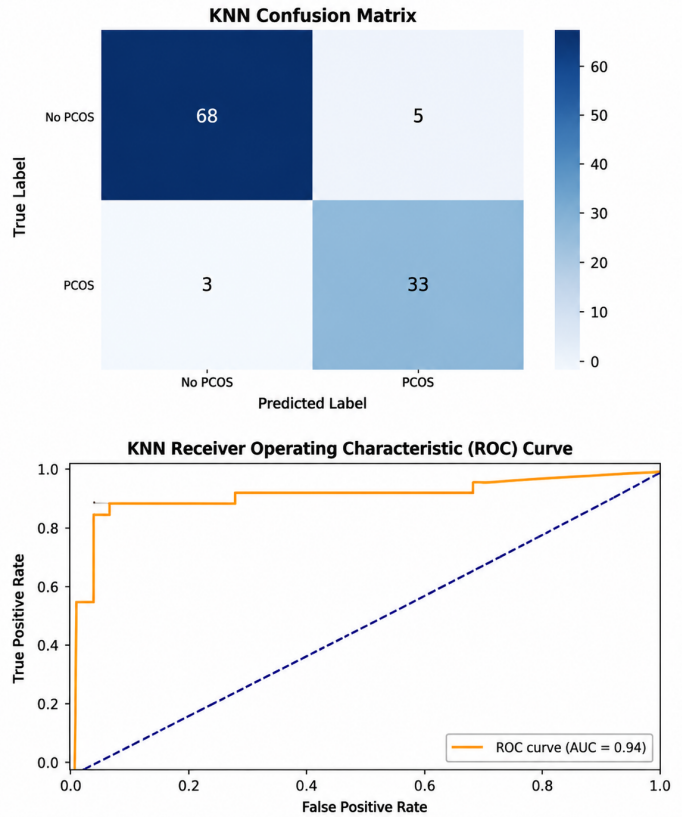

Figure 7. Confusion Matrix and ROC Curve of KNN Model for PCOS Classification.

accuracy of 92.66%, with a precision of 86.84%, recall of 91.67%, and F1 score of 89.19%, indicating a strong balance between identifying PCOS cases and limiting incorrect positive predictions. The ROC–AUC value of 0.9382 further demonstrates excellent discriminative capability. The high recall suggests that the model effectively minimized missed PCOS cases, highlighting its potential utility for PCOS screening and early detection.

XGBoost was employed as an advanced ensemble learning technique based on gradient boosting. The model builds a sequence of decision trees in which each successive tree is trained to correct the errors of preceding trees, thereby improving predictive accuracy and enhancing overall classification performance.

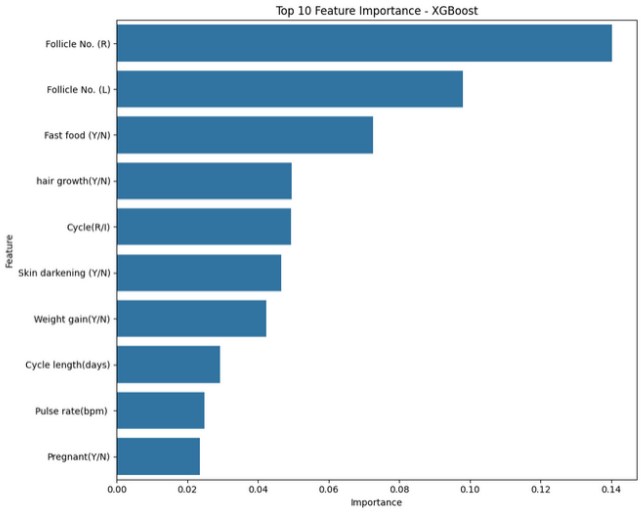

Figure 8. Top 10 Feature Importance Rankings Identified by the XGBoost Model.

Figure 8 presents the top ten features ranked by importance in the XGBoost model. Right and left ovarian follicle counts were the most influential predictors, emphasizing the importance of ovarian morphology in PCOS classification. Fast food consumption, hair growth, menstrual cycle regularity, skin darkening, cycle length, and weight gain were also identified as key contributors, reflecting the combined influence of lifestyle, metabolic, and reproductive factors. Pulse rate and pregnancy status exhibited comparatively lower importance but remained among the top predictive variables. The feature rankings demonstrate that the XGBoost model captured clinically relevant predictors associated with PCOS.

Figure 9 presents the SHAP summary plot of the XGBoost model, providing a global interpretation of feature contributions to PCOS prediction. Consistent with the feature importance analysis, follicle count in the right and left ovaries emerged as the most influential predictors. Additional important features included hair growth, weight gain, skin darkening, menstrual cycle regularity, AMH concentration, and fast food consumption. Positive SHAP values indicate increased contribution toward PCOS prediction, whereas negative values indicate reduced contribution.

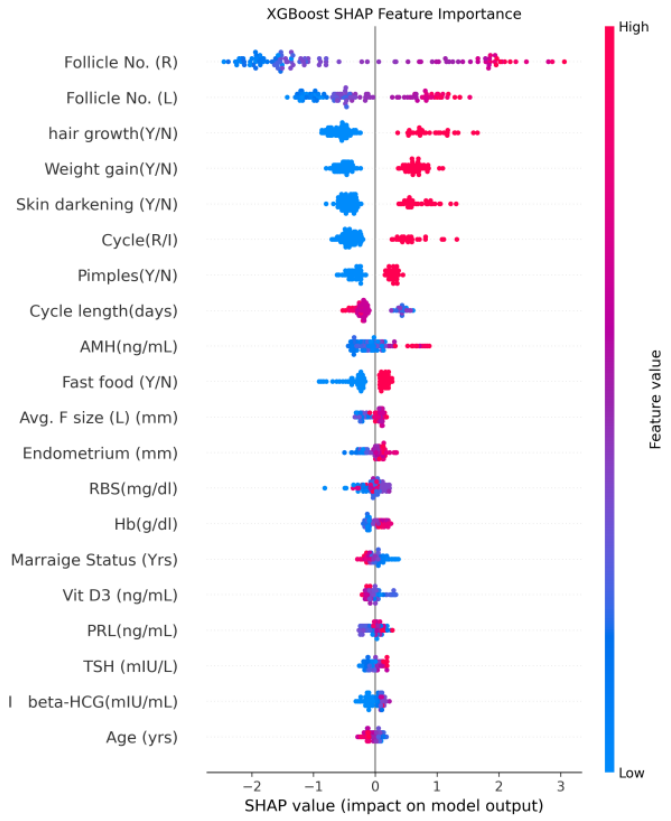

Figure 9. SHAP Summary Plot of the XGBoost Model for PCOS Classification.

The SHAP analysis further supports the clinical relevance and interpretability of the XGBoost classifier.

Figure 10 presents the confusion matrix and ROC curve of the XGBoost classifier. The model correctly classified 70 non-PCOS cases and 30 PCOS cases, with only 3 false positives and 6 false negatives, resulting in 100 correct predictions out of 109 test samples and an accuracy of 91.74%. The low number of misclassifications indicates effective identification of both PCOS and non-PCOS patients. The ROC curve further demonstrates excellent discriminative performance, achieving an AUC of 0.95 and maintaining high true positive rates across different decision thresholds. These findings confirm the strong predictive capability of XGBoost and support its potential use as a decision support tool for PCOS screening and risk assessment.

Table 8 summarizes the performance of the XGBoost classifier on the test dataset. The model achieved an accuracy of 91.74%, with a precision of 90.91%, recall of 83.33%, and F1 score of 86.96%, indicating strong and balanced classification performance. The ROC–AUC value of 0.9540 further demonstrates excellent discriminative capability across different

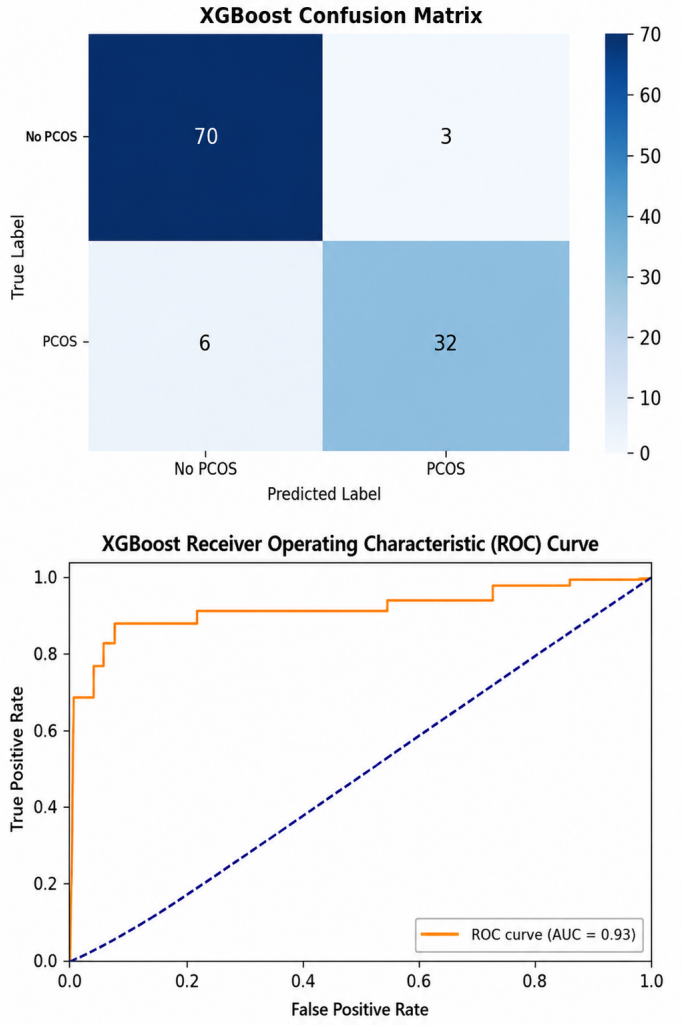

Figure 10. Confusion Matrix and ROC Curve of XGBoost Model for PCOS Classification.

Table 8. Performance Evaluation Metrics of the XGBoost Model for PCOS Classification.

Metric	Value
Accuracy	0.9174
Precision	0.9091
Recall	0.8333
F1 Score	0.8696
ROC–AUC	0.954

classification thresholds. Overall, these results confirm the effectiveness of XGBoost for PCOS classification and highlight its superior discriminative performance among the evaluated machine learning models.

Figure 11 illustrates the top ten features ranked according to their importance in the RF model. The right and left ovarian follicle counts emerged as the most influential predictors, followed by weight

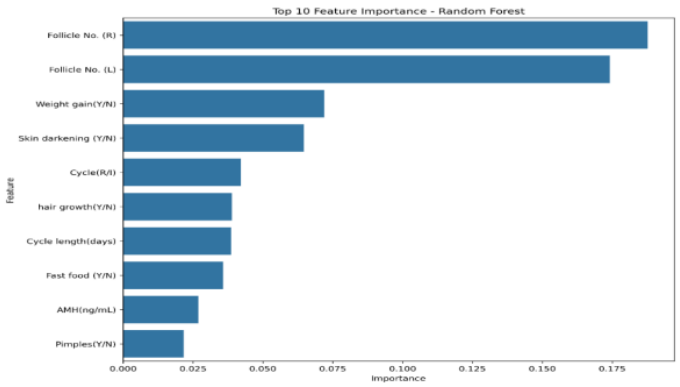

Figure 11. Top 10 Feature Importance Rankings Identified by the RF Model.

gain and skin darkening. Other important features included menstrual cycle regularity, hair growth, cycle length, fast food consumption, AMH concentration, and pimples. The identified predictors demonstrate that the RF model captured clinically relevant patterns associated with PCOS.

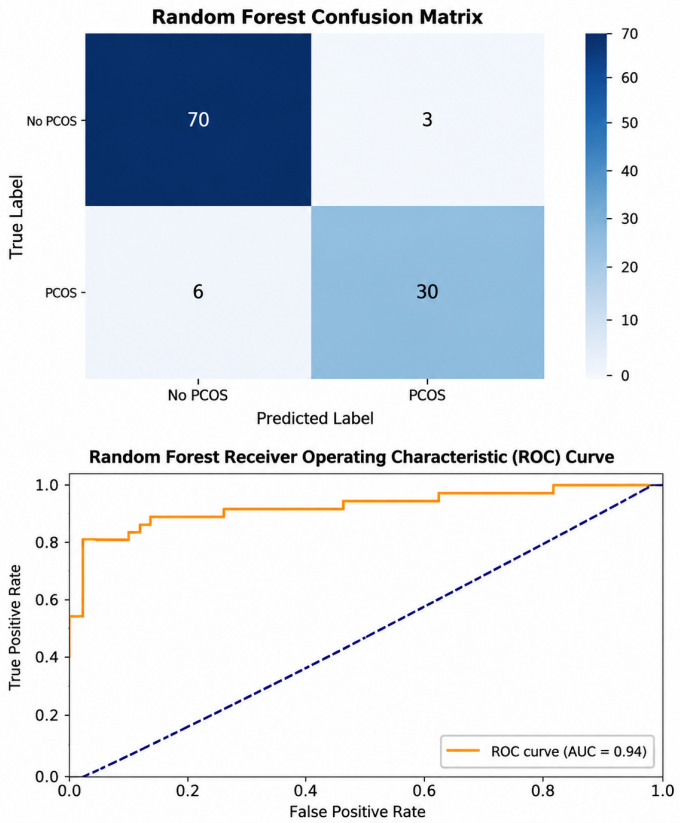

Figure 12. Confusion Matrix and ROC Curve of RF Model for PCOS Classification.

Figure 12 presents the confusion matrix and ROC curve of the RF model. The classifier correctly identified 70 non-PCOS cases and 30 PCOS cases, with only 3 false positives and 6 false negatives, resulting in 100 correct predictions out of 109 test samples. The low number of misclassifications indicates strong classification performance across both classes. The ROC curve further demonstrates excellent discriminative capability, achieving an AUC of approximately 0.94 and remaining well above the diagonal reference line across different classification thresholds. These findings confirm the effectiveness of the RF model for distinguishing between PCOS and non-PCOS patients.

Table 9. Performance Metrics of the RF Model for PCOS Classification.

Metric	Value
Accuracy	0.9174
Precision	0.9091
Recall	0.8333
F1-Score	0.8696
ROC-AUC	0.9433

Table 9 presents the performance of the RF model for PCOS classification. The model achieved an accuracy of 91.74% with a precision of 90.91%, indicating that most predicted PCOS cases were correctly identified. A recall of 83.33% demonstrates good sensitivity in detecting affected patients, while the F1-score of 86.96% reflects a balanced tradeoff between precision and recall. The ROC-AUC value of 0.9433 further confirms the model's strong discriminative ability and reliable classification performance.

SVM identifies an optimal decision boundary that maximizes the separation between PCOS and non PCOS classes, enabling effective classification in high dimensional feature spaces. Feature importance analysis was not conducted for the SVM model because SVM does not inherently provide interpretable feature importance scores. Additional explainability techniques would be required to obtain reliable feature contribution estimates and were not considered within the scope of this study.

Figure 13 shows the confusion matrix and ROC curve of the SVM model. The classifier correctly identified 72 non-PCOS cases and 21 PCOS cases, with only 1 false positive but 15 false negatives, resulting in 93 correct predictions out of 109 test samples. While the model demonstrated excellent specificity and effectively distinguished non-PCOS patients, a proportion of PCOS cases remained undetected. The ROC curve further highlights the model's strong discriminative

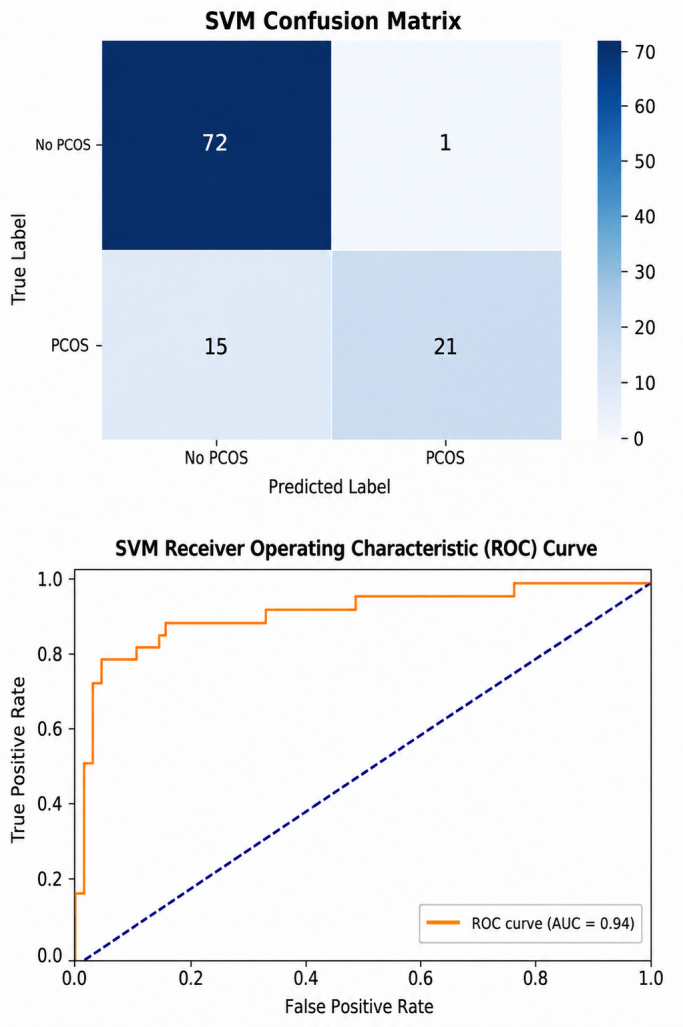

Figure 13. Confusion Matrix and ROC Curve of SVM Model for PCOS Classification.

capability, achieving an AUC of approximately 0.94 and remaining well above the diagonal reference line across different classification thresholds. These findings indicate that the SVM model provided good overall classification performance for PCOS prediction despite its comparatively lower sensitivity.

Table 10. Performance Metrics of the SVM Model for PCOS Classification.

Metric	Value
Accuracy	0.8532
Precision	0.9545
Recall	0.5833
F1-Score	0.7241
ROC-AUC	0.9387

Table 10 summarizes the performance of the SVM model. The classifier achieved an accuracy of 85.32%, precision of 95.45%, recall of 58.33%, and F1 score of 72.41%. While the high precision indicates that most predicted PCOS cases were correct, the lower recall suggests that some positive cases remained undetected. Despite this limitation, the ROC–AUC value of 0.9387 demonstrates strong discriminative capability, confirming the model's effectiveness in distinguishing between PCOS and non-PCOS patients.

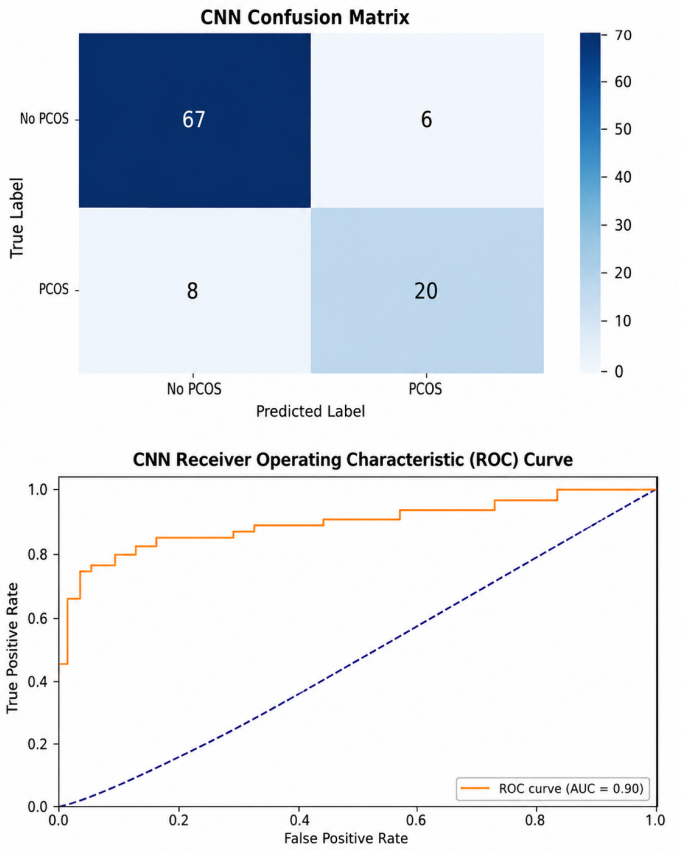

Figure 14. Confusion Matrix and ROC Curve of CNN-ResNet Model for PCOS Classification.

Figure 14 describes the confusion matrix and ROC curve of the CNN-ResNet model. The classifier correctly identified 67 non-PCOS cases and 28 PCOS cases, with 6 false positives and 8 false negatives, resulting in 95 correct predictions out of 109 test samples. Although a small number of misclassifications were observed, the model demonstrated balanced performance in distinguishing between PCOS and non-PCOS patients. The ROC curve further confirms its classification capability, achieving an AUC of approximately 0.90 and remaining above the diagonal reference line across most decision thresholds. These findings indicate that the CNN-ResNet model achieved good discriminative

performance while maintaining a favorable balance between sensitivity and specificity.

Table 11. Performance Metrics of the CNN-ResNet Model for PCOS Classification.

Metric	Value
Accuracy	0.8716
Precision	0.8235
Recall	0.7778
F1-Score	0.8000
ROC-AUC	0.9018

Table 11 summarizes the performance of the CNN-ResNet model for PCOS classification. The model achieved an accuracy of 87.16%, with a precision of 82.35%, recall of 77.78%, and F1 score of 80.00%, indicating balanced classification performance. The ROC–AUC value of 0.9018 further demonstrates good discriminative capability, confirming the model's effectiveness in distinguishing between PCOS and non-PCOS patients.

Figure 15 reveals the confusion matrix and ROC curve of the LSTM model. The classifier correctly identified 67 non-PCOS cases and 30 PCOS cases, with 6 false positives and 6 false negatives, resulting in 97 correct predictions out of 109 test samples. The equal number of false positive and false negative classifications indicates balanced performance in detecting both PCOS and non-PCOS cases. The ROC curve further demonstrates excellent discriminative capability, achieving an AUC of approximately 0.95 and remaining well above the diagonal reference line across different classification thresholds. These findings highlight the strong predictive performance of the LSTM model for PCOS classification.

Table 12. Performance Metrics of the LSTM Model for PCOS Classification.

Metric	Value
Accuracy	0.8899
Precision	0.8333
Recall	0.8333
F1-Score	0.8333
ROC-AUC	0.9452

Table 12 presents the performance metrics of the LSTM

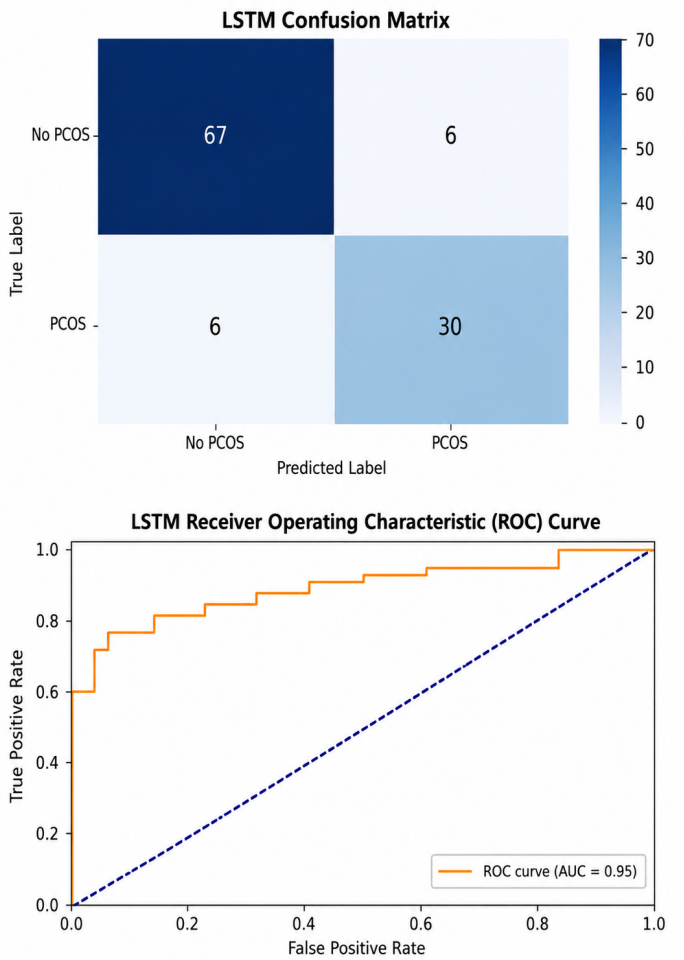

Figure 15. Confusion Matrix and ROC Curve of LSTM Model for PCOS Classification.

model for PCOS classification. The model achieved an accuracy of 88.99%, indicating that the majority of test instances were classified correctly. Precision and recall were both 83.33%, resulting in an F1 score of 83.33%, which reflects a balanced ability to identify PCOS cases while limiting incorrect classifications. Furthermore, the ROC-AUC value of 0.9452 demonstrates excellent discriminative capability, indicating that the LSTM model was effective in distinguishing between PCOS and non-PCOS patients.

4.1 Comparative Performance Analysis of Machine Learning and Deep Learning Models

Table 13 compares the performance of all evaluated models for PCOS prediction. KNN achieved the highest accuracy (92.7%), recall (91.7%), and F1 score (89.2%), indicating the strongest overall classification performance. XGBoost attained the highest ROC–AUC (95.4%), demonstrating superior discriminative capability, while SVM achieved the highest precision (95.5%) but the lowest recall (58.3%). Among the deep learning models, LSTM consistently

outperformed CNN-ResNet. Overall, KNN, XGBoost, RF, and LSTM demonstrated the most competitive results, with KNN providing the best balance between sensitivity and overall predictive effectiveness.

Table 13. Overall Performance Comparison of Machine Learning and Deep Learning Models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
SVM	85.3	95.5	58.3	72.4	93.9
KNN	92.7	86.8	91.7	89.2	93.8
XGBoost	91.7	90.9	83.3	87	95.4
RF	91.7	90.9	83.3	87	94.3
Decision Tree	90.8	90.6	80.6	85.3	93.1
LSTM	89	83.3	83.3	83.3	94.5
CNN-ResNet	87.2	82.4	77.8	80	90.2

4.2 Confusion Matrix Analysis

To further evaluate model performance, confusion matrix outcomes were analyzed for all classifiers. Table 14 summarizes the confusion matrix results for all evaluated models, including the numbers of true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP) for each classifier. KNN correctly identified the highest number of PCOS cases (TP = 33), followed by RF, XGBoost, and LSTM (TP = 30 each). SVM achieved the lowest number of false positives (FP = 1) but also missed the greatest number of PCOS cases (FN = 15). CNN-ResNet and LSTM produced the highest false positive counts (FP = 6), whereas KNN achieved the lowest false negative count (FN = 3), indicating strong sensitivity for PCOS detection. Overall, KNN and XGBoost demonstrated a favorable balance between correctly identifying PCOS cases and minimizing misclassifications.

Table 14. Confusion Matrix Summary of Evaluated Models.

Model	TN	FP	FN	TP
Decision Tree	70	3	7	29
KNN	68	5	3	33
RF	70	3	6	30
XGBoost	70	3	6	30
SVM	72	1	15	21
CNN-ResNet	67	6	8	28
LSTM	67	6	6	30

4.3 Comparison with Previous Studies

To better position the proposed framework within the existing body of literature, a comparative analysis was conducted against representative PCOS prediction studies that utilized the same publicly available Kaggle clinical dataset. Table 15 summarizes the datasets, predictive models, and best reported classification accuracies of previous studies alongside the results obtained in the present work. This comparison provides a broader perspective on the proposed framework's relative performance while highlighting its distinctive methodological characteristics.

Table 15 compares the proposed framework with representative studies that employed the same publicly available PCOS dataset. Previous studies have reported classification accuracies ranging from 86.76% to 98.89%, with several investigations achieving higher predictive performance through ensemble learning, feature selection, or specialized deep learning architectures. The KNN model developed in the present study achieved a competitive accuracy of 92.66%, while XGBoost produced the highest ROC AUC (95.4%), indicating excellent discriminative capability. Although the proposed framework does not achieve the highest reported classification accuracy, its primary contribution lies in establishing a standardized and reproducible benchmarking framework that evaluates five machine learning algorithms and two deep learning architectures under identical preprocessing, hyperparameter optimization, and evaluation settings. This unified experimental design enables a fair comparison of model performance, which is often difficult to achieve across independent studies because of differences in preprocessing strategies, validation protocols, feature engineering techniques, and evaluation methodologies. Furthermore, the incorporation of SHAP based explainability and feature importance analysis improves model transparency by identifying clinically meaningful predictors, thereby increasing the potential applicability of the framework in clinical decision support.

4.4 Practical Implications

The proposed framework has potential applications as a clinical decision-support tool for early PCOS screening in gynecology clinics and primary healthcare settings. By utilizing routinely collected demographic, hormonal, lifestyle, and ultrasound-related variables, the framework can assist clinicians in identifying women at high risk of PCOS before comprehensive

Table 15. Comparison of the proposed framework with representative PCOS prediction studies.

Study	Dataset	Best Model	Best Reported Accuracy (%)	Distinguishing Characteristic
[20]	Kaggle PCOS (541 patients)	HRFLR	91.01	Compared multiple ML classifiers with hybrid RF and LR approach
[29]	Kaggle PCOS (541 patients)	Stacking Ensemble	98.87	Ensemble based classification
[32]	Kaggle PCOS	Feature Selection + ML	98.89	Feature selection driven prediction
[33]	Kaggle PCOS	Quantum LSTM	94.5	Quantum inspired deep learning
[34]	Kaggle PCOS (541 patients)	SVM	86.76	Early prediction using selected indicators
[31]	Kaggle PCOS (541 patients)	Stacking Ensemble	94	Traditional and ensemble ML comparison with advanced feature selection
This Study	Kaggle PCOS (541 patients)	KNN	92.66	Standardized benchmarking of five ML and two DL models with SHAP explainability

diagnostic evaluation. Rather than replacing clinical judgment, the proposed models are intended to provide objective risk assessment that supports timely referral, additional laboratory investigations, and early intervention. Such decision-support systems may be particularly beneficial in resource-constrained healthcare environments where access to specialist expertise is limited.

From an implementation perspective, the proposed framework relies on structured clinical variables that are routinely collected during standard PCOS assessment and therefore does not require specialized hardware or expensive imaging technologies. The best-performing machine learning models also require relatively modest computational resources, making deployment feasible within hospital information systems or cloud-based clinical decision-support platforms. Nevertheless, prospective clinical validation and integration with electronic health record systems remain necessary before routine clinical adoption.

5 Study Limitations and Future Work

The findings of this study should be interpreted in light of several limitations. First, the experimental analysis was conducted using a single publicly available clinical dataset. Although the dataset contains a diverse set of demographic, hormonal, lifestyle, and ultrasound related attributes, it may not fully represent the broader population of women affected by PCOS. Second, the dataset consisted of 541 patient records, which represents a moderate sample size for machine learning based prediction studies. Larger cohorts may provide additional variability and improve the generalizability of predictive models.

Another limitation is the absence of external validation. Model performance was evaluated using a stratified train-test split from the same dataset, and independent validation using data collected from different institutions was not performed. Consequently, the generalization capability of the developed models across different clinical settings remains to be established. In addition, the deep learning models were evaluated solely using structured tabular clinical data. Other clinically relevant data modalities, such as ultrasound images, laboratory reports, and electronic health records, were not incorporated into the current framework.

Future research may address these limitations through the use of larger multicenter datasets collected from diverse populations and healthcare environments. External validation on independent cohorts should also be conducted to assess model stability and generalizability. Furthermore, multimodal learning frameworks integrating clinical, hormonal, and imaging data may provide a more comprehensive representation of PCOS characteristics and improve predictive performance. Additional efforts may also focus on the integration of explainable artificial intelligence techniques to enhance transparency, facilitate clinical interpretation, and support greater confidence in model assisted decision making.

6 Conclusion

This study presented a standardized framework for automated PCOS prediction using structured clinical data by evaluating five machine learning algorithms and two deep learning architectures under identical preprocessing, hyperparameter optimization, and evaluation conditions. Among the evaluated

models, K-Nearest Neighbors achieved the highest classification accuracy (92.66%), recall (91.67%), and F1-score (89.19%), while XGBoost achieved the highest ROC-AUC (95.40%). SHAP-based explainability identified ovarian follicle counts, menstrual cycle characteristics, hair growth, weight gain, skin darkening, and anti-Müllerian hormone (AMH) levels as the most influential predictors, providing clinically meaningful insights for PCOS diagnosis. Unlike previous studies that primarily emphasized predictive performance, this study provides a unified, reproducible, and interpretable benchmarking framework for objectively comparing machine learning and deep learning models. By combining standardized evaluation with explainable artificial intelligence, the proposed framework improves the transparency and reliability of automated PCOS prediction and offers a useful foundation for clinical decision-support systems. Future research should validate the framework using larger multicenter datasets and multimodal clinical data to improve its generalizability and real-world applicability.

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 utilized a publicly available anonymized clinical dataset from Kaggle. The original data collection was conducted in accordance with relevant ethical guidelines and institutional review board approvals. No additional ethical approval was required as the study involved secondary analysis of de-identified data.

References

[1] Nasim, S., Almutairi, M. S., Munir, K., Raza, A., & Younas, F. (2022). A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics. *IEEE Access, 10*, 97610-97624. [CrossRef]

[2] Raj, A., Joshi, P., Devi, S., & Rawat, S. (2023). Comparative analysis of machine learning algorithms to diagnose polycystic ovary syndrome. In *Combating Women's Health Issues with Machine Learning* (pp. 136-153). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003378556-9

[3] Mridul, A. H., Ahsan, N., Alam, S. S., Afrose, S., & Sultana, Z. (2024). Polycystic ovary syndrome (pcos) disease prediction using traditional machine learning and deep learning algorithms. *International Journal of Computer Information Systems and Industrial Management Applications, 16*(3), 25-25. https://cspub-ijcisim.org/index.php/ijcisim/article/view/712

[4] Govindharajan, G., Subramanian, S., Doraipandian, M., & Rajendran, S. (2026). Innovative AI-driven models for predicting polycystic ovarian syndrome: An extensive review of machine learning and deep learning frameworks. *Archives of Computational Methods in Engineering, 33*(2), 2115-2140. [CrossRef]

[5] Ghaderzadeh, M., Garavand, A., & Salehnasab, C. (2025). Artificial intelligence in polycystic ovary syndrome: A systematic review of diagnostic and predictive applications. *BMC Medical Informatics and Decision Making, 25*(1), 427. [CrossRef]

[6] Uddin, K. M. M., Chowdhury, A., Druvo, M. M. R., Jaima, M. T., Bhuiyan, M. T. A., & Islam, M. M. (2025). Improving Diagnostic Accuracy for PCOS: A Hybrid Machine Learning Architecture with Feature Selection, Data Balancing, and Explainable AI Techniques. *Results in Control and Optimization*, 100647. [CrossRef]

[7] Dhanka, S., Kumar, A., Sharma, A., Vundavilli, H., Maini, S., & Rajasekar, E. (2026). Advances in machine learning and deep learning for hormonal disorder diagnosis: an exhaustive review on PCOS, thyroid, and optimization techniques. *Archives of Computational Methods in Engineering, 33*(2), 2785-2829. [CrossRef]

[8] Ahmed, S., Rahman, M. S., Jahan, I., Kaiser, M. S., Hosen, A. S., Ghimire, D., & Kim, S. H. (2023). A review on the detection techniques of polycystic ovary syndrome using machine learning. *IEEE Access, 11*, 86522-86543. [CrossRef]

[9] Ahmad, R., Maghrabi, L. A., Khaja, I. A., Maghrabi, L. A., & Ahmad, M. (2024). SMOTE-based automated PCOS prediction using lightweight deep learning models. *Diagnostics, 14*(19), 2225. [CrossRef]

[10] Sen, P., Alina, T. A., Arefin, M. S., Reza, A. W., & Hussain, M. M. (2025, July). Machine Learning-Based Predictive Modeling for Polycystic Ovary Syndrome Diagnosis: A Multi-Algorithm Comparative Analysis. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE. [CrossRef]

[11] Zad, Z., Jiang, V. S., Wolf, A. T., Wang, T., Cheng, J. J., Paschalidis, I. C., & Mahalingaiah, S. (2024).

Predicting polycystic ovary syndrome with machine learning algorithms from electronic health records. *Frontiers in endocrinology, 15*, 1298628. [CrossRef]

[12] Faris, N. N., & Miften, F. S. (2022). An intelligence model for detection of PCOS based on k-means coupled with LS-SVM. *Concurrency and computation: practice and experience, 34*(21), e7139. [CrossRef]

[13] Devi, V. S., Nivedha, S., Harisha, V., Mol, D., & Janaranjini, J. R. (2026). Enhanced prediction of PCOS and PCOD using deep learning for early diagnosis and clinical risk stratification. *International Journal of Advanced Research in Computer Science & Technology (IJARCST), 9*(3), 783-793. [CrossRef]

[14] Ghadekar, P., Tekade, S., Sakharwade, D., Tripathi, A., Tiwadi, S., & Zanzane, S. (2024, July). Multimodal PCOS detection: Combining XG boost for images with zero shot learning for textual data. In *2024 Asia Pacific Conference on Innovation in Technology (APCIT)* (pp. 1-8). IEEE. [CrossRef]

[15] Hdaib, D., Almajali, N., Alquran, H., Mustafa, W. A., Al-Azzawi, W., & Alkhayyat, A. (2022, May). Detection of polycystic ovary syndrome (PCOS) using machine learning algorithms. In *2022 5th international conference on engineering technology and its applications (IICETA)* (pp. 532-536). IEEE. [CrossRef]

[16] Agirsoy, M., & Oehlschlaeger, M. A. (2025). A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features. *Scientific Reports, 15*(1), 33638. [CrossRef]

[17] Yadav, N., Ranjith, K. A., & Pande, S. D. (2024). Revolutionizing PCOS diagnosis: Deep learning image classification with LSTM-integrated artificial neural networks. In *Artificial intelligence and information technologies* (pp. 381-386). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9781032700502-61

[18] Pulluparambil, S. J., & B, S. B. (2025). Detection and prediction of polycystic ovary syndrome using attention-based CNN-RNN classification model. *International Journal of Advanced Computer Science and Applications, 16*(2). [CrossRef]

[19] Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications, 177*, 114905. [CrossRef]

[20] 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]

[21] 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]

[22] 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]

[23] Umaa, M. G., & Uma, M. P. (2024). SmartScanPCOS: A feature-driven approach to cutting-edge prediction of Polycystic Ovary Syndrome using Machine Learning and Explainable Artificial Intelligence. *Heliyon, 10*(20). [CrossRef]

[24] Suha, S. A., & Islam, M. N. (2022). An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. *Scientific Reports, 12*(1), 17123. [CrossRef]

[25] Abouhawwash, M., Sridevi, S., Sundararajan, S. C. M., Pachlor, R., Karim, F. K., & Khafaga, D. S. (2023). Automatic Diagnosis of Polycystic Ovarian Syndrome Using Wrapper Methodology with Deep Learning Techniques. *Computer Systems Science and Engineering, 47*(1), 239-253. [CrossRef]

[26] Khanna, V. V., Chadaga, K., Sampathila, N., Prabhu, S., Bhandage, V., & Hegde, G. K. (2023). A distinctive explainable machine learning framework for detection of polycystic ovary syndrome. *Applied System Innovation, 6*(2), 32. [CrossRef]

[27] Kottarathil, P. (2020). *Polycystic ovary syndrome (PCOS)* [Data set]. Kaggle. https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-PCOS

[28] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing, 151*, 107398. [CrossRef]

[29] Elmannai, H., El-Rashidy, N., Mashal, I., Alohali, M. A., Farag, S., El-Sappagh, S., & Saleh, H. (2023). Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. *Diagnostics, 13*(8), 1506. [CrossRef]

[30] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.

[31] Mohi Uddin, M., Bhuiyan, M. T. A., Rahman, M. M., Islam, M. M., & Uddin, M. A. (2025). Early PCOS detection: A comparative analysis of traditional and ensemble machine learning models with advanced feature selection. *Engineering Reports, 7*(2), e70008. [CrossRef]

[32] 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]

[33] Sarath, T., & Brindha, K. (2026). AI-quantum framework for accurate infertility risk classification in PCOS patients using EHR data. *Frontiers in Artificial Intelligence, 9*, 1776461. [CrossRef]

[34] Xu, Z. (2025). *Early detection of polycystic ovary syndrome: A systematic evaluation of machine learning models using clinical parameters*. Research Square. [CrossRef]

Sana Rubab is a Post RN student at the Department of Nursing, Quaid-e-Azam College of Nursing and Allied Health Sciences, Sahiwal, Pakistan. She is actively engaged in academic and clinical research with a focus on evidence-based nursing practice, patient care, and healthcare innovation. Her research interests include public health, nursing education, clinical decision-making, and the application of emerging technologies to improve healthcare outcomes. She is committed to advancing nursing knowledge through interdisciplinary collaboration and high-quality research. (Email: sanahrubab99@gmail.com)

Musarrat Shaheen is a nursing professional and researcher with a strong commitment to advancing healthcare quality and patient outcomes. She has extensive experience in clinical practice, nursing education, and evidence-based healthcare research. Her research interests include public health, nursing management, patient safety, healthcare innovation, and the integration of emerging technologies in clinical settings. She actively contributes to academic and research initiatives aimed at improving nursing practices and promoting excellence in healthcare delivery. (Email: musarratshaheen847@gmail.com)

Zohra Tabassum is a dedicated nursing professional and researcher committed to promoting excellence in healthcare practice and patient-centered care. She is actively involved in academic and clinical research, with interests in evidence-based nursing, public health, healthcare management, and nursing education. Her work focuses on improving healthcare outcomes through innovative approaches, interdisciplinary collaboration, and the application of research findings in clinical settings. She strives to contribute to the advancement of nursing knowledge and the delivery of high-quality healthcare services. (Email: Zohratabassum475@gmail.com)

Samreen Liaqat is a dedicated nursing professional and researcher with a strong interest in advancing evidence-based healthcare practices. She is actively engaged in academic and clinical research, focusing on nursing education, patient safety, public health, and quality improvement in healthcare services. Her research interests include healthcare innovation, clinical decision-making, and the integration of emerging technologies to enhance patient outcomes. Through interdisciplinary collaboration and continuous professional development, she aims to contribute to the advancement of nursing knowledge and the improvement of healthcare delivery. (Email: simrenmughal041@gmail.com)

Memoona Saleem is a dedicated nursing professional and researcher committed to improving patient care and advancing evidence-based nursing practice. She actively participates in academic and clinical research with interests in public health, nursing education, healthcare management, and quality improvement initiatives. Her work focuses on enhancing healthcare outcomes through innovative approaches, interdisciplinary collaboration, and the effective application of research in clinical settings. She is passionate about contributing to the development of nursing knowledge and promoting excellence in healthcare delivery. (Email: moonsaleem444@gmail.com)

Aamir Ali is a dedicated researcher in artificial intelligence, healthcare analytics, and IoT environments. His expertise includes machine and deep learning applications in medical diagnosis, speech recognition, and facial emotion recognition. He has also worked on IoT-focused projects such as malware classification, anomaly detection, and cyberattack prediction. Aamir is passionate about developing AI-driven solutions for early disease detection and improving patient care. He actively contributes to interdisciplinary research across clinical and image data domains. (Email: amirali4436823@gmail.com)