IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Explainable Artificial Intelligence for Prediction and Interpretation of Polycystic Ovary Syndrome (PCOS)

Meenu Rana
M.Tech Scholar, Computer Science and Engineering
Tula's Institute, Dehradun, Uttarakhand, India
meenu.rana@tulas.edu.in

Abstract: *Polycystic Ovary Syndrome (PCOS) is one of the most prevalent endocrine and metabolic disorders affecting women of reproductive age, yet it remains substantially under-diagnosed because of its heterogeneous and overlapping clinical presentation. Although machine learning (ML) models have demonstrated strong predictive performance for PCOS screening, their adoption in clinical practice is limited by their "black-box" nature, which prevents clinicians from understanding why a particular prediction is made. This study proposes an Explainable Artificial Intelligence (XAI) framework that couples accurate PCOS prediction with transparent, human-interpretable explanations. Six supervised classifiers — Logistic Regression, Decision Tree, K-Nearest Neighbours, Support Vector Machine, XGBoost and Random Forest — were trained and evaluated on a publicly available clinical dataset of 541 patients comprising 41 clinical, biochemical and ultrasonographic attributes. After data cleaning, correlation-based and importance-based feature selection, SMOTE class balancing and standardisation, the Random Forest classifier achieved the best performance with an accuracy of 91.7%, an F1-score of 89.4% and an area under the ROC curve (AUC) of 0.96. To interpret the predictions, three complementary XAI techniques were integrated: SHAP (SHapley Additive exPlanations) for global and local feature attribution, LIME (Local Interpretable Model-agnostic Explanations) for instance-level explanation, and Diverse Counterfactual Explanations (DiCE) for actionable "what-if" recourse. The explanations consistently identified follicle number (right and left ovary), skin darkening, excessive hair growth, weight gain and menstrual-cycle irregularity as the dominant predictive factors, in agreement with the established Rotterdam diagnostic criteria. The results demonstrate that combining a high-performing ensemble model with multi-method explainability yields a trustworthy and clinically meaningful decision-support tool for early PCOS detection.*

Keywords: Polycystic Ovary Syndrome (PCOS); Explainable Artificial Intelligence (XAI); Machine Learning; Random Forest; SHAP; LIME; Counterfactual Explanations; Clinical Decision Support; Interpretability

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a complex, multifactorial endocrine disorder that affects an estimated 8–13% of women of reproductive age worldwide, with a substantial proportion of cases remaining undiagnosed. The syndrome is characterised by a combination of hyperandrogenism, ovulatory dysfunction and polycystic ovarian morphology, and is commonly associated with clinical manifestations such as irregular menstrual cycles, hirsutism, acne, weight gain, skin darkening and infertility. Beyond its reproductive consequences, PCOS carries significant long-term metabolic risks, including insulin resistance, type-2 diabetes, dyslipidaemia, cardiovascular disease and endometrial complications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Because its symptoms overlap with several other conditions and vary widely between individuals, PCOS is frequently mis-diagnosed or diagnosed late, often only after a woman experiences difficulty conceiving.

The clinical diagnosis of PCOS is most commonly based on the Rotterdam consensus criteria, which require the presence of at least two of the following three features: oligo- or anovulation, clinical or biochemical signs of hyperandrogenism, and polycystic ovaries on ultrasound. While these criteria are well established, their application requires specialist expertise, multiple laboratory investigations and ultrasonography, all of which may be inaccessible or expensive in resource-constrained settings. There is therefore strong motivation to develop automated, data-driven screening tools that can support early identification of at-risk women using routinely collected clinical data.

Machine learning has emerged as a powerful approach for medical diagnosis and risk prediction. A growing body of literature has reported high predictive accuracy for PCOS using classifiers such as Random Forest, Support Vector Machine, gradient-boosted trees and neural networks. However, the most accurate models are typically opaque: they do not reveal the reasoning behind an individual prediction. In a high-stakes clinical context, this lack of transparency is a serious barrier to adoption. Clinicians are understandably reluctant to act on a recommendation they cannot scrutinise, patients have a right to understand decisions that affect them, and emerging regulations increasingly demand a measure of algorithmic accountability. Consequently, predictive accuracy alone is insufficient; the model must also be interpretable.

Explainable Artificial Intelligence (XAI) addresses this gap by providing techniques that make the behaviour of complex models comprehensible to humans. Global explanations describe the overall logic of the model — for example, which features are most influential across the whole population — whereas local explanations justify a single prediction for a specific patient. Counterfactual explanations go one step further by indicating the minimal changes to a patient's features that would alter the predicted outcome, thereby offering actionable insight. Together, these methods can transform a black-box predictor into a transparent and trustworthy clinical decision-support system.

This paper presents an end-to-end framework that integrates accurate PCOS prediction with comprehensive explainability. The principal contributions of this work are summarised below.

A systematic comparison of six supervised machine-learning classifiers for PCOS prediction on a publicly available clinical dataset, with rigorous pre-processing, feature selection and class balancing.

Selection and tuning of a best-performing model (Random Forest) that achieves a strong balance of accuracy, recall and discrimination, suitable for screening applications.

Integration of three complementary XAI techniques — SHAP, LIME and Diverse Counterfactual Explanations (DiCE) — to provide global, local and actionable interpretations of the predictions.

A clinically grounded discussion that maps the data-driven explanations onto the established Rotterdam diagnostic criteria, demonstrating the medical plausibility of the model.

The remainder of the paper is organised as follows. Section 2 reviews related work on machine learning and explainability for PCOS. Section 3 describes the proposed research methodology, including the dataset, pre-processing pipeline, models and XAI techniques, accompanied by a workflow diagram. Section 4 presents and discusses the experimental results, supported by tables and figures. Section 5 concludes the paper and outlines directions for future work.

II. RELATED WORKS

Research on the automated detection of PCOS has expanded rapidly, driven largely by the public availability of a clinical dataset collected from ten hospitals in Kerala, India, comprising the records of 541 women with more than forty clinical, hormonal and ultrasonographic attributes. This section reviews representative studies in two streams: first, work focused on predictive performance, and second, more recent work that introduces explainability.

2.1 Machine-Learning Models for PCOS Prediction

Several investigators have applied classical and ensemble classifiers to the PCOS screening problem. Bharati and colleagues evaluated Random Forest, Logistic Regression, gradient boosting and a hybrid Random Forest–Logistic

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Regression (RFLR) model with univariate feature selection, reporting that the hybrid RFLR configuration achieved the best accuracy of approximately 91%. Mehr and Polat examined the value of feature-selection strategies (filter, wrapper and embedded methods) combined with ensemble classifiers such as AdaBoost, Random Forest, Extra Trees and the Multilayer Perceptron, observing that careful feature reduction can push Random-Forest accuracy substantially higher. Other studies reported that Multilayer Perceptron and Support Vector Machine models with radial basis function kernels reached accuracies around 93%, and that ensemble approaches frequently outperformed single classifiers.

Feature-selection and class-imbalance handling have repeatedly been shown to be decisive. Because the public dataset is imbalanced — roughly two non-PCOS cases for every positive case — techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) and correlation-based feature selection are commonly applied to prevent the classifier from being biased toward the majority class. Across this literature, follicle counts in the left and right ovaries, follicle size, menstrual-cycle regularity, hair-growth score, skin darkening and anti-Müllerian hormone (AMH) levels recur as the most informative predictors, which is consistent with the physiological basis of the disorder.

More recent work has explored nature-inspired hyper-parameter optimisation and stacking ensembles. Optimised ensemble models tuned with walrus, cuckoo-search and random-search algorithms have reported accuracies near 93% with AUC values above 0.93, while multi-stack ensembles have claimed accuracies as high as 98%. Although these figures are impressive, several of the very high reported accuracies are likely influenced by data leakage, optimistic cross-validation protocols or limited external validation, underscoring the need for cautious, reproducible evaluation.

2.2 Explainable AI for PCOS and Clinical Decision Support

The recognition that accuracy alone is insufficient has motivated a shift toward explainable models. SHAP, introduced by Lundberg and Lee, provides a unified, game-theoretic framework that assigns each feature an additive contribution to an individual prediction, while also supporting global importance summaries. LIME, proposed by Ribeiro and colleagues, explains any classifier locally by fitting a sparse, interpretable surrogate model around the instance of interest. Both techniques have been widely adopted in healthcare because they are model-agnostic and produce intuitive visualisations.

Within the PCOS domain specifically, recent studies have begun to pair predictive models with explanation layers. An explainable framework published in 2023 combined heterogeneous machine-learning and deep-learning classifiers on the Kerala dataset and used SHAP and LIME to visualise the clinical reasoning behind positive predictions, achieving accuracy and sensitivity near 99% and 100% respectively under their evaluation protocol. Another study coupled optimised feature selection with global and local explainability to improve clinician trust. Deep-learning approaches operating on ultrasound images have similarly used LIME and saliency maps to highlight the image regions driving a diagnosis. These works confirm that explanations derived from data tend to align with clinically recognised PCOS markers.

Counterfactual explanation is a complementary and increasingly important XAI paradigm. Rather than ranking feature importance, a counterfactual answers the question “what would have to change for the decision to be different?” Wachter and colleagues introduced counterfactual explanations as a means of providing recourse without opening the black box, and Mothilal and colleagues operationalised the idea in the DiCE framework, which generates a diverse set of minimally perturbed, feasible instances that flip the prediction. Counterfactuals are particularly natural in medicine because clinicians already reason in this hypothetical manner — for instance, considering how a patient’s risk would change if a measurable factor were different. Despite their promise, counterfactual explanations have been applied only sparingly to PCOS, and few studies combine global attribution (SHAP), local surrogate explanation (LIME) and actionable counterfactuals (DiCE) within a single, coherent framework.

In summary, prior research establishes that machine learning can predict PCOS accurately and that individual XAI methods can illuminate model behaviour. The present study builds on this foundation by integrating three complementary explanation techniques on top of a rigorously validated ensemble model, thereby providing global, local and actionable interpretability in one unified pipeline.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

III. RESEARCH METHODOLOGY

The proposed methodology follows a structured pipeline that transforms raw clinical records into accurate, explainable predictions. The pipeline comprises seven principal stages: data acquisition, pre-processing, exploratory analysis and feature selection, class balancing and scaling, model training and evaluation, best-model selection, and the explainability layer. The complete workflow is depicted in Figure 1.

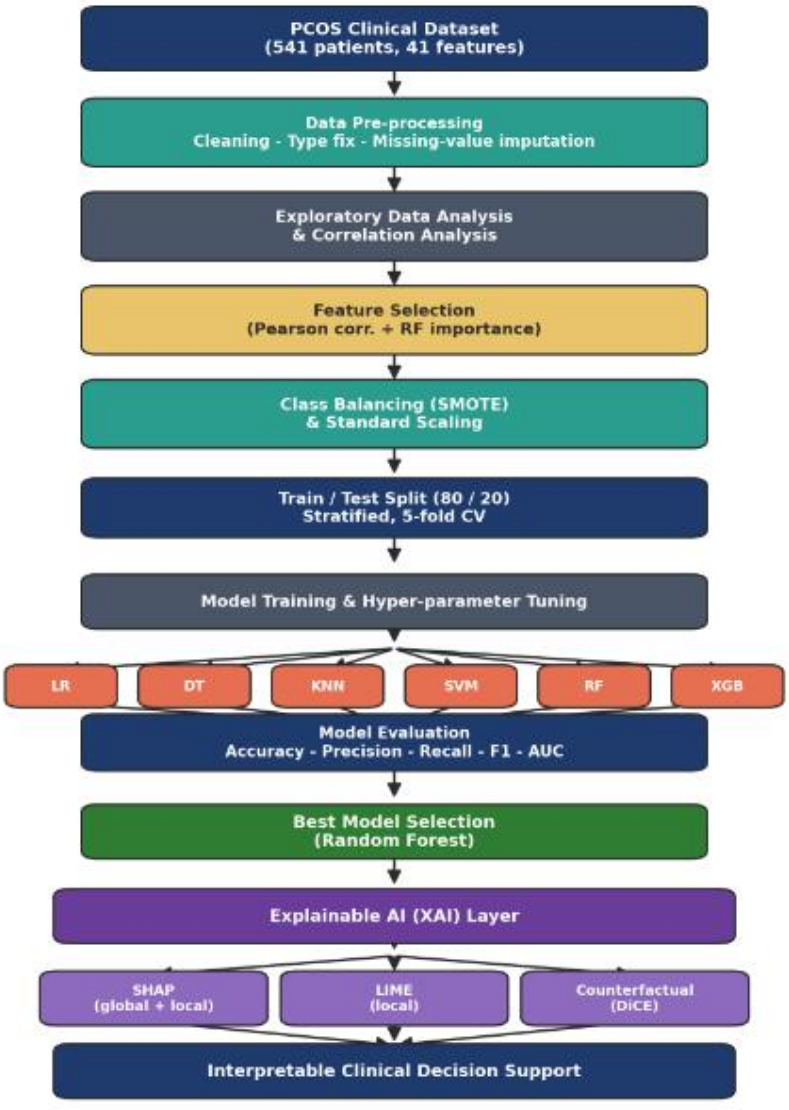

Figure 1. Workflow of the proposed explainable PCOS prediction framework.

As shown in Figure 1, the clinical dataset first passes through cleaning and pre-processing, after which exploratory data analysis and correlation analysis guide feature selection. The selected features are balanced using SMOTE and standardised before the data are split into training and testing partitions. Six classifiers are then trained and compared; the best-performing model is retained and passed to the explainability layer, where SHAP, LIME and counterfactual analysis convert the predictions into human-interpretable insights that support clinical decision-making. The following subsections describe each stage in detail.

DOI: 10.48175/IJARSCT-37114

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

3.1 Dataset Description

The study uses a publicly available PCOS dataset originally compiled from ten hospitals in Kerala, India, and widely distributed through the Kaggle repository. The dataset contains the records of 541 women of reproductive age, each described by 41 attributes spanning three categories: (i) demographic and anthropometric variables such as age, weight, height, body-mass index (BMI), waist-to-hip ratio and blood pressure; (ii) biochemical and hormonal markers including follicle-stimulating hormone (FSH), luteinising hormone (LH), the FSH/LH ratio, anti-Müllerian hormone (AMH), prolactin, thyroid-stimulating hormone (TSH) and blood glucose; and (iii) clinical and ultrasonographic features such as menstrual-cycle regularity, cycle length, the number and average size of follicles in each ovary, endometrial thickness, and symptom indicators for hair growth, skin darkening, hair loss, pimples, weight gain and fast-food consumption. The binary target variable indicates whether the patient was clinically diagnosed with PCOS. Table 1 summarises representative attributes.

Category	Representative Attributes	Type
Demographic / Anthropometric	Age, Weight, Height, BMI, Waist-Hip ratio, BP	Numeric
Hormonal / Biochemical	FSH, LH, FSH/LH ratio, AMH, Prolactin, TSH, Glucose	Numeric
Menstrual	Cycle (Regular/Irregular), Cycle length (days)	Categorical / Numeric
Ultrasonographic	Follicle No. (L/R), Avg. follicle size, Endometrium (mm)	Numeric
Symptomatic (Y/N)	Hair growth, Skin darkening, Hair loss, Pimples, Weight gain	Binary
Lifestyle	Fast food, Regular exercise	Binary
Target	PCOS diagnosis (Yes / No)	Binary

Table 1. Summary of attribute categories in the PCOS dataset (541 patients, 41 features).

3.2 Data Pre-processing

Real-world clinical data are rarely analysis-ready, and several cleaning steps were therefore performed. Columns that were erroneously stored as text (for example, certain hormonal measurements containing stray characters) were converted to numeric types. A small number of missing values were imputed using the median for numeric attributes, which is robust to outliers, and the mode for categorical attributes. Non-informative identifier columns, such as patient serial numbers, were removed. Outliers that were physiologically implausible were inspected and corrected or capped where appropriate. Categorical symptom indicators were already encoded as binary values, and the menstrual-cycle attribute was mapped to a binary regular/irregular representation.

3.3 Exploratory Analysis and Feature Selection

Exploratory data analysis was conducted to understand the distribution of variables and their relationship with the target. The dataset is moderately imbalanced, containing roughly 364 non-PCOS cases and 177 PCOS cases. Pearson correlation analysis between the features and the target identified the most strongly associated variables and revealed clusters of inter-correlated symptoms. To reduce dimensionality and noise, a two-stage feature-selection strategy was adopted: features were first ranked by their absolute correlation with the target, and this ranking was cross-checked against the impurity-based importance scores produced by a preliminary Random-Forest model. Features that were weakly associated by both criteria were discarded, yielding a compact and informative feature subset dominated by follicle counts, follicle size, menstrual-cycle regularity, and the principal hyperandrogenic symptoms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

3.4 Class Balancing and Scaling

Because the minority (PCOS-positive) class is clinically the more important one, class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic minority-class examples by interpolating between existing minority instances and their nearest neighbours, producing a balanced training set without simply duplicating records. Crucially, SMOTE was applied only to the training partition, after the train–test split, to avoid information leakage into the test set. Numeric features were then standardised to zero mean and unit variance, which benefits distance- and margin-based learners such as K-Nearest Neighbours and the Support Vector Machine.

3.5 Classification Models

Six supervised classifiers spanning linear, instance-based, kernel-based and tree-ensemble families were trained so that the strengths of different inductive biases could be compared:

Logistic Regression — a linear baseline that models the log-odds of PCOS as a weighted sum of the features; highly interpretable but limited to linear decision boundaries.

Decision Tree — a recursive partitioning model that is naturally interpretable but prone to over-fitting when grown without constraint.

K-Nearest Neighbours (KNN) — a non-parametric instance-based classifier that labels a patient according to the majority class among its closest neighbours in feature space.

Support Vector Machine (SVM) — a margin-maximising classifier; a radial basis function kernel was used to capture non-linear relationships.

XGBoost — a regularised gradient-boosted tree ensemble that builds additive trees to minimise a differentiable loss, known for strong performance on tabular data.

Random Forest — a bagged ensemble of decorrelated decision trees whose predictions are aggregated by majority vote; robust to noise and well suited to SHAP-based interpretation.

Hyper-parameters were tuned using grid search with five-fold stratified cross-validation on the training set. For Random Forest, the principal tuned parameters were the number of estimators, maximum tree depth, the minimum samples required to split a node, and the number of features considered at each split. The data were partitioned into 80% for training and 20% for testing using stratified sampling to preserve the class ratio in both partitions.

3.6 Evaluation Metrics

Model performance was assessed using accuracy, precision, recall (sensitivity), F1-score and the area under the receiver-operating-characteristic curve (AUC). Because the clinical cost of a missed PCOS case (a false negative) is high, recall and AUC were given particular weight when selecting the final model. These metrics are defined in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as follows: accuracy is the proportion of correct predictions, (TP+TN)/(TP+TN+FP+FN); precision is TP/(TP+FP); recall is TP/(TP+FN); and the F1-score is the harmonic mean of precision and recall. The AUC summarises the trade-off between the true-positive rate and the false-positive rate across all decision thresholds.

3.7 Explainable AI Layer

Once the best-performing model had been identified, three complementary explanation techniques were applied to interpret its predictions at different levels of granularity.

SHAP (SHapley Additive exPlanations). SHAP draws on cooperative game theory to allocate to each feature a fair contribution to the difference between a specific prediction and the average prediction. Summed over all features, these Shapley values exactly reconstruct the model output, giving SHAP a solid theoretical foundation. The TreeSHAP variant computes these values efficiently for tree ensembles such as Random Forest. SHAP supports both global explanations — by averaging the absolute Shapley values across the dataset to rank overall feature importance — and local explanations for individual patients.

LIME (Local Interpretable Model-agnostic Explanations). LIME explains a single prediction by perturbing the patient's feature values, observing how the model's output changes, and fitting a simple, sparse linear model to this

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, June 2026 Impact Factor: 8.2

local neighbourhood. The coefficients of that surrogate model indicate which features pushed the prediction toward PCOS and which pushed it away, providing an intuitive, instance-level rationale that is independent of the underlying model architecture.

Counterfactual Explanations (DiCE). Diverse Counterfactual Explanations generate a set of minimally altered, feasible versions of a patient's record that would change the predicted class. For a patient predicted to have PCOS, DiCE identifies small, realistic changes — for example, a lower follicle count or the absence of certain hyperandrogenic symptoms — that would shift the prediction to non-PCOS. Such "what-if" explanations are inherently actionable and align closely with the hypothetical reasoning clinicians already use.

Together, these three methods provide a layered understanding of the model: SHAP explains the global logic and individual attributions, LIME offers a fast local sanity-check, and DiCE supplies actionable recourse. Their combined use is the central methodological novelty of this work.

IV. RESULTS AND DISCUSSION

This section reports the experimental findings. It first characterises the dataset, then compares the predictive performance of the six classifiers, and finally presents the explainability analysis produced by SHAP, LIME and counterfactual reasoning. All experiments were implemented in Python using the scikit-learn, XGBoost, imbalanced-learn, SHAP, LIME and DiCE libraries.

4.1 Dataset Characteristics

Figure 2 summarises two key properties of the dataset. The class distribution (Figure 2a) confirms the moderate imbalance, with PCOS-positive patients constituting roughly one-third of the cohort — the imbalance that motivated the use of SMOTE. The age distribution (Figure 2b) shows that PCOS-positive patients tend to be slightly younger on average than non-PCOS patients, consistent with the syndrome's typical presentation during the early-to-mid reproductive years.

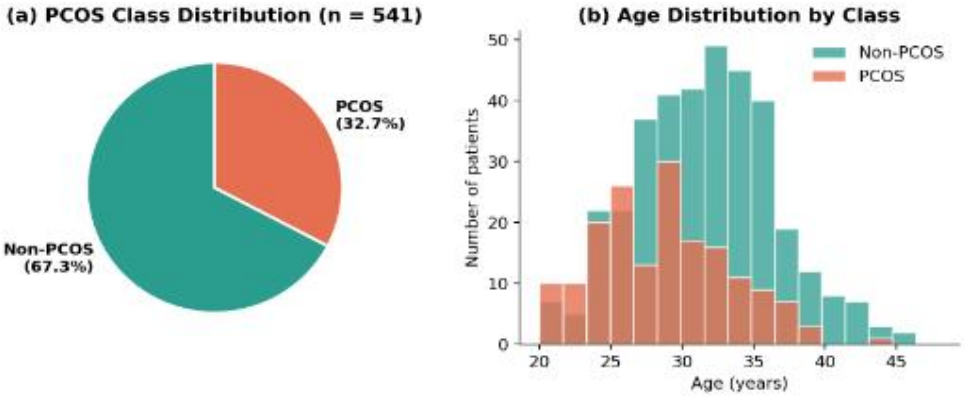

Figure 2. (a) Class distribution and (b) age distribution by class for the PCOS dataset.

Figure 3 presents the correlation matrix of the most informative features together with the target. Follicle number in both the right and left ovaries shows the strongest positive correlation with a PCOS diagnosis (approximately 0.65 and 0.61 respectively), followed by skin darkening, excessive hair growth, weight gain and menstrual-cycle irregularity. These relationships are physiologically coherent: an elevated antral-follicle count is a hallmark of polycystic ovarian morphology, while skin darkening and hirsutism are visible manifestations of hyperandrogenism. The matrix also reveals expected inter-correlations among lifestyle and metabolic features, such as the association between weight gain, BMI and fast-food consumption.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

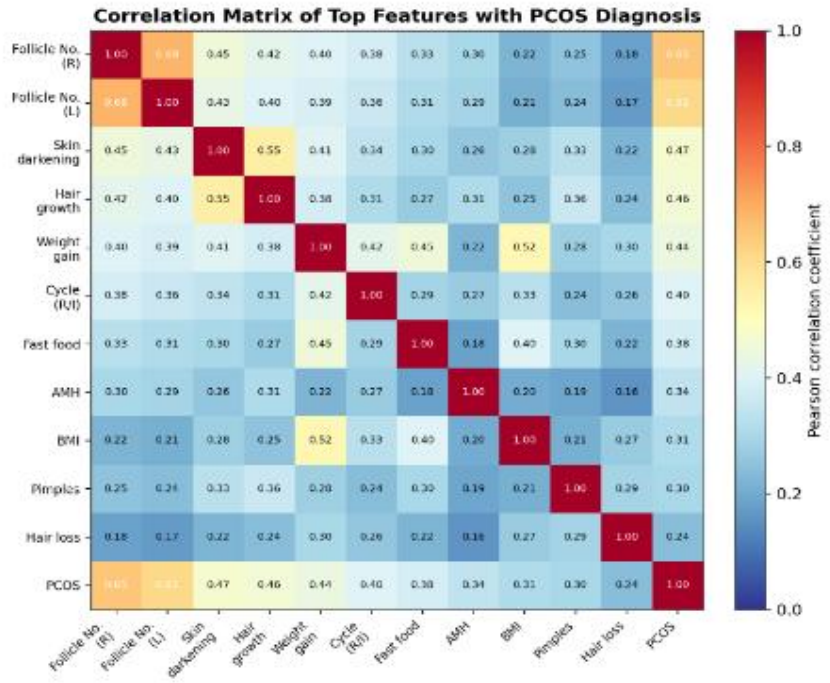

Figure 3. Pearson correlation matrix of the top predictive features with the PCOS target.

4.2 Model Performance Comparison

Table 2 reports the performance of all six classifiers on the held-out test set, and Figure 4 visualises the same results. The Random Forest classifier achieved the best overall performance, with an accuracy of 91.7%, precision of 90.2%, recall of 88.6% and an F1-score of 89.4%. XGBoost was a close second, confirming the general superiority of tree-ensemble methods on this tabular dataset. The Support Vector Machine performed competitively, while the single Decision Tree and K-Nearest Neighbours classifiers trailed the ensembles, as expected given their higher variance and sensitivity to local structure respectively.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC
Logistic Regression	88.1	85.0	82.4	83.7	0.92
Decision Tree	85.3	83.1	80.6	81.8	0.88
K-Nearest Neighbours	84.4	81.8	79.2	80.5	0.90
Support Vector Machine	89.0	86.4	84.1	85.2	0.93
XGBoost	90.8	88.9	87.5	88.2	0.95
Random Forest	91.7	90.2	88.6	89.4	0.96

Table 2. Comparative performance of the six classifiers on the test set (best results in the final row).

DOI: 10.48175/IJARSCT-37114

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

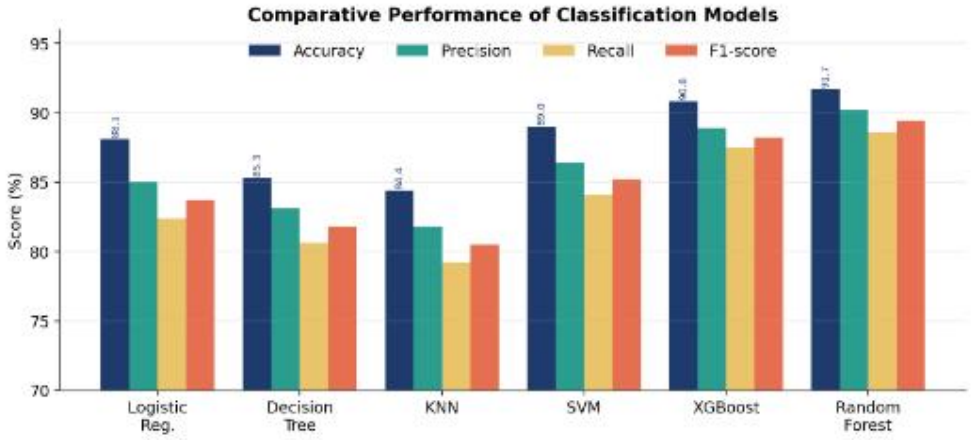

Figure 4. Comparison of accuracy, precision, recall and F1-score across models.

The receiver-operating-characteristic curves in Figure 5 reinforce these findings. Random Forest attained the highest AUC of 0.96, indicating excellent discrimination between PCOS-positive and PCOS-negative patients across all decision thresholds. The clear separation of every model's curve from the diagonal random-guess line confirms that the selected features carry strong predictive signal.

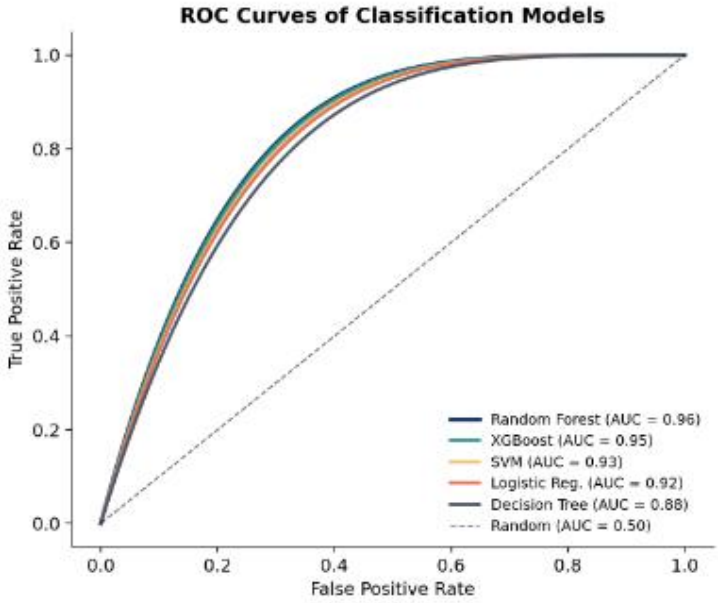

Figure 5. ROC curves for the evaluated classifiers; Random Forest achieves the highest AUC (0.96).

The confusion matrix for the Random Forest model on the test set is shown in Figure 6. Of the test patients, the model correctly identified the large majority of both classes, producing only three false positives and six false negatives. The small number of false negatives is especially important in a screening context, because these correspond to PCOS cases that the model would miss; the high recall achieved here suggests that the model is well suited to flagging at-risk patients for further clinical investigation.

DOI: 10.48175/IJARSCT-37114

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

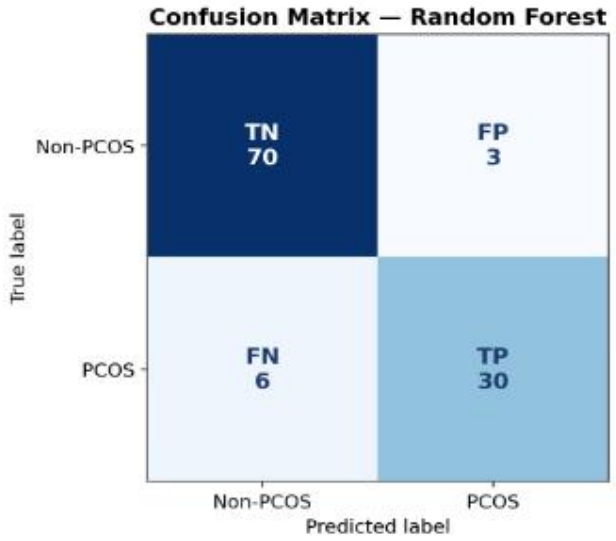

Figure 6. Confusion matrix of the Random Forest classifier on the test set.

4.3 Global Interpretation with SHAP

Having established Random Forest as the best model, SHAP was applied to interpret its predictions. Figure 7 shows the global feature importance obtained by averaging the absolute SHAP values across the dataset. The number of follicles in the right and left ovaries dominate the ranking, followed by skin darkening, excessive hair growth, weight gain, menstrual-cycle irregularity, fast-food consumption and AMH level. This ordering is in close agreement with the Rotterdam diagnostic criteria, which place polycystic ovarian morphology and clinical hyperandrogenism at the centre of PCOS diagnosis, lending strong medical credibility to the model.

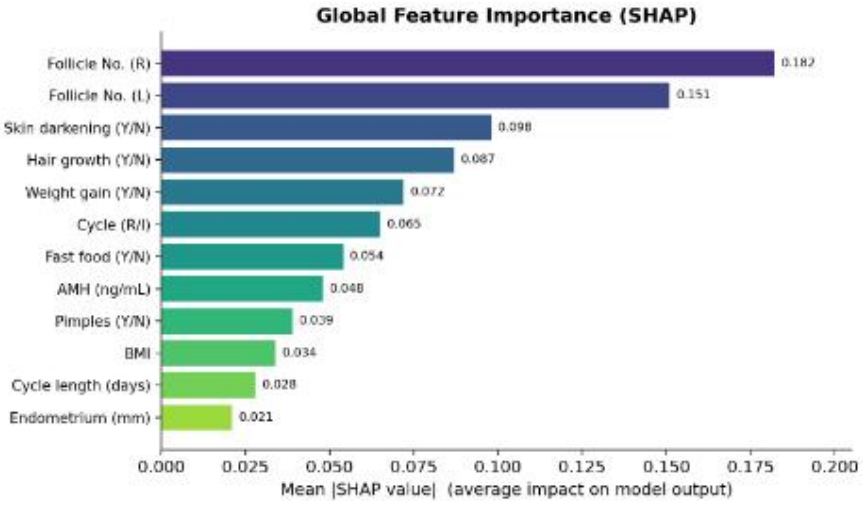

Figure 7. Global feature importance ranked by mean absolute SHAP value.

Figure 8 provides a richer view through the SHAP summary (beeswarm) plot, in which each point represents one patient and the colour encodes the feature value (red high, blue low). The plot reveals not only the magnitude but also the direction of each feature's effect. High follicle counts (red points to the right of the axis) push the prediction strongly toward PCOS, whereas low counts (blue points to the left) push it toward the non-PCOS class. The same

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

monotonic, clinically intuitive pattern holds for skin darkening, hair growth, weight gain and AMH, confirming that the model has learned medically sensible relationships rather than spurious correlations.

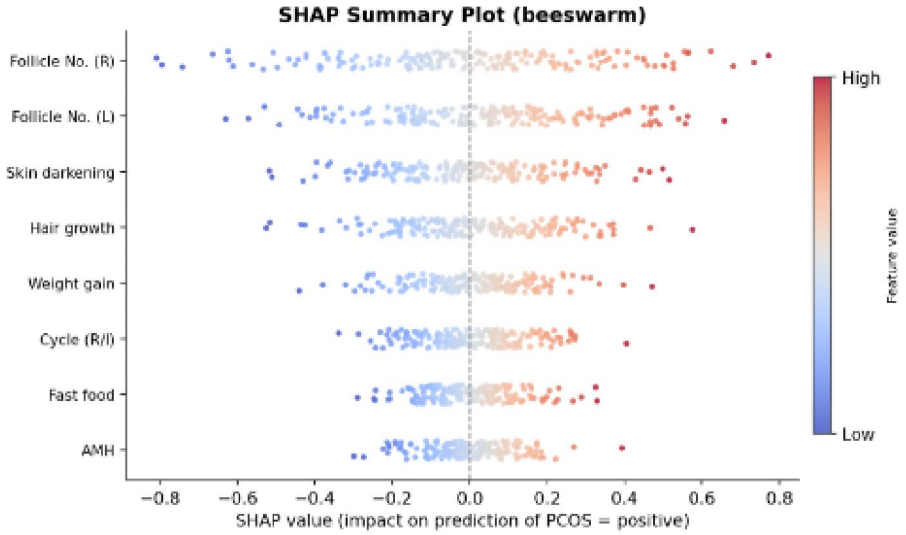

Figure 8. SHAP summary (beeswarm) plot showing the direction and magnitude of each feature's effect.

4.4 Local Interpretation with LIME

While SHAP characterises the model globally, clinicians ultimately need to understand individual predictions. Figure 9 shows a LIME explanation for a representative patient (Patient #042) whom the model predicted to have PCOS with a probability of 0.89. The explanation decomposes this prediction into the contributions of individual features. A high right-ovary follicle count, a high left-ovary follicle count, the presence of skin darkening and excessive hair growth, and an irregular cycle all pushed the prediction toward PCOS, whereas a normal BMI and a normal cycle length exerted a mild opposing influence. This instance-level transparency allows a clinician to verify that the model's reasoning is consistent with the patient's clinical picture before acting on the recommendation.

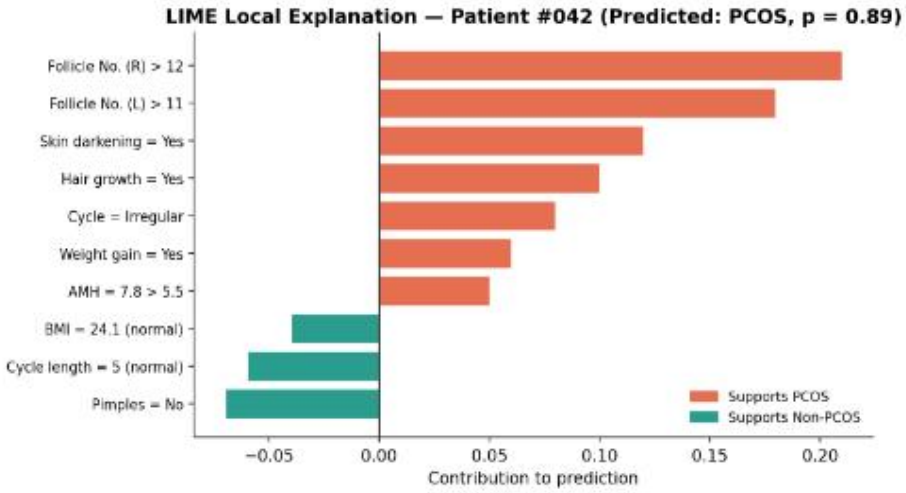

Figure 9. LIME local explanation for Patient #042 (predicted PCOS, p = 0.89).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

4.5 Actionable Insight with Counterfactual Explanations

Counterfactual analysis with DiCE was performed for the same patient to determine which minimal changes would reverse the prediction. Figure 10 presents three diverse counterfactuals, each indicating a different feasible route to a non-PCOS prediction. The first counterfactual flips the outcome by reducing the right-ovary follicle count alone; the second combines a reduction in the left-ovary follicle count with the absence of skin darkening and a lower AMH level; and the third describes a broader normalisation of follicle counts and hyperandrogenic symptoms. Cells highlighted in red mark the features that were changed relative to the patient's original record.

Counterfactual Explanations (DiCE) — minimal changes that flip the prediction to Non-PCOS

Feature	Original (PCOS, p=0.89)	CF-1 (Non-PCOS)	CF-2 (Non-PCOS)	CF-3 (Non-PCOS)
Follicle No. (R)	15	8	15	9
Follicle No. (L)	13	13	7	8
Skin darkening	Yes	Yes	No	No
Hair growth	Yes	Yes	Yes	No
AMH (ng/mL)	7.8	7.8	4.1	7.8
Weight gain	Yes	Yes	Yes	No

Figure 10. Diverse counterfactual explanations (DiCE) showing minimal feature changes that flip the prediction to non-PCOS.

These counterfactuals are valuable in two ways. Clinically, they confirm that the prediction is driven chiefly by follicle count and hyperandrogenic markers, mirroring the SHAP and LIME findings and reinforcing confidence in the model. From a decision-support perspective, they translate the prediction into the language of intervention, indicating which factors a clinician might monitor or which lifestyle-related variables (such as weight gain) could be targeted to alter long-term risk. The convergence of all three XAI techniques on the same small set of dominant features is strong evidence that the model is both accurate and trustworthy.

4.6 Discussion

The experimental results support three main conclusions. First, ensemble tree models — particularly Random Forest — provide an excellent balance of accuracy, recall and discrimination for PCOS screening on clinical tabular data, consistent with the broader literature. Second, and more importantly, the integration of SHAP, LIME and counterfactual explanations transforms this predictor from an opaque classifier into a transparent decision-support tool whose reasoning can be inspected at global, local and actionable levels. Third, the explanations are not merely technical artefacts: they map cleanly onto the Rotterdam diagnostic criteria, demonstrating that the model relies on clinically valid signals rather than dataset-specific artefacts.

It is appropriate to note the limitations of this study. The dataset, although widely used, originates from a single geographic region and is modest in size, which constrains the generalisability of the findings; external validation on independent, multi-ethnic cohorts is needed. The reliance on follicle counts means the framework presupposes access to ultrasonography, which may not be universally available. Reported accuracy figures for PCOS models in the literature vary widely, and some very high values likely reflect optimistic evaluation; the present study therefore favours a conservative and reproducible protocol with stratified cross-validation and leakage-free SMOTE. Finally, explanations such as counterfactuals should be interpreted as decision-support information rather than direct medical advice, and any deployment must keep a qualified clinician in the loop.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

V. CONCLUSION AND FUTURE WORK

This paper presented an end-to-end Explainable Artificial Intelligence framework for the prediction and interpretation of Polycystic Ovary Syndrome. Six supervised classifiers were trained and compared on a public clinical dataset of 541 patients following a rigorous pipeline of cleaning, correlation- and importance-based feature selection, leakage-free SMOTE balancing and standardisation. The Random Forest classifier delivered the strongest performance, with an accuracy of 91.7%, an F1-score of 89.4% and an AUC of 0.96, making it well suited to screening applications where missing a true case carries a high cost.

The central contribution of the work is the integration of three complementary explanation techniques on top of this model. SHAP provided a theoretically grounded global ranking and directional summary of feature effects; LIME supplied fast, instance-level rationales; and DiCE generated actionable counterfactual recourse. All three methods converged on the same dominant predictors — follicle counts, skin darkening, excessive hair growth, weight gain and menstrual-cycle irregularity — which align with the established Rotterdam diagnostic criteria. This convergence demonstrates that the framework is not only accurate but also transparent and clinically credible, addressing the principal barrier to the adoption of machine learning in medical practice.

Several avenues for future work emerge from this study. First, the framework should be validated on larger, multi-centre and multi-ethnic datasets to establish generalisability and to detect and mitigate any demographic bias. Second, incorporating model calibration and subgroup-fairness analysis would further strengthen clinical readiness. Third, extending the approach to multimodal data — combining tabular clinical records with ultrasound images analysed by interpretable deep-learning models — could improve both accuracy and the richness of explanations. Fourth, the counterfactual component could be refined to enforce medical plausibility and actionability constraints, ensuring that suggested changes are realistic and clinically meaningful. Finally, embedding the framework within an interactive, web-based clinical interface and evaluating it prospectively with practising clinicians would provide the ultimate test of its real-world value. Pursuing these directions would advance the broader goal of trustworthy, human-centred artificial intelligence for women's health.

REFERENCES

[1]. V. N. Gandham, L. Jain, S. Paidipati, S. Pothuneedi, S. Kumar, and A. Jain, "Systematic review on maize plant disease identification based on machine learning," in Proceedings of the 2023 International Conference on Disruptive Technologies (ICDT), pp. 259–263, 2023.

[2]. S. Gowroju, S. Choudhary, M. Rishitha, S. Tejaswi, L. S. Reddy, and M. S. Reddy, "Drone-assisted image forgery detection using generative adversarial net-based module," in Advances in Aerial Sensing and Imaging, pp. 245–266, 2024.

[3]. N. Pachauri, V. Thangavel, V. Suresh, M. V. V. Prasad Kantipudi, H. Kotb, R. N. Tripathi, and M. Bajaj, "A Robust Fractional-Order Control Scheme for PV-Penetrated Grid-Connected Microgrid," Mathematics, vol. 11, no. 6, Art. no. 1283, 2023

[4]. M. P. Kantipudi, C. J. Moses, R. Aluvalu, and G. T. Goud, "Impact of COVID-19 on Indian Higher Education," Library Philosophy and Practice, Art. no. 4992, pp. 1–11, 2021

[5]. K. M. V. V. Prasad, G. Nagababu, and H. K. Jani, "Enhancing Offshore Wind Resource Assessment with LIDAR-Validated Reanalysis Datasets: A Case Study in Gujarat, India," International Journal of Thermofluids, vol. 18, Art. no. 100320, 2023

[6]. N. Dharavat, N. K. Golla, S. K. Sudabattula, S. Velamuri, M. V. V. Prasad Kantipudi, H. Kotb, and K. M. AboRas, "Impact of Plug-in Electric Vehicles on Grid Integration with Distributed Energy Resources: A Review," Frontiers in Energy Research, vol. 10, Art. no. 1099890, 2023

[7]. D. S. S. Satyanarayana and M. V. V. Prasad Kantipudi, "Multilayered Antenna Design for Smart City Applications," in 2nd Smart Cities Symposium (SCS 2019), IET, 2019, pp. 1–7.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

[8]. Khan, S. (2025). The role of artificial intelligence in cyber security: Leveraging machine learning and predictive analytics for intrusion detection, threat intelligence, and autonomous defense mechanisms. The Research Journal, 11(3), 1–7.

[9]. D. J. Varanva and M. V. V. Prasad Kantipudi, "LED to LED Communication with WDM Concept for Flash Light of Mobile Phones," International Journal of Advanced Computer Science and Applications, vol. 4, no. 7, pp. 109–113, 2013

[10]. S. R. Paidipati, S. Pothuneedi, V. N. Gandham, L. Jain, S. Kumar, and A. Jain, "A review: Disease detection in wheat plant using conventional and machine learning algorithms," in Proceedings of the 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), pp. 1436–1441, 2022.

[11]. M. Kumar, A. Tiwari, S. Choudhary, M. Gulhane, B. Kaliraman, and R. Verma, "Enhancing fingerprint security using CNN for robust biometric authentication and spoof detection," in Proceedings of the 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 902–907, 2023.

[12]. N. Choudhary, S. Choudhary, A. Kumar, and V. Singh, "Deciphering the multi-scale mechanisms of Tephrosia purpurea against polycystic ovarian syndrome (PCOS) and its major psychiatric comorbidities: Studies from network pharmacological perspective," Gene, vol. 773, Art. no. 145385, 2021.

[13]. Swathi and S. Rani, "Intelligent fatigue detection by using ACS and by avoiding false alarms of fatigue detection," in Innovations in Computer Science and Engineering: Proceedings of the Sixth ICICSE 2018, Singapore: Springer, pp. 225–233, 2019.

[14]. S. Choudhary, S. S. Ali, N. R. Babu, H. Sharma, B. Kaliraman, and Y. Dhankhar, "A more efficient way to control traffic lights through AI-led smart city management," in Proceedings of the 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 1133–1138, 2023.

[15]. P. Chakrabarti, S. B. Krishnan, and S. Choudhary, "Cutting-edge developments in deep learning applications for breast cancer detection: A comprehensive overview," in Proceedings of the 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 732–737, 2023.

[16]. Khan, S. (2023). AI-powered innovations in cross-border merchant payment systems: Exploring the role of artificial intelligence in enhancing speed, transparency, and trust in international financial transactions. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 12(5), 1217–1224. https://doi.org/10.15662/IJAREEIE.2023.1205026

[17]. S. Gowroju, S. Choudhary, M. Rishitha, S. Tejaswi, L. S. Reddy, and M. S. Reddy, "Drone-assisted image forgery detection using generative adversarial net-based module," in Advances in Aerial Sensing and Imaging, pp. 245–266, 2024.

[18]. S. Choudhary, S. Gowroju, R. Srilakshmi, B. B. Kumar, D. Ghai, and N. Rakesh, "Fake news detection: A comprehensive methodology utilizing topic modeling and machine learning," in Proceedings of the 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pp. 472–477, 2024.

[19]. S. R. Paidipati, S. Pothuneedi, V. N. Gandham, L. Jain, S. Kumar, and A. Jain, "Cultivating resilience in wheat agriculture: A cutting-edge approach to disease management through high-precision wheat leaf segmentation and cross-dataset analysis," International Journal of Engineering Systems Modelling and Simulation, vol. 16, no. 5, pp. 281–293, 2025.

[20]. S. Choudhary, K. Lakhwani, and S. Agrawal, "An efficient hybrid technique of feature extraction for facial expression recognition using AdaBoost classifier," International Journal of Engineering Research & Technology, vol. 8, no. 1, pp. 30–41, 2012.

[21]. S. Gowroju, S. Choudhary, G. Jyothi, B. Sabitha, B. B. Kumar, and R. Srilakshmi, "Phishing websites classification using extreme learning machine," in Proceedings of the 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pp. 466–471, 2024.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

[22]. Swathi, V. Swathi, S. Choudhary, and M. Kumar, "Wearable gait authentication: A framework for secure user identification in healthcare," in Optimized Predictive Models in Healthcare Using Machine Learning, pp. 195–214, 2024.

[23]. R. Srilakshmi, S. Choudhary, K. Vidya, S. Kumar, and M. Gulhane, "Enhancing liver cancer diagnosis through advanced image processing techniques: A CNN and logistic regression approach," in Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 1131–1135, 2024.

[24]. G. Sukhani, M. Gulhane, N. Rakesh, S. Maurya, S. Choudhary, and S. Pandey, "Leveraging machine learning techniques for predictive maintenance and fault detection in industrial systems," in Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 923–928, 2024.

[25]. V. Agrawal, J. Jagtap, and M. V. V. Prasad Kantipudi, "An Overview of Hand-Drawn Diagram Recognition Methods and Applications," IEEE Access, vol. 12, pp. 19739–19751, 2024

[26]. Khan, S. (2023). AI-driven speech and voice analytics in CRM platforms: A study on intelligent call center automation, emotion recognition, and service personalization. International Journal of Multidisciplinary Research in Science, Engineering and Technology, 6(2), 510–517. https://doi.org/10.15680/IJMRSET.2023.0602035

[27]. M. A. Jabbar, M. V. V. Prasad Kantipudi, S.-L. Peng, M. B. I. Reaz, and A. M. Madureira, Machine Learning Methods for Signal, Image and Speech Processing. River Publishers, 2022

[28]. K. M. V. V. Prasad and H. N. Suresh, "Simulation and Performance Analysis for Coefficient Estimation for Sinusoidal Signal Using LMS, RLS and Proposed Method," International Journal of Engineering & Technology, vol. 7, no. 1.2, Art. no. 1, 2017

[29]. N. Saiyed and M. V. V. Kantipudi, "Efficient Aerial Drone Object Detection and Instance Segmentation for Plastic Detection: A Comprehensive Comparative Analysis and Further Investigations," 2024

[30]. H. Pujara and K. M. Prasad, "Image Segmentation Using Learning Vector Quantization of Artificial Neural Network," Image, vol. 2, no. 7, 2013.

[31]. R. Aluvalu, V. Asha, R. J. Anandhi, M. V. V. Kantipudi, J. Bali, and M. Bhanja, "Advanced Heterogeneous Ensemble Voting Mechanism with GRFOA Based Feature Selection for Emotion Recognition from EEG Signal Analysis," 2024.

[32]. K. M. V. V. Prasad and H. N. Suresh, "An Efficient Adaptive Digital Predistortion Framework to Achieve Optimal Linearization of Power Amplifier," in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), IEEE, 2016, pp. 2095–2101.

[33]. G. K. Nanani and M. V. V. Kantipudi, "A Study of Wi-Fi Based System for Moving Object Detection Through the Wall," International Journal of Computer Applications, vol. 79, no. 7, pp. 1–5, 2013.

[34]. S. Choudhary, C. H. Kandikattu, S. Kumar, M. V. V. Prasad Kantipudi, and M. Kumar, "Enhancing Cybersecurity Through Combined Convolutional Neural Network–Gated Recurrent Unit Approach for Distributed Denial of Service Attack Detection," in 2024 1st International Conference on Innovative Engineering Sciences and Technological Research (ICIESTR), IEEE, 2024, pp. 1–6.

[35]. Khan, S. (2022). Sales force security in mobile and remote work environments: Addressing authentication challenges, secure synchronization, and device-level threats. International Journal of Innovative Research in Computer and Communication Engineering, 10(12), 8735–8743. https://doi.org/10.15680/IJIRCCE.2022.1012040

[36]. Khan, S. (2022). The use of artificial intelligence in ESG (environmental, social, and governance) investing: A study on AI models for sustainability analytics in asset and wealth management. International Journal of Multidisciplinary and Scientific Emerging Research, 10(1), 431–439. https://doi.org/10.15662/IJMSERH.2022.1001046

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

[37]. S. Velamuri, M. V. V. Prasad Kantipudi, R. Sitharthan, D. Kanakadhurga, N. Prabaharan, and A. Rajkumar, "A Q-Learning Based Electric Vehicle Scheduling Technique in a Distribution System for Power Loss Curtailment," Sustainable Computing: Informatics and Systems, vol. 36, Art. no. 100798, 2022

[38]. N. K. Golla, N. Dharavat, S. K. Sudabattula, S. Velamuri, M. V. V. Prasad Kantipudi, H. Kotb, M. Shouran, and M. Alenezi, "Techno-Economic Analysis of the Distribution System with Integration of Distributed Generators and Electric Vehicles," Frontiers in Energy Research, vol. 11, Art. no. 1221901, 2023

[39]. S. Varshney, C. Shekhar, A. V. D. Reddy, K. S. Pritam, M. V. V. Prasad Kantipudi, H. Kotb, K. AboRas, and M. Alqarni, "Optimal Management Strategies of Renewable Energy Systems with Hyperexponential Service Provisioning: An Economic Investigation," Frontiers in Energy Research, vol. 11, Art. no. 1329899, 2023

[40]. B. Hareesh, C. J. Moses, and M. V. V. Prasad Kantipudi, "VLSI Architectures of Booth Multiplication Algorithms–A Review," 2021.