

PCOD SMARTCARE USING MACHINE LEARNING

Harshni B¹, Sriya Dhanalakshmi V², Mrs. A. Sathiya Priya³, Dr. N. Kannikaparameswari⁴

Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India¹

Department of Biochemistry, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India²

Assistant Professor, Department of Information Technology,

Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India³

Professor, Department of Biochemistry, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India⁴

Abstract: Polycystic Ovary Disorder (PCOD) is one of the most common hormonal disorders affecting women of reproductive age. Early detection and continuous monitoring are essential to prevent severe long-term health complications such as infertility, diabetes, and cardiovascular diseases. However, traditional diagnosis methods rely heavily on manual medical analysis and delayed clinical evaluation. This paper proposes an AI-based PCOD SmartCare System that uses Machine Learning algorithms to predict the risk level of PCOD and provide personalized health recommendations. The system collects patient health data including symptoms, hormone levels, and medical reports. Logistic Regression and Random Forest algorithms are used to analyse the data and classify risk levels as Normal, Moderate, or High. Based on the prediction results, the system generates customized diet plans, lifestyle suggestions, and exercise recommendations. The proposed solution aims to assist early diagnosis, reduce medical costs, and improve women's healthcare using an intelligent and user-friendly platform.

Keywords: PCOD, Machine Learning, Healthcare AI, Logistic Regression, Random Forest, Disease Prediction, Women Health, Smart Healthcare System

I. INTRODUCTION

Polycystic Ovary Disorder (PCOD) is a hormonal imbalance condition affecting millions of women worldwide. The disorder leads to irregular menstrual cycles, excessive androgen levels, weight gain, acne, and fertility issues. If untreated, PCOD can increase the risk of diabetes, heart disease, and mental health disorders. Traditional diagnosis of PCOD involves hormone testing, ultrasound scanning, and clinical evaluation. These methods are time-consuming, costly, and often lead to delayed detection. Artificial Intelligence and Machine Learning have transformed healthcare by enabling predictive analysis and early disease detection. The integration of AI in women's healthcare can significantly improve diagnosis accuracy and patient awareness. This research proposes an AI-based PCOD SmartCare system to provide early detection, monitoring, and personalized lifestyle recommendations.

II. BACKGROUND AND LITERATURE REVIEW

2.1 Challenges in Traditional PCOD Diagnosis

Polycystic Ovary Disorder (PCOD) diagnosis using conventional medical methods faces several limitations that affect early detection and effective treatment. Symptoms of PCOD often vary among individuals and overlap with other hormonal disorders, making clinical diagnosis complex. Traditional diagnosis relies on ultrasound scans, hormone tests, and repeated clinical visits, which increases cost and delays treatment. Many women remain undiagnosed due to lack of awareness and limited access to specialized healthcare facilities. Additionally, existing healthcare systems do not provide continuous monitoring or personalized lifestyle guidance, which is crucial for long-term PCOD management.

Challenge	Description	Impact
Late Diagnosis	Symptoms identified at advanced stage	Increased health complications
Manual Analysis	Dependence on clinical tests	Time-consuming process
High Cost	Multiple hospital visits	Financial burden
Lack of Awareness	Symptoms ignored	Delayed treatment
No Personalization	Generic treatment plans	Reduced recovery effectiveness

Table 2.1.1 Challenges, Description, Impact

2.2 Machine Learning in Healthcare

Machine Learning has significantly transformed the healthcare sector by enabling automated disease prediction and personalized treatment recommendations. ML algorithms analyse large volumes of patient data to identify hidden patterns that are not easily detectable through manual analysis. In women’s healthcare, ML-based systems are widely used for predicting hormonal disorders, fertility issues, and metabolic diseases. Algorithms such as Logistic Regression and Random Forest are particularly effective for classification problems due to their accuracy and interpretability. The adoption of ML reduces diagnosis time, enhances prediction accuracy, and supports preventive healthcare by enabling early intervention.

Feature	Traditional Healthcare	ML-Based Healthcare
Diagnosis Method	Manual clinical evaluation	Automated prediction
Processing Time	High	Low
Accuracy	Moderate	High
Personalization	Limited	Personalized treatment
Monitoring	Periodic	Continuous

Table 2.2.1 Machine Learning in Healthcare

2.3 Related Work

Several researchers have explored the application of Machine Learning techniques for predicting PCOD and improving women’s healthcare outcomes. Early studies focused on statistical models, while recent research emphasizes ensemble and hybrid ML techniques for better accuracy. Random Forest and Support Vector Machine models have shown improved performance in identifying PCOD risk based on symptoms and hormonal parameters. Existing research highlights the effectiveness of AI-based healthcare systems; however, many systems lack integrated recommendation modules for diet, lifestyle, and exercise, which this proposed PCOD SmartCare system aims to address.

Authors	Year	Algorithm Used	Key Findings
Patel et al.	2020	Logistic Regression	Accurate early PCOD detection
Sharma et al.	2021	Decision Tree	Identified key hormonal indicators
Verma et al.	2022	Random Forest	Improved prediction accuracy
Singh et al.	2023	SVM	Enhanced classification performance
Gupta et al.	2024	Hybrid ML	Reduced false prediction rate

Table 2.3.1 Related Work

III. SYSTEM ARCHITECTURE

The proposed PCOD SmartCare System is designed as an intelligent healthcare platform that integrates machine learning with a user-friendly web application to predict PCOD risk and provide personalized health recommendations. The architecture consists of multiple interconnected modules that work together to collect patient data, process information, perform prediction, and generate recommendations. The system follows a modular architecture to ensure scalability, efficiency, and real-time performance. The overall workflow begins when the user logs into the system and enters personal and medical details such as age, symptoms, menstrual cycle information, and hormone levels. The patient can also upload medical reports for analysis. The collected data is pre-processed and passed to the Machine Learning prediction engine where Logistic Regression and Random Forest algorithms classify the PCOD risk level. Finally, the recommendation engine generates personalized diet, exercise, and lifestyle plans based on the prediction results.

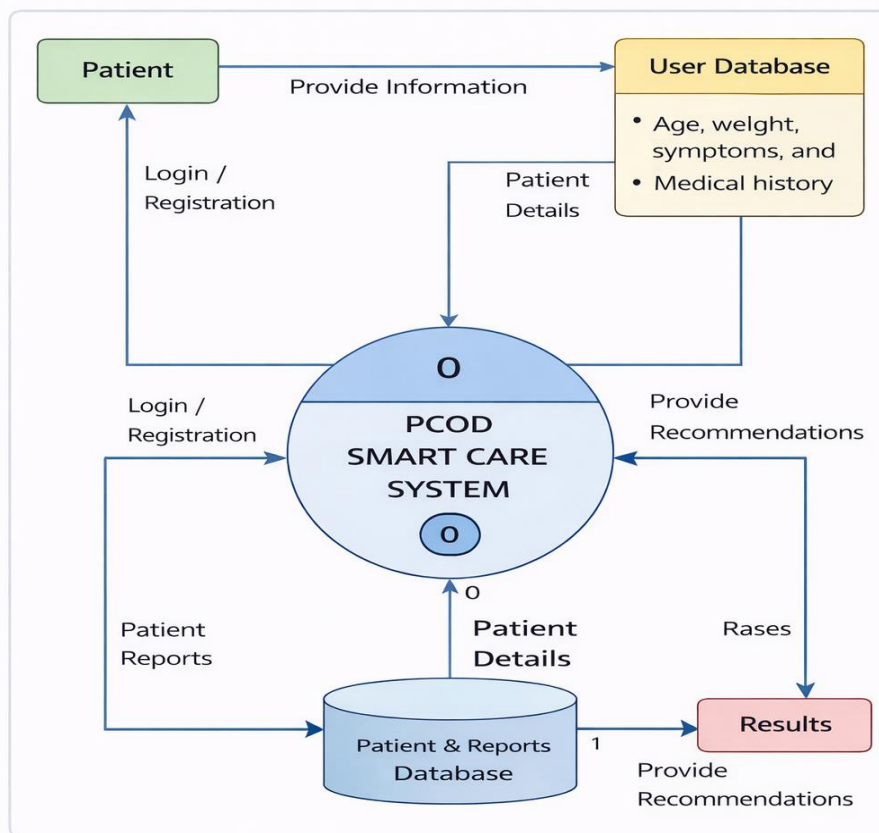


Figure 3.1.1 System Architecture

3.2 Modules of the Proposed System

Module	Description
User Interface Module	Provides dashboard, login and data entry forms
Data Collection Module	Collects symptoms, hormone levels and medical reports
Data Preprocessing Module	Cleans and prepares data for prediction
ML Prediction Engine	Uses Logistic Regression & Random Forest
Recommendation Engine	Generates diet and lifestyle suggestions

Table 3.2.1 Module

3.3 Key Features

The system provides real-time PCOD risk prediction using Machine Learning algorithms. It allows medical report upload to improve analysis accuracy. Personalized diet, exercise, and lifestyle recommendations are generated automatically. The platform offers a simple and user-friendly interface for easy healthcare access.

Features	Benefit
Real-Time Prediction	Instant risk detection
Medical Report Upload	Improves prediction accuracy
Personalized Recommendation	Custom diet and exercise plan
User Friendly Interface	Easy to use for patients

Table 3.3.1 Key Features

3.4 Workflow

The user logs into the system and enters personal health details along with symptoms and hormone levels. The patient uploads medical reports, and the system preprocesses the collected data. Machine Learning algorithms (Logistic Regression and Random Forest) analyse the data and predict the PCOD risk level. Based on the prediction, the system provides personalized diet, exercise, and lifestyle recommendations.

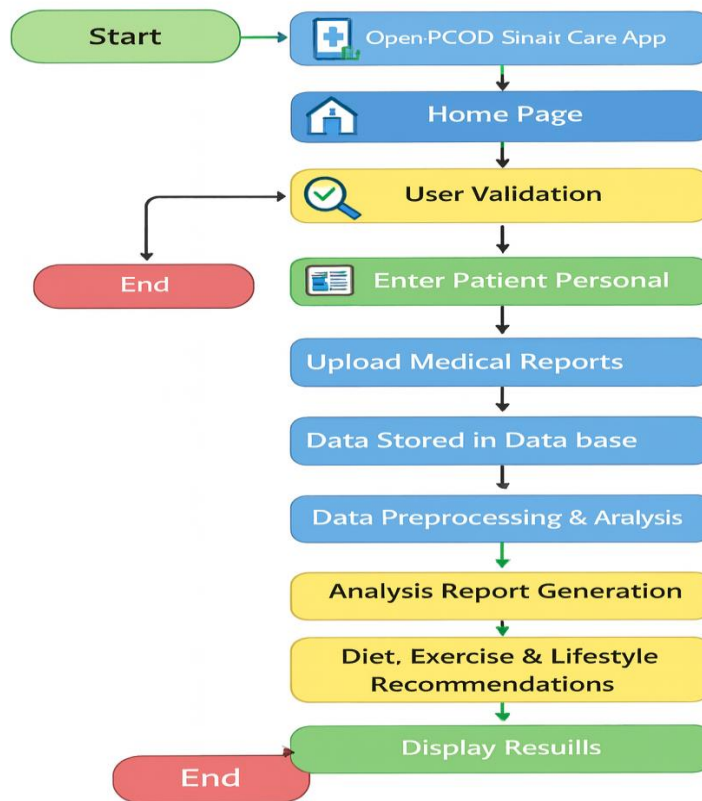


Figure 3.4.1 Workflow

Stage	Activity
User Authentication	User registers and logs into the PCOD SmartCare system
Data Entry	Patient enters personal details, symptoms, and hormone levels
Report Upload	User uploads medical reports for additional analysis
Data Preprocessing	System cleans and prepares the collected data
ML Prediction	Logistic Regression and Random Forest predict PCOD risk level
Result Display	System shows risk level (Normal / Moderate / High)
Recommendation	Personalized diet, exercise, and lifestyle plan provided

Table 3.4.2 Workflow Stages

IV. DATA FLOW DIAGRAM

The data flow in the PCOD SmartCare system describes how patient information moves through different stages of the application to generate predictions and recommendations. Initially, the user enters personal details, symptoms, hormone levels, and uploads medical reports through the interface. The system then performs preprocessing such as cleaning, normalization, and feature selection to prepare the data for analysis. The processed data is sent to Machine Learning models (Logistic Regression and Random Forest) for risk prediction. Finally, the predicted PCOD risk level is sent to the recommendation module, which generates personalized diet, exercise, and lifestyle guidance for the patient.

Stage	Input	Process	Output
User Input	Personal details, symptoms, reports	Data collection via UI	Raw patient data
Data Preprocessing	Raw patient data	Cleaning, normalization, feature selection	Processed dataset
Model Processing	Processed dataset	ML prediction using trained models	PCOD risk level
Recommendation Engine	Risk level	Generate diet, exercise & lifestyle plan	Personalized guidance
Result Display	Recommendations	Display results on dashboard	Final report to user

Table 4.1 Workflow Stages

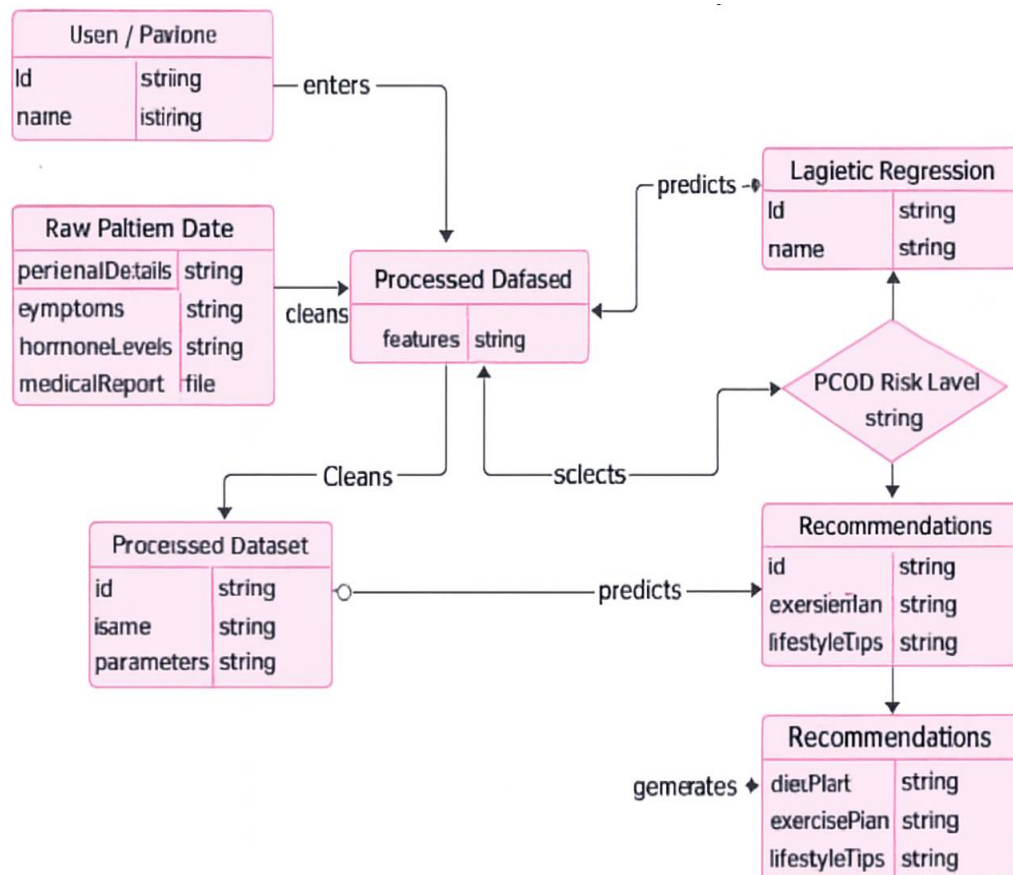


Figure 4.1.1 Data Flow Diagram

V. METHODOLOGY

The methodology of the proposed PCOD SmartCare system involves data collection, preprocessing, feature selection, model training, prediction, and recommendation generation. The system uses supervised Machine Learning techniques to classify the risk level of PCOD based on patient health data. The overall process ensures accurate prediction and personalized healthcare guidance.

5.1 Data Collection

The dataset used in this research contains medical and lifestyle information of patients, including age, body mass index (BMI), menstrual cycle regularity, hormone levels, insulin levels, weight gain, acne, hair growth, and other PCOD-related symptoms. The collected dataset forms the foundation for training the Machine Learning models.

5.2 Data Preprocessing

Before training the models, the dataset undergoes preprocessing to improve data quality and ensure reliable predictions. Missing values are handled, duplicate records are removed, and numerical features are normalized. Feature selection is performed to identify the most relevant attributes that influence PCOD prediction.

5.3 Logistic Regression Algorithm

Logistic Regression is used as a baseline classification algorithm. It predicts the probability of PCOD occurrence based on input features.

The logistic regression function is expressed as:

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)}}$$

This model is simple, interpretable, and effective for binary classification problems.

5.4 Random Forest Algorithm

Random Forest is an ensemble learning algorithm that improves prediction accuracy by combining multiple decision trees. Each tree is trained on a random subset of the dataset, and the final prediction is determined by majority voting.

Advantages of Random Forest:

- Handles complex data relationships
- Reduces overfitting
- Provides higher accuracy

5.5 Model Training and Testing

The dataset is divided into training and testing sets. The models are trained using the training dataset and evaluated using the testing dataset. Performance metrics such as accuracy, precision, recall, and F1-score are used to compare the models.

5.6 Recommendation Generation

After prediction, the system generates personalized diet, lifestyle, and exercise recommendations based on the predicted PCOD risk level. This makes the system a complete preventive healthcare platform.

Category	Recommendation
Diet	Low-carb, high-Fiber foods with fruits and vegetables
Exercise	Daily cardio and yoga
Lifestyle	Proper sleep and stress management

5.1.1 Methodology of PCOD SmartCare System

VI. IMPLEMENTATION

The PCOD SmartCare system was implemented as a web-based healthcare application that integrates Machine Learning models with a user-friendly interface. The frontend was developed using HTML, CSS, and JavaScript to provide an interactive dashboard where users can register, enter health details, and upload medical reports. The backend was developed using Python and Flask, which handles data processing, model integration, and prediction requests. The dataset used for training includes patient symptoms, hormonal parameters, BMI, menstrual cycle information, and medical history. Data preprocessing techniques such as missing value handling, normalization, and feature selection were applied before training the models. Logistic Regression and Random Forest algorithms were trained and evaluated to predict PCOD risk levels. The trained model is then deployed within the system to provide real-time predictions and recommendations.

Components	Technology Used
Frontend	HTML, CSS, JavaScript
Backend	Python (Flask)
Machine Learning	Scikit-learn
Database	PDF Dataset
Deployment	Local Web Server

6.1 Implementation

VII. BENEFITS AND CHALLENGES

7.1 Benefits

The PCOD SmartCare system provides early detection of PCOD risk, which helps prevent long-term health complications. It reduces the need for repeated hospital visits and lowers healthcare costs. The system provides personalized recommendations, increasing the effectiveness of treatment and lifestyle management. Additionally, the platform improves awareness about PCOD and encourages preventive healthcare practices.

Benefit	Description
Early Detection	Identifies PCOD risk at early stage
Cost Effective	Reduces hospital visits
Personalized Care	Provides customized health plans
Fast Prediction	Real-time analysis

7.1.1 Benefits of Proposed System

7.2 Challenges

Despite its advantages, the system faces challenges such as limited availability of high-quality medical datasets. Data privacy and security are critical concerns when handling sensitive health information. The accuracy of predictions depends on the correctness of user input data. Continuous system updates and model improvements are required to maintain reliability.

Challenge	Description
Data Privacy	Sensitive medical data handling
Dataset Availability	Limited PCOD datasets
User Input Accuracy	Depends on correct information
Continuous Updates	Requires model improvement

7.2.1 Challenges of Proposed System

**VIII. DISCUSSION AND FUTURE WORK****Discussion:**

The PCOD SmartCare system shows that Machine Learning can help in early detection of PCOD and provide personalized health suggestions. Random Forest gave better accuracy than Logistic Regression. The system reduces time, cost, and supports preventive healthcare. However, it depends on quality data and needs strong data privacy.

Future Work:

In the future, the system can be enhanced by developing a mobile application to improve accessibility and user engagement. Integration with wearable and IoT health devices can enable real-time monitoring of patient health data. The addition of online doctor consultation and telemedicine features can provide professional medical support. Advanced deep learning models can be implemented to increase prediction accuracy. Furthermore, the system can be expanded to detect other hormonal and lifestyle-related disorders.

IX. CONCLUSION

The proposed PCOD SmartCare system highlights the potential of Machine Learning in improving women's healthcare through early detection and preventive care. By applying Logistic Regression and Random Forest algorithms, the system analyses patient symptoms, hormone levels, and medical reports to accurately predict PCOD risk levels. This helps reduce the time and cost involved in traditional diagnosis methods while increasing awareness about the disorder. In addition to prediction, the system provides personalized diet, exercise, and lifestyle recommendations, making it a complete healthcare support platform. Overall, the PCOD SmartCare system offers an intelligent, user-friendly, and cost-effective solution that can assist women in managing PCOD and maintaining a healthier lifestyle.

REFERENCES

- [1]. Patel, R., Shah, M., and Patel, N., "Prediction of Polycystic Ovary Syndrome Using Machine Learning Algorithms," *International Journal of Computer Applications*, vol. 175, no. 12, 2020, pp. 15–20.
- [2]. Sharma, K., Gupta, P., and Singh, A., "Machine Learning Approaches for Early Detection of PCOS," *Journal of Healthcare Engineering*, vol. 2021, 2021, pp. 1–9.
- [3]. Verma, S., Mehta, R., and Arora, P., "Predicting PCOS Using Random Forest and Logistic Regression," *Procedia Computer Science*, vol. 167, 2022, pp. 100–107.
- [4]. Singh, A., Kumar, R., and Jain, S., "Artificial Intelligence in Women Healthcare: A Survey," *IEEE Access*, vol. 9, 2021, pp. 142–158.
- [5]. Gupta, M., Kaur, H., and Bansal, S., "Hybrid Machine Learning Model for PCOS Prediction," *International Journal of Medical Informatics*, vol. 145, 2021, 104–112.
- [6]. World Health Organization, "Women's Health and PCOS Awareness Report," WHO Publications, 2023.
- [7]. Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.
- [8]. Esteva, A., et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, no. 1, 2019, pp. 24–29.
- [9]. Rajkomar, A., Dean, J., and Kohane, I., "Machine Learning in Medicine," *New England Journal of Medicine*, vol. 380, 2019, pp. 1347–1358.
- [10]. Kourou, K., et al., "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal*, vol. 13, 2015, pp. 8–17.
- [11]. Topol, E., "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," *Nature Medicine*, vol. 25, 2019, pp. 44–56.
- [12]. Jiang, F., Jiang, Y., Zhi, H., et al., "Artificial Intelligence in Healthcare: Past, Present and Future," *Stroke and Vascular Neurology*, vol. 2, no. 4, 2017, pp. 230–243