



PARSHWANATH CHARITABLE TRUST'S

## A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering  
Data Science



# FemAI ( PCOD & PCOS Risk Assessment System)

**Harsh Lad - 24207002**

**Siddhesh Surve - 24207004**

**Aryan Shelar - 24207012**

**Sujal Jain - 24207015**

**Project Guide  
Ms. Hardiki Patil**

# Outline

- Introduction
- Literature Survey of the existing systems
- Limitations of the existing systems
- Problem statement
- System Design
- Technologies and methodologies
- Implementation
- Conclusion
- References

# Sustainable Development Goals (SDG) mapped

## SDG 3: Good Health and Well-being

- **Mapping:** FEMAI promotes women's reproductive health by enabling **early detection and preventive care** for PCOD/PCOS.
- Ensures accessibility of healthcare information and support, improving **quality of life and well-being**.

# Introduction

- PCOD and PCOS are among the most prevalent women's health challenges, yet many cases remain undiagnosed for years due to vague, overlapping symptoms (like irregular periods, weight gain, acne, and infertility).
- Diagnosis today mostly relies on expensive, time-consuming tests not accessible to all; many women turn to unreliable online sources or self-diagnosis, risking misinformation.
- Missed or late diagnosis leads to increased stress, anxiety, and escalating lifestyle challenges.
- No comprehensive system currently integrates symptom tracking, medical history, and AI-based prediction in a unified platform.
- Recognizing these gaps inspired us to build FEMAI—a support and prediction platform empowering women to detect risk early, receive preventive advice, and seek timely care.

# Introduction

## Objectives of FEMAI

- Develop a predictive model for early detection of PCOD and PCOS.
- Apply machine learning techniques to improve diagnostic accuracy.
- Integrate voice-to-text functionality for user inputs.
- Enable interactive visualization of results and trends and Generate summarized reports with generalized lifestyle guidance

# Literature Survey of the existing system

Sr. No.	Title	Author(s)	Year	Outcomes	Methodology	Result
1	PCOScare: Conventional Machine Learning Classifiers for Diagnosing and Prevention	Vaibhav C. Gandhi, Khyati R. Nirmal, Uma Maheswari, Sudha Rajesh, P. Tharcis, Dhruvi Thakkar	2024	An optimized feature selection approach for early PCOS detection using ML; identifies key predictors and improves model interpretability, achieving higher accuracy.	A variety of feature selection methods (filter, wrapper, embedded), ML classifiers (SVM, RF, XGBoost), and ensemble techniques are applied. Data from clinics is preprocessed, class imbalance is addressed with ENN, and performance metrics are compared.	Feature selection boosts accuracy; best results with Random Forest (95%) and stacking models; important predictors include follicle count, weight gain, and cycle irregularities.

# Literature Survey of the existing system

Sr. No.	Title	Author(s)	Year	Outcomes	Methodology	Result
2	Optimized Machine Learning for Early Detection of Polycystic Ovary Syndrome for women	Bharti Panjwani, Jyoti Yadav, Vijay Mohan, Neha Agarwal, Saurabh Agarwal	2025	Proposes an optimized ensemble learning model (WaOEL) for early PCOS detection using only basic, non-invasive clinical indicators and symptoms, enabling cost-effective screening	Datasets amalgamated and refined to a 12-feature symptomatic dataset; ensemble model stacks seven base learners with deep learning as meta-classifier; hyperparameters optimized with metaheuristic algorithms (WaO, CSO, RSO); feature importance assessed via RF and SHAP values	WaOEL achieved 92.8% accuracy (AUC 0.93); key features: obesity, cholesterol, waist-to-hip ratio; highlights feasibility of PCOS screening from routine, low-cost indicators.

# Literature Survey of the existing system

Sr. No.	Title	Author(s)	Year	Outcomes	Methodology	Result
3	SmartScanPCOS : A feature-driven approach to prediction using Machine Learning and explainable AI	Umaa Mahesswari G, Uma Maheswari P	2024	Developed an XAI-based PCOS prediction tool using a hierarchical two-level Random Forest ensemble, integrating optimized feature selection to improve performance and transparency	Three feature selectors (PCA, Salp Swarm Optimization, Mutual Information) evaluated; nine traditional classifiers and two stacking ensemble models compared; Explainable AI tools (SHAP, LIME, DALEX) are utilized for interpretability	Two-level RF ensemble with 17 selected features achieves 99.31% accuracy and robust explainability; Smart Predictor aligns with clinical expert review and offers real-time feature contribution insights.

# Literature Survey of the existing system

Sr. No.	Title	Author(s)	Year	Outcomes	Methodology	Result
4	A stacked learning framework for accurate classification of polycystic ovary syndrome	Heba M. Emara, Walid El-Shafai, Naglaa F. Soliman, Abeer D. Algarni, Reem Alkanhel, Fathi E. Abd El-Sami	2025	Proposes stacked learning with ADASYN oversampling and BORUTA feature selection for PCOS and cervical cancer classification, enhancing data balance and interpretability	Employs ADASYN, SMOTE, and random undersampling; feature selection by BORUTA; train/test split; stacking uses LR, RF, KNN as base learners and XGBoost meta-learner; metrics: accuracy, recall, ROC, F1 Score	Stacked learning with ADASYN/BORUTA achieves 97% accuracy for PCOS, 99% for cervical cancer; methodology generalizes well to screening more health conditions.

# Limitations of existing systems

- Diagnosis frequently depends on advanced hormonal and imaging tests, making early detection inaccessible and unaffordable for many women, especially in low-resource settings.
- Most publicly available datasets lack demographic and symptom diversity, leading to machine learning models that cannot accurately generalize across different age groups, ethnicities, or comorbid conditions.
- Symptoms of PCOD/PCOS are highly variable and overlapped with other disorders (e.g., thyroid, insulin resistance), complicating accurate labeling for both clinicians and AI systems.

# Limitations of existing systems

- Manual data annotation and verification for ground truth are time-consuming, increasing the risk of inconsistent or noisy training data.
- Data privacy and ethical concerns hinder cross-institutional data sharing, limiting the scale and robustness of models trained on multi-center records.
- Real-time tracking and longitudinal monitoring are rarely supported in existing systems, preventing effective follow-up and assessment of treatment efficacy.
- Many models ignore lifestyle, environmental, and genetic factors which are critical in personalized prediction and risk estimation.

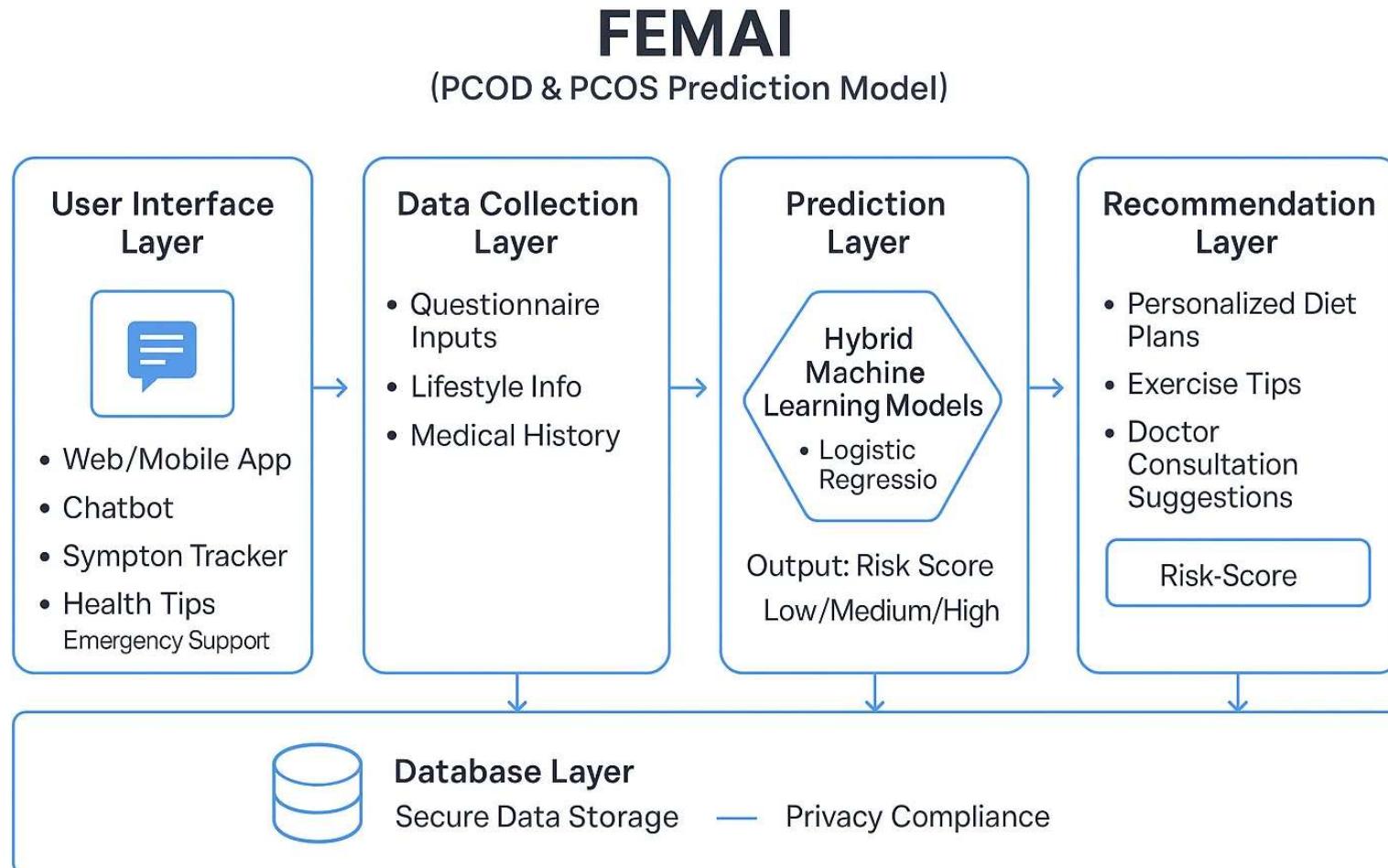
# Problem statement

- Problem: Women often remain undiagnosed due to vague symptoms, costly tests, and lack of awareness.

## Proposed Solution (FEMAI):

- Build a hybrid ML-based model to detect risk of PCOD & PCOS.
- Use questionnaires, medical history, and basic test reports as inputs. Provide early prediction + awareness for preventive care.

# System Design



**Fig 1: System Design**

# **Technologies and methodologies**

## **Technologies Used**

### **1) Frontend (User Interaction):**

- HTML, CSS, Javascript → Responsive web interface.
- Bootstrap for styling.
- Chatbot UI integrated.

### **2) Backend :**

- Python (Flask) for server-side logic.
- REST APIs to connect UI ↔ ML model.

# **Technologies and methodologies**

## **3) Database**

- MySQL (structured data).

## **4) Machine Learning Stack**

- Libraries: Pandas, NumPy, Scikit-learn
- Preprocessing: Data normalization, feature selection
- ML Algorithms:
  - KMeans , Random Forest , Logistic Regression

## **5) Deployment & Hosting**

- .Github or Netlify for Hosting

# **Technologies and methodologies**

## **6) Methodologies :**

- Ensemble Learning Approach → combine multiple models for robust.
- Scaler approach → categorize the data from training dataset.
- Predictions Explanation → show which features influenced the prediction.

# Implementation

The image displays two side-by-side screenshots of a mobile application interface titled "PCOD Prediction Chatbot" with the subtitle "Your Health Companion". The top navigation bar is orange with the app's logo and name. The first screenshot on the left shows the initial part of the questionnaire with fields for Age (28), BMI (27.5), Irregular Periods (Yes), Weight Gain (Yes), and Excess Hair Growth (Yes). The second screenshot on the right shows the continuation of the questionnaire with fields for Acne (Yes), Hair Loss (Yes), Family History (Number of family members with PCOD) (1), and Pain (Yes). A large blue "Predict" button is at the bottom of the second screen.

PCOD Prediction Chatbot  
Your Health Companion

Age:  
28

BMI:  
27.5

Irregular Periods:  
Yes

Weight Gain:  
Yes

Excess Hair Growth:  
Yes

Acne:  
Yes

Hair Loss:  
Yes

Family History (Number of family members with PCOD):  
1

Pain:  
Yes

Predict

fig1. Questionnaire for detecting PCOD

# Implementation

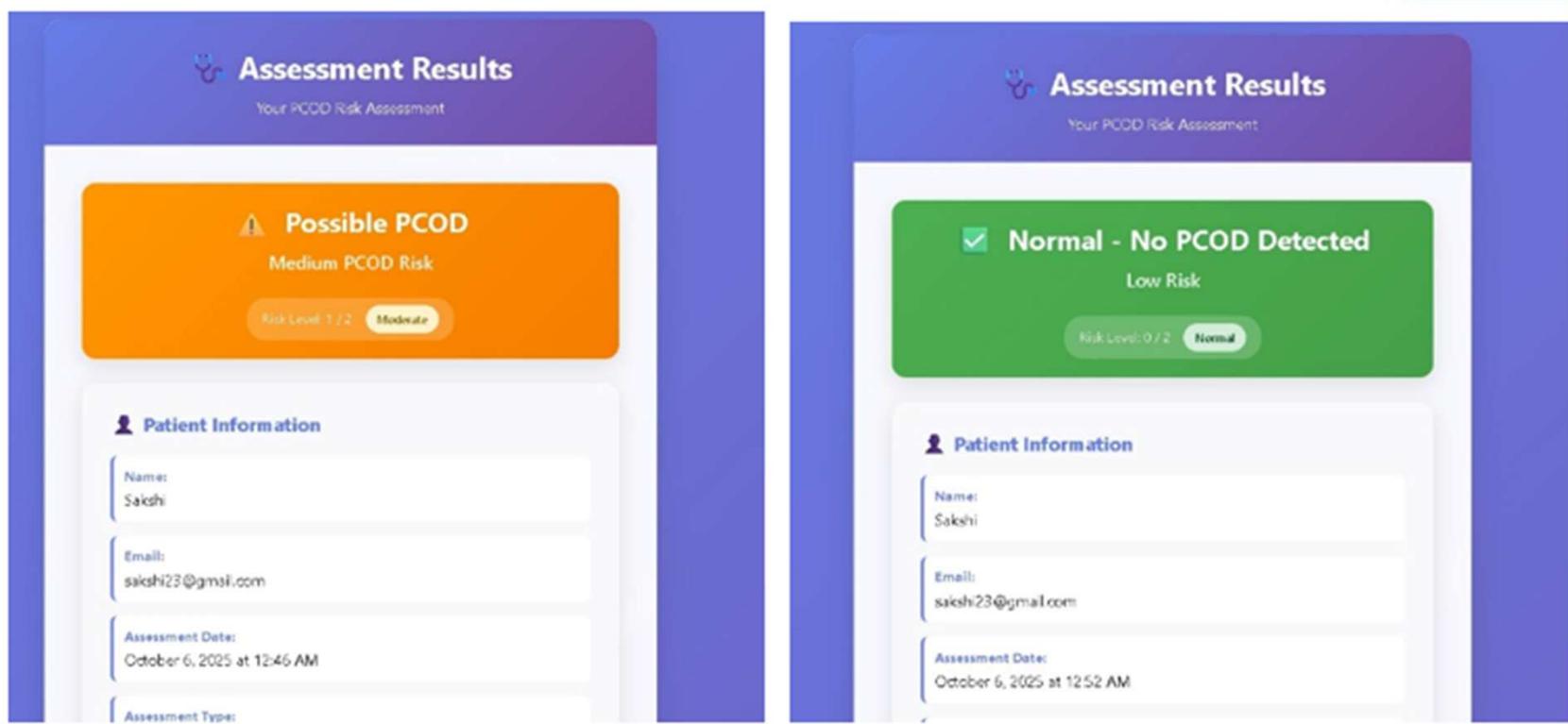


fig2. Assessment Report for PCOD

# Implementation

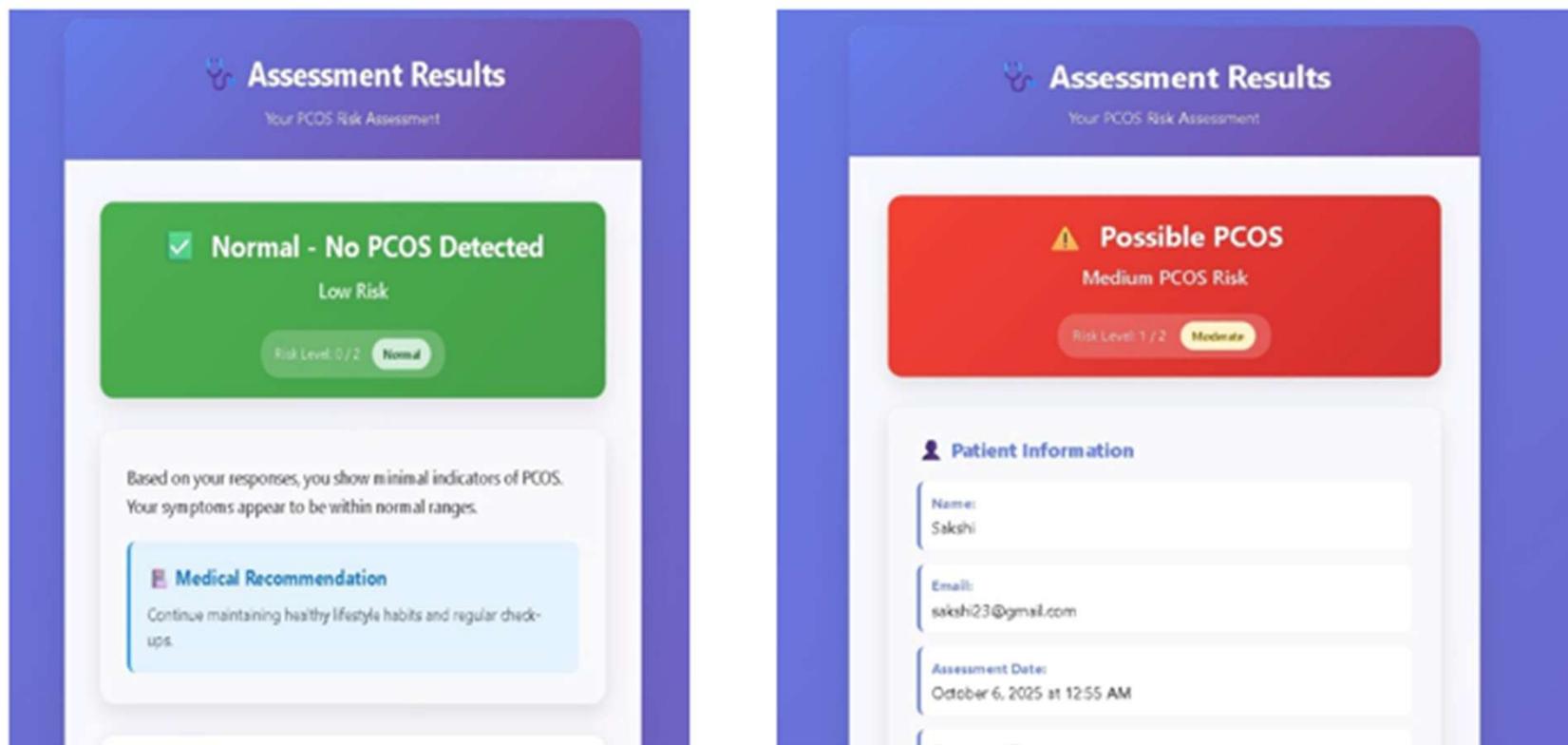


fig3. Assessment Report for PCOS

# Implementation

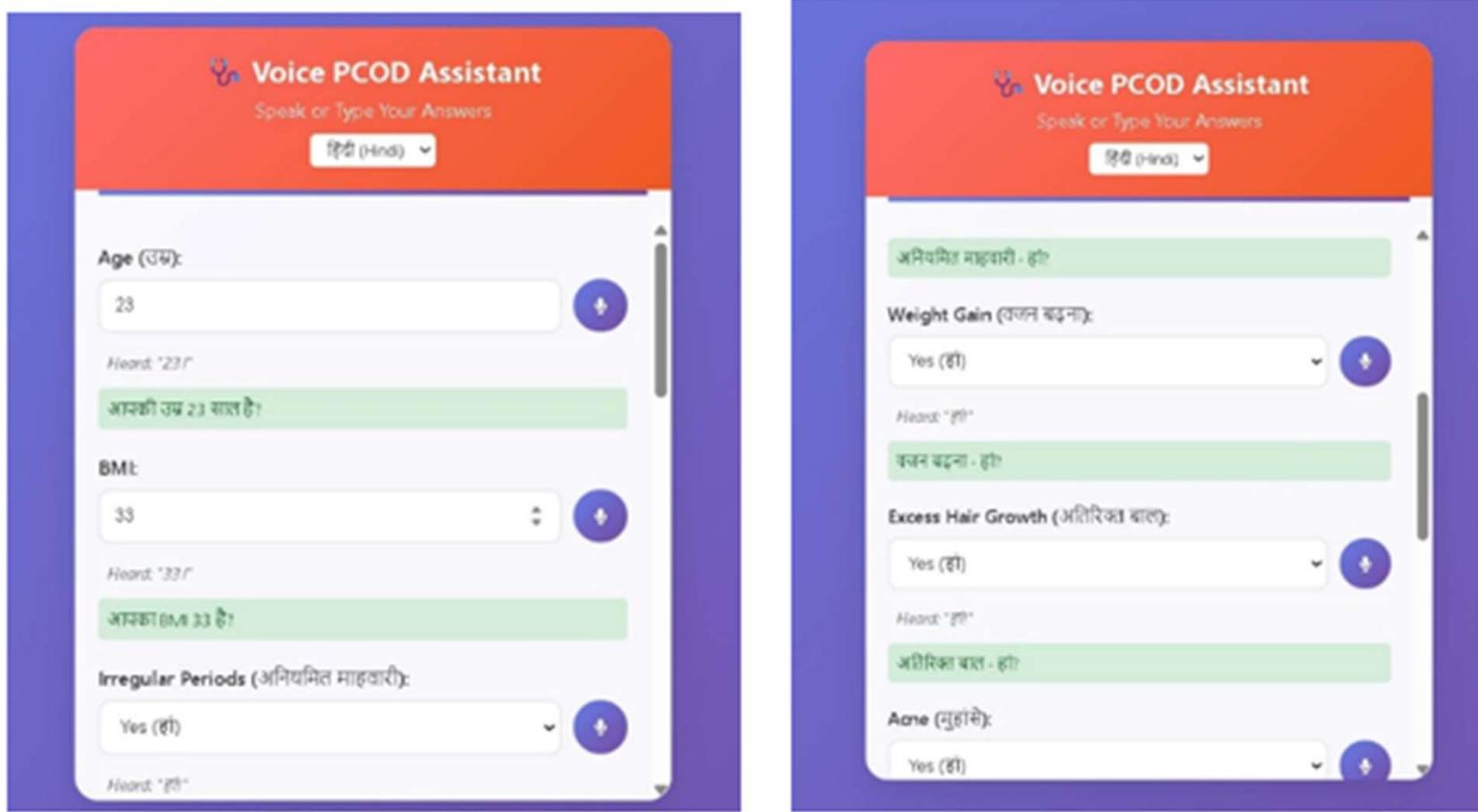


fig4. Voice Assistant for data input.

# Conclusion

- FEMAI provides an accessible, affordable, AI-driven solution.
- Predicts both PCOD & PCOS with better accuracy than existing tools.
- Helps early awareness, reduces dependency on costly tests.
- Future scope:
  - Larger datasets for higher accuracy.
  - Mobile app for rural healthcare access.
  - Integration with wearable health devices.

# References

- [1] V. C. Gandhi, K. R. Nirmal, U. Maheswari, S. Rajesh and P. Tharcis, "PCOScare: Conventional Machine Learning Classifiers for Diagnosing and Prevention," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), vol. 12, no. 3, pp. 3247-3255, 2024  
[https://ijisae.org/index.php/IJISAE/article/view/5930.](https://ijisae.org/index.php/IJISAE/article/view/5930)
- [2] Panjwani, J. Yadav, V. Mohan, N. Agarwal and S. Agarwal, "Optimized Machine Learning for the Early Detection of Polycystic Ovary Syndrome in Women," Sensors, vol. 25, no. 1166, 2025, <https://www.mdpi.com/1424-8220/25/4/1166> doi: 10.3390/s25041166.
- [3] U. G. Mahesswari and P. Maheswari, "SmartScanPCOS: A Feature-Driven Approach to Cutting-Edge Prediction of Polycystic Ovary Syndrome using Machine Learning and Explainable Artificial Intelligence," Computers in Biology and Medicine, vol. 145, pp. 105-117, 2024  
[https://www.sciencedirect.com/science/article/pii/S240584402415236X.](https://www.sciencedirect.com/science/article/pii/S240584402415236X)

# References

- [4] H. M. Emara et al., "A Stacked Learning Framework for Accurate Classification of Polycystic Ovary Syndrome with Advanced Data Balancing and Feature Selection Techniques," *Frontiers in Physiology*, vol. 16, 2025, Art no. 1435036 doi: 10.3389/fphys.2025.1435036.  
<https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2025.1435036/full>.
- [5] L. Morris, T. Qiu and N. Raghuraman, "Federated Learning on Patient Data for Privacy-Protecting Polycystic Ovary Syndrome Treatment," *Proceedings of Medical AI Conferences*, pp. 103-115, 2025  
[https://www.researchgate.net/publication/365174780\\_Federated\\_Learning\\_on\\_Patient\\_Data\\_for\\_Privacy-Protecting\\_Polycystic\\_Ovary\\_Syndrome\\_Treatment](https://www.researchgate.net/publication/365174780_Federated_Learning_on_Patient_Data_for_Privacy-Protecting_Polycystic_Ovary_Syndrome_Treatment).
- [6] K. Vora, A. Shah, N. Shah and P. Verma, "Prediction of Polycystic Ovary Syndrome Using Optimized Machine Learning Classifiers," *NMIMS University Research Journal*, vol. 19, no. 3, pp. 107-120, 2023 <https://www.mdpi.com/1990-8220/25/4/1166>.

**Thank You...!!**