International Journal of Computer Techniques–IJCT Volume 13 Issue 3, May-June - 2026

Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

RESEARCH ARTICLE OPEN ACCESS

OVARIA – A Smart Approach for Early PCOD Prediction

Sushmitha Suresh[1], Meena G[2], Pragna P S[3], Niswana N Swamy[4]

[1]Department of Computer Science & Engineering, K S School of Engineering and Management, Bengaluru, Karnataka
Email: sushmitha.suresh7@gmail.com

[2]Department of Computer Science & Engineering, K S School of Engineering and Management, Bengaluru, Karnataka
Email: meenag@kssem.edu.in

[3]Department of Computer Science & Engineering, K S School of Engineering and Management, Bengaluru, Karnataka
Email: 22pspragna@gmail.com

[4]Department of Computer Science & Engineering, K S School of Engineering and Management, Bengaluru, Karnataka
Email: niswananarayanswamy@gmail.com

----------**********----------

Abstract:

OVARIA is a web-based system designed to estimate the risk of Polycystic Ovarian Disease in women by analyzing nonclinical health and lifestyle information. The application collects data like age, body mass index, menstrual history, and general symptoms through a structured questionnaire and preprocesses these inputs by standard cleaning, encoding, and scaling. A Random Forest classifier trained on this dataset categorizes users into three groups: likely PCOD, no PCOD, and at risk while providing lifestyle changes to maintain PCOD.

This model is integrated on a Flask-based web interface that enables users to sign up, submit their details, view their risk predictions, and review educational content on PCOD. Additionally, the system provides basic lifestyle guidance on diet, physical activity, and self-monitoring to help users balance PCOD. OVARIA is designed to support the early identification of PCOD risk through machine learning integrated into a user-friendly web platform, in particular for young women with little access to specialist healthcare services.

Keywords — PCOD (Polycystic Ovarian Disease), Random Forest Classifier, Flask web-interface

----------**********----------

I. INTRODUCTION

PCOD is among the most common hormonal disorders among women within the childbearing age. It mainly shows up as an irregular menstrual cycle, with development of small cysts in the ovaries and symptoms like acne, excessive hair growth, weight gain, and mood swings. Many women remain unaware of PCOD due to a lack of accessible screening tools and overlapping symptoms with other health issues. This mostly results in late diagnosis, increasing the chances of complications arising, such as infertility, diabetes, and cardiovascular problems.

Early detection and timely lifestyle changes may help manage PCOD effectively and prevent health issues in the long run. The conventional approach towards detection, however, requires clinical tests and specialist visits, which may not be feasibly available to everyone, especially in rural or otherwise underserved areas. To bridge this gap, Ovaria comes into being-a smart, web-based application developed for the prediction of PCOD based on non-clinical health data. Collected through a

user-friendly questionnaire, the system acquires knowledge on age, body mass index, menstrual history, and common symptoms. This collected data is further processed and analyzed in the core of the system by machine learning algorithms, classifying the users as a person likely to have PCOD, not having PCOD, or at risk of developing it.

Ovaria focuses on easy accessibility and convenience of screening for PCOD, especially among young women who do not have regular access to healthcare facilities. Early risk assessment, coupled with lifestyle recommendations through the application, essentially promotes proactive health management and timely medical consultation. The use of machine learning behind a simple web interface gives Ovaria its practicality in creating awareness and promoting early intervention in the management of PCOD.

II. RELATED WORK

C.Venkatachalam et al. [1] Herein, a hybrid model using Random Forest and Extreme Learning Machine was proposed for the improvement of early prediction of PCOS/PCOD and infertility risk. The model achieves 93.5% accuracy through advanced preprocessing and feature selection, outperforming individual methods. It emphasizes early diagnosis to prevent complications and advocates mobile deployment for enhanced rural healthcare access. This work informs the design of ensemble-based PCOD screening applications like OVARIA.

Dhruvi Kapadia et al. [2] This paper implements multinomial logistic regression for PCOS risk prediction using demographic and symptom data from online surveys. The model achieved 82% accuracy, highlighting AI's screening potential. Inclusion of features like acanthosis nigricans was recommended for improved results. The study supports integrating AI assistants and personalized models for early, non-invasive detection and broader awareness, validating digital methodologies similar to OVARIA.

Dhananjay Patel et al. [3] This work describes a telemedicine solution leveraging Random Forest classifiers and integrated video consultations to address PCOS/PCOD diagnosis in rural areas. The system features symptom tracking, educational content, ASHA worker support, and secure communication, enabling reliable early detection and healthcare access. The holistic, socially-aware design aligns with OVARIA's approach and demonstrates the technical and practical value of machine learning in women's health.

Sulekha Kaushik et al. [4] This paper benchmarks multiple algorithms—including Random Forest, CatBoost, Decision Tree, and Support Vector Classifier—for PCOD prediction from clinical data. Random Forest and CatBoost yielded highest accuracy with minimal features. The study underscores gene-environment influences, the diagnostic gap, and supports using AI-driven apps and optimal features for early PCOD screening, validating OVARIA's design choices.

Ms. Srinithi V et al. [5] The study presents some of the explored machine learning algorithms, trained on Kaggle PCOS datasets, using SMOTE for data balancing and CNNs for ultrasound image analysis. Random Forest delivered the highest tabular prediction accuracy, while CNNs excelled at cyst detection. The research recommends hybrid numeric/image systems for improved results, supporting OVARIA's choice of Random Forest and suggesting future extensions for image-based diagnosis.

III. PROBLEM STATEMENT

Many women suffer from Polycystic Ovarian Disease (PCOD), but it often goes undiagnosed due to lack of awareness and delayed screening. There is a need for an accessible tool that can help in early identification of PCOD risk based on basic health and lifestyle data.

IV. OBJECTIVES

The main objectives of this research are:

1. To predict the risk of PCOD using user input and health data and build a simple and easy-to-use application for women.
2. To help users understand the common symptoms of PCOD.
3. To provide basic lifestyle tips to manage PCOD.
4. To encourage users to seek medical advice when risk is detected.
5. To support early screening to avoid long-term health issues.

V. SYSTEM ARCHITECTURE

A. System Components

The proposed system consists of the following components:

1. **User Interface (Frontend)**
   - Web-based user interface built using HTML, CSS, and JavaScript that offers users access to an interactive interface that is responsive and accessible.
   - It includes forms for user registration, login, and questionnaire submission to collect relevant health and lifestyle data.
   - Users can view their risk prediction results, educational content about PCOD, and personalized lifestyle recommendations.

2. **Web Application Server (Backend)**
   - Implemented with Flask framework in Python, the backend handles client requests, processes user inputs, and communicates with machine learning model.
   - It manages user authentication, session management, and data validation to ensure security and data integrity.
   - The backend also stores user details and questionnaire responses in database for model training and evaluation purposes.

3. **Database**
   - A relational database stores users profiles and tracking entries.
   - It supports efficient data retrieval and management for both users and system administrators.

B. Workflow

1. **User Registration**
   - New users register and sign in by providing basic credentials such as name, email, and password.

2. **Questionnaire Data Collection**
   - The users log in and access the questionnaire to input various health, lifestyle, and demographic particulars required for risk assessment of PCOD.
   - Data points include age, BMI, menstrual cycle patterns, symptoms such as acne or hair thinning, family history, and lifestyle factors like diet and exercise habits.

3. **Preprocessing of Input Data**
   - The backend receives the raw questionnaire data and passes it through the preprocessing module.

4. **Machine Learning Prediction**
   - The preprocessed data is fed into a Random Forest classification model.
   - The model analyzes the input features and predicts one of the three risk classes: 'Low/No Risk of PCOD', 'Moderately /Likely to have PCOD', and 'High Risk of PCOD'.

5. **Result Presentation**
   - The prediction outcome is sent back to the frontend and displayed to the user in a clear and visually using donut chart along with risk percentage.

6. **Lifestyle Change Recommendations and educational content**
   - Based on the prediction, the system advises on lifestyle changes, sleep hygiene, etc.
   - It also offers an educational section explaining PCOD, its symptoms, risks, and causes of PCOD.

VI. METHODOLOGY

1. **Data Preprocessing:** Handled missing values by imputation, encoded categorical labels numerically, and normalized input features to ensure consistent data for machine learning analysis.
2. **Model Training:** Employed a Random Forest classifier to learn patterns in the processed dataset and predict PCOD risk categories based on user inputs.
3. **Evaluation:** Assessed model accuracy and overall performance using metrics such as accuracy, precision, recall and confusion matrix analysis.
4. **Deployment:** Implemented a web-based user interface using Flask, allowing users to register, submit questionnaire responses, and receive risk predictions along with lifestyle recommendations.

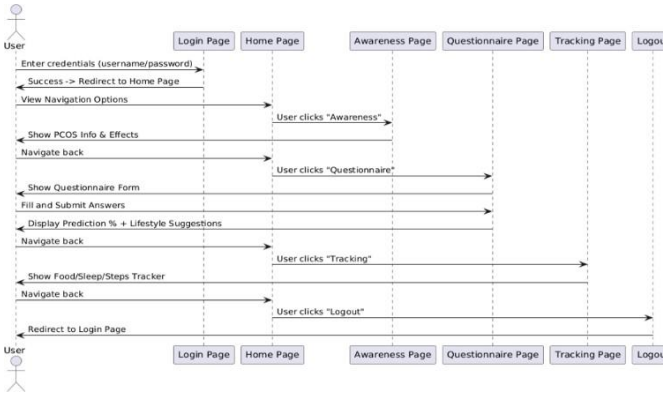

Fig. 1 High-Level Design

1. **User Authentication and Navigation**: Users start by entering their login credentials, and upon successful authentication, they are redirected to the Home Page where all main navigation options are available.

2. **Awareness Information Access**: It will allow the user to view the educational content about PCOS/PCOD, its symptoms, and effects from the Home Page. They can navigate back to continue using the other features.

3. **Health Questionnaire and Risk Prediction:** Users proceed to the Questionnaire Page, answer a set of health and lifestyle questions, and submit their responses. The system displays a prediction percentage for PCOD risk along with personalized lifestyle suggestions based on the model output.

4. **Tracking Features**: Users can access the Tracking Page, where they can log and track daily metrics such as food intake, sleep, and physical activity, helping them monitor lifestyle factors linked to PCOD.

5. **Session Management and Logout**: From any page, users can choose to log out, which securely ends their session and redirects them back to the Login Page, ensuring user data privacy and security.

VII. RESULTS AND DISCUSSIONS

The system was tested on few women who are known to have PCOD, do not have PCOD, and women having an imbalanced lifestyle:

1. **Model Performance:** The Random Forest classifier correctly classified users into 'Low/No PCOD', 'Moderately/Likely to have PCOD', and 'At High Risk of PCOD' categories.

2. **Accuracy Metrics:** The Random Forest classifier achieved 90% accuracy on the PCOD test dataset, demonstrating strong and reliable prediction.

3. **Navigation and Accessibility:** The frontend uses a navigation bar and clear page layouts, allowing users to move easily between login, home, questionnaire, tracking, and logout sections. Information is organized so users can quickly find and use all core system features.

4. **Responsive Data Interaction:** Interactive forms enable users to enter their health and lifestyle information for PCOD risk assessment. Results and recommendations are displayed instantly.

5. **Tracking Page:** Tracking page allows the users to log daily habits, step count or also maintain notes on anything.

VIII. SNAPSHOTS

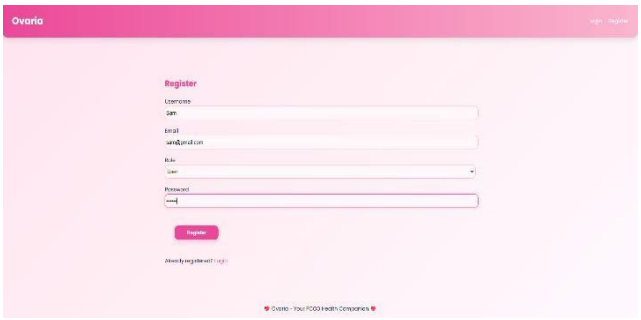

Fig. 2 Register Page

It provides a page for registration where users can sign up by providing their required credentials.

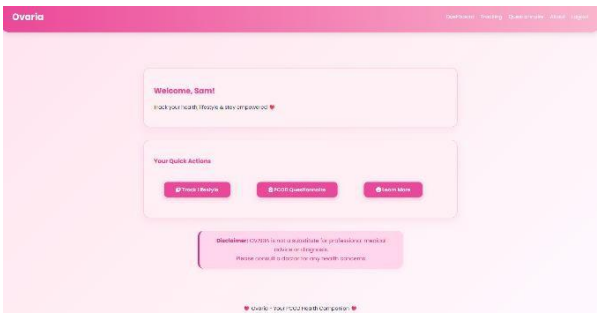

Fig. 3 Home Page

- The home page provides a welcoming interface, greeting the user by name and encouraging active health tracking and empowerment through friendly messaging.
- Key actions are made easily accessible with prominent buttons for tracking lifestyle, accessing the PCOD questionnaire, and exploring more educational content, streamlining user navigation and engagement.
- There is a persistent disclaimer, stating that OVARIA is not a replacement for a medical professional, and users should consult a doctor for serious health concerns.
- The header navigation bar will provide easy access to the dashboard, tracking, questionnaire, about information, and logout for smooth navigation throughout the site.

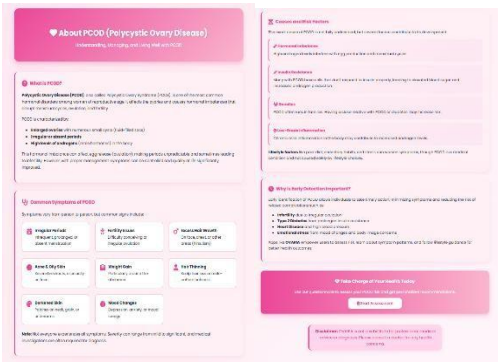

Fig. 4 About Page

- The About page offers a comprehensive and accessible overview of PCOD (Polycystic Ovary Disease), summarizing its definition, hormonal causes, and impact on menstrual cycles, ovulation, and fertility.
- Key clinical features displayed include enlarged ovaries with cysts, irregular periods, and increased androgens. Common symptoms such as fertility issues, acne, darkened skin, excess hair, mood changes, weight gain, and hair thinning are clearly communicated for user awareness.
- Detailed sections explain risk factors (hormonal imbalance, insulin resistance, genetics, low-grade inflammation) and stress the role of lifestyle choices, while emphasizing PCOD is a medical condition requiring proper management.
- Early detection is important to minimize risks such as infertility, diabetes, heart disease, and emotional stress; OVARIA is therefore a helpful tool for risk assessment and education.

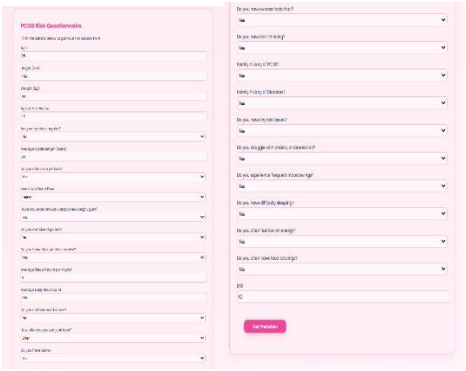

Fig. 5 Questionnaire Page

- The questionnaire page provides a comprehensive form for users to enter demographic, menstrual, lifestyle, and symptom-related information necessary for PCOD risk assessment.
- Input fields include age, height, weight, cycle details, dietary habits, exercise patterns, skin/hair issues, family history, and other clinical indicators, enabling precise prediction using machine learning.
- Drop-down menus and free-form fields ensure quick, accurate data collection and improve usability for users with varied backgrounds.

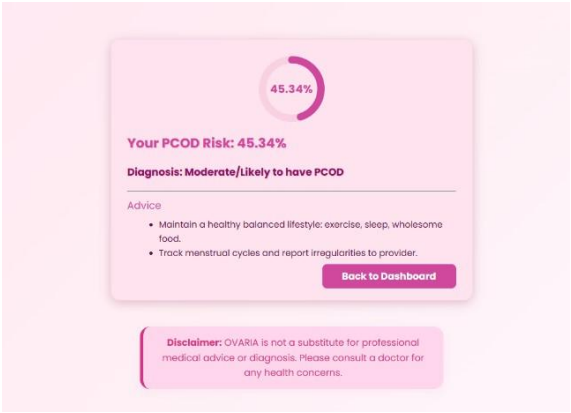

Fig. 6 Result Page

- The web application calculates and visually displays the user's PCOD risk percentage, providing quantifiable results; for example, the result shown is 45.34% risk, categorized as moderate or likely to have PCOD.
- The system delivers based on the predicted risk, recommending healthy lifestyle changes

and prompting users to monitor their menstrual cycles and report irregularities to healthcare providers.

- Results are presented in a user-friendly format with clear color coding and actionable next steps, enhancing understanding and engagement for non-expert users.
- The interface includes a clear disclaimer, emphasizing that the platform is not a replacement for medical diagnosis and encouraging users to consult professionals for any health concerns.

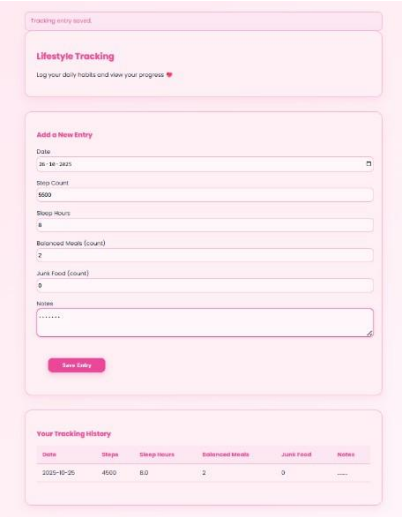

Fig. 7 Tracking Page

- The tracking page enables users to record daily lifestyle metrics including step count, sleep hours, balanced meal intake, and junk food consumption, promoting self-monitoring and healthy habits.
- A user-friendly form allows easy entry of data for each day, with options to add notes for additional context or observations.
- This feature encourages regular engagement with health behaviors and supports personalized lifestyle management as part of PCOD risk reduction.

IX. CONCLUSIONS

OVARIA illustrates the benefit of the integration of machine learning techniques within an easy-to-use web platform for early risk prediction of PCOD. The collection of basic health and lifestyle data and the use of a Random Forest classifier allow the system to screen individuals efficiently, even outside clinical settings. OVARIA offers users personalized risk assessments, along with practical lifestyle suggestions and educational resources to facilitate proactive health choices for women. This digital method fills gaps in awareness and access, which otherwise might not be sought in a timely manner, especially for young women. Further refinement and wider dissemination of such technology-driven innovations have the potential to enhance preventive healthcare and ensure early intervention in conditions like PCOD.

ACKNOWLEDGMENT

The authors acknowledge with thanks the guidance and support by the faculty throughout the project. Thanks to the Department of Computer Science and Engineering for the resource facilities, constructive suggestions, and feedback through all stages of development. The authors are grateful to all their peers and family members for their encouragement and valuable support. The authors would like to thank K. S. School of Engineering and Management for providing a research environment and necessary infrastructure. Healthcare professionals and contributors of data are also thanked for providing anonymized PCOD datasets to develop models. Lastly, the authors extend heartfelt thanks to all concerned for directly or indirectly helping them to complete the project successfully.

REFERENCES

[1] [1] Srinithi, V. & Rekha, R., (2023) “Machine Learning for Diagnosis of Polycystic Ovarian Syndrome (PCOS/PCOD)”, Proc. Int. Conf. Intelligent Systems for Communication, IoT and Security (ICISCoIS).
[2] [2] Kapadia, D. & Jain, R., (2023) “PCOS Prediction: Advancements in Medical Informatics Using Artificial Intelligence”, Proc. 2nd Int. Conf. Futuristic Technologies (INCOFT).
[3] [3] Kaushik, S. & Mishra, S. K., (2023) “Prediction of PCOD using Machine Learning Algorithms”, Proc. 14th Int. Conf. Computing Communication and Networking.
[4] [4] Venkatachalam, C. & Rani, V., (2024) “Detection of PCOS/PCOD and Prediction of Infertility at the Early Stage Based on RF-ELM Technique”, Proc. Asian Conf. Intelligent Technologies (ACOIT).
[5] [5] Patel, D., Mhaskar, M., Shendage, R., Modi, R., Sarvade, R. & Ram, A., (2024) “Machine Learning-Enhanced Solutions for Mitigating PCOS/PCOD-Related Women’s Health Disparities in Rural Areas: A Holistic Telemedicine System Design and Implementation”, Proc. Int.

Conf. Brain Computer Interface & Healthcare Technologies (iCon-BCIHT).

[6] [6] Dutta, P., Paul, S. & Majumder, M., (2021) "An Efficient SMOTE Based Machine Learning Classification for Prediction & Detection of PCOD", Research Square.

[7] [7] Bharati, S., Podder, P. & Mondal, M. R. H., (2020) "Diagnosis of Polycystic Ovary Disease Using Machine Learning Algorithms", 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh.

[8] [8] Denny, A., Raj, A., Ashok, A., Ram, C. M. & George, R., (2019) "i-HOPE: Detection and Prediction System for Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques", TENCON 2019 – 2019 IEEE Region 10 Conference (TENCON), Kochi, India.

[9] [9] Saravanan, A. & Sathiamoorthy, S., (2018) "Detection of Polycystic Ovarian Syndrome: A Literature Survey", Asian Journal of Engineering and Applied Technology.

[10] [10] Satish, C. R., Nandipati, C. R., Chew, X. Y. & Khaw, K. W., (2020) "Polycystic Ovarian Syndrome (PCOD) Classification and Feature Selection by Machine Learning Techniques".