ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1463-1474
Shiva Srinitha *et. al.,* / International Journal of Engineering & Science Research

PCOS Detection System

K.Shireesha, Shiva Srinitha, Thungaturthy Tanmai
[1] Associate Professor, Department Of Cse, Bhoj Reddy Engineering College For Women, India.
[2,3]B. Tech Students, Department Of Cse, Bhoj Reddy Engineering College For Women, India.

ABSTRACT
Polycystic Ovary Syndrome is one of the familiar hormonal diseases which occurs during reproductive age of women life. It is increasing rapidly among women due to a destructive lifestyle. About 1 in 10 women of childbearing age are affected by PCOS. Polycystic Ovary Syndrome is mainly occurring of having cysts in the ovary. PCOS is caused in women because of hormonal imbalance.

1-INTRODUCTION

Polycystic Ovary Syndrome is one of the familiar hormonal diseases which occurs during reproductive age of women life. It is increasing rapidly among women due to a destructive lifestyle. About 1 in 10 women of childbearing age are affected by PCOS. Polycystic Ovary Syndrome is mainly occurring of having cysts in the ovary. PCOS is caused in women because of hormonal imbalance. The aim of this project is to develop a system that utilizes machine learning techniques to detect and predict the risk of Polycystic Ovary Syndrome (PCOS) in individuals.

It helps in providing actionable insights for personalized prevention strategies.

This approach can help in initiating timely interventions to manage symptoms and prevent long-term health issues like infertility, diabetes, and heart disease.

Existing System

In the existing system for PCOS detection, healthcare professionals, such as gynecologists and endocrinologists, manually assess a patient's symptoms, medical history, and test results to diagnose the condition. The current diagnostic process typically involves physical examinations, blood tests, ultrasound scans, and symptom analysis. However, this approach can be time-consuming, subjective, and prone to inconsistencies due to the reliance on human judgment.

Proposed System

The proposed system for the detection and prevention of PCOS using machine learning (ML) aims to automate the process of identifying the risk of PCOS, diagnosing it early, and providing personalized recommendations for management.

2-LITERATURE REVIEW

Polycystic Ovary Syndrome (PCOS) is one of the most common endocrine disorders affecting women of reproductive age, characterized by hormonal imbalances, irregular menstrual cycles, and the presence of ovarian cysts. Early and accurate diagnosis is critical to prevent long-term complications such as infertility, type 2 diabetes, cardiovascular diseases, and endometrial cancer.

Several studies have focused on developing systems for the detection and diagnosis of PCOS using a range of clinical, hormonal, and imaging parameters. Traditional diagnostic methods rely on the Rotterdam criteria, which include at least two of the following: oligo/anovulation, clinical or biochemical signs of hyperandrogenism, and polycystic ovaries visible on ultrasound. However, these criteria can be ambiguous and subject to interpretation, leading to inconsistent diagnoses.

Recent research has explored the use of machine learning (ML) and artificial intelligence (AI) to enhance PCOS detection. Algorithms such as

Decision Trees, Support Vector Machines (SVM), Random Forests, and Neural Networks have been employed to analyze patient datasets and classify PCOS cases with improved accuracy. For instance, studies utilizing datasets from the UCI Machine Learning Repository have demonstrated that ML models can achieve accuracy rates above 90% when trained on relevant features like BMI, insulin levels, glucose levels, and menstrual cycle data.

Furthermore, researchers have investigated the integration of ultrasound imaging and image processing techniques for automated detection of ovarian cysts. Techniques such as edge detection, texture analysis, and deep learning-based segmentation have shown promise in identifying cystic structures with minimal human intervention.

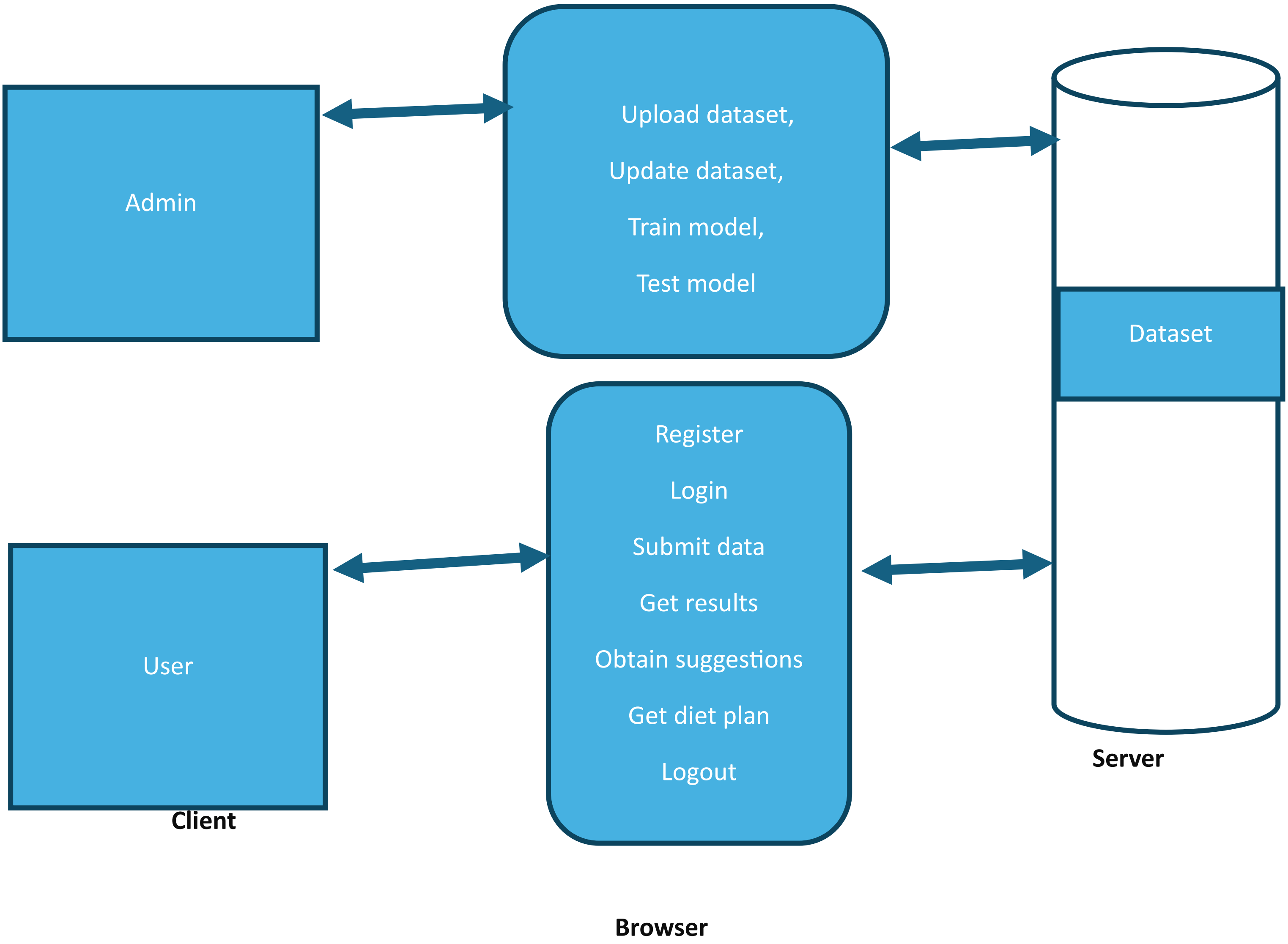
Additionally, some works have explored mobile health (mHealth) applications and wearable technologies to track physiological parameters and symptoms over time, offering a more personalized approach to PCOS management and detection.

Despite advancements, challenges remain in terms of data availability, generalizability of models, and interpretability of results. Many models are trained on small, region-specific datasets and lack the diversity needed for broader clinical application. As such, ongoing efforts aim to build more robust, interpretable, and clinically integrated PCOS detection systems.

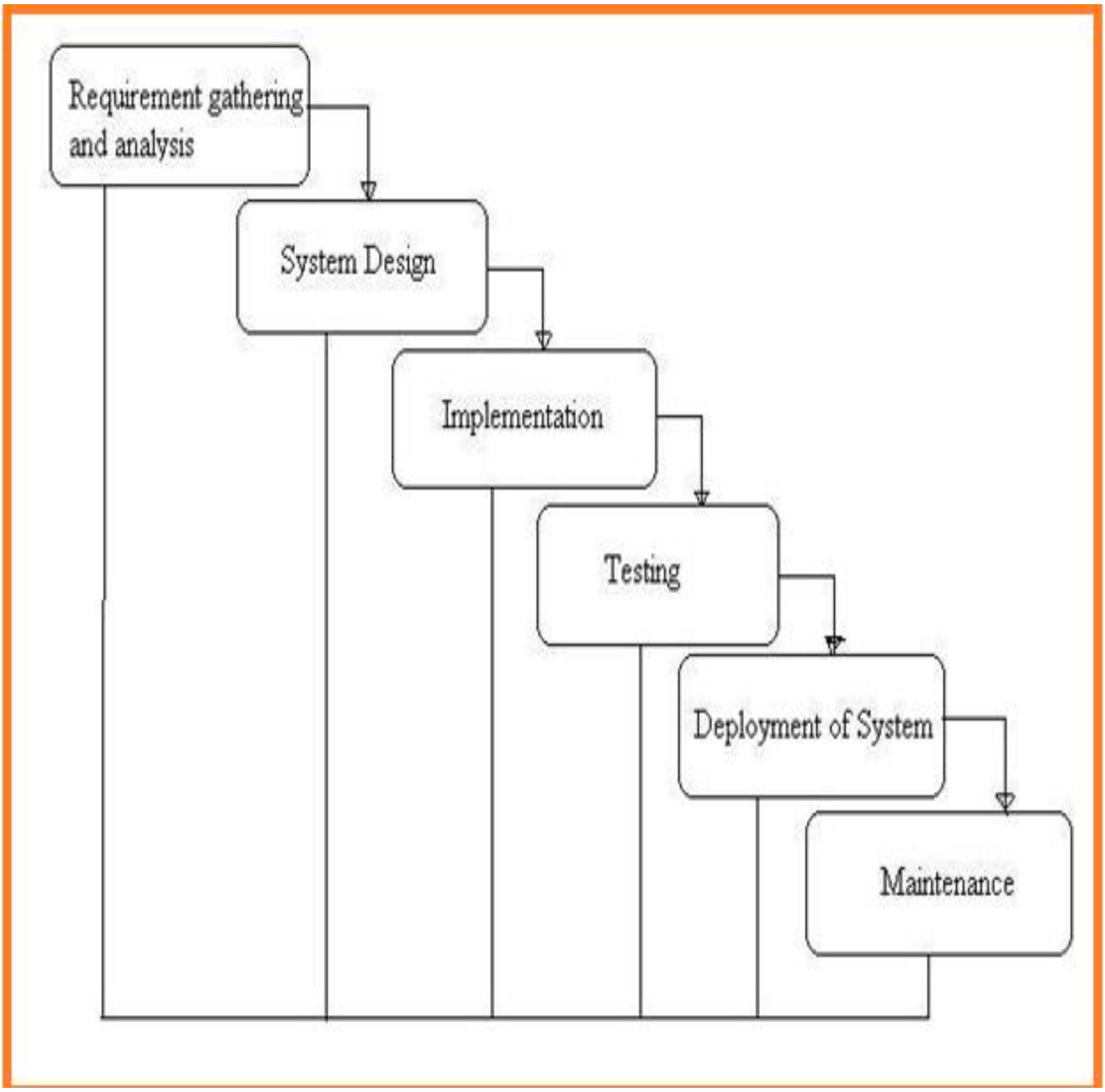
3-DESIGN

- Design represents the number of components we are using as a part of this project and the flow of request processing i.e., what components are processing the request and in which order.
- Architecture in software is a formal description and representation of a system organized in a way that supports reasoning about the structure of the system. It's about making fundamental design decisions that will impact the system's characteristics like performance, scalability and maintainability.

SOFTWARE ARCHITECTURE:



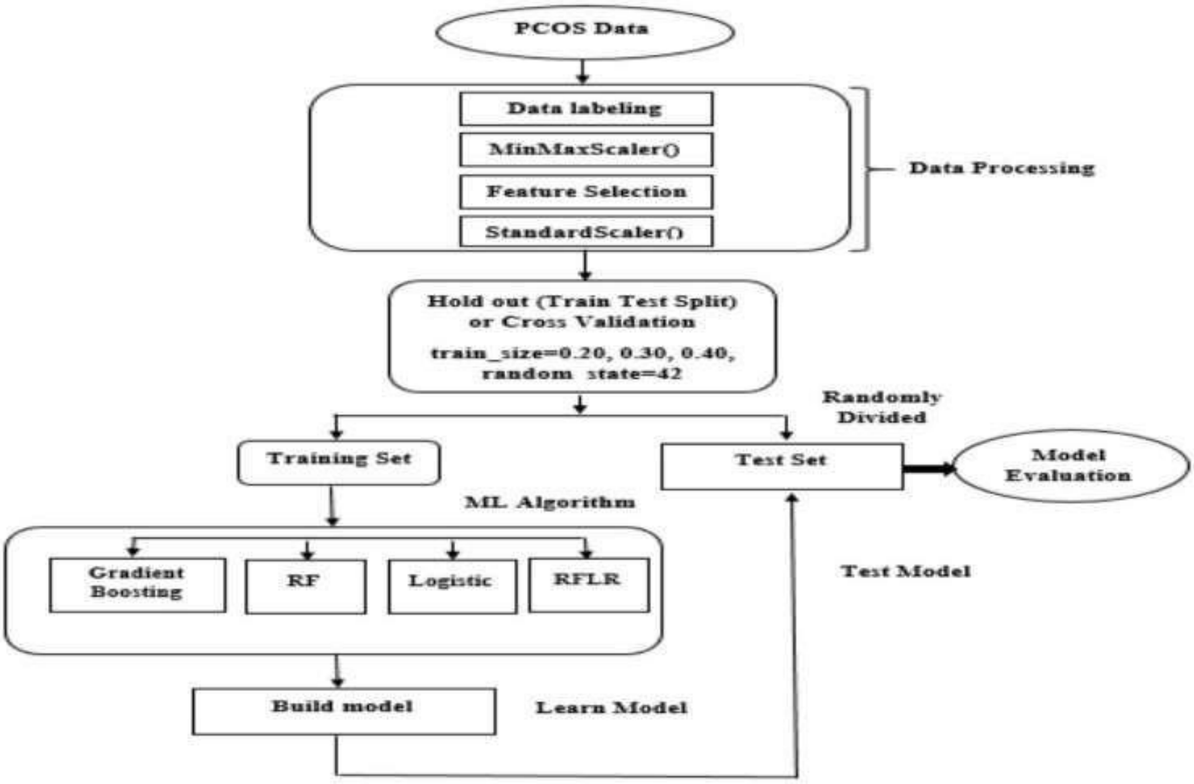
SOFTWARE PROCESS MODEL:



DATA FLOW DIAGRAM

The Data Flow Diagram (DFD) of a PCOS Detection System outlines the flow of data between various components involved in diagnosing PCOS. The system begins with user input, where patients or clinicians enter medical and clinical data such as BMI, hormone levels, and menstrual history. This data is processed by a diagnostic engine that may use machine learning algorithms to analyze patterns and predict PCOS likelihood. The results are then stored in a database and presented to the user through a user interface. Feedback and reports can also be generated for further consultation or treatment planning.

ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1463-1474
Shiva Srinitha *et. al.,* / International Journal of Engineering & Science Research

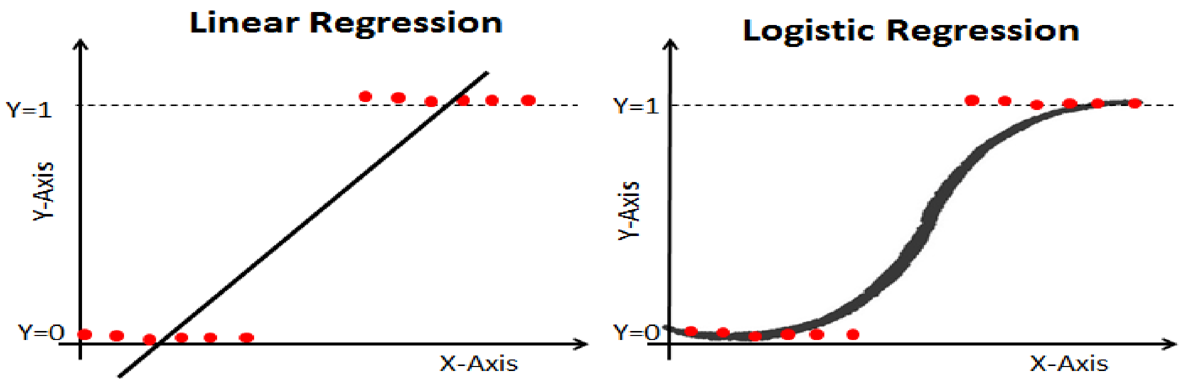


4-IMPLEMENTATION

ALGORITHMS

Logistic Regression:

In linear regression, the linear regression hyperplane that is obtained cannot be used to predict the dependent variable by using the independent variable.

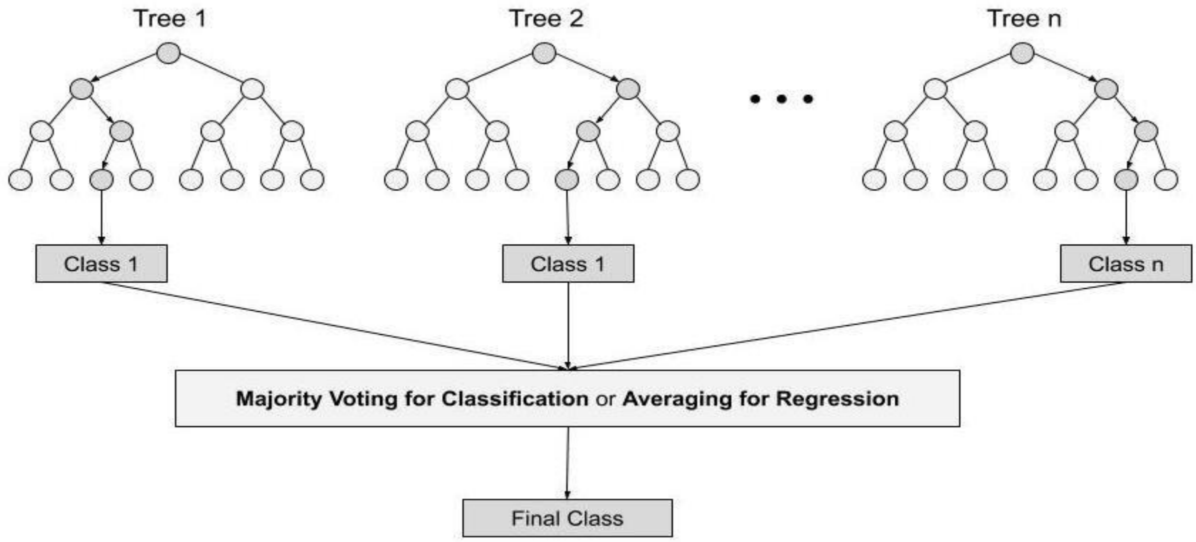


- Hence, when there is categorical data, logistic regression is used. Logistic Regression predicts whether something is true or false instead of predicting something continuous. It is used for classification.
- The sigmoid function is used to convert the independent variable into an expression of probability which ranges from 0 and I concerning the dependent variable.
- The ability to provide probabilities and classify new samples using continuous and discrete measurements makes it a popular Machine Learning algorithm. A drawback of Logistic Regression is the assumption of linearity between the dependent and independent variables.

- Using cross-validation helps ensure that the model works well on new data and doesn't just memorize the training data. It's also important to look at various performance metrics like precision and recall rather than just accuracy, especially in noisy datasets.
- Lastly, more advanced algorithms, like neural networks, can offer additional flexibility in capturing complex patterns.

Random Forest

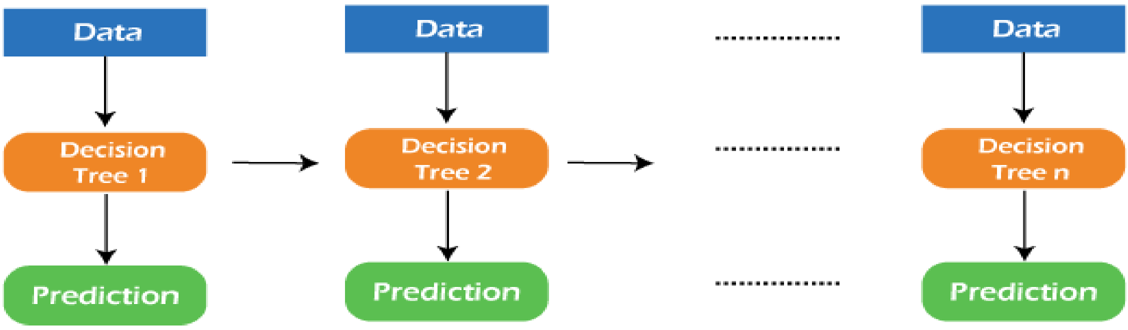
- Random forest is a supervised learning algorithm. It is a collection of Decision Trees. Decision Tree is hierarchical in nature in which nodes represent certain conditions on a particular set of features, and branches split the decision towards the leaf Nodes.
- Leaf determines the class labels.



- Random Forest is an ensemble learning technique, which means it combines the predictions of multiple models (in this case, decision trees) to improve performance and reduce overfitting.
- By averaging (in regression) or taking the majority vote (in classification) from several decision trees, the overall result is more robust and accurate than a single decision tree.
- In this method, each model tries to correct the errors made by the previous models. Each model is trained on a modified version of the dataset, the instances that were misclassified by the previous models are given more weight. The final prediction is made by weighted voting.
- It is suitable for high dimensional data modelling as it can handle missing values, continuous, categorical and binary data but for very data sets, the size of the trees can take up a lot of memory. It can tend to over-fit, so there is a need to tune the hyper-parameter

Gradient Boosting

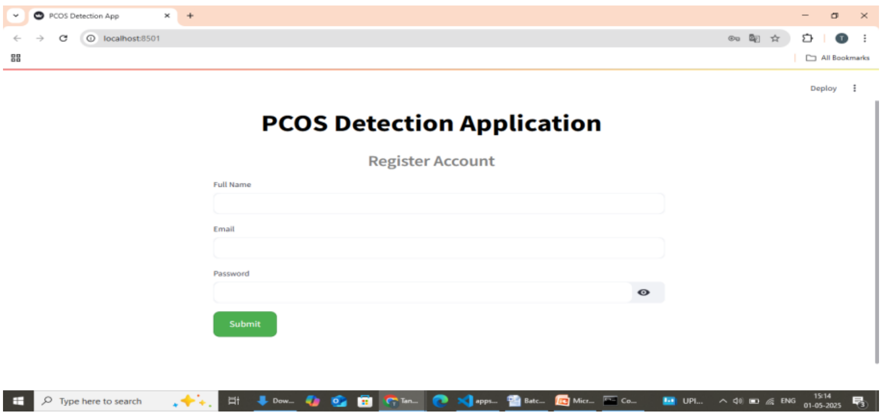
- Gradient boosting is a machine learning **ensemble technique** that sequentially combines the predictions of multiple weak learners, typically decision trees. It aims to improve overall predictive performance by optimizing the model's weights based on the errors of previous iterations, gradually reducing prediction errors and enhancing the model's accuracy. This technique is most commonly used for **linear regression**.

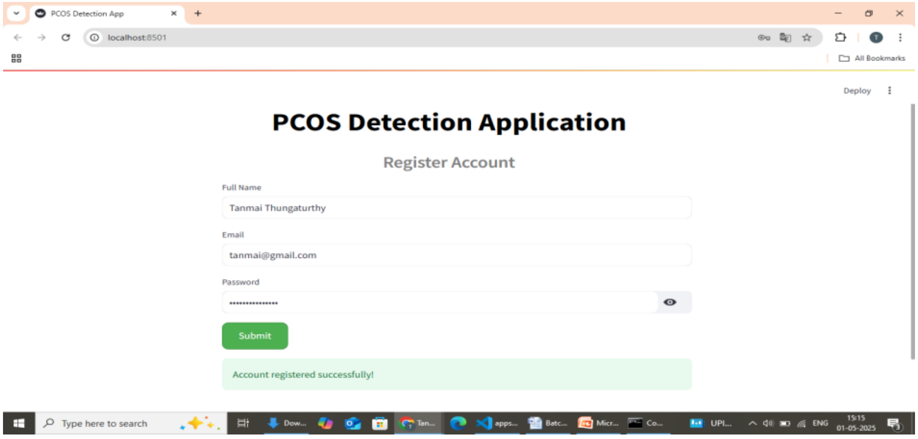


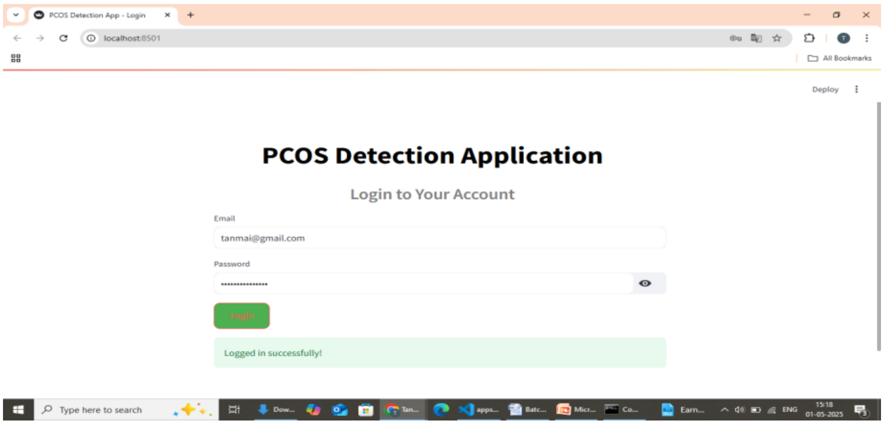
- A **weak learner** is a machine learning model that is slightly better than a random guessing model. For example, let's say we are classifying mushrooms into edible and inedible. If a random guessing model is 40% accurate, a weak learner would be just above that: 50-60%.
- Boosting combines dozens or hundreds of these weak learners to build a strong learner with the potential for over 95% accuracy on the same problem.
- The most popular weak learner is a decision tree, chosen for their ability to work with almost any dataset.
- Errors play a major role in any machine learning algorithm. There are two main types of errors: bias error and variance error. The gradient boost algorithm helps us minimize the model's bias error.
- The main idea behind this algorithm is to build models sequentially, and these subsequent models try to reduce the errors of the previous model.
- When the target column is continuous, we use a Gradient Boosting Regressor; when it is a classification problem, we use a Gradient Boosting Classifier.

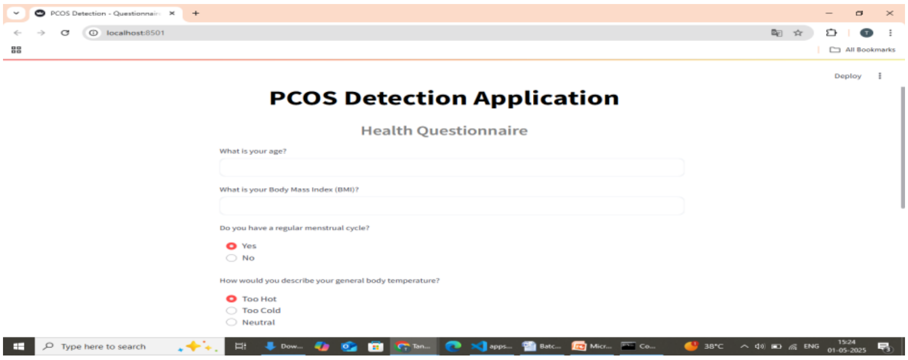
The only difference between the two is the "Loss function". The objective is to minimize this loss function by adding weak learners using gradient descent.

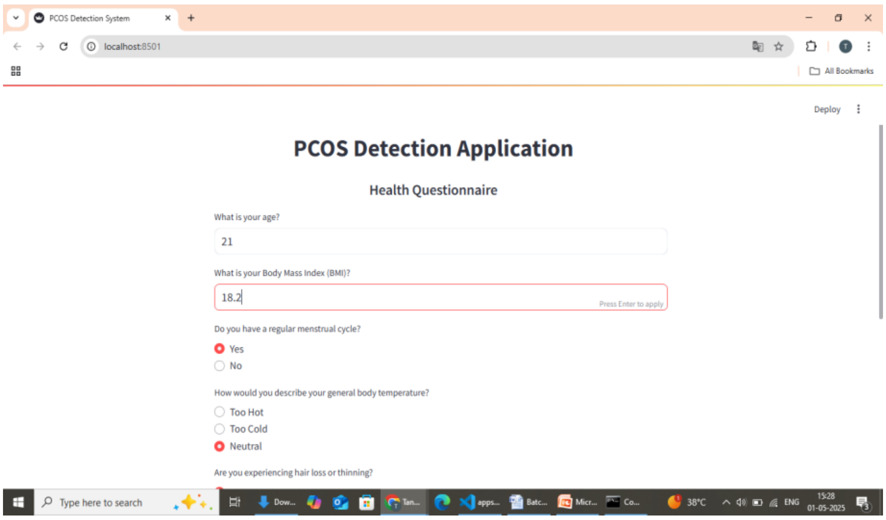
5-SCREENSHOTS:

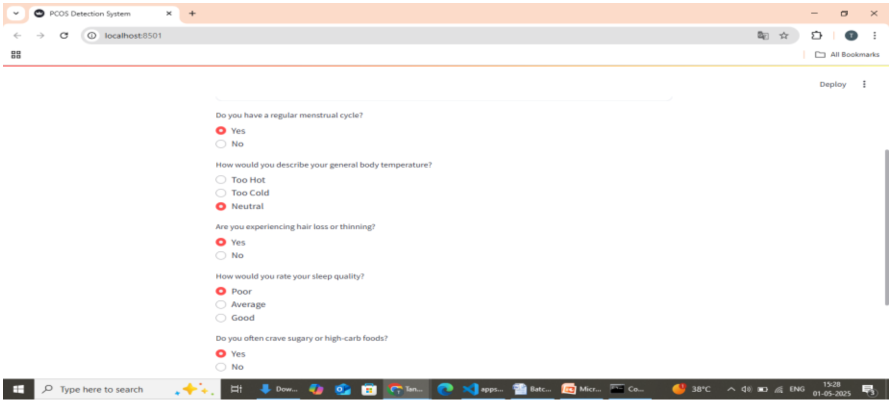


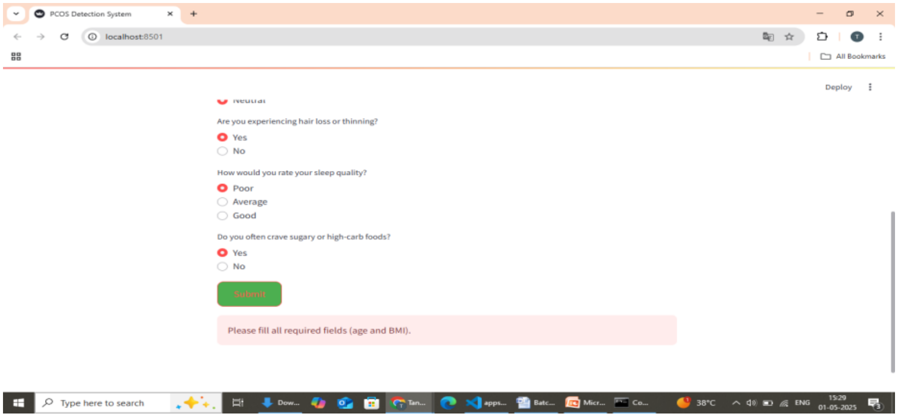

PCOS Detection App
localhost:8501
Deploy
PCOS Detection Application
Register Account
Full Name
Tanmai Thungaturthy
Email
tanmai@gmail.com
Password
Submit
Account registered successfully!
Type here to search

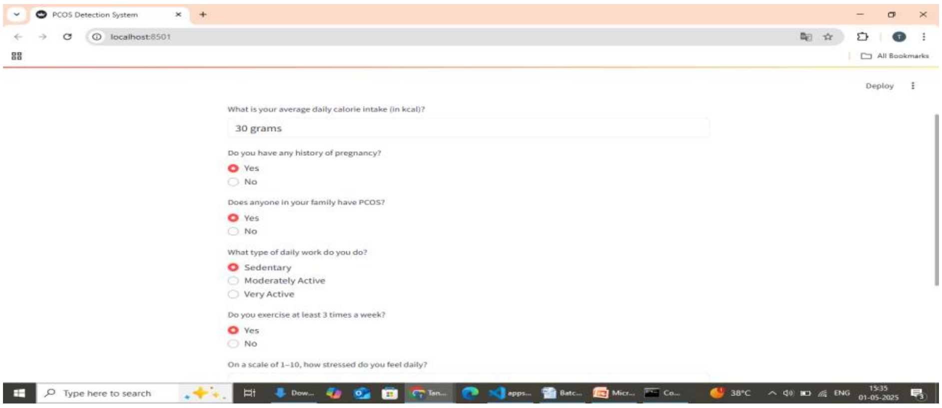

PCOS Detection App - Login
localhost:8501
Deploy
PCOS Detection Application
Login to Your Account
Email
tanmai@gmail.com
Password
Logged in successfully!
Type here to search

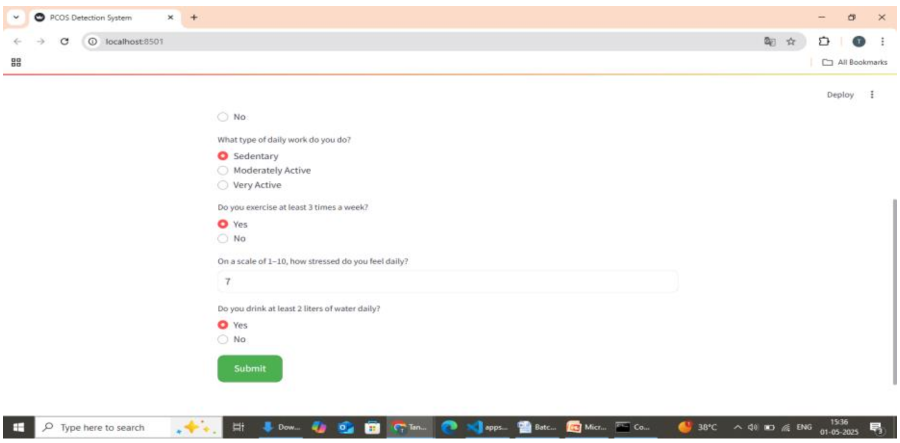

localhost:8501
Deploy
PCOS Detection Application
Health Questionnaire
What is your age?
What is your Body Mass Index (BMI)?
Do you have a regular menstrual cycle?
Yes
No
How would you describe your general body temperature?
Too Hot
Too Cold
Neutral
Type here to search

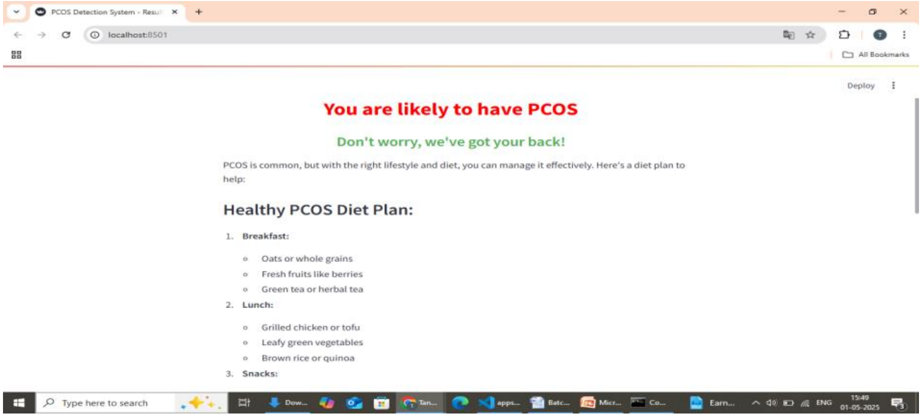

PCOS Detection System
localhost:8501
Deploy
PCOS Detection Application
Health Questionnaire
What is your age?
21
What is your Body Mass Index (BMI)?
18.2
Do you have a regular menstrual cycle?
Yes
No
How would you describe your general body temperature?
Too Hot
Too Cold
Neutral
Are you experiencing hair loss or thinning?
Type here to search

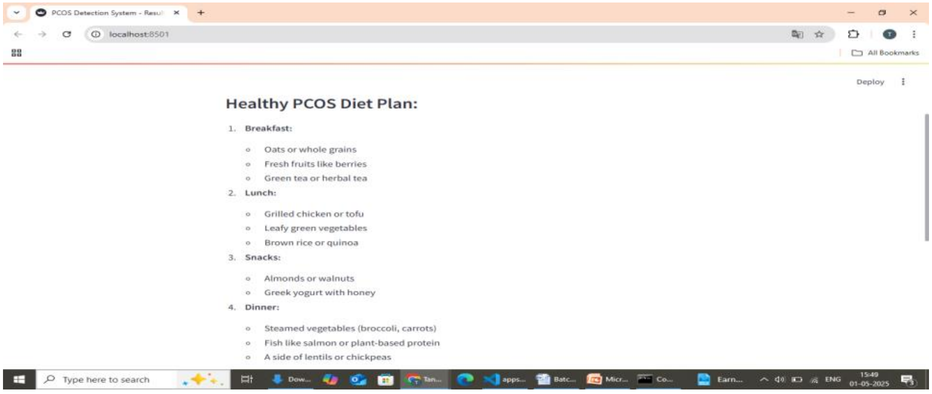

PCOS Detection System
localhost:8501
Deploy
Do you have a regular menstrual cycle?
Yes
No
How would you describe your general body temperature?
Too Hot
Too Cold
Neutral
Are you experiencing hair loss or thinning?
Yes
No
How would you rate your sleep quality?
Poor
Average
Good
Do you often crave sugary or high-carb foods?
Yes
No

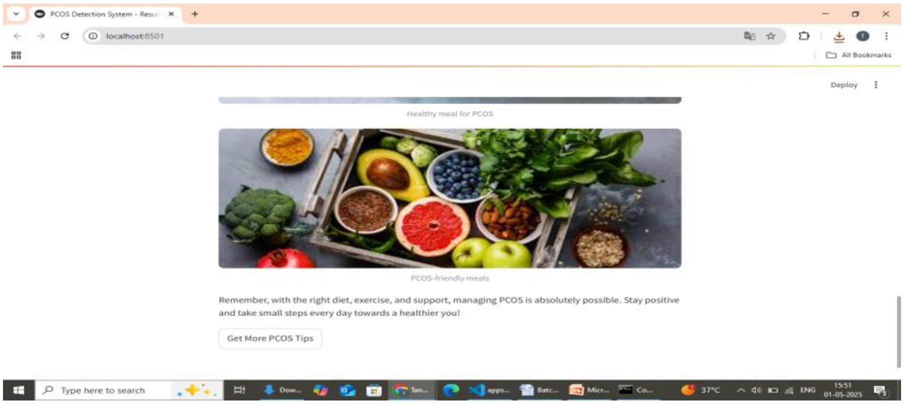

PCOS Detection System
localhost:8501
Deploy
Are you experiencing hair loss or thinning?
Yes
No
How would you rate your sleep quality?
Poor
Average
Good
Do you often crave sugary or high-carb foods?
Yes
No
Please fill all required fields (age and BMI).

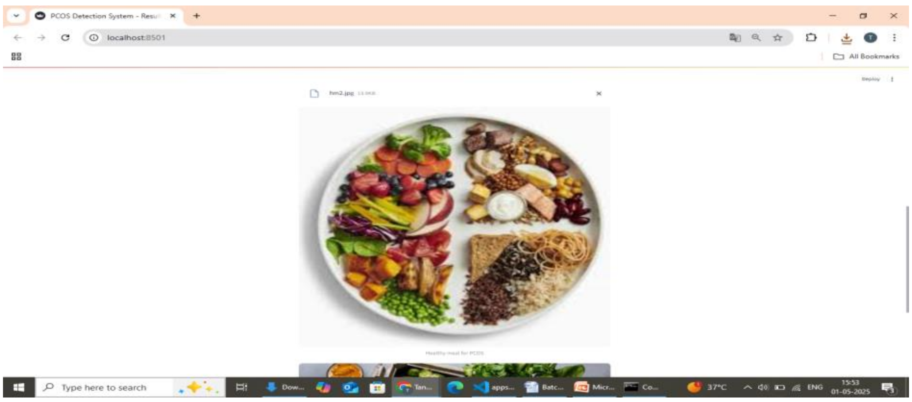

PCOS Detection System
localhost:8501
Deploy
What is your average daily calorie intake (in kcal)?
30 grams
Do you have any history of pregnancy?
Yes
No
Does anyone in your family have PCOS?
Yes
No
What type of daily work do you do?
Sedentary
Moderately Active
Very Active
Do you exercise at least 3 times a week?
Yes
No
On a scale of 1–10, how stressed do you feel daily?

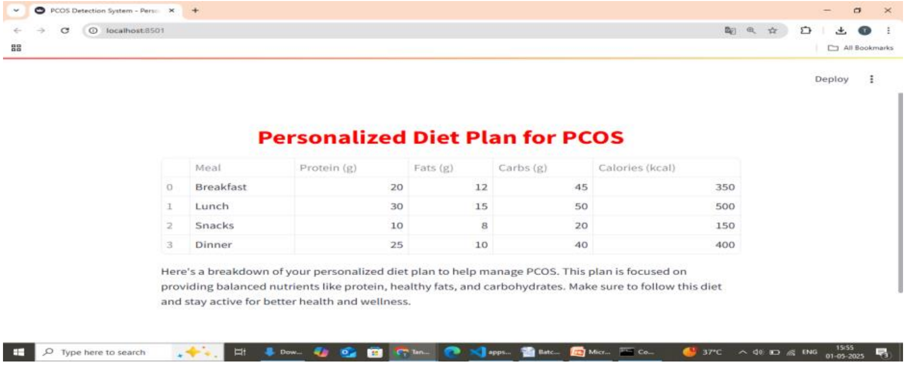

PCOS Detection System
localhost:8501
Deploy
No
What type of daily work do you do?
Sedentary
Moderately Active
Very Active
Do you exercise at least 3 times a week?
Yes
No
On a scale of 1–10, how stressed do you feel daily?
7
Do you drink at least 2 liters of water daily?
Yes
No
Submit

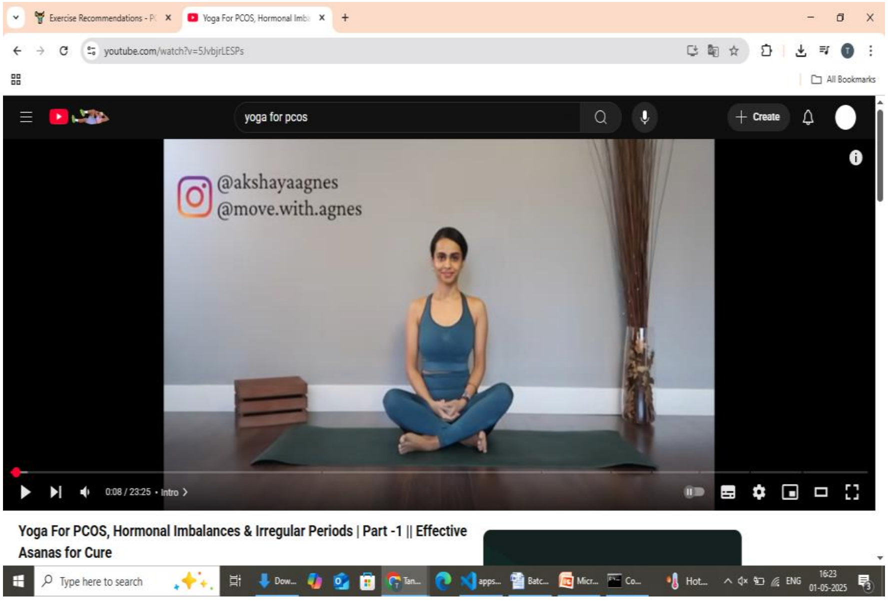

PCOS Detection System - Result
localhost:8501
Deploy
You are likely to have PCOS
Don't worry, we've got your back!
PCOS is common, but with the right lifestyle and diet, you can manage it effectively. Here's a diet plan to help:
Healthy PCOS Diet Plan:
1. Breakfast:
Oats or whole grains
Fresh fruits like berries
Green tea or herbal tea
2. Lunch:
Grilled chicken or tofu
Leafy green vegetables
Brown rice or quinoa
3. Snacks:
Type here to search

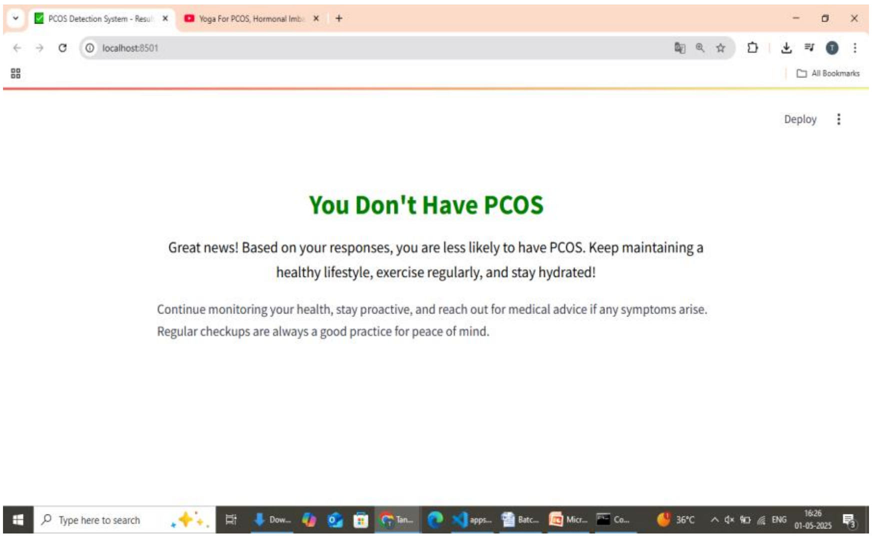

PCOS Detection System - Result
localhost:8501
Deploy
Healthy PCOS Diet Plan:
1. Breakfast:
Oats or whole grains
Fresh fruits like berries
Green tea or herbal tea
2. Lunch:
Grilled chicken or tofu
Leafy green vegetables
Brown rice or quinoa
3. Snacks:
Almonds or walnuts
Greek yogurt with honey
4. Dinner:
Steamed vegetables (broccoli, carrots)
Fish like salmon or plant-based protein
A side of lentils or chickpeas
Type here to search


PCOS Detection System - Result
localhost:8501
Deploy
Healthy meal for PCOS
PCOS-friendly meals
Remember, with the right diet, exercise, and support, managing PCOS is absolutely possible. Stay positive and take small steps every day towards a healthier you!
Get More PCOS Tips
Type here to search


PCOS Detection System - Result
localhost:8501
Type here to search







6-CONCLUSION

PCOS detection system leverages machine learning models to analyze patient data, identifying key patterns and symptoms associated with the condition. By utilizing models like Logistic Regression, Random Forest, and Gradient Boosting, the system processes various inputs such as sleep cycles, food cravings, and medical parameters to predict the likelihood of PCOS. The best-performing model is selected to provide timely and precise diagnostic support, helping in early detection and personalized treatment planning for patients.

FUTURE SCOPE

The future scope for PCOS detection includes integrating deep learning models for improved accuracy and early diagnosis. Incorporating real-time patient data from wearable devices and mobile apps can enhance monitoring and prediction. Expanding datasets to include ultrasound and genetic markers will make models more comprehensive. Telemedicine integration can support remote consultations and personalized treatment plans.

REFERENCES

1.R. Agarwal, N. Tiwari, and P. S. Bedi, "Predictive Analysis of PCOS Using Data Mining Techniques," International Journal of Computer Science and Mobile Computing, Vol. 8, No. 8, pp. 78-85, 2020..

2. M. K. Gupta, A. Jain, and P. Jain, "A Survey on Machine Learning Algorithms," International Journal of Computer Applications, Vol. 173, No. 10, pp. 6-12, 2017.

3. B. Xie, Y. Li, and X. Zhou, "An Overview of Supervised Learning Algorithms," IEEE Access, Vol. 7, pp. 153935-153950, 2019. DOI: 10.1109/ACCESS.2019.2957260.

4. H. Anwar, F. Arshad, and A. Bashir, "Polycystic Ovary Syndrome Prediction Using Machine Learning Algorithms," IEEE Access, Vol. 8, pp. 123456–123467, 2020. DOI: 10.1109/ACCESS.2020.3031234.