Detection of Endocrine Disorder Based on Machine Learning using Physical and Clinical Database

Utkarshani Tiwari (2102840100177)
Muskan Singh (2102840100106)
Ashutosh Kumar Singh (2102840100050)
Ashish Singh (2102840100048)

Department of Computer Science Engineering
Under Supervision: Mr. Prafull Pandey (Asst. Prof)

Contents - Part 1

Contents - Part 2

Introduction

- Polycystic Ovary Syndrome (PCOS) is a hormonal disorder commonly affecting women of reproductive age.
- It can lead to irregular menstrual cycles, infertility, acne, weight gain, and other metabolic issues.
- Early detection is crucial, but diagnosis is often delayed due to varied and overlapping symptoms.
- Machine Learning offers a promising solution by analyzing large datasets to identify patterns linked to PCOS.
- This project focuses on using ML algorithms to enhance the accuracy and efficiency of PCOS detection.
- The goal is to support healthcare professionals with data-driven, reliable diagnostic tools.

Problem Statement

- PCOS is a complex and underdiagnosed condition due to its wide range of symptoms and lack of standardized diagnostic criteria.
- Traditional diagnosis relies heavily on manual interpretation of clinical and ultrasound data, which can be time-consuming and subjective.
- Many patients remain undiagnosed or misdiagnosed, leading to delayed treatment and long-term health complications.
- There is a need for a reliable, data-driven approach to assist in the early and accurate detection of PCOS.
- Machine Learning can be utilized to analyze medical data and uncover patterns that are not easily visible to the human eye.

Literature Review

YEAR	AUTHOR	ALGORITHM	PROBLEM STATEMENT	ACCURACY
2024	Md Mahbubur Rahman (Corresponding Author) Ashikul Islam Forhadul Islam	Random Forest AdaBoost	This project uses machine learning to detect PCOS early by analyzing clinical and hormonal data through a smart web-based system.	94 94
2023	Hela Elmannai Nora El-Rashidy Ibrahim Mashal	Recursive Feature Elimination (RFE)	This study builds an accurate, explainable ML model for early PCOS detection to prevent health risks.	93.7
2022	Dana Hdaib Noor Almajali Hiam Alquaran	Linear Discriminant	PCOS diagnosis is complex due to symptom overlap and limited tools. This project uses machine learning on clinical data to improve accuracy.	92.6
2020	Vaidehi Thakre Shreyas Vedpathak Kalpana Thakre Shilpa Sonawani	Random Forest Classifier	This project aims to detect Polycystic Ovary Syndrome (PCOS) using machine learning algorithms. It analyzes clinical and hormonal data to assist in early and accurate diagnosis.	90.9

Research Gaps

- Limited use of automated systems in current PCOS diagnostic procedures.
- Most existing studies focus on individual symptoms rather than a holistic data-driven diagnosis.
- Lack of integration between clinical, biochemical, and ultrasound data in predictive modeling.
- Inadequate use of Machine Learning models for large-scale, accurate PCOS detection.
- Few open-access datasets and standardized benchmarks for comparing diagnostic models.
- Limited tools available to assist doctors in early-stage detection and decision-making.

Proposed Solution

- Develop a Machine Learning-based model to predict the likelihood of PCOS in patients using medical data.
- Use a dataset containing clinical, hormonal, and ultrasound features for comprehensive analysis.
- Apply data preprocessing techniques to handle missing values, outliers, and feature scaling.
- Train and evaluate various ML algorithms to identify the best-performing model.
- Provide an interpretable output to assist healthcare professionals in understanding the prediction results.
- Aim to improve early diagnosis, reduce human error, and support faster clinical decision-making.

Software Development Life Cycle (SDLC)

- Planning
- Requirement Analysis
- Design
- Implementation
- Testing
- Deployment
- Maintenance

SDLC Models

- Waterfall
- V-Model
- Iterative
- Spiral
- Big Bang
- Agile

Iterative Model

- A development approach where the project is broken down into smaller, manageable cycles (iterations).
- Develop incrementally with each iteration improving on the previous one.
- Early feedback and adjustments are integrated.
- Continuous testing and refinement.

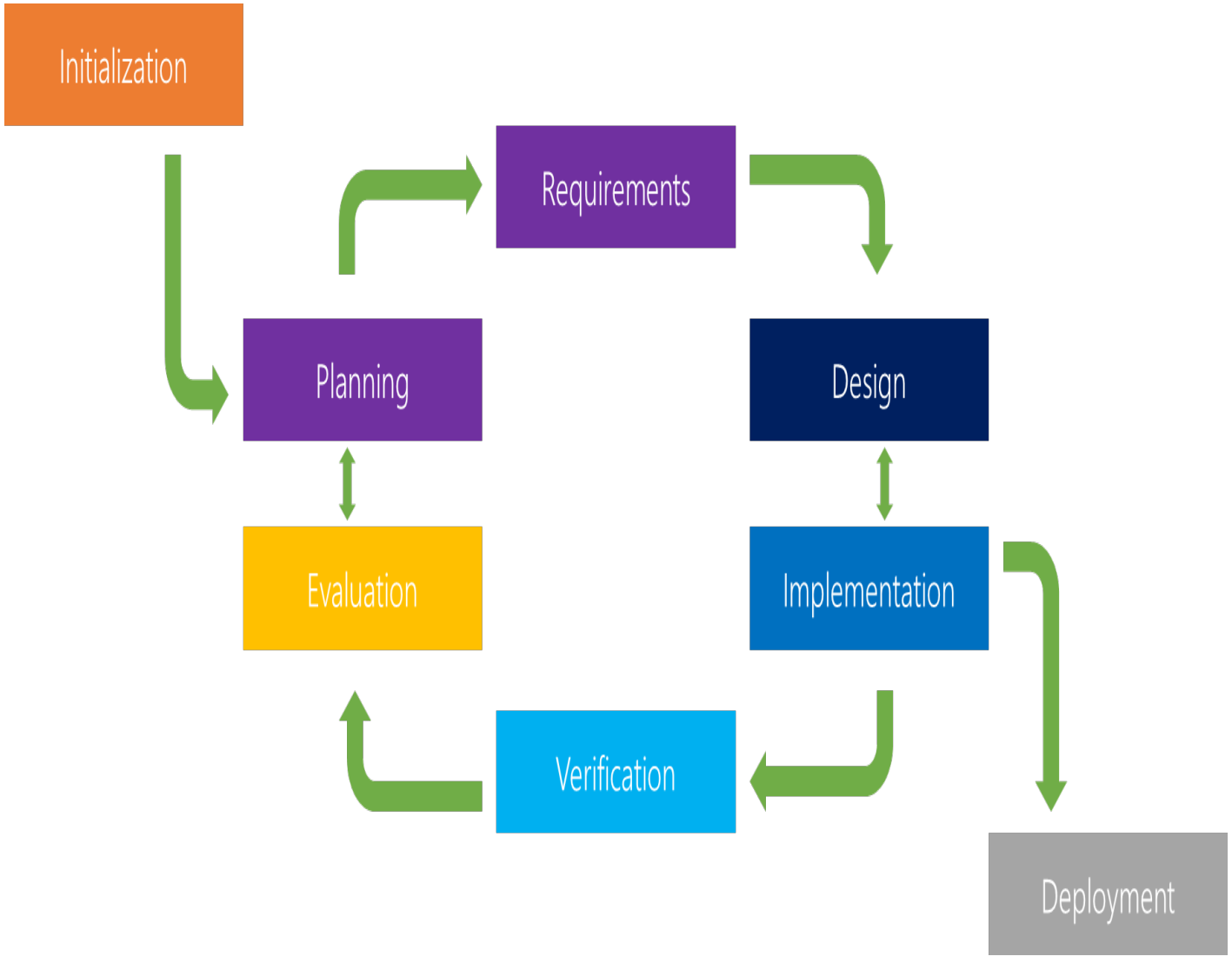
Iterative Workflow

- Requirement Gathering: Understand the problem and define core functionalities.
- Design: Plan UI, architecture, and ML model inputs/outputs.
- Implementation: Develop in parts, starting with ML model, then web and database.
- Testing: Test each module, identify, and fix bugs.
- Evaluation Feedback: Gather feedback and assess model/UI performance.
- Refinement: Improve accuracy, adjust inputs, or enhance UX.
- Repeat: Cycle through design $\rightarrow$ implement $\rightarrow$ test $\rightarrow$ refine.

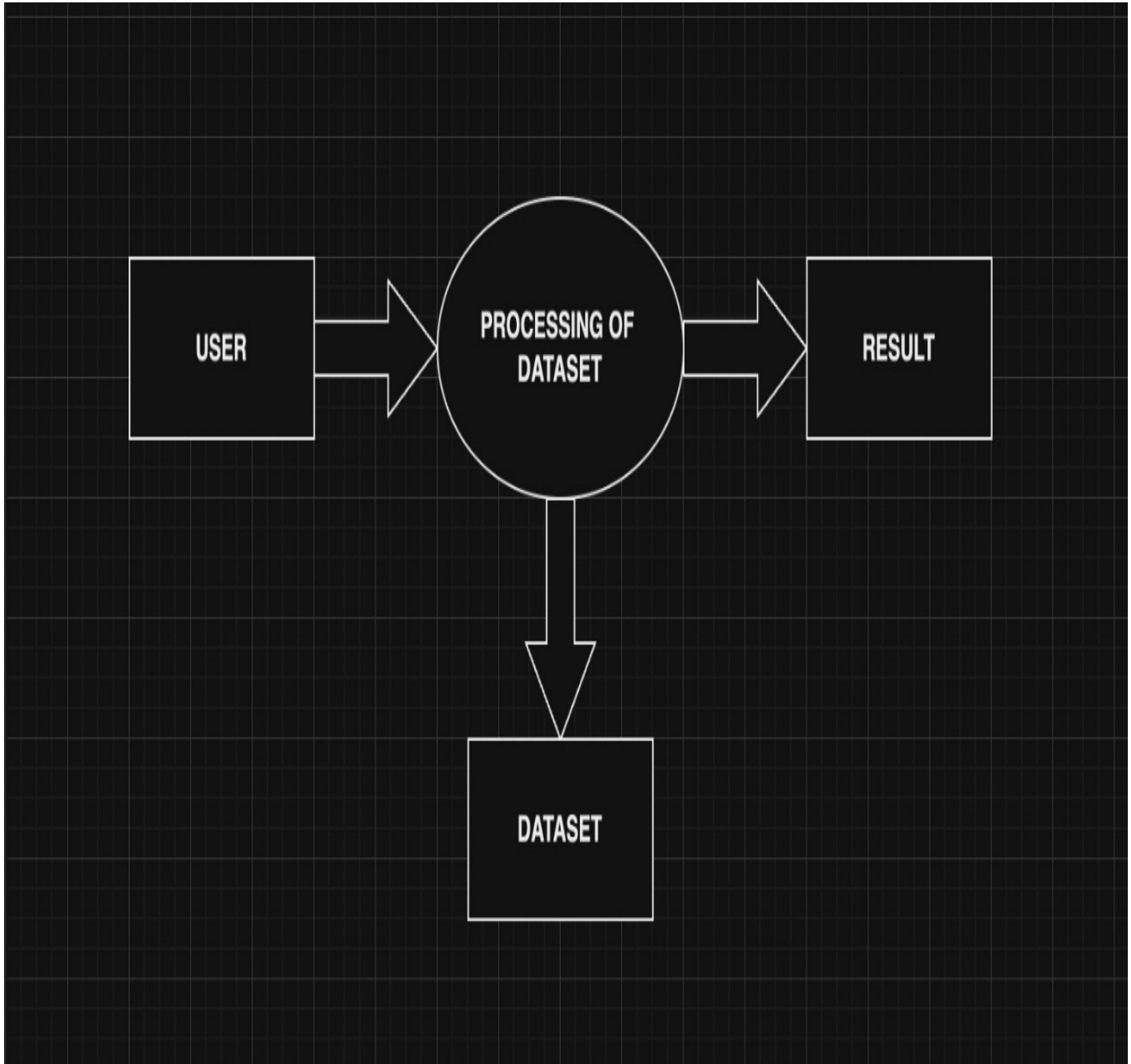
Why Iterative Model?

- Progressive Development: Build in phases—data, UI, and database.
- Flexibility: Easily adjust features without restarting.
- Early Issue Detection: Test each module early to fix issues fast.
- User Feedback: Integrate feedback at each stage.
- Efficient Resource Use: Focus on manageable components.
- ML-Friendly: Matches the iterative nature of ML model development.

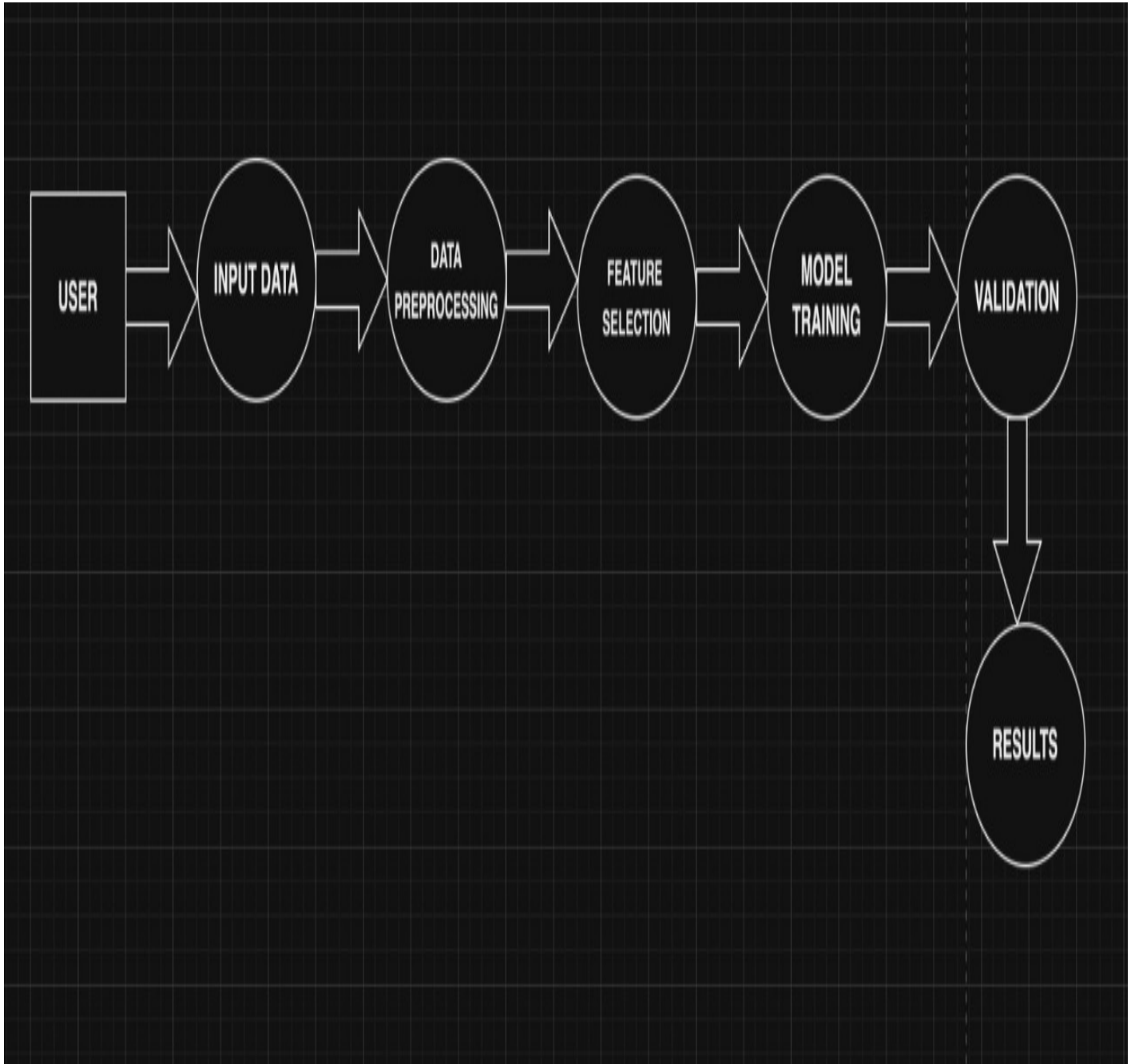
Iterative Model Diagram



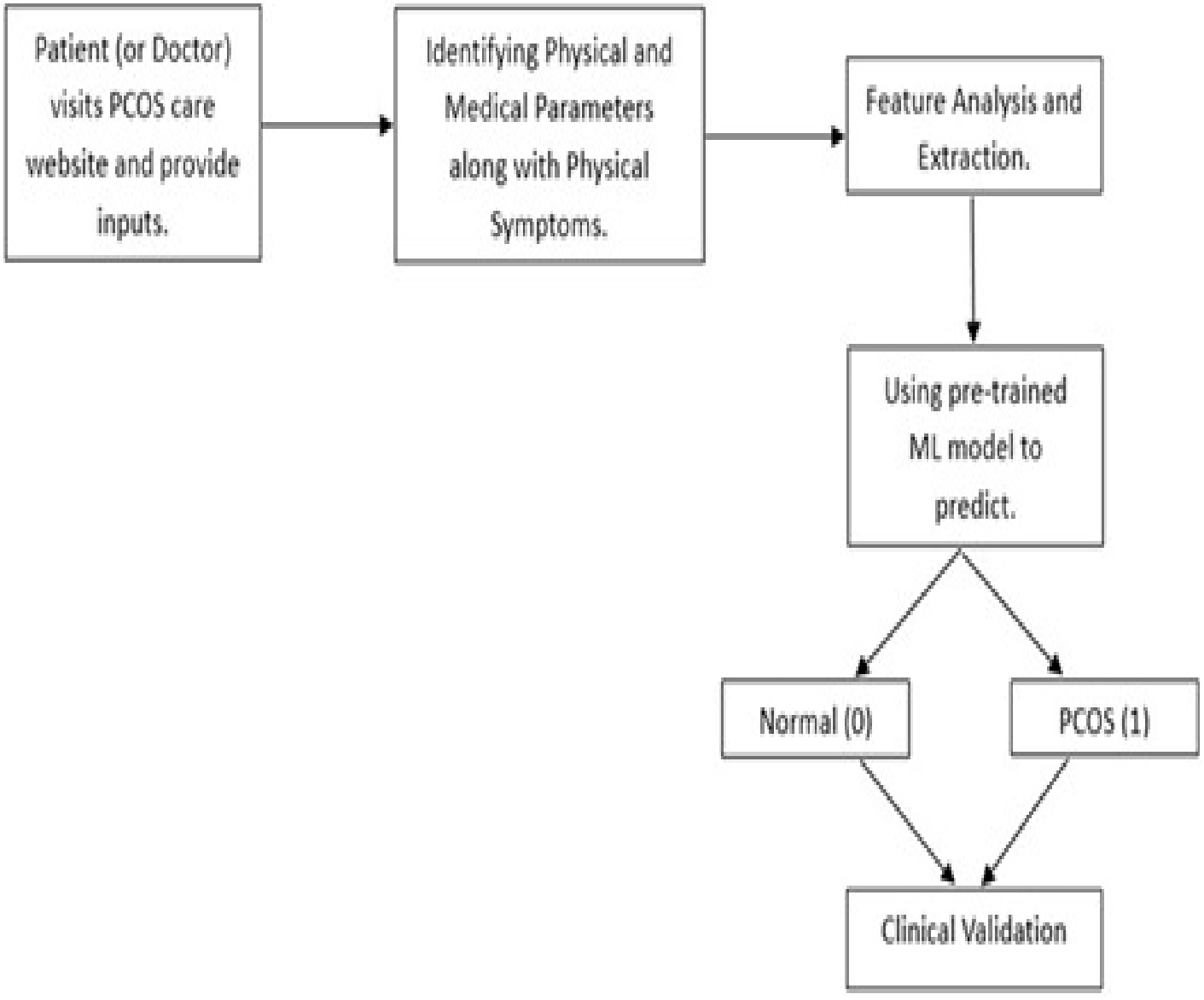
System Design - 0 Level DFD



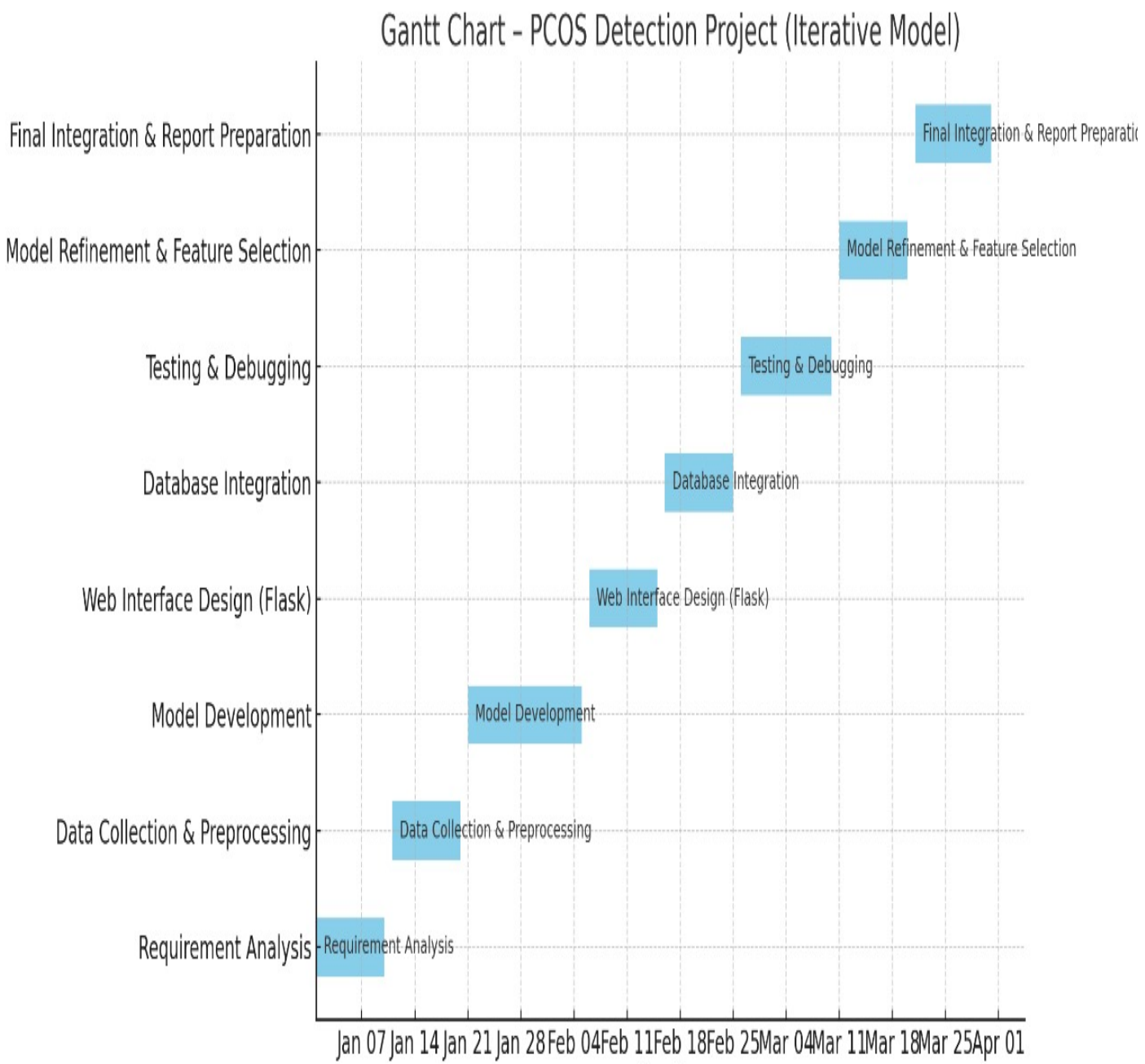
System Design - 1 Level DFD



System Design - Activity Diagram



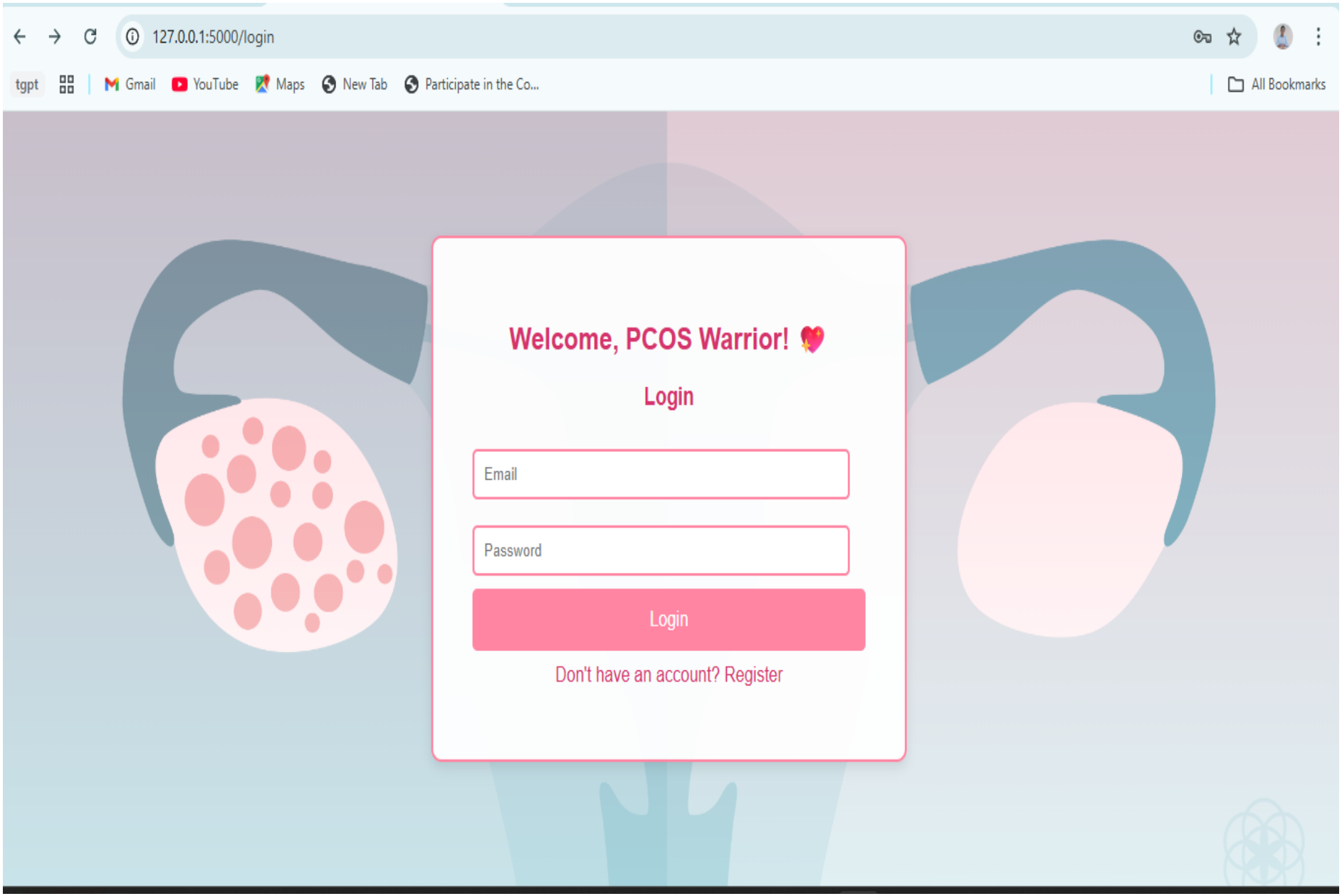
System Design - Gantt Chart



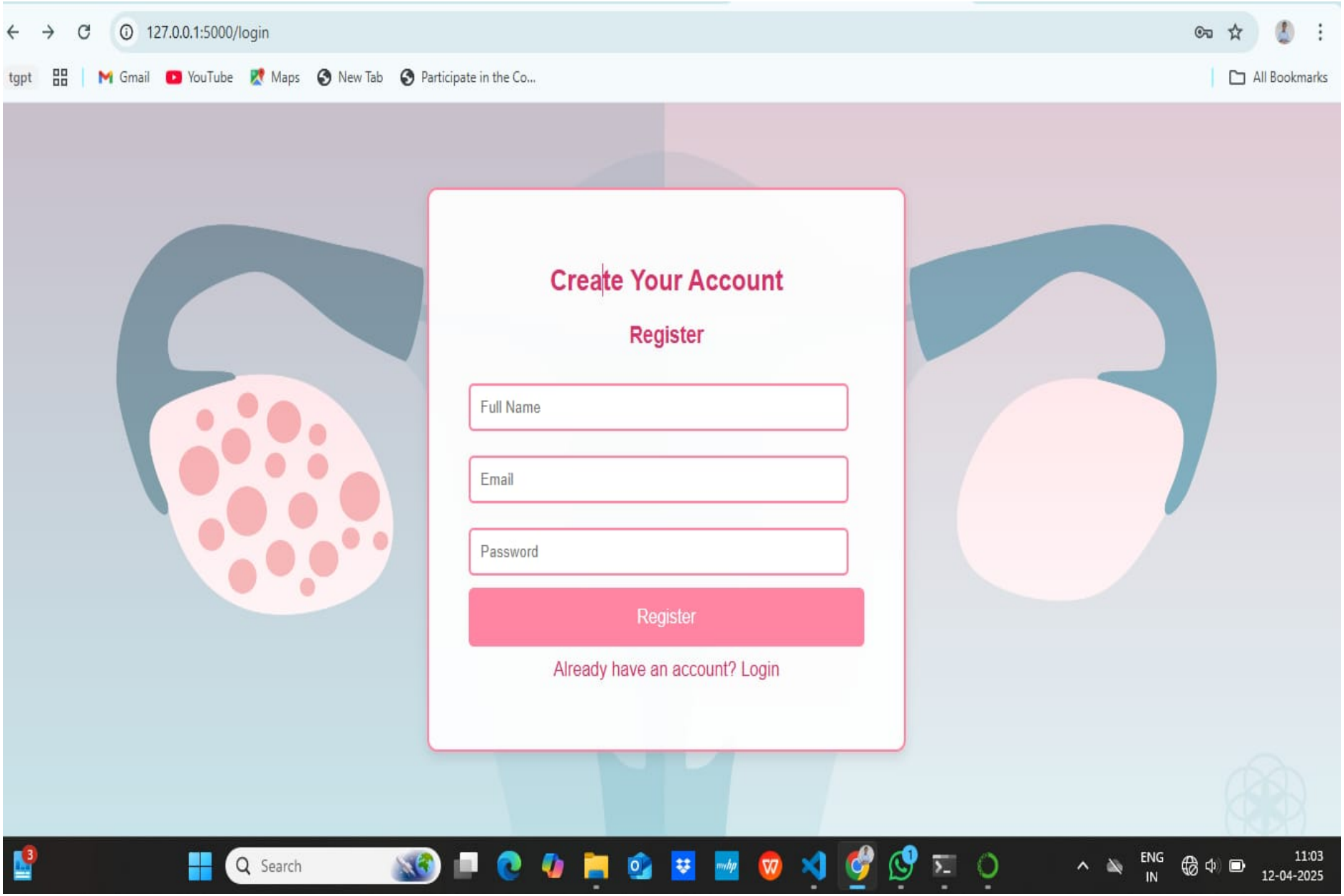
Coding - Libraries and Tools

- Core Libraries:
  - Numpy
  - Pandas
  - Seaborn
  - Matplotlib
  - Web Development: Flask
  - Database: flask-mysqldb

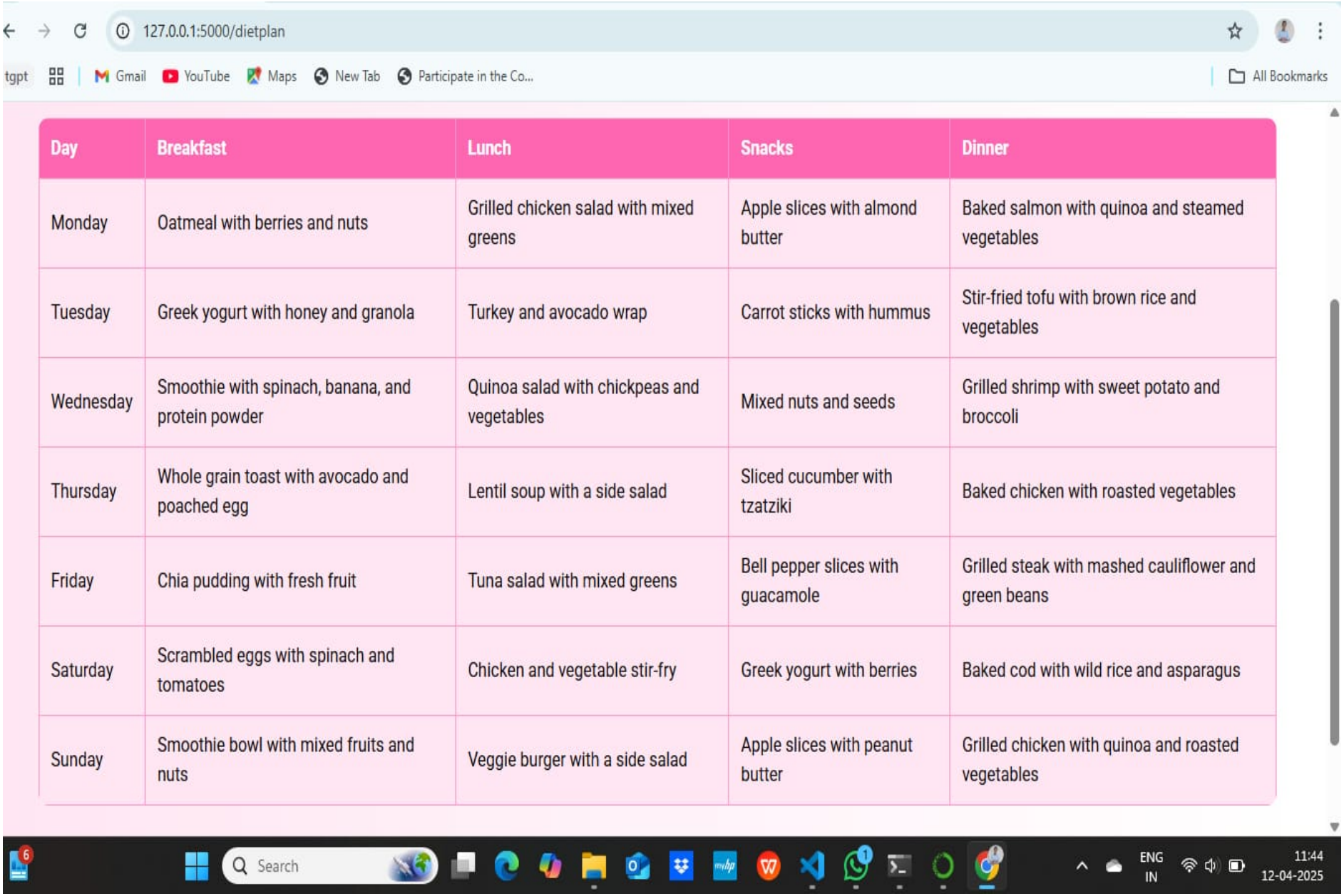
Implementation - Screenshot 1-Login Page



Implementation - Screenshot 2-Registration Page

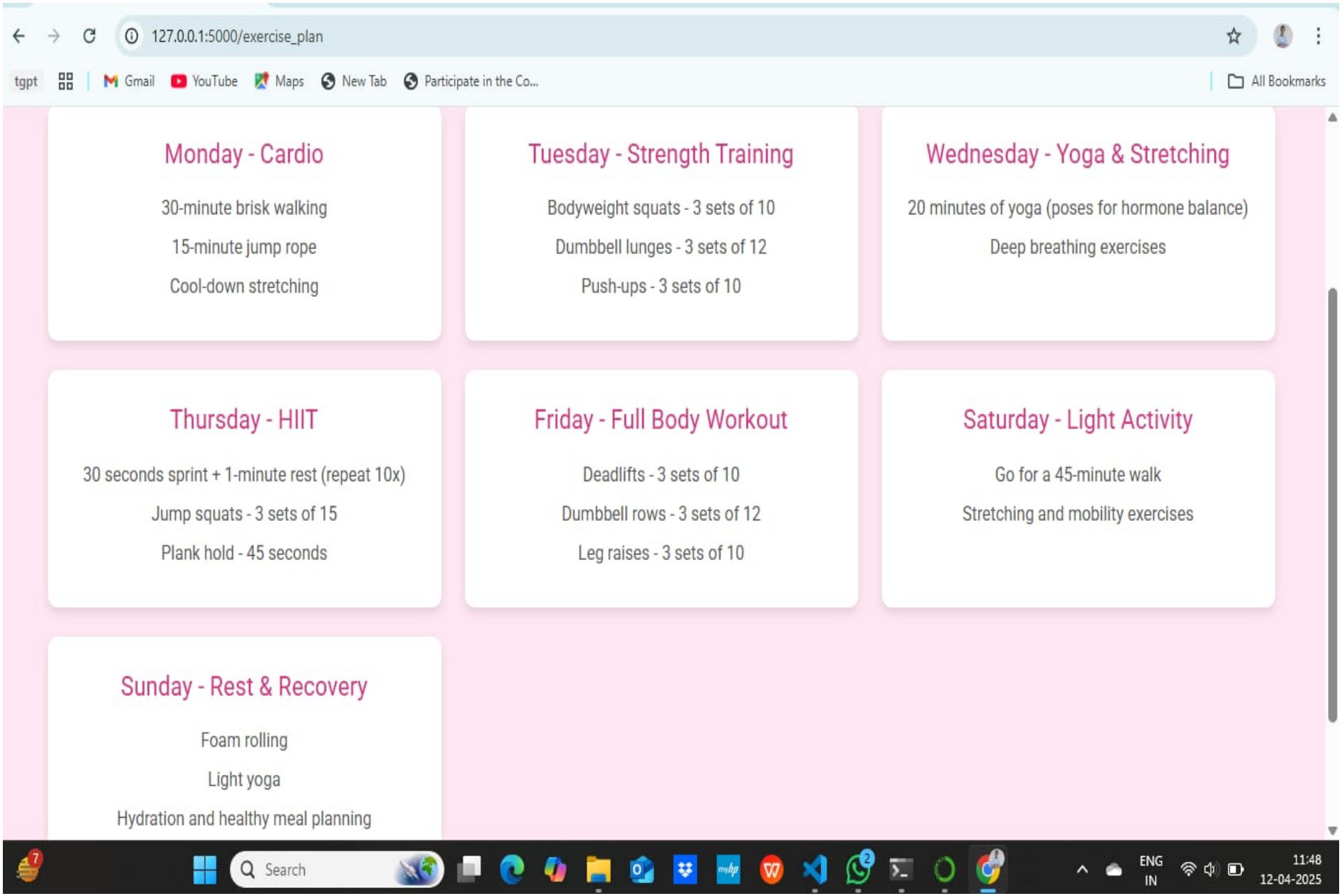


Implementation - Screenshot 2-Diet Chart



Day	Breakfast	Lunch	Snacks	Dinner
Monday	Oatmeal with berries and nuts	Grilled chicken salad with mixed greens	Apple slices with almond butter	Baked salmon with quinoa and steamed vegetables
Tuesday	Greek yogurt with honey and granola	Turkey and avocado wrap	Carrot sticks with hummus	Stir-fried tofu with brown rice and vegetables
Wednesday	Smoothie with spinach, banana, and protein powder	Quinoa salad with chickpeas and vegetables	Mixed nuts and seeds	Grilled shrimp with sweet potato and broccoli
Thursday	Whole grain toast with avocado and poached egg	Lentil soup with a side salad	Sliced cucumber with tzatziki	Baked chicken with roasted vegetables
Friday	Chia pudding with fresh fruit	Tuna salad with mixed greens	Bell pepper slices with guacamole	Grilled steak with mashed cauliflower and green beans
Saturday	Scrambled eggs with spinach and tomatoes	Chicken and vegetable stir-fry	Greek yogurt with berries	Baked cod with wild rice and asparagus
Sunday	Smoothie bowl with mixed fruits and nuts	Veggie burger with a side salad	Apple slices with peanut butter	Grilled chicken with quinoa and roasted vegetables

Implementation - Screenshot 3-Exercise Chart



Results

SR. No	Models	Accuracy
0	LR	82.56%
1	SVM	70.64%
2	RF	91.32%
3	GB	82.56%

Results


Performance Comparison of Models (Without XGBoost)
F1 Score
Precision
Recall
Score
1.0
0.8
0.6
0.4
0.2
0.0
0.50
0.57
0.45
0.53
0.57
0.50
0.50
0.52
0.48
0.56
0.56
0.55
Logistic Regression
Random Forest
SVC
Gradient Boosting
Models

Conclusion

- Developed a Machine Learning-based system for early detection of PCOS.
- Early diagnosis is essential to prevent long-term reproductive health issues.
- System analyzes clinical data to identify PCOS patterns and support timely medical intervention.
- Encourages a healthier lifestyle through early awareness.

Future Work

- Multilingual Support
- Expanded Dataset
- Incorporate Hormonal and Ultrasound Features
- Integration with Mobile Platforms

References

- Rohit Mishra et al," Comparative Analysis of Conventional IP Network and MPLS Network over VoIP Application" / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 4496-4499
- Silva I, et al. Polycystic ovary syndrome: clinical and laboratory variables related to new phenotypes using machine-learning models. J Endocrinol Invest 2022:1–9.
- Rohit Mishra et. al, "Protein Structure Prediction Using Needle Man Wunsch Algorithm" Journal of Harbin Engineering University ISSN: 1006-7043
- M. Sumathi "Study and detection of PCOS related diseases using CNN,' IOP Conf. Ser., Mater. Sci. Eng., vol. 1070, Nov. 2021, Art. no. 012062.
- Rohit Mishra et. al, "ENHANCED IMAGE ANALYSIS FOR GASTRIC CANCER USING CNN" DOI:10.48047/ecb/2023.12.9.51

Thank You!