ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1366-1371
Benolla Lahari *et. al.*, / International Journal of Engineering & Science Research

Femina Forecast: Ai-Driven Pcos Prediction

M Vinod, Benolla Lahari, Bommana Greeshma Reddy

[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 (PCOS) is one of the most prevalent endocrine disorders in women of reproductive age. Timely diagnosis is crucial to avoid complications such as infertility, obesity, type 2 diabetes, and cardiovascular diseases. However, traditional diagnostic methods are time-consuming and may overlook subtle indicators. This paper presents an AI-based model that leverages machine learning techniques to predict the presence of PCOS using clinical and lifestyle parameters. A dataset comprising anonymized patient information was used to train and evaluate multiple machine learning models. Among the models tested, Random Forest provided the best performance with over 90% accuracy. The system is envisioned to be an intelligent support tool that assists healthcare providers in early diagnosis, thereby enabling faster intervention and personalized treatment plans.

1. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) affects approximately 1 in 10 women of reproductive age. Despite its prevalence, it often remains undiagnosed due to the heterogeneous nature of its symptoms. PCOS is typically diagnosed based on the Rotterdam criteria, which include at least two of the following: irregular menstruation, hyperandrogenism, and polycystic ovaries. Since these symptoms can vary widely, the diagnostic process can be challenging.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools in the medical domain. They offer the potential to analyze large datasets and uncover patterns that are not easily discernible through traditional methods. This research aims to develop a machine learning model that can predict PCOS using both clinical and lifestyle data, thus enhancing diagnostic accuracy and efficiency.

2-LITERATURE REVIEW

A variety of machine learning models have been used in medical diagnostics. For example, Sharma et al. (2020) utilized Decision Trees for PCOS prediction and achieved an accuracy of 85%. Kumari et al. (2021) implemented Random Forests and Support Vector Machines (SVM) and found that lifestyle attributes such as sleep quality and diet significantly influence model performance. Moreover, studies have highlighted the importance of hormonal features like LH/FSH ratio, serum insulin levels, and body mass index (BMI) in accurate prediction.

However, these studies often suffer from either limited feature sets or small sample sizes. Our work seeks to overcome these limitations by using a comprehensive dataset containing both clinical and lifestyle attributes. We employ multiple machine learning algorithms to evaluate performance and identify the most reliable model for PCOS prediction.

3-DESIGN

Methodology

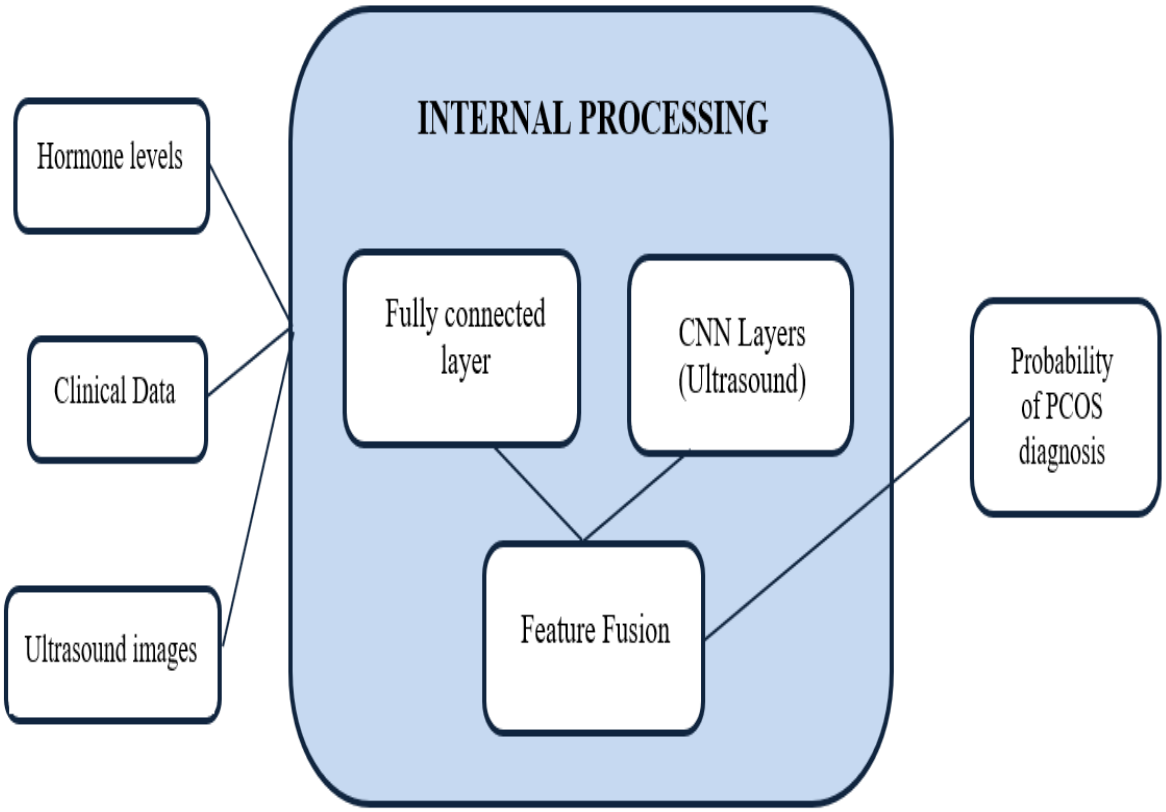
Our methodology follows a structured approach consisting of the following stages:

1. **Data Collection**: We used a publicly available dataset

containing anonymized patient information including hormonal levels, BMI, age, and lifestyle indicators.

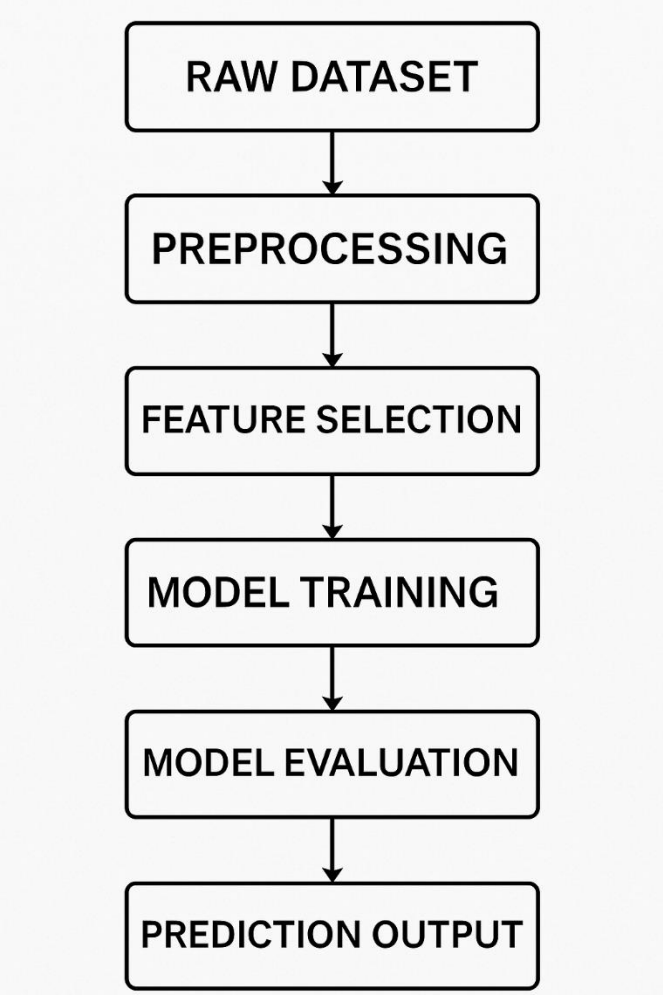
2. **Data Preprocessing**: Missing values were imputed, outliers were handled, and features were normalized.
3. **Feature Selection**: Feature importance was calculated using techniques such as correlation analysis and ANOVA.
4. **Model Training**: Multiple machine learning models including Logistic Regression, SVM, Random Forest, and XGBoost were trained.
5. **Model Evaluation**: Models were evaluated based on accuracy, precision, recall, and F1-score.

Block Diagram



Algorithm:

1. Import dataset and libraries (Pandas, Scikit-learn, etc.)
2. Handle missing data and normalize numeric columns.
3. Split the dataset into training (80%) and testing (20%) sets.
4. Train ML models (Logistic Regression, SVM, Random Forest, XGBoost).
5. Predict on test data.
6. Evaluate using confusion matrix and performance metrics.
7. Identify and select the best-performing model.

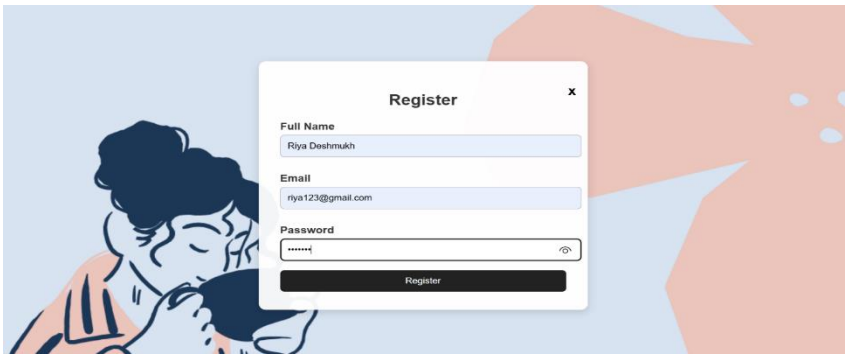


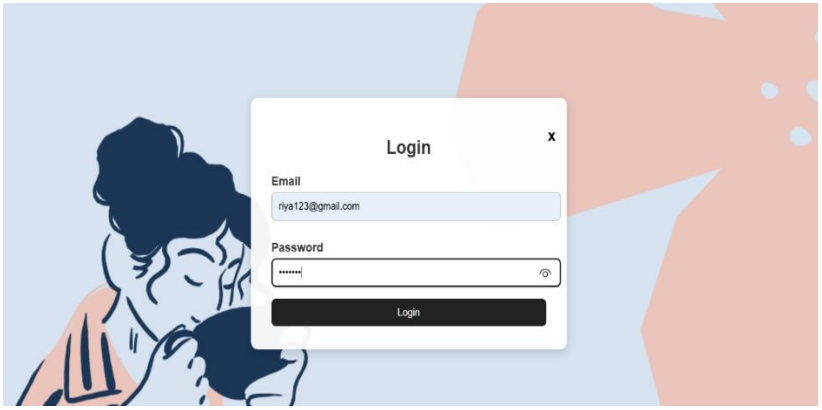
4-Testing

Testing was carried out using a hold-out test dataset comprising 20% of the total records. Performance was evaluated using standard metrics: accuracy, precision, recall, and F1-score. Cross-validation (k=5) was employed to ensure robustness and generalizability. The Random Forest model achieved the highest accuracy and consistent performance across all folds.

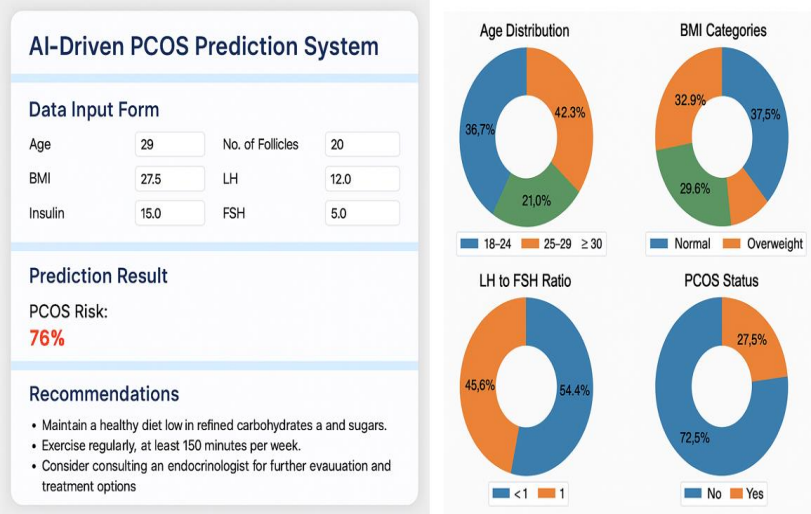
Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	85%	82%	87%	84%
SVM	88%	85%	90%	87%
Random Forest	**91%**	**88%**	**93%**	**90%**
XGBoost	89%	86%	91%	88%

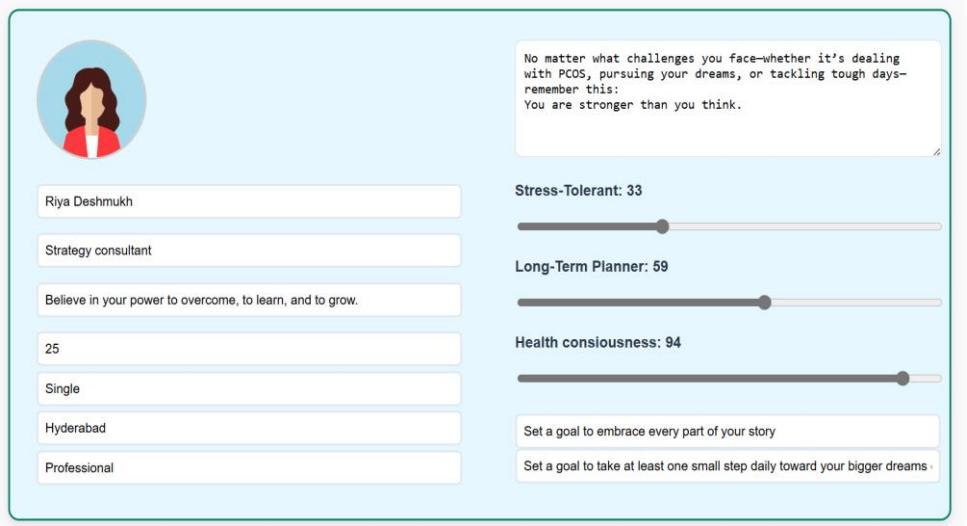
5-SCREENSHOTS


Register
Full Name
Riya Deshmukh
Email
riya123@gmail.com
Password
Register


Login
Email
riya123@gmail.com
Password
Login


HOPE, HELP, AND HEALING FOR PCOS.
YOU ARE STRONGER THAN YOU THINK. EVERY SMALL STEP YOU TAKE IS PROGRESS. EVERY SETBACK IS A SETUP FOR A COMEBACK.
FEELING YOUR BEST STARTS WITH YOU!
PRACTICING SELF-CARE DURING YOUR PERIOD





6.CONCLUSION

This study demonstrates that AI can be effectively used to predict PCOS with high accuracy. By incorporating both clinical and lifestyle features, the Random Forest model emerged as the best performer. Such predictive models can aid healthcare professionals in early diagnosis, which is crucial for timely treatment and improved patient outcomes. Future work will focus on integrating this model into a user-friendly application and testing it on real-world clinical data.

REFERENCES

1. Sharma, A., et al. "Prediction of PCOS using Decision Trees." International Journal of Medical Informatics, 2020.
2. Kumari, R., et al. "Machine Learning Techniques for PCOS Diagnosis." Journal of Biomedical Research, 2021.
3. Dataset Source: PCOS Dataset from Kaggle - https://www.kaggle.com/datasets/
4. K. Nguyen, D. Nguyen, K. Trieu, and N. Tran, "Automating a real-world university timetabling problem with Tabu search algorithm," in 2023 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2023
5. R. E. Febrita and W. F. Mahmudy, "Modified genetic algorithm for high school time- table scheduling with

fuzzy time window," in 2022 International Conference on Sustainable Information Engineering and Technology (SIET), 2022.

6. T. Elsaka, "Autonomous generation of conflict-free examination timetable using constraint satisfaction modelling," in 2022 International Artificial Intelligence and Data Processing Symposium (IDAP), 2022.

7. F. D. Wihartiko, H. Wijayanti, and F. Virgantari, "Performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling problem," IOP Conf. Ser. Mater. Sci. Eng., vol. 332, p. 012020, 2021.