Grenze International Journal of Engineering and Technology, June Issue

Personalized Lifestyle Guidance for Women with PCOS: A Machine Learning Approach to Sustainable Health

Renuka Sagar[1], S Sarah Azeez[2], Sana Shaik[3], Zoya Tabassum[4] and Shaik Mohammad Zaheed Hussain[5]
[1-5]Department of Artificial Intelligence and Machine Learning, Ballari Institute of Technology and Management, Ballari, India
Email: renukasagar83@gmail.com; sarahazeez04@gmail.com; sanaajeelan@gmail.com; zoya.tabassum1611@gmail.com; zhussain9753@gmail.com

***Abstract*— Polycystic Ovary Syndrome (PCOS) is a common hormonal condition that affects many women of reproductive age. It often goes unnoticed because its symptoms can appear in many different ways, making early detection difficult. Early identification of the specific type of PCOS and tailored lifestyle interventions can reduce long-term health risks such as infertility, metabolic complications, and cardiovascular disease. This study introduces PCOS Management and Guidance, a web-based application designed to classify PCOS types using a Random Forest classifier trained on clinical and symptomatic data. The system incorporates optical character recognition (OCR) to process ultrasound images and supplements the prediction with personalized guidance on diet, exercise, and stress management. The model achieved a high level of accuracy, demonstrating that ML-based methods can be effectively applied in women's health diagnostics. By combining type detection with lifestyle recommendations, PCOS Management and Guidance offers a practical tool to support early intervention and patient self-management.**

***Keywords*— PCOS Management and Guidance, Machine Learning, AI-powered Assistant, AI in Healthcare, Lifestyle Recommendation System, Random Forest Classifier, Optical Character Recognition, Python, AI-driven Support, Patient-Centric Design, PCOS, Women's Health Tech, Web Development, and Predictive Analytics. Women's Reproductive Health, Healthcare.**

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a widespread hormonal condition that impacts many women during their reproductive years, with studies estimating its prevalence between 8% and 13% of the global female population [1]. It manifests through a range of symptoms such as irregular menstrual cycles, the presence of ovarian cysts, elevated androgen levels, and metabolic imbalances [2].

Despite being so common, PCOS often goes undiagnosed because its symptoms vary greatly from person to person and current diagnostic techniques have several limitations [3]. When left untreated, the condition can contribute to serious long-term health issues, including infertility, obesity, cardiovascular problems, and type 2 diabetes [4]. Current diagnostic approaches largely aim to determine the presence of PCOS without classifying its subtypes.

Grenze ID: 01.GIJET.12.2.143_21

However, research suggests that PCOS manifests in distinct forms namely insulin-resistant, inflammatory, adrenal, and post-pill types and that treatment outcomes improve significantly when management is tailored to the subtype [5]. Lifestyle interventions such as dietary regulation, structured physical activity, and stress management remain the first-line treatment strategies, but their effectiveness varies depending on the individual's condition [6]. Therefore, there is a pressing need for intelligent systems that not only support early detection of PCOS but also recommend lifestyle modifications based on personalized characteristics.

In response to this gap, we present PCOS Management and Guidance, an AI-driven framework designed for both type detection and lifestyle guidance. The system employs a Random Forest classifier trained on clinical and symptomatic data, complemented by optical character recognition (OCR) to process ultrasound reports. The solution is deployed as a Flask-based web application, enabling user-friendly access for both patients and healthcare practitioners. By integrating detection and personalized recommendations, PCOS Management and Guidance addresses a critical unmet need in the digital health ecosystem for women's health.

The main objectives of the current research are:

- To build and train a machine learning model capable of accurately identifying and classifying various PCOS subtypes based on individual patient data.
- To integrate Optical Character Recognition (OCR) techniques for analyzing ultrasound images, thereby improving the precision of diagnostic results.
- To design a Flask-powered web application that enables real-time PCOS type prediction and ensures easy accessibility for users.
- To deliver customized lifestyle recommendations covering areas such as diet, exercise, and stress management tailored to each user's predicted PCOS subtype.

Main contributions of this paper include:

- Subtype Detection with Random Forest model to distinguish among insulin-resistant, inflammatory, adrenal, and post-pill PCOS subtypes.
- Use of OCR for ultrasound reports, enabling semi-automated feature extraction to complement clinical data.
- Evidence-based diet, exercise, and stress management suggestions aligned with the predicted subtype.
- A scalable Flask application that provides accessibility for patients and healthcare practitioners.

The rest of this paper is organized as follows. Section II reviews previous studies on PCOS detection, machine learning approaches in healthcare, and lifestyle management systems. Section III describes the methodology of the proposed work, including dataset preprocessing, model development, OCR integration, and web application design. Section IV presents the experimental results and discussion, highlighting model performance, system validation, and usability of the web platform. Finally, Section V concludes the paper by summarizing the contributions and outlining directions for future research and development.

II. Related Works

Numerous studies have explored various aspects of Polycystic Ovary Syndrome (PCOS), including its diagnosis, classification, and management, with growing attention on the application of artificial intelligence (AI), genetic insights, and evidence-based clinical practices.

A. Clinical Guidelines and Diagnosis

Teede HJ et al. [1] describe Polycystic Ovary Syndrome (PCOS) as one of the most prevalent endocrine disorders among women of reproductive age, with its occurrence estimated between 8% and 13%, varying according to population and diagnostic standards. It presents with complex reproductive, metabolic, and psychological features, leading to significant delays in diagnosis and dissatisfaction with care among patients. Existing guidelines are often inconsistent or outdated, lacking adequate consumer input. To address these issues, the first international evidence-based guideline was established, standardizing the use of Rotterdam diagnostic criteria for adult women.

B. Artificial Intelligence and Machine Learning in PCOS Detection

Barrera FJ et al. [2] produced A systematic review of 31 studies out of 135 screened demonstrated the potential of AI and machine learning (ML) in PCOS diagnosis [2]. Data sources included clinical records, biochemical, imaging, and genetic information. Support Vector Machine (42% of studies), K-nearest neighbor (26%), and regression models (23%) were most commonly applied. Reported performance metrics were promising, with diagnostic accuracy ranging between 89% and 100%, specificity between 75% and 100%, and sensitivity between 41% and 100%.

C. Data-Driven Diagnosis with Public Datasets

Bharati S et al. [3] have also explored data-driven approaches using publicly available datasets. One study employed the Kaggle PCOS dataset with 43 attributes across 541 women, of which 177 had confirmed PCOS [3].

Feature selection highlighted the FSH/LH ratio as the most significant attribute. Multiple classifiers, including gradient boosting, random forest, logistic regression, and hybrid random forest–logistic regression (RFLR), were applied. The RFLR model demonstrated the highest performance, achieving 91.01% testing accuracy and 90% recall with 40-fold cross-validation.

D. Genetic Studies and Subtyping of PCOS

Dapas M et al. [4] Genetic investigations have further uncovered phenotypic subtypes of PCOS. A genome-wide association study (GWAS) involving 893 genotyped PCOS cases and 263 replication cases identified two distinct subtypes [4]: a "reproductive" group (21–23%) with elevated LH and SHBG levels, and a "metabolic" group (37–39%) with higher BMI, glucose, and insulin. GWAS highlighted four loci associated with the reproductive subtype and one with the metabolic subtype, particularly clustering in families carrying rare DENND1A variants. However, the study was limited by a small sample size and the inclusion of only European ancestry women.

The literature reviewed presents encouraging trends in the use of artificial intelligence, machine learning, and genetic studies for the diagnosis and classification of PCOS. A number of limitations are however present.

One of the primary concerns is the lack of large-scale clinical validation is most studies rely on small datasets or retrospective data, which limits their applicability in real-world healthcare settings. Furthermore, there is an absence of standardized diagnostic criteria across studies, resulting in inconsistencies when comparing results and reducing generalizability.

Another constraint is the restricted diversity of study populations, as many genetic studies focus primarily on women of European ancestry, leaving significant gaps in applicability for other ethnic groups. Similarly, most AI-driven approaches remain at the experimental stage with limited translation into clinical workflows.

There is also a tendency to focus on achieving higher accuracy scores without addressing practical integration challenges such as interpretability, patient privacy, and clinician acceptance. Lastly, long-term studies examining the impact of AI-driven PCOS management on patient outcomes and lifestyle improvements are lacking.

In conclusion, while existing research highlights the promise of AI/ML techniques and genetic insights in improving PCOS detection and management, subsequent work must prioritize empirical validation, inclusivity of diverse populations, ethical adequacy, and integration into clinical practice to ensure meaningful real-world impact.

III. Proposed Framework

A. Framework Purposes

PCOS Type Detection and Lifestyle Guidance system integrates Machine Learning and OCR in a user-friendly web interface to deliver accurate, accessible diagnostics and personalized support. The primary goals of this framework include:

- Accurate PCOS Type Prediction: Using a Random Forest Classifier, the system analyzes clinical, biochemical, and lifestyle parameters to classify the patient into one of four PCOS types (Insulin-Resistant, Inflammatory, Post-Pill, or Adrenal), enabling early detection and targeted intervention.
- Automated Data Extraction via OCR: The platform integrates the Tesseract OCR engine to extract key medical parameters directly from uploaded clinical reports, reducing manual data entry and improving accuracy.
- User-Friendly Web Interface: Developed with Flask, the web application provides a structured input form that allows patients or clinicians to either manually enter medical details or upload a report for automated analysis. This configuration ensures accessibility across devices and promotes ease of use.
- Personalized Lifestyle Guidance: Based on the predicted PCOS type, the system generates a customized plan that includes diet recommendations, exercise routines, stress-management practices, and supplement suggestions, guiding users toward evidence-based lifestyle modifications.
- Efficient Processing and Interaction: The architecture ensures low-latency predictions by combining pre-processed datasets, median imputation for missing values, and optimized Random Forest training, resulting in fast and reliable outputs.

The system is built for scalability, supporting future additions like wearable device integration, multilingual support, and advanced analytics for broader reproductive health. Fig. 1 illustrates the system's flow from user input to ML-based prediction and lifestyle recommendations.

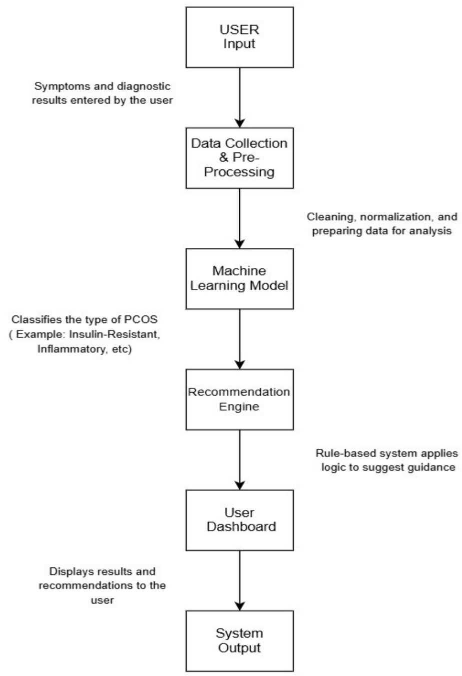

Fig.1. Block Diagram

B. System Architecture

The proposed PCOS Type Detection and Lifestyle Guidance system is structured around five major components that ensure seamless integration between user input, machine learning-based prediction, and personalized lifestyle recommendations.

- User Input Module: This module serves as the initial interaction point where users can manually enter clinical details such as age, BMI, and hormonal levels or upload diagnostic reports. The system incorporates Optical Character Recognition (OCR) technology to automatically extract and digitize relevant data from uploaded medical reports, minimizing manual entry errors and improving efficiency.
- Data Pre-processing Module: Once the input is received, it undergoes several pre-processing steps, including data cleaning, encoding of categorical attributes, and imputation of missing values. These processes ensure that the dataset is structured and standardized for accurate model interpretation and analysis.
- Machine Learning Classification Module: The pre-processed data is then fed into a Random Forest Classifier, which categorizes each case into one of the four PCOS subtypes—Insulin-Resistant, Inflammatory, Post-Pill, or Adrenal. The use of an ensemble-based model enhances accuracy and minimizes overfitting, providing consistent diagnostic predictions across diverse datasets.
- Recommendation Engine: Based on the predicted PCOS subtype, this module generates personalized lifestyle recommendations. These include tailored suggestions on diet, physical activity, supplement intake, and stress management, helping users adopt healthier habits and manage their condition more effectively.
- Backend Integration and Synchronization: The underlying system architecture ensures smooth coordination between all modules—OCR, pre-processing, classification, and recommendation. This synchronized backend infrastructure allows the system to deliver accurate predictions and recommendations in real time, improving accessibility and user experience.

Fig. 2 illustrates the Data Flow Diagram (DFD) of the proposed framework, highlighting the interaction among system modules and the step-by-step flow from user data acquisition to predictive analysis and generation of customized lifestyle guidance.

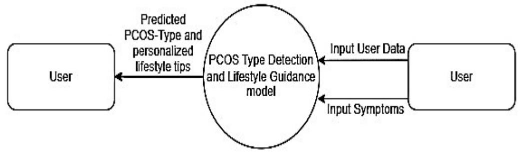

Fig. 2. Data Flow Diagram

C. Dataset Description

The dataset utilized in this study for Polycystic Ovary Syndrome (PCOS) type detection and lifestyle guidance was constructed from two primary sources: (i) a publicly available clinical dataset, and (ii) synthetically generated records.

- Clinical Dataset: Approximately 50% of the data was derived from the publicly available PCOS dataset hosted on Kaggle (Kottarathil, 2020) [11]. This dataset contains anonymized medical records of women undergoing diagnostic evaluation for PCOS. It includes demographic details (e.g., age, weight, BMI), clinical features (e.g., cycle regularity, follicle count, AMH levels), and biochemical parameters (e.g., LH, FSH, fasting insulin). These features have been widely reported in the literature as significant markers for PCOS diagnosis.
- Synthetic Augmentation: To balance the dataset and ensure sufficient representation across PCOS subtypes, additional records were generated synthetically. A controlled randomization strategy was applied to simulate realistic variations in clinical and lifestyle features while preserving domain relevance. The synthetic data was generated using ChatGPT, where feature distributions (e.g., hormonal ranges, anthropometric measures) were constrained to medically plausible intervals derived from existing literature and clinical reports. This approach prevented the introduction of extreme or biologically implausible values while expanding the dataset's diversity.
- This study acknowledges certain limitations related to the use of synthetic data and the absence of clinical validation. Synthetic augmentation using ChatGPT was employed only to address class imbalance and improve model generalization, particularly for underrepresented PCOS subtypes such as Adrenal and Post-Pill, without replacing authentic clinical data. All generated samples followed medically validated patterns from peer-reviewed literature under human supervision to ensure consistency and ethical compliance. The dataset used was fully anonymized and publicly available, safeguarding patient privacy and adhering to data protection guidelines such as GDPR and HIPAA principles. While synthetic data enhanced model training, it may introduce bias or limit physiological variability, and thus requires validation with real-world clinical datasets to confirm reliability. Future work aims to collaborate with healthcare professionals for clinical trials and user feedback, ensuring that predictions and lifestyle recommendations align with real patient outcomes.
- Overall, the project emphasizes transparency, accountability, and responsible AI practices in health care research to support ethical and privacy-conscious innovation

D. Workflow

The proposed PCOS Type Detection and Lifestyle Guidance system is developed to offer an intuitive and accessible experience for both patients and healthcare professionals. The overall workflow of the system involves the following key stages:

- User Onboarding: Users access the web based application where they can either fill out a structured medical form or upload diagnostic reports for automated extraction through OCR.
- Data Input and Extraction: Clinical, biochemical, and lifestyle parameters are either manually entered or extracted using the integrated OCR engine, ensuring efficiency and reducing manual errors.
- Pre-Processing and Validation: The input data undergoes cleaning, encoding, and imputation to handle missing or inconsistent values. Each attribute is validated against defined medical ranges before analysis.
- Machine Learning Prediction: The Random Forest Classifier processes the refined dataset and predicts the patient's PCOS type (Insulin-Resistant, Inflammatory, Post-Pill, or Adrenal).
- Personalized Lifestyle Guidance: Based on the predicted type, the system generates tailored recommendations including dietary plans, exercise routines, supplement suggestions, and stress-management techniques.

The sequence diagram (Fig. 3) illustrates the interaction between the user input module, the OCR and preprocessing module, the machine learning classifier, and the recommendation engine

Fig. 4 provides a flowchart outlining the logical steps from user data entry through to PCOS type prediction and lifestyle recommendation generation.

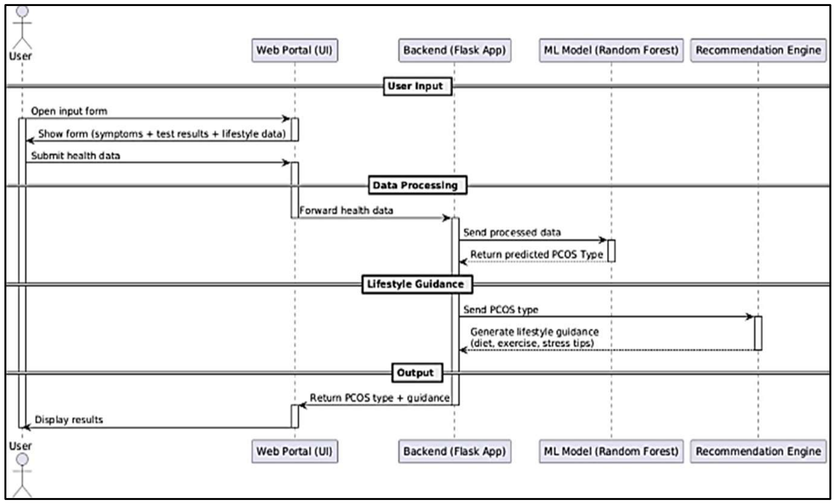

Fig .3. Sequence Diagram

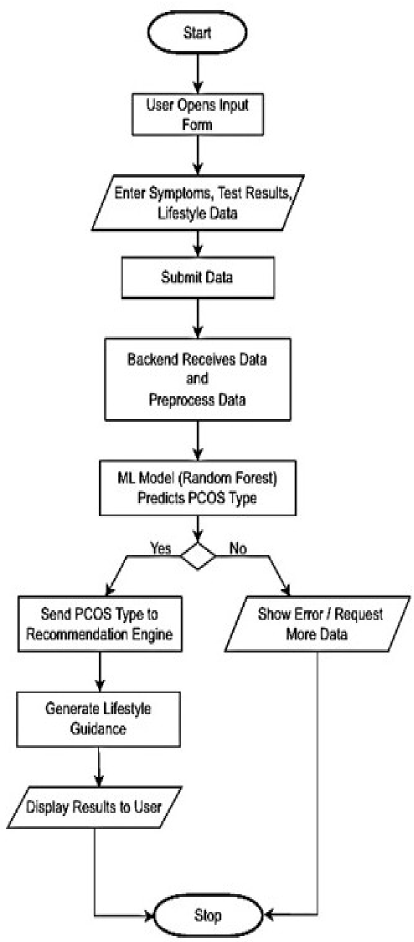

Fig. 4. Activity Diagram

IV. Results and Discussions

A. System and Software Requirements

To ensure smooth execution and reliable performance, the PCOS-Type Detection and Lifestyle Guidance system was developed with the following requirements:

Software Requirements

- Python 3.11: Primary environment for developing the machine learning model.
- Machine Learning Libraries: Scikit-learn, Pandas, NumPy, and Matplotlib for data preprocessing, model training, and evaluation.
- HTML, CSS, JavaScript: Used to design the interactive web-based interface for user input and displaying results.

Hardware Requirements

- Standard PC: Minimum specifications: Intel i5 processor, 8GB RAM.
- Internet Connection: Required for deployment and access to the web-based interface.

B. Comparative Analysis

The proposed PCOS detection model shows enhanced accuracy and reliability when compared to traditional manual diagnostic techniques. In contrast to conventional methods that depend primarily on clinical assessment, the system integrates multiple symptoms, biochemical data, and lifestyle factors to classify PCOS types more accurately.

Furthermore, the system not only identifies the PCOS subtype but also generates personalized lifestyle recommendations (e.g., diet, exercise, and wellness strategies). This dual functionality enhances patient engagement and provides guidance that extends beyond mere diagnosis.

PCOS Subtype	Precision	Recall	F1-Score
Adrenal	1.0000	0.5000	0.6667
Inflammatory	1.0000	0.9583	0.9787
Insulin-Resistant	0.8987	0.9861	0.9404
Post-Pill	0.7500	0.4286	0.5455
Unknown	0.0000	0.0000	0.0000
Overall Accuracy	0.9167		
Macro Average	0.7297	0.5746	0.6262
Weighted Average	0.9070	0.9167	0.9045

Figure. 5. Evaluation Metrics

As shown in figure 5, the Random Forest model attained an overall accuracy of 91.67%, with a weighted F1-score of 0.90, demonstrating strong and consistent performance across the major PCOS subtypes. High predictive performance was observed for the Insulin-Resistant subtype (F1 = 0.94) and the Inflammatory subtype (F1 = 0.98), indicating reliable classification for these dominant categories. The Adrenal and Post-Pill subtypes showed comparatively lower recall values (0.50 and 0.43, respectively), which can be attributed to limited sample representation and overlapping clinical features. The Unknown category could not be classified effectively due to the absence of sufficient training samples. The macro-average F1-score of 0.63 reflects performance variation across classes caused by dataset imbalance, highlighting the need for improved subtype representation in future work.

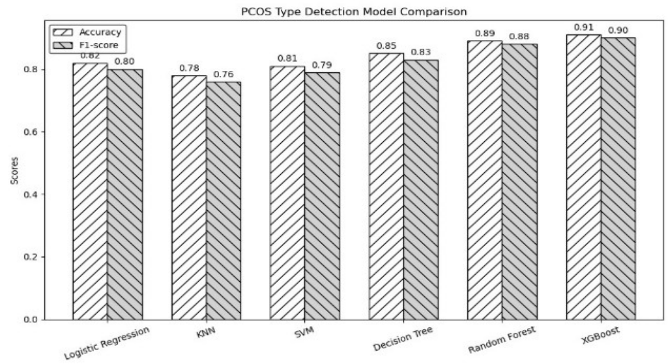

Figure. 6. Model Comparison

After comparing all the models as illustrated in figure 6, Random Forest showed strong and steady results with around 89% accuracy. Even though XGBoost performed slightly better, Random Forest was chosen because it's easier to interpret, more stable, and faster to train. In a health-related system like PCOS detection, it's important to understand why the model makes certain predictions. Random Forest gives reliable results without needing heavy tuning, making it a practical and trustworthy choice for this project. As noted, the following Figs. 7-8, features snapshots of PCOS Management and Guidance. The inclusion of a user-friendly interface ensures accessibility, allowing patients and clinicians to interact with the system seamlessly.

The developed system utilizes a Random Forest Classifier as its primary predictive model. This algorithm was selected due to its strong capability to process complex, high-dimensional clinical datasets, efficiently handle both categorical and numerical variables, and minimize overfitting through the use of multiple decision trees. The model was trained using a carefully curated clinical dataset comprising demographic, anthropometric, biochemical, and lifestyle parameters. Evaluation results demonstrated high accuracy and stability, confirming the model's effectiveness in accurately classifying different PCOS subtypes.

The system generates two key outputs:

PCOS Type Prediction

Once the user enters the clinical and lifestyle information manually or uploads a medical report, the system classifies the patient into one of the four PCOS categories:

- Insulin-Resistant
- Inflammatory
- Post-Pill
- Adrenal

This classification is displayed clearly on the user dashboard, enabling patients and healthcare providers to understand the specific type of PCOS the individual is likely experiencing.

Personalized Lifestyle Guidance

Beyond prediction, the system provides personalized recommendations tailored to the predicted PCOS type.

The guidance is structured into four domains:

- Diet Plan – customized meals based on low glycemic index foods, anti-inflammatory diets, or hormone-balancing nutrients.
- Exercise Plan – activity suggestions ranging from strength training to yoga or low-intensity cardio, depending on the PCOS type.
- Supplements – evidence-based supplements like Inositol, Omega-3, Zinc, or Ashwagandha recommended with indicative dosages.
- Stress Management – mindfulness, journaling, sleep hygiene, or relaxation techniques for holistic health improvement.

This personalization aspect ensures that every user receives type-specific and actionable guidance rather than generic advice, which enhances the system's practical utility.

The figure 7 shows the form-based input interface, where the user provides health parameters or uploads a report.

Figure 8 presents the prediction result, displaying the classified PCOS type and personalized lifestyle guidance module, offering a structured health plan including diet, exercise, supplements, and stress management tips.

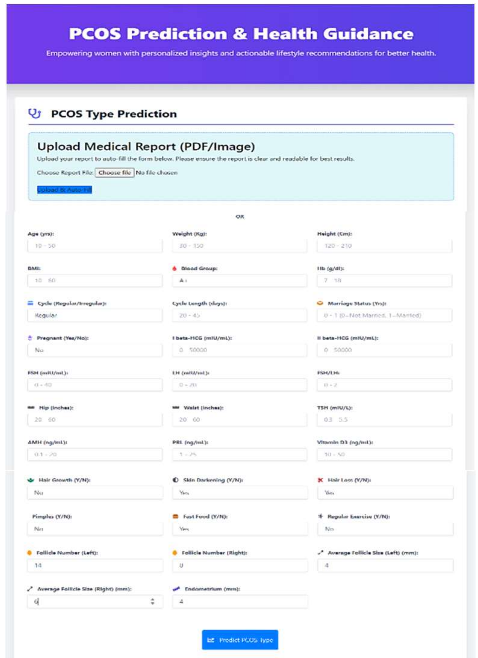

Fig. 7. Interface

By integrating machine learning with clinically validated parameters, this system enables early detection and classification of different PCOS types while improving access to healthcare through technology.

The framework also offers personalized lifestyle recommendations, helping women make informed choices regarding their reproductive and overall health, which can enhance their quality of life.

In addition, the system supports global development goals, particularly the United Nations Sustainable Development Goals: SDG 3 (Good Health and Well-being) by encouraging preventive care and health awareness, and SDG 5 (Gender Equality) by focusing on women's health issues that are often overlooked in research and digital health initiatives. This demonstrates that the impact of the system goes beyond individual health, contributing to wider societal and global benefits.

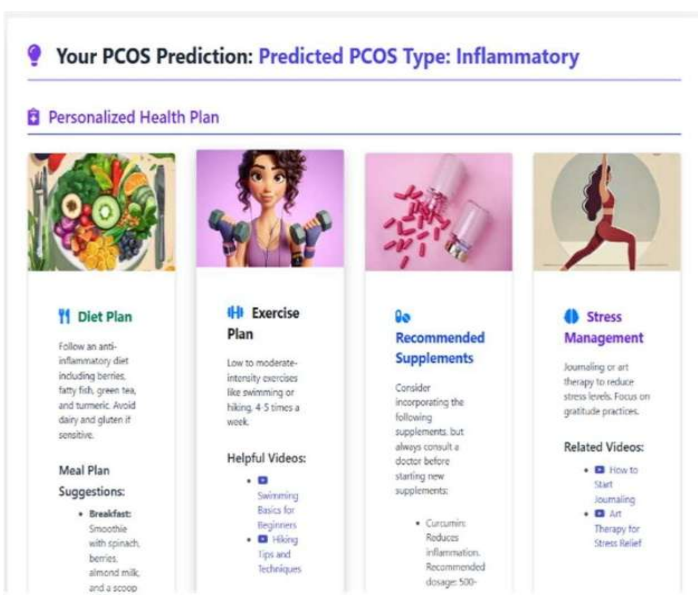

Fig. 8. Output

C. Accuracy of the model

The confusion matrix in figure 9 illustrates the classification performance of the Random Forest model across the five PCOS subtypes: Adrenal, Inflammatory, Insulin-Resistant, Post-Pill, and Unknown.Cells along the diagonal show the instances that were correctly predicted, whereas the off-diagonal cells indicate where misclassifications occurred.

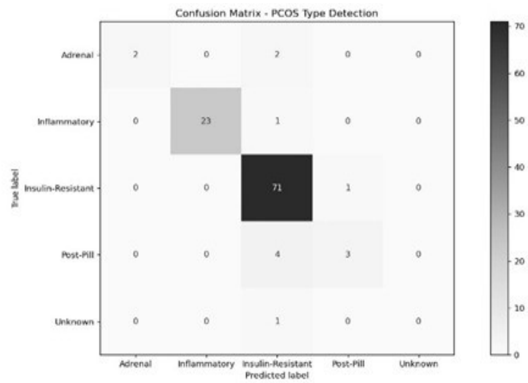

Fig. 9. Confusion Matrix

The matrix shows that the model performs strongly for the Insulin-Resistant and Inflammatory subtypes, with very high correct prediction counts. A few misclassifications are observed for Adrenal and Post-Pill cases, which could be due to overlapping clinical features between subtypes.

This visualization provides clear evidence of the model's ability to differentiate between PCOS phenotypes, highlighting both strengths and areas for improvement.

The classification metrics in figure 5 presents precision, recall, and F1-scores for each PCOS subtype.

The Insulin-Resistant and Inflammatory subtypes achieve the highest overall metrics, reflecting the robustness of predictions for these groups.

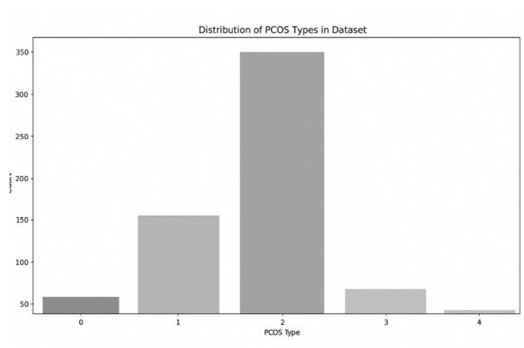

Fig .10. Dataset Distribution Chart

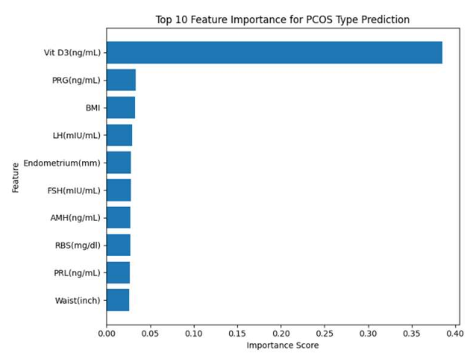

Fig 11

The Adrenal and Post-Pill types show lower recall and F1-scores, indicating that these cases are more challenging to classify correctly.

The overall accuracy of the model is 91%, which demonstrates high predictive reliability. The dataset distribution chart in figure 10 displays the relative frequency of different PCOS types within the collected data.

The majority of patients fall under the Insulin-Resistant subtype, followed by the Inflammatory group.

The Adrenal, Post-Pill, and Unknown types occur less frequently, which partly explains the model's lower performance in predicting these categories due to class imbalance.

This visualization is crucial for contextualizing the predictive results, as it shows how dataset imbalance influences model accuracy and underscores the clinical importance of focusing on the most prevalent PCOS phenotypes.

The feature importance analysis in Figure 11 shows that Vitamin D (Vit D3) is the most influential Fig .11. Feature Importance Analysis parameter for PCOS subtype prediction. Key hormonal markers such as

Progesterone (PRG), LH, FSH, AMH, and Prolactin (PRL) also contribute significantly, highlighting the central role of hormonal imbalance in PCOS. Additionally, BMI, Random Blood Sugar (RBS), and Waist circumference appear among the important predictors, emphasizing the strong association between PCOS and metabolic dysfunction. Endometrium thickness shows a moderate contribution, indicating its relevance in reproductive health assessment markers.

This visualization not only validates the biological relevance of the model's predictions but also demonstrates the potential of AI-driven approaches to prioritize clinically significant features, assisting doctors in both diagnosis and treatment planning.

D. Evaluation Metrics

The system was validated through multiple test cases, as summarized in Table 12.1. The results confirm that the model effectively handles valid inputs while also detecting invalid or empty entries, ensuring robustness in real-world applications:

- Error Handling: Proper error messages were generated for invalid or missing inputs, ensuring system robustness.
- Accuracy: The model achieved high diagnostic accuracy across multiple test runs, confirming reliability in predicting PCOS types.
- Personalization: Lifestyle recommendations were dynamically updated when users modified their input symptoms, highlighting adaptability.

The system functionality was validated through multiple test cases, as summarized in Table 1. The results indicate that the system excels not only in accurate detection but also in providing personalized guidance for each patient.

TABLE I: TEST CASES

T.id	INPUT	EXPECTED OUTCOMES	PASS / FAIL
1	Input valid relevant symptoms like irregular periods, weight gain etc.	Detected PCOS type is displayed to the user.	PASS
2	Input invalid data (e.g., random text) in the symptoms field.	Error message "Invalid input. Please enter valid symptoms".	PASS
3	Submit empty fields for symptoms and concerns.	Error message "Please enter symptoms for analysis".	PASS
4	Input valid hormone levels and request personalized lifestyle guidance.	Detected PCOS type with tailored lifestyle recommendations (diet, exercise, etc.) is displayed.	PASS
5	Generate recommendations after making changes to input.	System recalculates and displays updated PCOS type and lifestyle. recommendations	PASS

The Random Forest model achieved an overall accuracy of 91.7% with a weighted F1-score of 0.90, indicating strong performance across most PCOS subtypes as illustrated in figure 5. However, the macro F1-score of 0.63 reveals that the model's performance varies across classes due to class imbalance. Insulin-Resistant and Inflammatory types were classified accurately, while Adrenal and Post-Pill types had lower recall, suggesting fewer data points or overlapping symptoms. The Unknown category was not correctly classified, possibly due to insufficient samples. Future work will focus on balancing the dataset and improving the detection of less represented subtypes.

E. Comparison with State-of-the-Art Systems:

Compared to existing PCOS diagnostic methods and AI-based models, the proposed system provides two key advantages:

- Integrated Guidance: Unlike most studies that focus only on detection, this system combines diagnosis with lifestyle advice, making it more patient-centric.
- Interactive Interface: The real-time input-output interaction via a web-based platform enhances usability for both medical professionals and patients.

This positions the system as a step ahead of traditional statistical approaches and static diagnostic models.

F. Analysis

The evaluation highlights the potential of AI-driven PCOS diagnosis and management systems. The proposed framework provides faster, more consistent, and user-friendly detection compared to conventional diagnostic methods. By offering lifestyle recommendations tailored to the detected subtype, it addresses the practical needs of patients, encouraging proactive self-care.

However, challenges remain. First, dataset limitations must be addressed by incorporating larger, more diverse populations to improve generalizability. Second, future development should include integration with clinical decision support systems for seamless adoption in healthcare settings. Finally, incorporating multi-language support and mobile deployment will ensure broader accessibility.

With these refinements, the PCOS Detection and Lifestyle Guidance system has the potential to serve as a comprehensive digital health assistant, bridging the gap between data-driven research and real-world patient care.

V. Conclusion

The proposed PCOS-Type Detection and Lifestyle Guidance system demonstrates how artificial intelligence and data-driven methods can be effectively used to support accurate and user-friendly diagnosis of PCOS. Unlike conventional clinical approaches that are often time-consuming and dependent on manual interpretation, the system enables quick identification of PCOS subtypes along with personalized lifestyle recommendations related to diet and exercise. The performance of the system was validated through multiple test cases, showing reliable accuracy, proper handling of invalid inputs, and dynamic updating of recommendations based on user data.

While the current results are promising, the study also recognizes the need for larger and more diverse datasets, improved clinical interpretability of model outputs, and wider accessibility across platforms. Future work will focus on strengthening clinical collaboration through real-world validation with healthcare professionals, integrating the system into clinical decision-support environments, and deploying the application on mobile platforms for broader reach. These improvements aim to support real-time clinical use and enhance patient care.

Overall, the work demonstrates the potential of AI-based systems to assist both patients and clinicians by providing timely, reliable, and practical support for PCOS diagnosis and management.

References

[1] H. J. Teede, M. L. Misso, M. F. Costello, A. Dokras, J. Laven, and L. Moran et al., "Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome," Human Reproduction, vol. 33, no. 9, pp. 1602–1618, 2018, doi: 10.1093/humrep/dey256.

[2] S. Shekhar, F. J. Barrera, E. Brown, A. Rojo, J. Obeso, and H. Plata et al., "Artificial intelligence and machine learning in the detection of polycystic ovarian syndrome: A systematic review and meta-analysis," Endocrine Practice, vol. 29, no. 5, pp. S88–S89, 2023, doi: 10.1016/j.eprac.2023.03.198.

[3] J. Balakrishnan and Y. K. Dwivedi, "Conversational commerce: Entering the next stage of AI-powered digital assistants," Annals of Operations Research, vol. 333, pp. 653–687, 2024, doi: 10.1007/s10479-021-04049-5.

[4] M. Dapas, F. T. J. Lin, G. N. Nadkarni, R. Sisk, R. S. Legro, and M. Urbanek et al., "Distinct subtypes of polycystic ovary syndrome with novel genetic associations: An unsupervised, phenotypic clustering analysis," PLOS Medicine, vol. 17, no. 6, 2020.

[5] Ş. Yaşar, "Machine learning-driven metabolomic biomarker discovery for PCOS: An interpretable approach using Random Forest and SHAP," Medical Records, vol. 7, no. 3, pp. 763–767, Sep. 2025, doi: 10.37990/medr.1718952.

[6] S. C. R. Nandipati and C. X. Ying, "Polycystic ovarian syndrome (PCOS) classification and feature selection by machine learning techniques," Applied Mathematics and Computational Intelligence, vol. 9, pp. 65–74, doi: 10.1038/s41377-021-00658-8.

[7] V. Gupta and P. V. Suresh, "A comprehensive review of predicting lifestyle-based disease specifically PCOS among women using data mining and machine learning approaches," in Soft Computing: Theories and Applications, R. Kumar, A. K. Verma, O. P. Verma, and T. Wadehra, Eds. Lecture Notes in Networks and Systems, vol. 971, Singapore: Springer, 2023.

[8] N. Yadav, S. D. Pande, and A. Ranjith Kumar, "Automated PCOS diagnosis and medical report generation using YOLO v11 and Pali-Gemma v2," Biomedical Materials & Devices, 2025, doi: 10.1007/s44174-025-00395-1.

[9] A. Kumar, A. Goyal, B. K. Rai, and S. Sharma, "OCR-based medical prescription and report analyzer," in AIP Conference Proceedings, vol. 2424, 2022, Art. no. 070006, doi: 10.1063/5.0081176.

[10] A. Ghazaryan, J. Hovhanisyan, A. Kirakosyan, A. Aramyan, and S. M. Moghadam, "An intelligent web application for drug information retrieval using optical character recognition and deep learning," Norwegian Journal of Development of the International Science, no. 152, 2025, doi: 10.5281/zenodo.15024998.

[11] P. Kottarathil, "Polycystic Ovary Syndrome (PCOS) Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos/data