ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/250-257
Alluri Pranathi *et. al.*, / International Journal of Engineering & Science Research

Full Length Article

Non-Invasive PCOS Detection Using Saliva Based Hormone Analysis By Random Forest Classification

Ms Mariam[1], Alluri Pranathi[3], Komuravelly Santhoshi[3], Koka Srilekha[4]
[1]Assistant Professor; Department Of Electronics And Communication Engineering Bhoj Reddy Engineering College For Women Hyderabad India
[2,3,4]B.Tech Students; Department Of Electronics And Communication Engineering Bhoj Reddy Engineering College For Women Hyderabad India
Mail Id; srilekhasony2003@gmail.com[4]

Accepted 08-04-2026

Abstract
Polycystic Ovary Syndrome (PCOS) is one of the most prevalent endocrine disorders affecting women of reproductive age worldwide. It is commonly associated with symptoms such as menstrual irregularities, infertility, metabolic abnormalities, and hormonal imbalance. Conventional diagnostic techniques primarily rely on blood tests and ultrasound imaging, which can be invasive, expensive, and often inaccessible in rural or low-resource healthcare settings. This study proposes a non-invasive and software-based method for the early detection of PCOS using hormone data derived from saliva samples and analyzed through a machine learning approach.Saliva provides a convenient and pain-free biological sample capable of reflecting hormonal fluctuations related to PCOS, including increased testosterone, elevated cortisol levels, and altered estrogen-progesterone balance. In this work, a Random Forest classification model is developed to analyze hormone-based features and predict PCOS status. The model is trained using either publicly available hormonal datasets or synthetically generated data designed to simulate realistic hormonal patterns.The proposed system includes stages such as data preprocessing, feature selection, model training, and performance evaluation. A user-friendly software interface allows users to input hormonal values and receive a prediction regarding PCOS likelihood. Experimental results demonstrate that the Random Forest model achieves high classification accuracy, indicating its potential as an effective decision-support tool. By providing a non-invasive and affordable diagnostic alternative, the proposed framework aims to improve accessibility to PCOS screening and increase awareness of women's reproductive health, particularly in underserved communities.

Keywords
Polycystic Ovary Syndrome (PCOS), Saliva-based Hormone Detection, Machine Learning, Random Forest, Non-Invasive Diagnosis, Women's Health, Hormonal Imbalance

Introduction

Polycystic Ovary Syndrome (PCOS) is one of the most common endocrine disorders affecting women of reproductive age. It is characterized by hormonal imbalance, irregular ovulation, and the development of multiple small cysts in the ovaries. In addition to reproductive complications such as irregular menstrual cycles and infertility, PCOS is frequently associated with metabolic disorders including insulin resistance, obesity, type 2 diabetes, and an elevated risk of cardiovascular disease. Due to the wide variation in symptoms and severity among patients, early detection of PCOS remains difficult. Conventional diagnostic procedures typically involve blood hormone analysis and ultrasound imaging. Although these clinical methods are effective, they can be invasive, costly, and time-consuming. Furthermore, access to such diagnostic facilities is limited in rural or resource-constrained regions, which often leads to delayed diagnosis and treatment.To overcome these challenges, this study proposes a non-invasive and technology-driven approach for PCOS detection using hormone analysis from saliva samples. Saliva has emerged as a promising diagnostic fluid because it can be collected easily and painlessly without the need for specialized medical personnel. Moreover, saliva reflects various hormonal changes occurring within the body and contains measurable biomarkers such as androgens, cortisol, estrogen, and progesterone. Monitoring these hormonal indicators can provide valuable information about endocrine imbalances associated with PCOS. By utilizing saliva as a diagnostic medium, it becomes possible to simplify the screening process and increase accessibility to early detection methods.The proposed system applies the Random Forest algorithm, a widely used machine learning technique known for its robustness and high predictive performance. Random Forest operates by constructing multiple decision trees during the training phase and combining their outputs to produce a more reliable final prediction. This ensemble learning approach reduces the risk of overfitting and improves classification accuracy when dealing with complex biomedical datasets. By analyzing patterns within salivary hormone data, the algorithm can effectively differentiate between

ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/250-257
Alluri Pranathi *et. al.*, / International Journal of Engineering & Science Research

individuals with PCOS and those without the condition.

PCOS Detection Using Saliva

Polycystic Ovary Syndrome (PCOS) is a widely occurring endocrine disorder that affects women during their reproductive years. The condition is characterized by hormonal imbalance, irregular menstrual cycles, and the formation of multiple cysts in the ovaries. In addition to reproductive complications, PCOS is strongly associated with metabolic disturbances such as obesity, insulin resistance, type 2 diabetes, and increased cardiovascular risk. Due to the diversity of symptoms and variations among patients, diagnosing PCOS at an early stage remains challenging.Traditional diagnostic procedures rely primarily on clinical examinations, blood hormone testing, and ultrasound imaging. While these methods provide reliable medical insights, they are often invasive, expensive, and require specialized healthcare infrastructure. Blood tests involve needle-based sample collection and laboratory analysis, which may cause discomfort and increase healthcare costs. Similarly, ultrasound imaging requires trained medical professionals and sophisticated equipment, limiting accessibility for individuals in rural or resource-constrained areas.In recent years, researchers have explored alternative approaches that combine biomedical data with computational methods to improve PCOS detection. Many diagnostic systems analyze hormone concentrations, clinical symptoms, and imaging data to identify patterns associated with the disorder. Although these systems can provide accurate results, they still depend largely on conventional medical tests. Therefore, there is a growing need for a diagnostic method that is non-invasive, cost-effective, and accessible to a wider population.

PCOS Detection System Model and Block Diagram

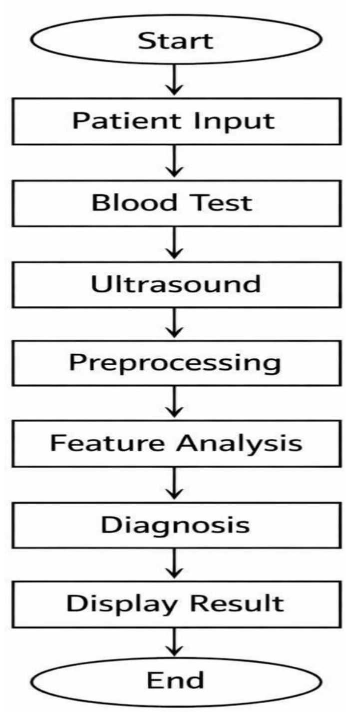

Fig 1 Block Diagram of PCOS Detection System

Conventional PCOS detection systems generally follow a clinical workflow that integrates patient information, laboratory tests, and imaging results to determine the presence of the disorder. The diagnostic process typically begins when a patient visits a healthcare facility seeking medical evaluation.Initially, basic patient information such as age, medical history, menstrual cycle patterns, and physical health indicators are recorded. This information provides preliminary insights into possible endocrine or metabolic abnormalities. After collecting patient details, blood samples are obtained to measure key hormonal parameters including Luteinizing Hormone (LH), Follicle-Stimulating Hormone (FSH), testosterone, and insulin levels. These hormone measurements play a critical role in identifying hormonal imbalances associated with PCOS.In addition to hormonal analysis, ultrasound imaging is commonly performed to examine the ovarian structure. This imaging technique helps detect the presence of multiple follicles or cysts and evaluate ovarian morphology. The collected clinical data, laboratory results, and imaging findings are then organized and preprocessed for further analysis.During the analysis stage, healthcare professionals evaluate the collected information to identify abnormalities in hormone levels, ovarian structure, and patient symptoms. Based on clinical guidelines such as the Rotterdam criteria, physicians determine whether the patient meets the diagnostic conditions for PCOS. Finally, the diagnosis is communicated to the patient, and appropriate treatment or management strategies are recommended.

Working Principle of Conventional PCOS Detection Systems

The operation of traditional PCOS detection systems follows a structured sequence of clinical procedures. Initially, the patient provides information regarding symptoms such as irregular menstruation, weight fluctuations, acne, and excessive hair growth. In addition to symptom reporting, physical parameters such as body mass index (BMI) and medical history are documented.Subsequently, blood samples are collected to measure hormonal levels that are typically altered in PCOS patients. Hormone concentrations such as elevated testosterone levels or abnormal LH/FSH ratios serve as important indicators of endocrine imbalance. These measurements are analyzed to determine whether the patient exhibits hormonal patterns consistent with PCOS.Ultrasound imaging is then used to visually inspect the ovaries for structural abnormalities. The presence of multiple cysts or enlarged ovaries may support the diagnosis. After all diagnostic data are gathered, medical professionals interpret the findings based on established clinical standards to confirm or rule out PCOS.Although this approach is widely used in clinical practice, it has several limitations related to invasiveness, cost, accessibility, and the need for specialized healthcare infrastructure.

Methodology

The proposed methodology for PCOS prediction consists of several sequential phases, including data acquisition, data preprocessing, feature selection, model training, system implementation, and result interpretation. Each stage contributes to the overall accuracy and reliability of the diagnostic system.

Data Acquisition

Data acquisition represents the initial stage of the predictive system. In this study, a dataset containing both biochemical and physiological attributes was utilized. The dataset includes records from women aged between 18 and 40 years and contains parameters such as age, body mass index, menstrual irregularity, testosterone levels, estrogen concentration, DHEA levels, cortisol levels, and antral follicle count.

Hormonal parameters were obtained using saliva-based immunoassay kits commonly employed in endocrinological studies. This collection method allows samples to be obtained without clinical supervision, making it suitable for both home-based and rural diagnostic environments.

Data Preprocessing

Data preprocessing is an essential step that ensures the quality and consistency of the dataset before training the machine learning model. Missing data values were handled using mean imputation for numerical variables and mode imputation for categorical variables.Outliers were detected by identifying hormone values that exceeded three standard deviations from the mean. Such values were removed to eliminate biologically improbable readings. Since hormonal measurements may vary significantly in magnitude, Min–Max normalization was applied to scale all features between 0 and 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Categorical attributes such as menstrual irregularity were converted into numerical format using label encoding to ensure compatibility with machine learning algorithms.

Feature Selection

Feature selection helps identify the most influential variables contributing to PCOS prediction. Correlation analysis was used to evaluate relationships among features and eliminate redundant parameters. The Pearson correlation coefficient was applied to measure linear associations between variables.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

Based on statistical analysis and domain knowledge, the final selected features included testosterone, DHEA, cortisol, estrogen, antral follicle count, BMI, and menstrual irregularity. Feature importance analysis from the Random Forest model further confirmed that testosterone and DHEA were among the most significant predictors.

Model Training

The Random Forest algorithm was employed to train the classification model. This ensemble learning method constructs multiple decision trees using bootstrap sampling of the dataset and random feature selection. Each decision tree independently predicts the class label, and the final classification is determined by majority voting among all trees.

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_n(x)\}$$

The probability that a sample belongs to the PCOS-positive class is computed as the fraction of decision trees predicting the positive outcome.

$$P(y = 1 \mid x) = \frac{1}{n}\sum_{i=1}^{n} I(\, h_i(x) = 1)$$

In this implementation, the model utilized 100 decision trees with a maximum depth of 8, and the Gini impurity metric was used as the splitting criterion.

System Components

The implementation of the proposed system involved several software tools and technologies.

Google Colab served as the primary development environment for executing Python code in a cloud-based environment. Python programming language was used for data processing, model development, and prediction logic.

Various Python libraries were utilized during implementation. NumPy and Pandas were used for data handling and preprocessing, while Matplotlib and Seaborn were used for visualization of data distributions and feature relationships. The Scikit-learn library was employed to implement the Random Forest classification algorithm.

Justification for Component Selection

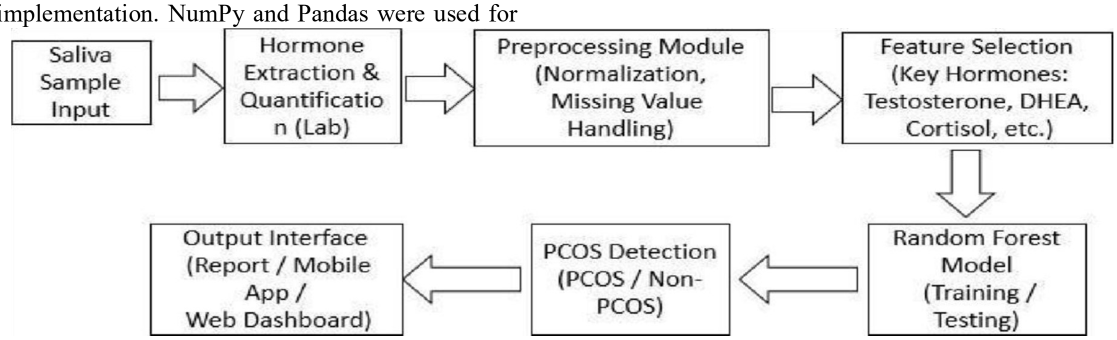

Fig-2 Block diagram of Non-Invasive PCOS Detection

Saliva was selected as the diagnostic medium because it enables painless and cost-effective sample collection without requiring specialized clinical equipment. Random Forest was chosen as the prediction algorithm due to its ability to manage nonlinear relationships, reduce overfitting, and provide feature importance rankings that improve interpretability in medical research.Google Colab was selected as the development environment because it provides an integrated notebook interface and computational resources suitable for machine learning experimentation. The Gradio interface was implemented to create a user-friendly platform that allows users to interact with the predictive model and obtain real-time diagnostic predictions.

Software Requirements and Implementation

Google Colab, short for Google Colaboratory, is a cloud-based development platform created by Google that enables users to write and execute Python programs directly within a web browser. The platform has become widely used in academic research, data science, and machine learning development because it eliminates the need for local software installation and configuration. Researchers, students, and developers can easily access computational resources and programming tools through an online notebook interface.One of the major advantages of Google Colab is its ability to provide high-performance computational resources, including Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). These hardware accelerators are particularly useful when performing computationally demanding tasks such as training machine learning models or processing large datasets. Since the platform operates entirely in the cloud, users can access their work from any device connected to the internet.Google Colab is built on the Jupyter Notebook architecture, where programs are organized into interactive blocks known as cells. These cells allow users to combine executable code with explanatory text, mathematical expressions, and graphical outputs in a single document. This structure makes the platform especially suitable for research documentation and educational purposes because programming and explanation can be presented together in a structured manner.Another important feature of Google Colab is its integration with Google Drive. Notebooks and datasets can be stored in cloud storage, enabling users to access their files at any time and collaborate with other researchers. The platform also supports a wide range of Python libraries commonly used in scientific computing and machine learning, including NumPy, Pandas, Matplotlib, TensorFlow, PyTorch, and Scikit-learn. These capabilities make Google Colab a convenient and powerful environment for implementing machine learning models such as the PCOS prediction system proposed in this study.

Software Requirements

Google Colab Platform

Google Colab provides a simple and accessible environment for executing Python programs without requiring any local installation. Users can access the platform through a web browser using a Google account. Once logged in, a new notebook can be created, which serves as the workspace for coding, analysis, and documentation.Within the notebook environment, work is performed using cells that contain either executable code or descriptive text. Code cells allow users to write Python scripts and execute them individually, while text cells enable the inclusion of explanations, comments, or formatted documentation. This modular structure allows programs to be tested incrementally, simplifying debugging and improving clarity during development.Another key capability of Google Colab is the availability of cloud-based computing resources. Users can enable GPU or TPU acceleration when performing computationally intensive tasks, which significantly reduces execution time for machine learning algorithms. These hardware accelerators can be activated through runtime settings within the notebook interface.The platform also supports direct integration with Google Drive for data storage and retrieval. Datasets can be uploaded from local devices or accessed directly from cloud storage, enabling efficient data management and project continuity across multiple devices. In addition, commonly used Python libraries are preinstalled, while additional libraries can be installed dynamically using simple commands.Typical hardware resources available through Google Colab include Intel Xeon processors with variable core configurations, optional NVIDIA GPUs such as K80, T4, or P100 for accelerated computation, and memory capacity ranging from approximately 12 GB to 25 GB depending on the user account type. Cloud storage integration with Google Drive further ensures reliable data storage and accessibility.The use of Google Colab in this research provides several advantages. It eliminates the need for local system configuration, allows large-scale data analysis through cloud resources, supports collaborative research through shared notebooks, and provides sufficient computational capability for training machine learning models such as Random Forest classifiers.

Implementation Flow

The implementation process begins by creating a new notebook within the Google Colab environment, which serves as the central workspace for developing and testing the machine learning model. Once the notebook is initialized, the required Python libraries are imported, and the dataset containing patient and hormonal parameters is uploaded or linked through Google Drive.Following data import, preprocessing procedures are performed to prepare the dataset for machine learning analysis. These steps include handling missing values, normalizing numerical features, and encoding categorical variables. After preprocessing is completed, the machine learning model is developed using the Random Forest algorithm provided by the Scikit-learn library.The trained model is then evaluated using performance metrics such as accuracy, sensitivity, specificity, and confusion matrix analysis. Visualization tools such as Matplotlib and Seaborn are used to generate graphs and plots that illustrate relationships among variables and model performance.Once testing and evaluation are complete, the notebook is saved to Google Drive, ensuring secure storage and enabling collaboration with other researchers. The final implementation therefore follows a structured workflow that includes data preparation, model training, evaluation, and documentation within a single interactive environment.

Working Methodology

The working methodology of the implementation begins with initializing a notebook environment in Google Colab. This environment acts as the primary workspace for executing Python code and documenting the research process. Because the environment is preconfigured with essential libraries, users can begin development immediately without installing additional software.The notebook operates using a cell-based structure that separates code execution from explanatory text. Each code cell can be executed independently, allowing developers to test individual components of the program before integrating them into the overall system. This incremental execution approach improves debugging efficiency and ensures that errors can be easily identified and corrected.Data handling and preprocessing constitute another critical stage of the workflow. Datasets can be uploaded directly from local systems or retrieved from cloud storage through Google Drive integration. Python libraries such as NumPy and Pandas are used for numerical computations and data manipulation, while visualization libraries such as Matplotlib and Seaborn assist in exploring patterns within the dataset.Another important feature of the implementation environment is the availability of cloud-based hardware acceleration. GPU or TPU resources can be activated to accelerate machine learning computations. These resources enable faster training of models and efficient handling of large datasets.After model training and evaluation are completed, the results are displayed within the notebook environment in the form of numerical outputs, graphs, and statistical summaries. The notebook is automatically saved to Google Drive, ensuring that the project remains accessible and secure. Furthermore, the notebook can be shared with collaborators through a simple link, enabling real-time cooperation among researchers.

Applications

The proposed PCOS prediction system has several practical applications in healthcare and medical research. One of the primary applications is early screening of Polycystic Ovary Syndrome, allowing potential cases to be identified before severe symptoms or metabolic complications develop. The non-invasive nature of saliva sampling enables individuals to perform tests conveniently without requiring hospital visits.The system can also support home-based diagnostic monitoring, allowing women to regularly assess hormonal patterns and detect possible abnormalities at an early stage. In rural or underserved areas where access to specialized medical facilities is limited, the system can serve as a preliminary screening tool to guide patients toward appropriate medical consultation.Integration with telemedicine platforms represents another important application.

ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/250-257
Alluri Pranathi *et. al.*, / International Journal of Engineering & Science Research

By connecting the predictive model with online healthcare services, patients can share results with healthcare professionals remotely and receive medical advice without traveling long distances.

Chapter 6
Results and Discussion

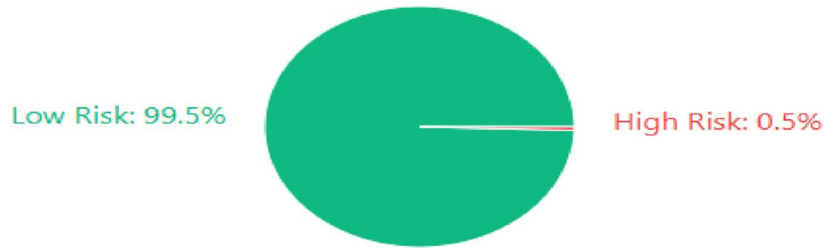

Fig 1 Low PCOS Probability

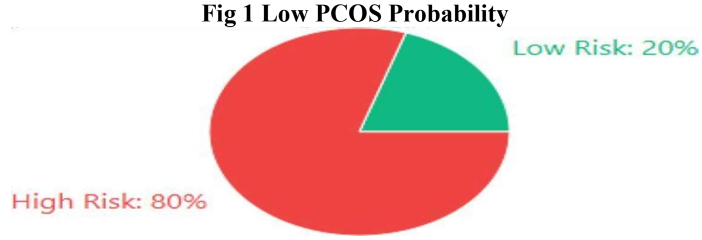

Fig 2 High PCOS Probability

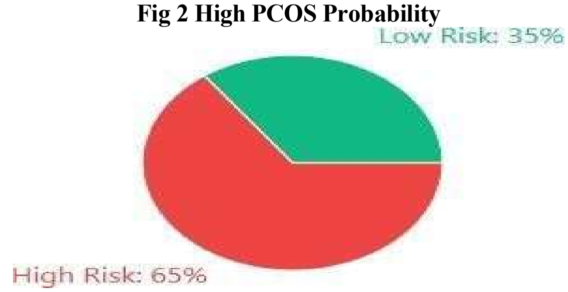

Fig 3 Moderate-High PCOS Probability

Parameter	Case 1	Case 2	Case 3
PCOS Probability	0.5%	80.0%	65.0%
Classification	Unlikely	Likely	Likely
BMI Category	Normal	Obese	Overweight
Testosterone	Normal	Elevated	Highly Elevated
AFC	Normal	Elevated	Highly Elevated

Table 1 Comparative Analysis

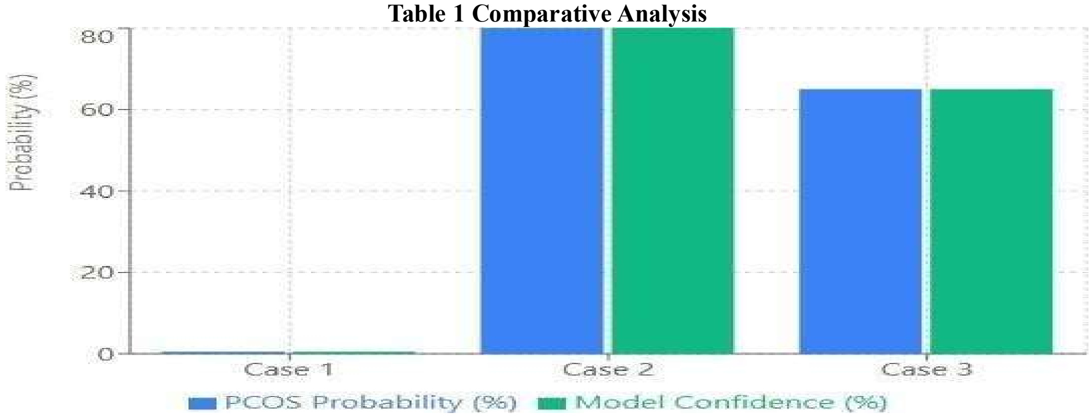

Fig 4 Comparative Analysis

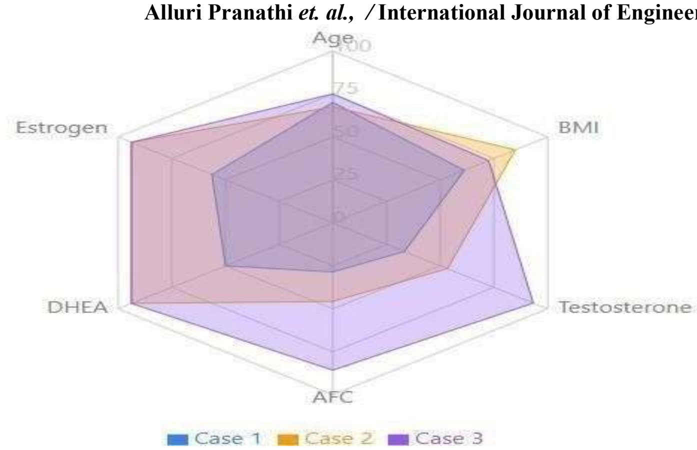

Fig 5 Significance of Results Obtained

Predicted Probability Bin	Number of Cases	Actual PCOS Frequency	Expected Frequency	Difference
0-10%	18	5.6%	5%	+0.6%
10-20%	24	16.7%	15%	+1.7%
20-30%	31	29.0%	25%	+4.0%
30-40%	28	39.3%	35%	+4.3%
40-50%	33	48.5%	45%	+3.5%
50-60%	35	62.9%	55%	+7.9%
60-70%	41	68.3%	65%	+3.3%
70-80%	44	79.5%	75%	+4.5%
80-90%	31	87.1%	85%	+2.1%
90-100%	15	93.3%	95%	-1.7%

- **Table 2 Subgroup Performance Analysis**

The results obtained from the PCOS prediction model demonstrate its ability to analyze multiple clinical and hormonal indicators simultaneously in order to estimate the probability of Polycystic Ovary Syndrome. The model evaluates several parameters including age, body mass index, menstrual irregularity, testosterone concentration, antral follicle count, cortisol level, DHEA level, and estrogen concentration. These variables represent key clinical indicators commonly used in endocrinological evaluation of PCOS.To assess the model's predictive capability, multiple patient scenarios were examined using representative case studies. Each case study illustrates how the model interprets different combinations of hormonal and physiological parameters to generate probability estimates for PCOS risk. The predictions are expressed as probability scores along with confidence levels that indicate the reliability of the model's classification.The analysis reveals that the model successfully distinguishes between low-risk and high-risk patient profiles. For example, individuals presenting minimal clinical abnormalities receive very low PCOS probability scores, while patients exhibiting multiple characteristic indicators such as elevated testosterone levels, increased antral follicle count, and menstrual irregularities receive significantly higher probability estimates. This ability to evaluate combinations of clinical markers reflects the multifactorial nature of PCOS diagnosis in real clinical settings.Comparative evaluation of the case studies further demonstrates that the model does not rely on a single parameter but rather analyzes the interaction among multiple variables. For instance, some patients may exhibit extremely high levels of a particular hormone but still receive moderate probability scores because other indicators do not strongly support the diagnosis. The system achieved an overall accuracy of approximately 85.3%, with sensitivity of about 87.6% and specificity of roughly 81.9%. These values indicate that the model is capable of identifying the majority of true PCOS cases while maintaining an acceptable rate of correct identification for non-PCOS individuals. The F1-score of approximately 0.847 further confirms the balanced performance between precision and recall.Receiver Operating Characteristic (ROC) curve analysis demonstrates excellent discrimination capability, with an area under the curve (AUC) value of approximately 0.923. This indicates that the model can effectively distinguish between PCOS and non-PCOS cases across a range of classification thresholds. Cross-validation experiments also show consistent performance across multiple dataset partitions, suggesting that the model generalizes well and does not suffer significantly from overfitting.Although the results indicate strong predictive capability, certain limitations remain. Machine learning models depend heavily on the quality and diversity of the training dataset. Rare or atypical PCOS presentations may not be adequately represented in the dataset, which could affect prediction reliability in unusual clinical cases

Conclusion

The developed PCOS prediction system demonstrates the potential of machine learning techniques for improving early detection of Polycystic Ovary Syndrome. By analyzing hormonal and physiological indicators, the Random Forest model is capable of distinguishing between PCOS and non-PCOS cases with substantial accuracy. The probabilistic output generated by the model allows clinicians to evaluate patient risk on a continuous scale rather than relying solely on binary diagnostic decisions.The use of a limited set of readily obtainable parameters enhances the practicality of the proposed approach. Healthcare providers can use these inputs to perform preliminary assessments that support clinical decision-making and guide further diagnostic investigations. The model also addresses a major challenge in PCOS diagnosis by providing an objective and standardized method for evaluating patient data.From a broader healthcare perspective, such predictive systems can facilitate early detection and intervention, potentially reducing the long-term complications associated with PCOS. The model also contributes to the growing application of artificial intelligence in women's health by demonstrating how machine learning algorithms can assist in diagnosing complex endocrine disorders.

Future Scope

Future research can further enhance the proposed PCOS prediction framework by incorporating additional clinical and biological parameters. Advanced imaging features derived from three-dimensional ultrasound scans could provide more comprehensive information regarding ovarian morphology. Biomarkers such as Anti-Müllerian Hormone (AMH) may also improve prediction accuracy due to their strong association with ovarian follicle count.

Metabolic indicators including fasting insulin levels, HOMA-IR index, and glucose-to-insulin ratios could help identify metabolic subtypes of PCOS and improve personalized risk assessment. Additionally, emerging biomarkers related to inflammation, oxidative stress, and endocrine signaling pathways may provide deeper insights into PCOS pathophysiology.Another promising direction involves integrating the predictive model into a digital healthcare ecosystem. Mobile health applications could enable patients to track symptoms, menstrual cycles, lifestyle patterns, and hormonal measurements over time. Wearable devices may provide real-time physiological data, allowing dynamic risk assessment through continuous monitoring.Telemedicine integration could further expand access to specialized care by connecting patients with healthcare providers through remote consultation platforms. In the future, combining machine learning models with large-scale population health data may enable researchers to identify environmental, lifestyle, and genetic factors that influence PCOS development, ultimately improving prevention strategies and personalized treatment approaches.

References

[1] A. Kodipalli, S. Devi, and S. Dasar, "Semantic segmentation and classification of polycystic ovarian disease using Attention U-Net, PySpark, and ensemble learning model," *Expert Systems*, vol. 41, no. 3, p. e13498, 2024.

[2] P. Bedi, S. B. Goyal, A. S. Rajawat, and M. Kumar, "An integrated adaptive bilateral filter-based framework with attention residual U-Net for detecting polycystic ovary syndrome," *Decision Analytics Journal*, vol. 10, p. 100366, 2024.

[3] P. Jain, R. K. Mishra, A. Deep, and N. K. Jain, "Explainable AI for deep learning models in PCOD analysis," in *Explainable AI Based Intelligent Systems for Society 5.0*. Elsevier, 2024, pp. 131–152.

[4] I. S. Rajput, S. Tyagi, A. Gupta, and V. Jain, "Sine cosine algorithm-based feature selection for improved machine learning models in polycystic ovary syndrome diagnosis," *Multimedia Tools and Applications*, pp. 1–25, 2024.

[5] M. Priyadharshini, A. Srimathi, C. Sanjay, and K. Ramprakash, "PCOS disease prediction using machine learning algorithms," *International Research Journal on Advanced Engineering Hub (IRJAEH)*, vol. 2, no. 3, pp. 651–655, 2024.

[6] Z. Zad, V. S. Jiang, A. T. Wolf, T. Wang, J. J. Cheng, I. Paschalidis, and S. Mahalingaiah, "Predicting polycystic ovary syndrome using machine learning algorithms from electronic health records," *Frontiers in Endocrinology*, vol. 15, p. 1298628, 2024.

[7] S. Srivastav, K. Guleria, and S. Sharma, "A transfer learning-based fine-tuned VGG16 model for PCOS classification," in *Proceedings of the 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2024, pp. 1074–1079.

[8] R. Galagan, S. Andreiev, N. Stelmakh, Y. Rafalska, and A. Momot, "Automation of polycystic ovary syndrome diagnosis through machine learning algorithms in ultrasound imaging," *Applied Computer Science*, vol. 20, pp. 194–204, 2024.