iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68601

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Smart Health Care: Machine Learning for PCOS Detection and Prediction

P Divitha[1], NS Vaishnavi[2], Mahajanam Sudeepthi[3], Yallari Sruthi[4], Mr. V. Raghubathy[5], Dr. S. Vijaya Kumar[6]
[1, 2, 3, 4]UG Scholar, [5]Assistant Professor, [6]Professor, Department of CSE (AI & ML), Sreenivasa Institute of Technology and Management Studies, Chittoor, India.

Abstract: Today, Polycystic Ovary Syndrome (PCOS) is found in numerous women, so it is a common issue. PCOS is a hormonal disorder that leads to delayed, irregular, or absent menstrual cycles in the female body. This syndrome may cause the growth of type 2 diabetes, gestational diabetes, weight gain, excess body hair, and many other complications. In advanced cases, PCOS may lead to infertility, which is a challenge for patients who are attempting to conceive. According to statistics, the rate of incidence of PCOS in recent years has greatly increased, which is alarming. If PCOS is diagnosed early enough, individuals can adhere to their physician's advice and live a healthier life. The data collection in this study includes records of 466 patients. The purpose of this study is to utilize machine learning models to determine patterns in this disorder. The learned information is then fed into different algorithms to determine accuracy, specificity, sensitivity, and precision using different ML models, including Logistic Regression (LR), Decision Tree (DT), XGBoost, Random Forest (RF), and Support Vector Machine (SVM) among others. The study used the Mutual Information model for feature selection and compared models to identify the most precise one. Using the Mutual Information model for feature engineering, AB and RF attained the highest accuracy of 94 %.
Keywords: Machine learning ,SVM,GradientBoosting,Polycystic ovary syndrome.

I. INTRODUCTION

Polycystic ovary syndrome more commonly known as PCOS. PCOS is the most common endocrine disease and entails imbalances sexhormones in a woman's body. The treatment period for PCOS isdependent on the symptoms of the patient. The cysts that develop on theovaries of certain affected individuals are the reason the syndromegets its name, though they are not a common symptom nor the reasonforthe illness. Medication, surgery, and lifestyle modification are some possible treatments for PCOS sufferers. PCOS is a common hormonal disorder in women of reproductive age who fall under the category of assigned female at birth (AFAB) [1]. PCOS can lead to ovulation cysts, hirsutism, acne, irregular menstruation, and elevated testosterone levels. Although the cause of PCOS is stillunknown, possible causesinvolve inflammation, genetics, or insulin resistance. PCOS may causean increased risk of diabetes, hypertension, depression, and other illnesses[2,3]. The female reproductive system includes the ovaries, fallopian tubes,uterus, and vagina [5]. The ovaries hold a lifelong supply of eggs that are stored in small, fluid-filled sacs known as follicles. Hormones regulating the activity of the ovaries are secreted by the pituitary gland sit lated at the base of the brain. Every month, the gland secretes follicle stimulating hormone (FSH) and luteinizing hormone (LH) into the bloodstream, which stimulates ovaries to mature eggs. When eggs mature, follicles secrete oestrogen. This causes ovulation when LH becomes elevated. Eggs not fertilized pass through the fallopian tube. Unused follicles break down, but in the event of PCOS, this process may be upset by menstrual cycle and possibly leading to infertility [4–6].Stein and Leventhal in 1935 found that PCOS is present in 5–10 % of women aged 12–45 [7].It affects more than 5 million women worldwide, with 70 % being commonly undiagnosed [8].Ethnicity also contributes, with diagnostic rates of 4.8 % among white women, 8 % among African American women, 6.8 % among Spanish women, and 31.3 % among Asian women [9]. PCOS occurs in 15.3 % of Indian women, considerably affecting their way of life and possibly resulting in disorders like anxiety, sleep apnea, and metabolic syndromes. This elevates the chances of developing diseases like diabetes, endometrial cancer, and heart diseases [10]. Early detection is important owing to its influence. Symptoms result from excessive androgen levels, leading to problems like weight gain, irregular menstruation, excess body hair, and infertility. Hormonal imbalance interferes with ovulation cycles [11,12]. Although the precise cause of PCOS is not yet known, insulin insensitivity is known to play a role, as illustrated by darkening of the skin in some areas. Lifestyle factors including diet and lack of exercise may also induce PCOS [13,14]. Treatment aimed at controlling symptoms: hormonal birth control controls cycles and minimizes testosterone; anti-androgens suppresses abnormal acne and hair growth; diabetes treatment such as metformin control insulin and cycles; fertility pills facilitateovulation. When meds do not work, laparoscopic ovarian drilling will decrease production of testosterone, possibly restoring the ability to ovulate [15].

With artificial intelligence (AI) help, we can currently identify and treat complicated sicknesses [16].By way of using machine learning algorithms, PCOS can now be identified earlier on, before adding up to constitute what it fully becomes [15].progress to a critical condition. Our research has designed a web-based interface coupled with machine learning models to predict PCOS at an early stage with a user-friendly interface. PCOS is a binary classification problem, which deals with the determination of whether a woman has PCOS or not. It gives a crisp prediction with a yes or no result. Different models like LR, GaussianNB, SVM, KNN,Bernoulli,Multinomial NB and XGBoost are used for this prediction. These models are able to give exact predictions, with each model giving different outputs, if they are provided with correct data. The system proposed preprocesses data by steps like string-to-integer conversion, fillling null values, and deleting unnecessary columns in initial stage.

II. LITERATURE SURVEY

The health and biological information collections are growing rapidly. In order to analyze such large and complex data, artificial intelligence and machine learning processes have become popular [17]. Several scholars have proposed AI-driven PCOS diagnosis algorithms in recent years using clinical parameters and vital signs as datasets.Silva et al. introduced the BorutaShap method and successively trained a random forest model in their paper. A dataset contained 73 healthy women and 72 PCOS patients. 58 features were ranked on the basis of their relevance and importance. Finally, the model wasable to attain an accuracy of 86 % [18]. V.V. Khanna et al.suggested a model for predicting PCOS in fertile patients. Theyutilized an open-source dataset of 541 patients from Kerala,India. The authors utilized ML models such as NB, DT, LR, KNN, RF, SVM, AdaBoost, Extra Trees andGradient Boost to predict PCOS. They also proposed multi-stacking ML.To render model predictions understandable, interpretable, and accurate,they used Explainable AI techniques.The result revealed that multi-stacking ML delivered the best accuracy of 98 % [19]. S.Bharati et al. concentrated on data-driven diagnosis of PCOS in women.They used an open-source dataset of 541 women, of which 177 were diagnosed with PCOS. The authors used a univariate feature selection method in combination with LR, RF, Gradient Boosting and a hybrid RFLR model. For model training and testing, they partitioned the dataset using cross-validation and holdout methods. Based on the results,the RFLR model with UFS had the highest accuracy of 91.01 % [20]. A.Zigarelli et al. presented the creation of a predictive model for the self-diagnosis of PCOS using machine learning methods.They used a publicly available Kaggle PCOS dataset with information from 466 patients, taken from 10 different hospitals in Kerala, including 44 features. The authors used the CatBoost classifier and tested its efficiency with K-fold validation, achieving an 82.5 %
accuracy rate for invasive procedures and 90.1 % accuracy rate for non-invasive clinical indicators [21]. Y. A. Abu Adla et al. created a model that uses machine learning methods to automate the diagnosis of PCOS. They used a dataset with information from 466 patients and 39 features, which were ranked according to their importance.In order to decrease the number of features, the authors proposed a hybrid feature selection strategy both with filters and wrappers. Apart from that,they employed some machine learning models with the identified features to predict PCOS. SVM proved to have maximum accuracy of 91.6 % [22].M. M. Hassan et al. proposed a model for the diagnosis of PCOS patients.Researchers employed different machine learning techniques for the identification of PCOS patients, namely logistic regression, random forest, SVM,CART, and naïve bayes classification. After examining the results, it was discovered that the random forest algorithm worked extremely well, attaining 96 % accuracy for PCOS diagnosis in the tested set [23]. A. Denny et al. discussed an approach for early identification and prediction of PCOS by utilizing minimal but promising clinical and metabolic features, which act as early indicators for the disease. Principal Component Analysis (PCA) was used by authors to reduce the number of features.

Table 1:

Comparison table of existing works.

Reference	Dataset	Method	Accuracy
Silva et al. [18]	73 Healthy Women and 72 PCOS Patients	XGBoost	86.00%
V.V. Khanna et al. [19]	Dataset of 541 Patients from Kerala, India.	Multi-Stacking ML	98.00%
S. Bharatiet al.[20]	Dataset of 541 Patients.	RF,LR	91.01%

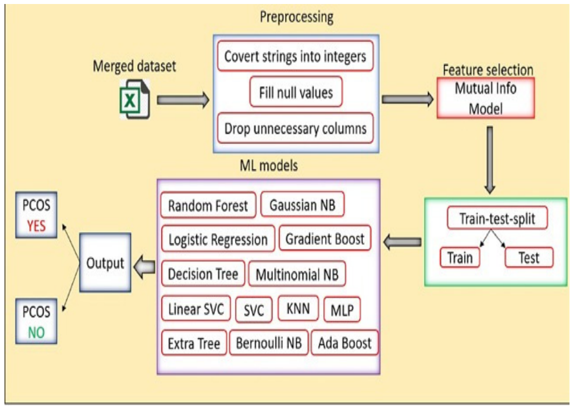


Fig. 1. Work flow of proposed model.

III. PROPOSED METHODOLOGY

The dataset obtained from Kaggle was curated correctly, performing essential preprocessing operations. These operations involved filling null values, deleting duplicate and deleting columns, and systematically converting all data from object type to numerical format. These operations provided a strong base for subsequent analyses. The dataset is now prepared for the implementation of different machine learning models. Prior to entering this phase, the dataset is divided into separate training and testing subsets. We added as part of our methodology a note-worthy feature selection method called Mutual Informationmethod,renowned for its effectiveness,permitted us to identify the most significant features. We particularly selected
the first 12 features, emphasizing their score. We aim to assess each model based on critical performance metrics such as accuracy, precision, F1 score, and roc score.This comprehensive comparative study will provide us with the model which shows optimal performance. The end result of our research will influence our decision-making process to choose the model which shows better performance. That selected model will be nicely integrated into our user interface so that it does not interfere with a smooth and effective user experience.Working process of our suggested method is illustrated in Fig. 1.

Table 2:

Description of the first dataset with hyperparameter values.

Column	Description	Mutual Information Score
Sl. No	Serial no of dataset	
Patienr File No	Patient Number	
PCOS(Y/N)	PCOS Status	
Age (Yrs)	Age of the patient	0.023806
Weight (Kg)	Weight in Kilograms	0.030802
Height (Cm)	Height in Centimetre	0.033325
BMI	BMI of patients	0.018678
Blood Group	Patients Blood Group	0.000000
Cycle Length(Days)	Number of days period lasts	0.069582
Marriage status	How long the patient have been married.	0.000000
Pregnant(Y/N)	Pregnant status	0.002934
No of abortions	Number of abortion	0.002345

FSH(mIU/mL)	FSH stands for Follicle-Stimulating Hormone quantity.	0.097867
LH(mIU/mL)	Luteinizing Hormone quantity	0.034567
Waist(Inch)	Hip size inches	0.008676
FSH/LH	Ration of FSH according quantity	0.009787
Hair Growth(Y/N)	hair Growth	0.115225
Hair loss(Y/N)	Hair loss	0.002739
Pimples(Y/N)	Pimples appearing	0.023504
Fast Food(Y/N)	Fast food Habit	0.578986
Weight gain(Y/N)	Weight gaining	0.082270

IV. PERFORMANCE MATRIX

A confusion matrix is an important measure in determining the effectiveness of a machine learning model when it is implemented on a given dataset for classification. It plays an important role in determining the performance of the model by presenting its accuracy in labelling categorical tags for input instances. In this research, when evaluating metrics such as Precision, Recall,False Positive Rate, True Negative Rate, and F1-Score, the confusion matrix is used. The matrix displays a detailed analysis of four important measures: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).These measures give a clear indication of the model's ability to correctly classify and identify positive and negative results, which helps in
Evaluating improving the model's classification accuracy
F1 Score=2 Sensitivity (Precision Sensitivity / Precision)

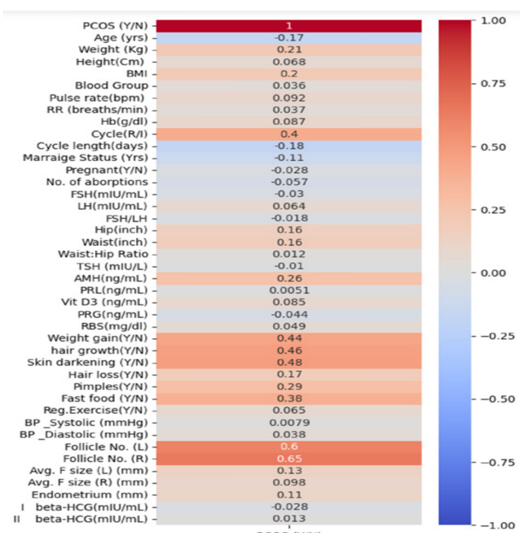


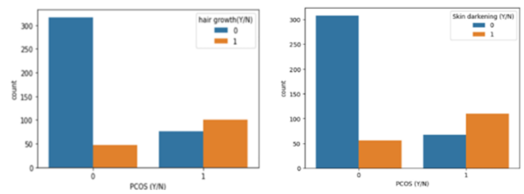
Fig.2. The heat map color density of used data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

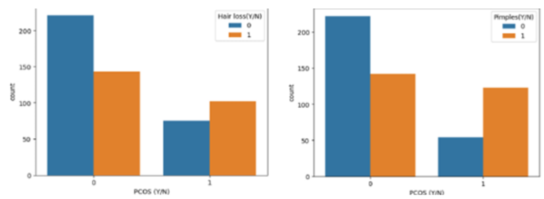
V. DATASET

This data is taken from the Kaggle, with 541 patient records and 43 distinct columns in both datasets together. There are 364 PCOS-negative cases and 177 PCOS-positive cases. The 'PCOS (Y/N)' column is the outcome, where 0 is a negative result and 1 is the patient having PCOS. In the whole dataset, 0 is a 'negative' or 'no' and 1 is 'positive' or 'yes'. Columns includeweight, height, number of abortions, blood group, pulse rate, cyclelength (days), pregnancy status (y/n), hair loss (y/n), pimples (y/n), etc.

Proposed research operated using two tables holding records of the samepatients. Presents another feature that has been included to the main tablein the preprocessing stage.



a) Hair Growth b) Skin Whitening



c) Hair loss d) Pimples

Fig.3.Different Symptoms of PCOS patients

A. *Data Preprocessing:*

In order to feed the dataset into various machine learning models,preprocessing of data must be done in order to obtain the best results. Preprocessing techniques involve null value filling, elimination of duplicate values, unnecessary columns removal, and encoding methods application, etc. The second dataset has 6 columns, whereas the first dataset consists of 43 columns. These datasets hold test results of different tests performed on the same 541 patients. There are four features common to both datasets, and hence duplicates were eliminated upon merging. Sincea consequence, the dataset had 466 patients and 16columns. A column with no name and columns 'Sl. No.' and 'Patient FileNo.' were dropped since they had no effect on the output. The last dataframe had 16 columns and 466 rows. Column 'AMH (ng/ mL)' had strings, therefore it was converted to numeric using the'tonumeric (.)' function from the panda’s library.

Moreover, the columns 'Marriage Status (yrs.)', 'Fast Food (Y/N)', and 'AMH (ng/mL)'had null values, which were replaced by each column's median value. Prior to fitting the machine learning models, the standardscaler was applied to the dataset except for the target column. Thestandard scaler scales the values into a numeric number range

B. *Data Splitting:*

Prior to performing the split, we combined two CSV files with the use of the Pandas library and ended up with a new data frame with two unique columns. We initially pulled out the "PCOS (Y/N)" target column and saved it as y, creating a data frame. The rest of the columns were put in x, resulting in two datasets: x for independent columns and y for the target or dependent column. We then used the train_test_split function of the model selection module of the scikit-learn library to split x and y into test and training sets. The data has been partitioned in 80:20 ratios, 80 % has been assigned for training and 20 % has been assigned for testing. The whole dataset is finally split for training and testing into four subsets: x_test, x_train, y_train, and y_test.

C. *HyperParameter Tuning:*

Hyperparameters are important parameters that are generally used prior to the usage of classification models. Unlike common hyper parameter tuning strategies like GridSearchCV, Mutual Information(Mutual info) is the best-suited strategy in the hyperparameter world.Mutual Information uses information-theoreticto evaluatethe relationship between hyperparameters and the model's performancemetrics. Mutual Information helps to identify the most critical hyperparameters and their optimal values, resulting in better model performance.

In hyper parameter tuning, Mutual Information provides a data-oriented approach to enable a more adaptive and informative adjustment of hyperparameter values for enhancing overall model effectiveness. Depending on the dataset, a value of 'k' is determined in order to identify the top features out of all values. In Mutual Information, features are chosen on the basis of input and output pairs. In this feature selection measure, the best features of the data are chosen'mutual_info_regression' and 'mutual_info_classif' functions. In theproposed study, the best 12 parameters are chosen out of 45 featureson the basis of this model's obtained value.

D. Data Visualization:

There are two prominent symptoms of PCOS: irregular menstrual cycle, with extended periods of time between cycles, and obesity or excess weight gain. Other symptoms may be determined through medical examinations performed in diagnostic centres. The data holds the outcomes of different diagnostic tests and surveys. A comparison bar chart in Fig. 2 depicts prominent symptoms of PCOS. Symptoms like hair growth,skin darkening, pimples, eating fast foods, exercising regularly, andweight gain are analysed. The chart presents the prevalence of symptoms in women with and without Postle bars reveal that women without PCOS tend to be less likely toexperience skin darkening, hair loss, and excessive hair growth, whilewomen with PCOS tend to frequently experience these symptoms. Furthermore, there is a significant correlation between PCOS and eatingfast foods, sinceas well as excess weight gain, with PCOS patients having a greaterprevalence of these factors than patients without PCOS.The heat map is of particular significance in detecting connections between the features. In this case, we can see colours from blueto red. The darker the red colour in a cell, the more intense the positiverelationship between the corresponding features. As a result, as onevalue goes up, the other also tends to go up. On the other hand, darkerblue colours show a more negative correlation between the pairedfeatures. Therefore, as one goes up, the other tends to go down. Cellsappearing in white have no visible correlation between them. Hence, cells near deep red or deep blue colours are our most significant features. Fig. 3 illustrates Follicle No. (L), Follicle No. (R),Weight gain, hair growth, and Skin darkening are the more responsible attributes for PCOS.

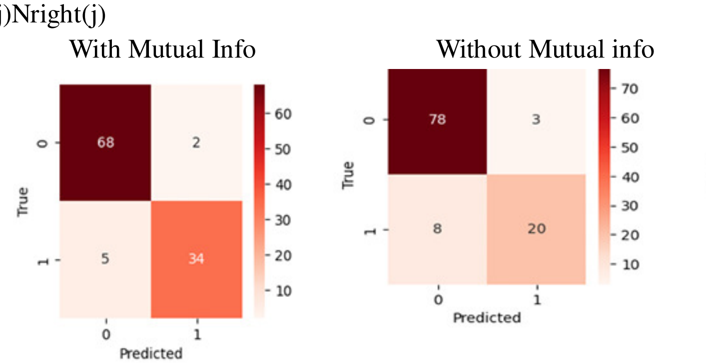
E. Machine Learning Models:

Machine learning algorithms enable forecasting systems to capitalize on patterns in data to provide precise predictions in any field,as well as improve decision-making and effectiveness [28,29]. The performance is measured based on 13 classifiers, namely Logistic Regression, Decision Tree, Ada Boost, XGBoost, Bernoulli NB, Support Vector Machines, Random Forest, K-Nearest Neighbor, Naïve Bayes, Gradient Boosting, and Multilayer perceptron, all explicated in this section. Next, the best-performing model, indicating the highest accuracy among the above classifiers, is evaluated. In this project we have used XGBoost and Random Forest.

1) Random Forest:

Random Forest consists of an array of decision trees. Random Forest chooses the optimal subset of attributes by ranking based on information gain. Through majority voting on sets of various decision trees, an instance gets classified. The final prediction comes as a function of aggregating the mean prediction in regression or mode of the classes for class membership predictions extracted from all the trees . That is to say that we can even find the node probabilityby dividing the number of total samples gathered by the number of totalsamples arriving at the node. The higher the value, the greater the importanceof the attribute. In Equation the impurity is denoted by the Pij.

Pij=LjNj Lleft(j)Nleft(j) Lright(j)Nright(j)



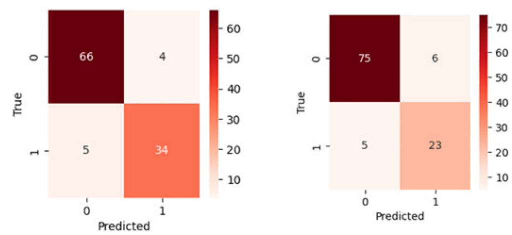
a) Random Forest

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

b) XGBoost(or) Gradient Boost:

Gradient Boosting is a general machine learning model that enhances the accuracy of predictions by combining the strengths of weak models. Primarily, Gradient Boosting is applied to decision tree classifiers [35]. Gradient Boosting is a strong boosting algorithm that can be used for both classification and regression tasks. It learns from its mistakes at each step. Using this boosting algorithm, we can get the maximum level of accuracy. It can perform both unstructured and structured data . It can be computationally expensiveand could demand a vast amount of memory and may be prone to problemslike overfitting, complexity, balanced data, and interpretability.

With Mutual Info Without Mutual info



b)XGBoost or GradientBoost

VI. RESULT ANALYSIS AND WEB INTERFACE

In this section, we will describe the results, discussions, and web interface of the proposed method. Here, we show the classificationresults for different machine learning algorithms, such as Logistic Regression, Decision Tree, AdaBoost, XGBoost, Bernoulli Naive Bayes, SupportVector Machines, K-Nearest Neighbor, Naive Bayes, Random Forest,XgBoost or Gradient Boosting, and Multilayer Perceptron. Each model was tested
individually, using the confusion matrix.

A. Result and discussion

The Mutual Information was used on the dataset, focusing on the output column PCOS (Y/N). The 12 highest ranked features were chosen,and all the ML models were used on these features. The results, withand without the Mutual Information model, for 13 machine learning algorithms using the chosen 12 features from the dataset.

With the Mutual Information model, the majority of the classifiersdelivered improving results depending on true positive values. RandomForest, Support Vector Classifier, Multinomial Naive Bayes, and Bernoulli Naive Bayes had improved performance with Mutual Information, while Linear SVC's performance with Mutual Informationwas poor. The other models performed average improvement relativeto the ones without Mutual Information. The true negative valueof PCOS significantly reduced for SVC classifiers, whereas the others exhibited an average advancement based on the Mutual Information model.

Table 3:
Performance measurement of the PCOS dataset without using Mutual Informa tion model.

Model	Accuracy	Precision	Recall	F-1 score	ROC curve
GB	92 %	87.10	84.38	0.86	89.59
RF	94 %	96.30	81.25	0.88	89.98

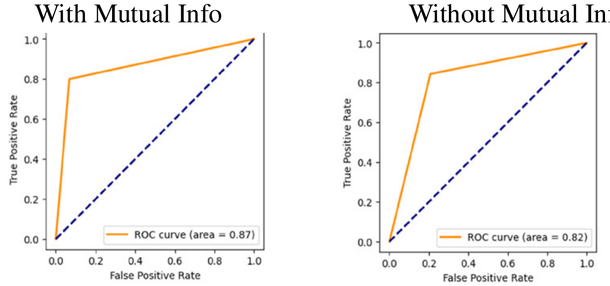
Table 4:
Performance measurement of the PCOS dataset with using Mutual Information model.

Model	Accuracy	Precision	Recall	F-1 score	ROC curve
GB	93%	93.55	82.86	0.88	90.08
RF	94 %	96.67	82.86	0.89	90.75

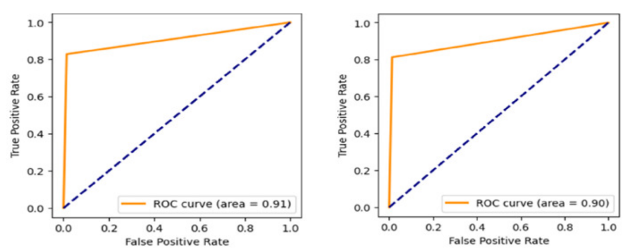
Mutual Information model is unique. On the accuracy front, Random Forest, Gradient Boosting, Extra Trees, and AdaBoost havesuccess rates of over 90 % without the Mutual Information featureengineering process. On the other hand, Logistic Regression, Random Forest,Multi-layer Perceptron, Linear SVC, Gradient Boosting, Extra Trees, andAdaBoost have the best success rate, over 90 %, when using the MutualInformation model. According to ROC scores, RandomForest and AdaBoost classifiers provide superior results. performance matrix measures like accuracy, precision, recall, F1 score, and ROC score with or without the employment of the Mutual Information model.

When employing Mutual Information and utilizing the selected features, the accuracy obtained from the models is significantly better compared to not using Mutual Information. Through Mutual Information, the highest accuracy of 94 % was achieved by two models, namely Random Forest and AdaBoost, XGBoost. Moreover, seven out of thirteen models achieved an accuracy of 90 % or higher. The lowest accuracy, at 77 %, was observed in the Bernoulli Naïve Bayes model, making it the only model that scored below 80 %. Conversely, without Mutual Information, the highest accuracy of 94 % was attained solely by the Random Forest model, and only four models achieved an accuracy of 90 % or above. The lowest accuracy recorded was 34 % in the
Multinomial Naïve Bayes model, with three models scoring below 73 %, as demonstrated.

With Mutual Info　　　Without Mutual Info



a) XG Boost



b) Random Forest

B. Web Interface:

The suggested approach resulted in the creation of a web-based systemfor Polycystic Ovary Syndrome detection. The user interface of the PCOScategorization system is built using the Django framework. First, users have to register, and then they have to log in tosuccessfully use the system. Fig. 8 shows the working steps of thesuggested web-based system. To improve user security andvalidity, the system designed registration and login forms illustrates the uploading process of a test image for PCOS identification using the system. The user feeds an image, which is then matched with a set of trainable images. After the end of the PCOS identification, the system makes predictions on the types of fruits in the given input. The predicted result is presented.

VII. CONCLUSION

PCOS patients can endure infertility, unable to give birth to children if the disease is left undiagnosed in the early stages. With limitations in early diagnosis, the incidence rate of a PCOS patient increase is greater than past years' records. In response to this, suggested research has come up with a machine learning-based web interface prediction system. Early diagnosis may empower patients to make necessaryfollow prescribed salary steps recommended by their physician to maintain a healthier life. The aim of our research is to employ the use of machine learning models in understanding patterns of this condition. The models are trained using data to exhibit accuracy, specificity, sensitivity, precision, and overall performance employing different ML algorithms like Random Forest, Logistic Regresion, Decision Tree Classifier, AdaBoost Classifier, XGBoostClassifier,Support Vector Machines, among others.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

In order to obtain feature selection,we utilized the Mutual Information method, presenting a comparisonof the best accuracy obtained when Mutual Information was not utilized. Interestingly, Random Forest and XGBoost Classifier obtained the topaccuracy of 94 %.

Table 5:
Comparison analysis of recent works and proposed system.

Title	Number of data	Number of parameters used	Hyperparameter tuning	Method	Accuracy
Polycystic ovary syndrome: Clinical and laboratory variables related to new phenotypes using machine-learning models	145	14	Scaling,mergingBorutaShap,	RF	86.00 %
Diagnosis of polycystic ovary syndrome using machine learning algorithms .	541	10	Univariate feature selection algorithm	RF,LR	91.01
Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: observational study.	541	10	Principal component analysis	XGBoost	90.10 %
Automated detection of polycystic ovary syndrome using machine learning techniques 91.60 %	466	12	Hybrid feature selection approach Linear	RF	91.60 %

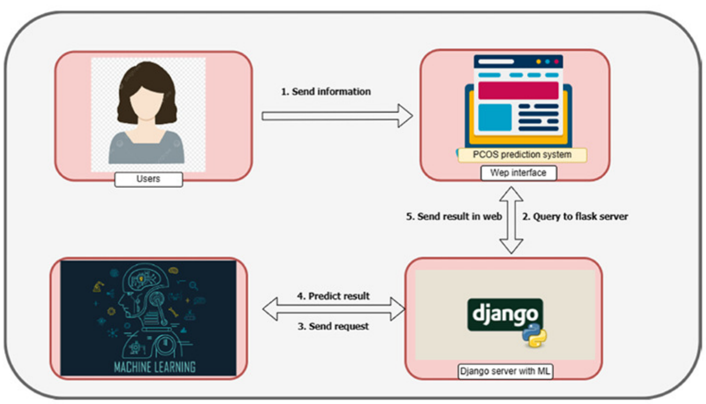


Fig.4.Working procedures of web-based interface.

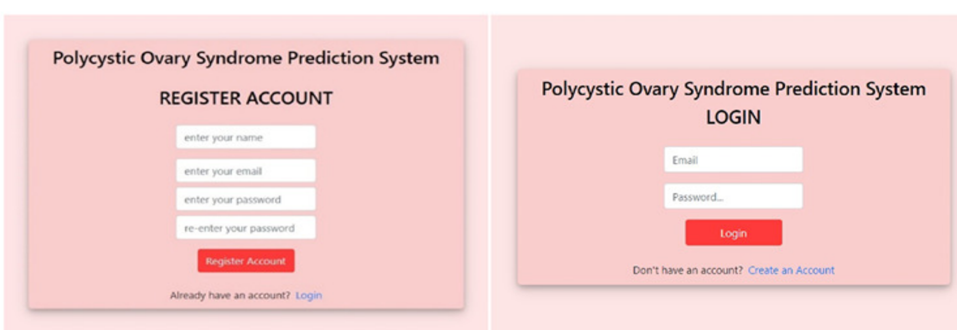
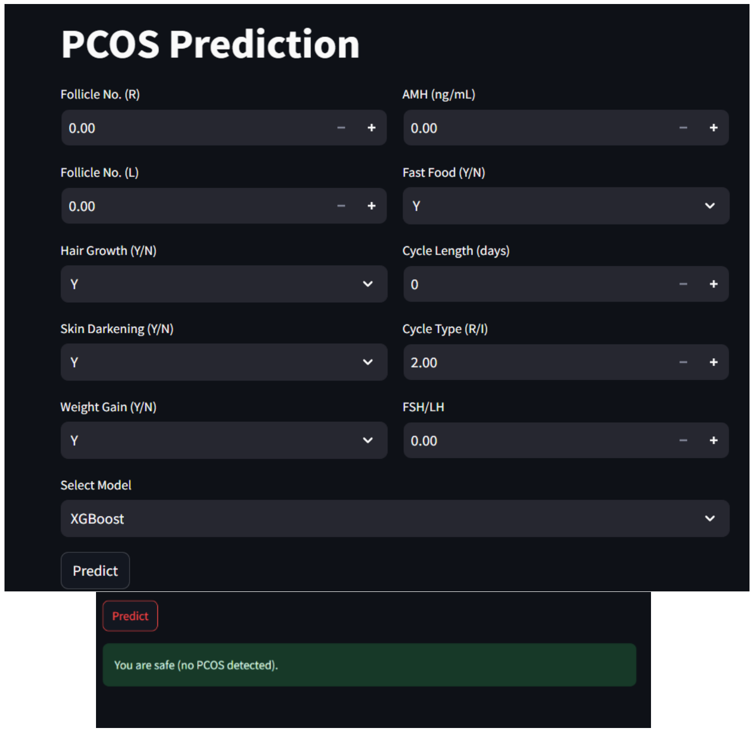


Fig.5.User registration and login form of proposed web interface.



Result of the working module

REFERENCES

[1] O˘ guz SH, et al. The prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome in treatment-naïve transgender people assigned female at birth. Endocrine 2023:1–6.

[2] Kulkarni S, et al. Polycystic ovarian syndrome: Current scenario and future insights. Drug Discov Today 2023:103821.

[3] Zhao H, et al. Insulin resistance in polycystic ovary syndrome across various tissues: an updated review of pathogenesis, evaluation, and treatment. J Ovarian Res 2023;16(1):9.

[4] Ma L, et al. The life cycle of the ovary. In: Ovarian aging. Springer; 2023. p. 7–33.

[5] Hajam YA, et al. Herbal medicine applications for polycystic ovarian syndrome. CRC Press; 2023.

[6] Wu T, et al. The cellular and molecular mechanisms of ovarian aging. In: Ovarian aging. Springer; 2023. p. 119–69.

[7] Adashi EY, et al. The polycystic ovary syndrome: the first 150 years of study. F&S Reports 2023;4(1):2–18.

[8] Miles KS. Our pearls matter: PCOS through the lens of women of color and white women. 2023.

[9] Bhat SA. Detection of polycystic ovary syndrome using machine learning algorithms. Dublin: National College of Ireland; 2021.

[10] Benjamin JJ, et al. Stress and polycystic ovarian syndrome-a case control study among Indian women. Clinical Epidemiology and Global Health 2023;22:101326.

[11] Karkera S, Agard E, Sankova L. The clinical manifestations of polycystic ovary syndrome (PCOS) and the treatment options. European Journal of Biology and Medical Science Research 2023;11(1):57–91.

[12] Shankar DY, et al. Overview of polycystic ovary syndrome (PCOS). World Journal of Advanced Engineering Technology and Sciences 2023;8(2):11–22.

[13] Dadoush SFM. Diagnosing and treating the causes of women's polycystic ovary syndrome: clinical and prospective study. African Journal of Advanced Pure and Applied Sciences (AJAPAS) 2023:401–7.