FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 1, January, 2025, pp 323 - 332
DOI: https://doi.org/10.33003/fjs-2025-0901-3088

LEVERAGING MACHINE LEARNING MODELS FOR PREDICTING THE LIKELIHOOD OF POLYCYSTIC OVARIAN SYNDROME IN WOMEN OF REPRODUCTIVE AGE

[1]Oliha, F. O, [2]Martins, K. M. and *[1]Okhuoya, J. O.

[1]Computer Science Department, University of Benin, Nigeria.
[2]Computer Science Department, Pan Atlantic University, Lagos, Nigeria

*Corresponding authors' email: joseph.okhuoya@uniben.edu

ABSTRACT

Conventional diagnostic approaches for polycystic ovarian syndrome (PCOS – a condition characterized by heterogeneity and the absence of a singular diagnostic test) are often invasive, time-consuming, and rely on varying criteria, resulting in inconsistencies in diagnosis. This study addresses the pressing challenge of improving the diagnosis of PCOS by exploring machine learning applications to bridge gaps in its prediction and diagnosis, offering a potential pathway toward greater accuracy and efficiency. The Cross-Industry Standard Process for Data Mining methodology was adopted for implementation using a comprehensive dataset from a public library – Kaggle. Results identified XGBoost algorithm as the most effective predictive model for diagnosing and predicting PCOS, achieving an accuracy of 98.7%. The results of the study indicated that the XGBoost algorithm is reliable with strong accuracy and dependability in diagnosing PCOS, establishing the PCOS Predictor as a valuable tool in clinical environments. This study thus represents a significant step forward in transforming the diagnostic landscape of PCOS, combining technological advancements with clinical insights to enhance women's healthcare.

Keywords: Diagnosis, Machine Learning, Machine Learning Models, PCOS

INTRODUCTION

Polycystic Ovarian Syndrome is a widespread condition that affects millions of women around the globe caused primarily by high androgen levels. Indications by Taieb & Feryel (2024) have shown that this hormonal imbalance at such levels leads to being able to reproduce as well as related complications in metabolism. While the medical literature deems hyperandrogenism as a significant area of focus in regard to PCOS, PCOS research needs to be broadened to cover its different roots. In the opinions by Mohamed et al. (2023) and Thorat et al. (2024), PCOS is an inherently genetic condition where environmental influences and lifestyle factor in its development and prevalence. It is now well established that the condition affects women's fertility and their metabolism, but little attention has been paid to significant damage – when the side effects are left unattended to, it becomes a burden for those diagnosed with PCOS (Mohamed et al., 2023). As an inherently multifaced condition, PCOS encompasses hormonal aberrations, metabolic irregularities, and problems associated with reproduction in women of age as pointed out by Siddiqui et al. (2022). This disorder is estimated to have a widespread rating 5-10% of reproductive-age women diagnostically; treatment of the disorder is a challenge since the symptoms are wide-ranging such as irregular periods, hirsutism, skin lesions, or even being unable to conceive (Poojitha and Talla, 2024; Rani et al., 2024).

The prevalence of PCOS indeed varies significantly, ranging from 6% to 20%, depending on the population studied and the diagnostic criteria applied (Rani et al., 2024). This variability is largely attributed to the different diagnostic criteria used, such as the national institute of health – NIH 1990, Rotterdam 2003, and Androgen Excess – AE-PCOS 2006, as well as the characteristics of the study populations, including age and ethnicity (Lazareva, 2023; Hatoum et al., 2024; Neven et al., 2024). According to the study by Hatoum et al. (2024), the prevalence rates using the NIH 1990 criteria are generally lower, with studies reporting around 4.98% in adolescents. Corroborating this assertion, the study by Lazareva (2023), had confirmed the Rotterdam 2003 criteria as the most inclusive set, often resulting in higher prevalence rates, such as 8.80% in adolescents and up to 21% in broader populations. In addition, the "AE-PCOS 2006 criteria", which is similar to the "NIH", was reported to have yielded lower prevalence rates, around 4.74% in adolescents (Hatoum et al., 2024). With references to Neven et al. (2024), a global meta-analysis indicates a prevalence of 9.8% using the Rotterdam criteria and 6.3% with the "International Evidence-based Guideline criteria", which exclude polycystic ovarian morphology to prevent over-diagnosis.

Diagnosing PCOS can feel like a frustrating journey for many women and healthcare providers. The variability in PCOS prevalence highlights the complexity of its diagnosis and the influence of diagnostic criteria. The process of diagnosing presents significant diagnostic challenges due to its heterogeneous nature and the absence of a singular diagnostic test. While the Rotterdam criteria often result in higher prevalence rates, the exclusion of certain features like polycystic ovarian morphology in newer guidelines aims to reduce over-diagnosis (Unfer et al., 2024). This complexity necessitates a multifaceted approach to diagnosis and understanding of the condition. As a result, the need for standardized diagnostic approaches to better understand and manage PCOS across different populations is highlighted in the present study. More so, a study suggests incorporating additional parameters to enhance diagnostic accuracy (Unfer et al., 2024). Another study noted that new markers, such as the triglyceride glucose (TyG) index, show promise in identifying metabolic risks associated with PCOS (Keyif et al., 2023). In addition, technological innovations such as artificial intelligence (AI) and machine learning (ML) are being explored as tools to improve diagnostic precision and reduce errors (Fahs et al., 2023). The conceptual approach and application of artificial language are paramount to the present study in lieu of understanding PCOS predictions and its health-related diagnosis.

In the domain of AI and ML, numerous models have been used in predicting the outcomes with varying levels of success in the healthcare profession. For related efforts, Bharati et al.

(2020) used "gradient boosting (GB), random forest (RF), linear regression (LR), and an ensemble RFLR model that combined RF and LR using a univariate feature selection (UFS) algorithm" on a dataset of PCOS. The dataset was split by using holdout and cross-validation techniques for training and testing. The outcome of the result showed that the RFLR model with UFS combined performed better. Tiwari et al. (2020) applied a feature selection method based on correlation to choose a subset of the relevant features in the dataset. They tested various machine learning models, including support vector machine (SVM), LR, RF, decision tree (DT), k-nearest neighbors (KNN), Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), GB, AdaBoost (AB), XGBoost (XB), and CatBoost. The RF model is considered the best under the correlation thresholds. Thakre et al. (2021) employed various machine-learning models to identify the presence or absence of PCOS in females. They applied "chi-square feature selection" and retained the top 30 features, keeping in mind that RF was the best-performing model for its high accuracy rate. Khanna et al. (2023) investigated machine learning and deep learning techniques: LR, DT, RF, SVM, Naïve Bayes (NB), KNN, AdaBoost, XGBoost, ExtraTrees and proposed a "multi-stacking machine learning framework for PCOS prediction using Explainable AI (XAI) methodologies" to enhance the interpretability, reliability, and understanding of model predictions. The results showed that the multi-stacking machine learning approach outperformed other models in terms of performance, as also evidenced by Kumar et al. (2024).

Despite the advancements in understanding PCOS, some argue that the reliance on established criteria may overlook unique presentations of the syndrome, suggesting a need for personalized diagnostic approaches that consider individual patient profiles and symptoms (Rani et al., 2024). A blur in diagnosis can lead to a poor diagnostic prediction for patients who might be saddled with a winding road filled with more doctor visits, misdiagnoses, and delays in receiving the right treatment. These delays can have grave consequences down the line, increasing the risk of long-term health problems. Poor diagnosis is a major challenge motivating this research. In justifying this research motivation, the widely accepted Rotterdam criteria, while useful, do not encompass all manifestations of the disorder, leading to potential misdiagnosis or underdiagnosis according to Unfer et al. (2024). Thus, keying into personalized diagnostic approach as opined by Rani et al. (2024), solutions from across literature and bio-laboratory have highlighted new ways in addition to the Rotterdam criteria including artificial intelligence under the capabilities of machine learning models for improving the diagnostic predictions. Consequently, the present study focuses on integrating advanced technologies like machine learning to complement existing diagnostic criteria while emphasizing the importance of leveraging innovation to mitigate the limitations of traditional methods. It further proposes an interface deployed on a machine learning model for seamlessly diagnosing PCOS presence and absence in females of reproductive age; trains and evaluates the model efficiency and presents the findings.

For advances in knowledge, the study projects the potential to improve healthcare outcomes for women with PCOS, emphasizing the need for early and more accurate adjustable diagnostic framework that accommodates the diverse presentations of PCOS. Significantly, it contributes to the field by:

Enhancing diagnostic precision with machine learning models to improve the accuracy and reliability of PCOS diagnosis. This represents a shift toward data-driven, personalized diagnostic approaches that consider individual patient profiles and symptoms via the interface developed. This integration broadens the diagnostic landscape, potentially reducing underdiagnosis and misdiagnosis.

Proposing an interface with a machine learning model that demonstrates the feasibility of using AI to predict and diagnose PCOS, setting a precedent for similar applications in reproductive health and beyond. This contributes to the ongoing discourse on developing more comprehensive and inclusive diagnostic tools for PCOS.

Highlighting that the findings have direct implications for improving patient experiences, reducing diagnostic delays, and mitigating the long-term health risks associated with poor diagnostic predictions.

MATERIALS AND METHODS

This study employs supervised machine learning classification techniques to predict the presence of PCOS in individuals; this approach aligns with the Cross-Industry Standard Process for Data Mining – CRISP-DM methodology. CRISP-DM is a structured approach used in data science projects to guide them from inception to completion (Lathifah et al., 2023; Oliha and Omobude, 2023; Kumar et al., 2024). As depicted in Figure 1, the process encompasses investigating the problem domain, data understanding, data preparation, modelling, evaluation, and deployment. However, the present study adapts the methodology to inaugurate with data understanding phase. This structured approach ensures thorough data analysis and robust model development, leading to practical and actionable insights for stakeholders in the health sector.

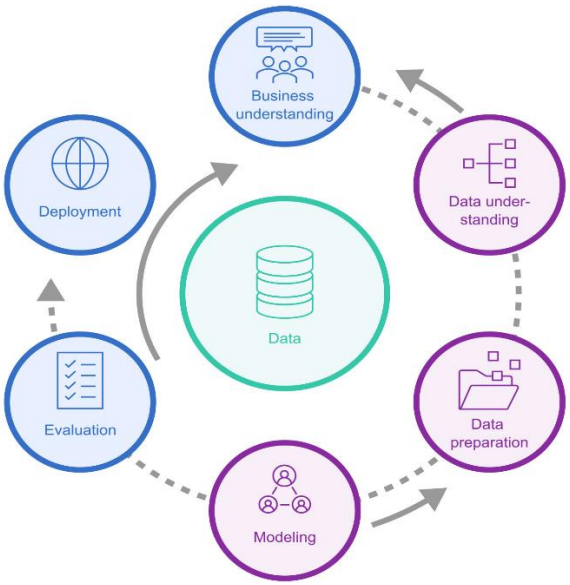


Figure 1: Crisp-DM Phases (Lathifah et al., 2023)

Data Understanding and Preparation.
The Kaggle dataset, which contained 2000 records was selected as the data source for the study. Each record represents an individual with PCOS or no PCOS with 44 distinct features. Among these, 11 are categorical indicating the presence or absence of features such as acne, skin darkening, weight gain, etc. The other 33 features are numerical detailing specifications such as body measurements and various hormone levels. This comprehensive dataset provides a robust foundation for developing and validating our PCOS prediction models. Table 1 below summarizes the features of the datasets.

Table 1: PCOS Dataset Summary

Attribute Name	Description	Type of Data
Sl. No	Serial number of the patient	numerical
Patient File No.	Unique file number for each patient	numerical
PCOS (Y/N)	Indicates whether the patient has PCOS or not	categorical
Age (yrs)	Age of the patient in years	numerical
Weight (Kg)	Patient's weight in kilograms	numerical
Height (Cm)	Patient's height in centimetres	numerical
BMI	Body Mass Index calculated from weight and height	numerical
Blood Group	Patient's blood group type	categorical
FSH(mIU/mL)	Follicle-Stimulating Hormone level	numerical
LH(mIU/mL)	Luteinizing Hormone level	numerical
TSH (mIU/L)	Thyroid-Stimulating Hormone level	numerical
AMH(ng/mL	Anti-Mullerian Hormone level	numerical
Weight gain(Y/N)	Indicates if the patient experienced weight gain	categorical
hair growth(Y/N)	Indicates if the patient has abnormal hair growth	categorical
Skin darkening (Y/N)	Indicates if the patient has skin darkening	categorical
Fast food (Y/N)	Indicates if the patient frequently consumes fast food.	categorical
Reg.Exercise(Y/N)	Indicates if the patient exercises regularly	categorical
BP _Systolic (mmHg)	Systolic blood pressure measurement	numerical
BP _Diastolic (mmHg)	Diastolic blood pressure measurement	numerical
Follicle No. (L)	Number of follicles in the left ovary	numerical
Follicle No. (R)	Number of follicles in the right ovary	numerical
Endometrium (mm)	Thickness of the endometrium in millimetres	numerical
Pulse rate(bpm)	Patient's heart rate measured in beats per minute	numerical
RR(breaths/min)	Respiratory rate in breaths per minute	numerical
Hb(g/dl)	Haemoglobin level in grams per deciliter	numerical
Cycle(R/I)	Type of menstrual cycle (Regular/Irregular)	categorical
Cycle length(days)	Duration of the menstrual cycle in days	numerical
Marriage Status(Yrs)	Duration of marriage in years	numerical
Pregnant(Y/N)	Indicates whether the patient is pregnant	categorical
No. Of Abortions	Number of abortions the patient has had	numerical
I beta-HCG(mIU/mL)	First beta-HCG hormone level	numerical
II beta-HCG(mIU/mL)	Second beta-HCG hormone level	numerical

FSH/LH	Ratio of FSH to LH levels	numerical
Hip(inch)	Hip measurement in inches	numerical

The study employed extensive exploratory data analysis (EDA) to gain insights into the dataset's features. Upon this, the categorical feature proportion was adopted and presented using pie charts to gain insights into the relative distribution of these features. The pie charts as depicted in Figure 2 provide a visual representation of the proportion of each category within the features with binary (Y/N) markings.

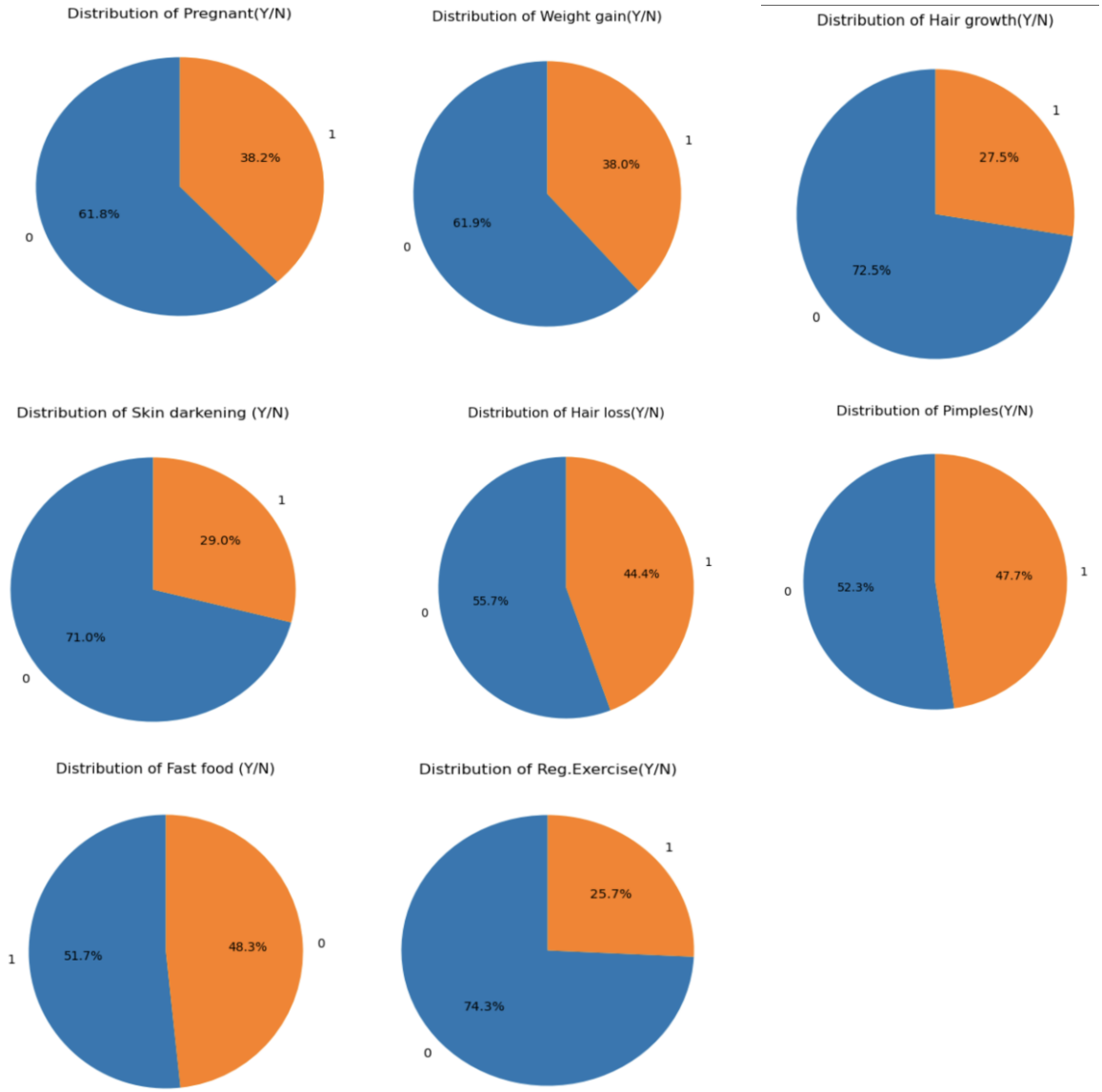
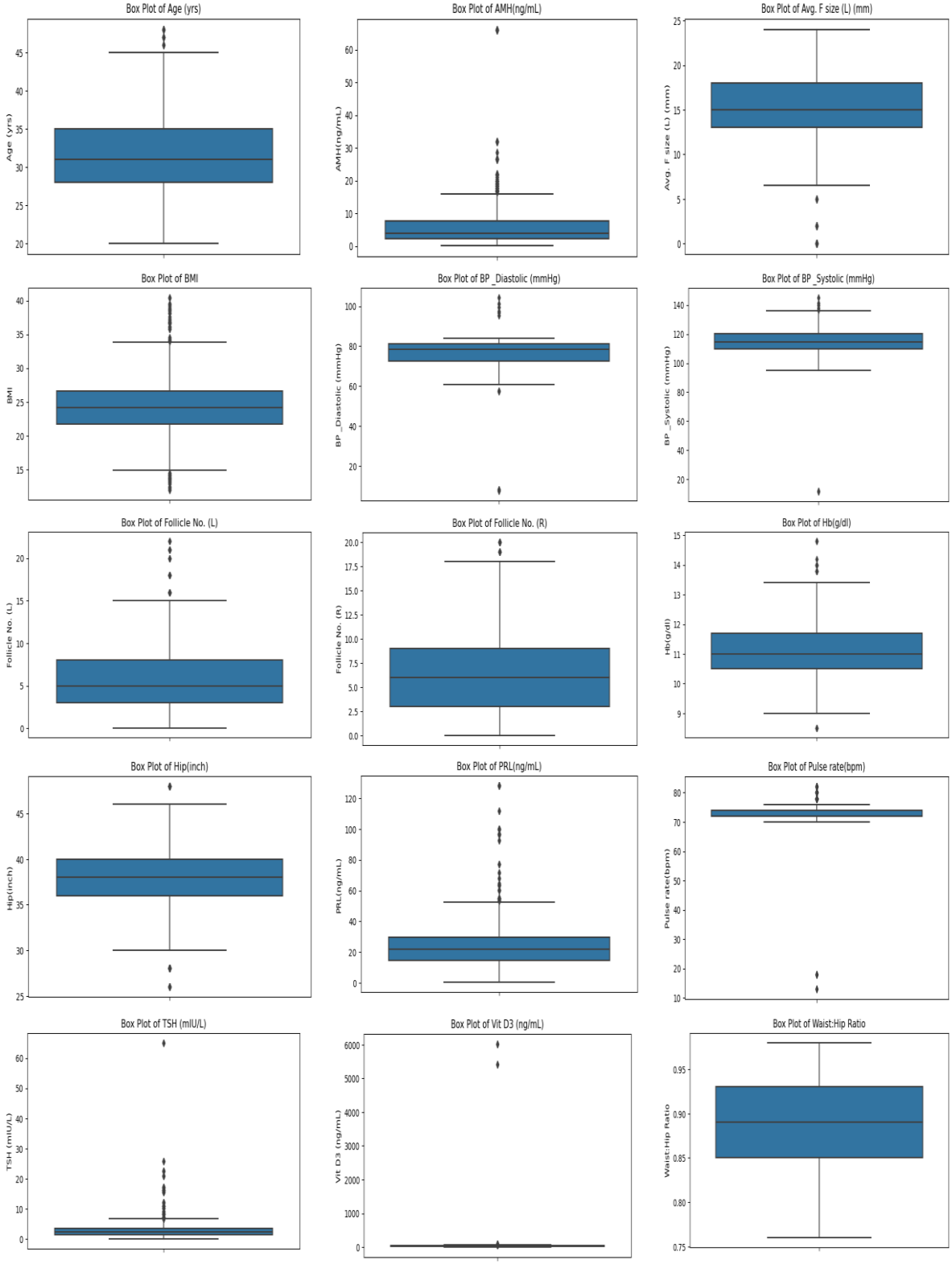


Figure 2: Pie chart of Categorical Features

Figure 2 explores the features with categorical distributions from Table 1 (with all Y/N), including "pregnancy, weight gain, hair growth, skin darkening, hair loss, pimples, fast food, and regular exercise". It is evident from Figure 2 that the features average over 51% positivity for each category, further validating the categorical veracity of the PCOS binary features. For numerical features, the boxplot was adopted for visualization. Figure 3 utilizes box plots to visualize the distribution, spread, and skewness of numerical features including hormone levels; body measurements such as height and weight, and measurements in the female reproductive organ such as endometrium thickness.


Box Plot of Age (yrs)
Age (yrs)
Box Plot of AMH(ng/mL)
AMH(ng/mL)
Box Plot of Avg. F size (L) (mm)
Avg. F size (L) (mm)
Box Plot of BMI
BMI
Box Plot of BP _Diastolic (mmHg)
BP _Diastolic (mmHg)
Box Plot of BP _Systolic (mmHg)
BP _Systolic (mmHg)
Box Plot of Follicle No. (L)
Follicle No. (L)
Box Plot of Follicle No. (R)
Follicle No. (R)
Box Plot of Hb(g/dl)
Hb(g/dl)
Box Plot of Hip(inch)
Hip(inch)
Box Plot of PRL(ng/mL)
PRL(ng/mL)
Box Plot of Pulse rate(bpm)
Pulse rate(bpm)
Box Plot of TSH (mIU/L)
TSH (mIU/L)
Box Plot of Vit D3 (ng/mL)
Vit D3 (ng/mL)
Box Plot of Waist:Hip Ratio
Waist:Hip Ratio

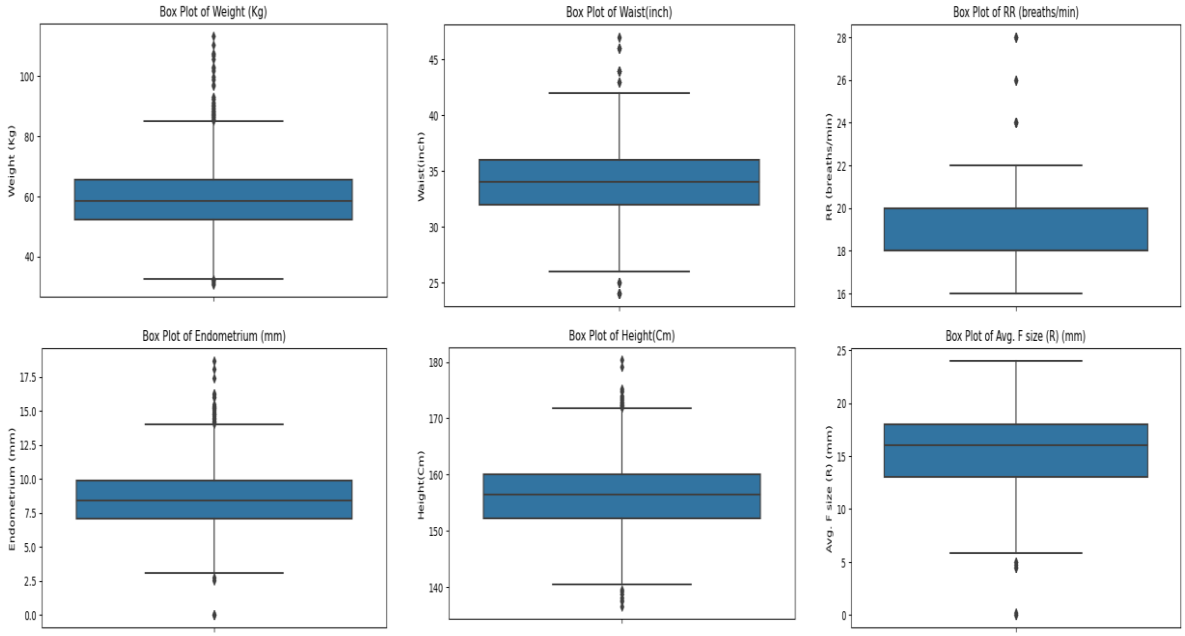


Figure 3: Boxplots of Numerical Features.

In the next phase of the CRISP-DM, the preparation of the PCOS dataset for predictive insights was next. This phase involved data cleaning – checking for and addressing missing values, duplicates, and incorrect data entries as well as dealing with outliers; feature selection – identifying and selecting relevant features that contribute significantly to PCOS diagnosis using SelectKBest and Chi-square test; data transformation – normalizing or standardizing the data and encoding categorical variables; splitting the dataset – dividing the data into training and testing sets to evaluate the model's performance. Data cleaning was achieved by addressing missing values and outliers. The transformation was achieved by normalizing, aggregating, and encoding variables and finally, features were selected categorically.

Model Evaluation

In this phase of the methodology, the task of selecting suitable ML algorithms for prediction models, training the selected models, and optimizing them for optimal performance. The following ML models were selected for evaluation based on their popularity and effectiveness in binary classification tasks: LR, RF, DT, SVM, KNN, NB, and XGB (Vinothini et al., 2024). Vinothini et al. (2024) discussed these ML models noting that the LR has a high utilization rate in medical diagnosis cases because it serves as a technique for estimating the likelihood of a binary outcome – producing binary values of true or false (yes or no). RF, serves as a classification and regression tool with a massive success rate, providing measures for changing relevance or large-scale challenges. DT is another ML technique with high significance regarding modelling circumstances of categorization method, providing classification criteria to human understanding. SVM adopts ML theory to enhance predictive accuracy while automatically avoiding overfitting to the dataset – creating a hyperliner that separates data into different classes. XGB is a classification technique that uses a gradient-boosting algorithm with regularization terms to avoid overfitting. It has become the de facto standard model in the Kaggle space used in healthcare, finance, and marketing (Tiwari et al., 2022). KNN is another ML technique commonly used in large-scale applications for anomaly detection. It is most effective when decision boundaries are irregular according to Vinothini et al. (2024). NB is a classification algorithm for classification tasks and works well with categorical data. It emphasizes feature independence and might be incompatible with imbalanced datasets. The study is drawn to these ML models according to the prepositions by Vinothini et al. (2024) on performance priority.

Due to the categorical approach – a binary classification task, categorizing individuals into groups of PCOS or no PCOS requires some degree of accuracy and precision. For the presence of PCOS prediction, the dataset was split into training and test subsets – a 70/30 split, to ensure a robust evaluation of the models. The training set was used to fit the models while the test set, which remained untouched during training, provided an unbiased estimate of the model's generalization performance. This approach ensured that our evaluation was not influenced by the test data during the model selection process. The three top-performing models were evaluated based on training time and qualitative factors before the final model was selected. Figure 4 depicts the training times for the models.

```
Random Forest Training Time: 0.23979973793029785 seconds
Decision Tree Training Time: 0.007878303527832031 seconds
XGBoost Training Time: 0.04027080535888672 seconds
```

Figure 4: Training times of the models

The models were evaluated to assess their performance using accuracy, precision, recall, and F1 score. Cross-validation was ensured for the generalization of data. After selecting the final models, hyperparameter tuning was performed using GridSearchCV to provide the best model parameters and metrics. An interface was developed to deploy the best performing model.

PCOS Predictor Interface and Deployment

The predictor interfaces are depicted in Figure 5 and Figure 6. The PCOS predictor is a user-friendly web-based interface to make the prediction model accessible and practical for health care providers. This dashboard allows healthcare providers to input key medical information to obtain PCOS predictions and the probability of the presence of disease in case patients want a second opinion. Its intuitive design ensures ease of use, even for those without technical expertise, thereby bridging the gap between sophisticated machine learning models and practical healthcare applications.

Figure 5: The deployed PCOS Predictor

Figure 6: PCOS Predictor Results Page

RESULTS AND DISCUSSION

The models evaluated for PCOS diagnosis demonstrated some degree of prediction using the metrics of accuracy, precision, recall, and F1 score as depicted in Figure 7.

	accuracy	precision	recall	f1_score
Logistic Regression	0.891122	0.887006	0.777228	0.828496
Random Forest	0.986600	0.994898	0.965347	0.979899
Decision Tree	0.978224	0.965517	0.970297	0.967901
SVM	0.944724	0.977401	0.856436	0.912929
KNN	0.904523	0.934132	0.772277	0.845528
Naive Bayes	0.867672	0.778281	0.851485	0.813239
XGBoost	0.984925	0.994872	0.960396	0.977330
AdaBoost	0.948074	0.957219	0.886139	0.920308

Figure 7: Model Evaluation Results

The logistic regression provides relatively modest performance compared to other algorithms. While its precision (88.7%) indicates it has a lower rate of false positives, the recall (77.7%) suggests it struggles to detect all true cases of PCOS. This algorithm might miss some cases, making it less ideal for clinical use where high sensitivity is crucial. Random Forest achieves the highest performance across all metrics, with near-perfect precision (99.48%) and recall (96.54%). This indicates that it can accurately identify PCOS cases and minimize both false positives and negatives, making it a strong candidate for PCOS diagnosis. Decision Tree performs well, with high recall (97.03%) and precision (96.55%). While slightly less effective than Random Forest, it remains a reliable choice, particularly for interpretable decision-making in a clinical setting. SVM demonstrates good performance, particularly with high precision (97.74%). However, its recall (85.64%) is relatively lower, suggesting it might miss some PCOS cases. SVM could be suitable when minimizing false positives is prioritized. KNN's performance is moderate, with high precision (93.41%) but lower recall (77.22%). Similar to Logistic Regression, it may fail to identify all cases, which could limit its utility in highly sensitive diagnostic tasks. Naive Bayes shows the lowest performance among the algorithms, with particularly low precision (77.83%). While recall (85.15%) is decent, the overall lower metrics suggest it may not be ideal for PCOS prediction. XGBoost performs almost on par with Random Forest, showcasing excellent metrics across the board. Its precision (99.49%) and recall (96.04%) highlight its reliability and robustness for PCOS prediction. AdaBoost achieves a balance between precision (95.72%) and recall (88.61%), making it a competitive choice. However, it falls short of Random Forest and XGBoost in overall performance. The comparison of their performance is illustrated in Figure 8 as well.

Analyzing the bar charts in Figure 8, the performance metrics for LR are visibly lower compared to other models. The Recall parameter was the lowest among the four metrics, indicating that Logistic Regression struggles to capture all true cases of PCOS, which is crucial for medical diagnostics. RF consistently achieves near-perfect scores across all metrics (accuracy, precision, recall, and F1 score). The height of the bars for Random Forest is nearly at the maximum (close to 1), reflecting its robustness and reliability for PCOS prediction. DT Performs slightly worse than Random Forest but still achieves high values across all metrics – its recall and precision are well-balanced, making it a reliable yet interpretable model. SVM precision is very high, but recall is noticeably lower than Random Forest and Decision Tree. This makes SVM suitable in scenarios where minimizing false positives is more important than capturing all cases of PCOS. KNN shows moderate performance with good precision but lower recall, similar to Logistic Regression. While its simplicity is appealing, the lower recall makes it less reliable for medical diagnostics.

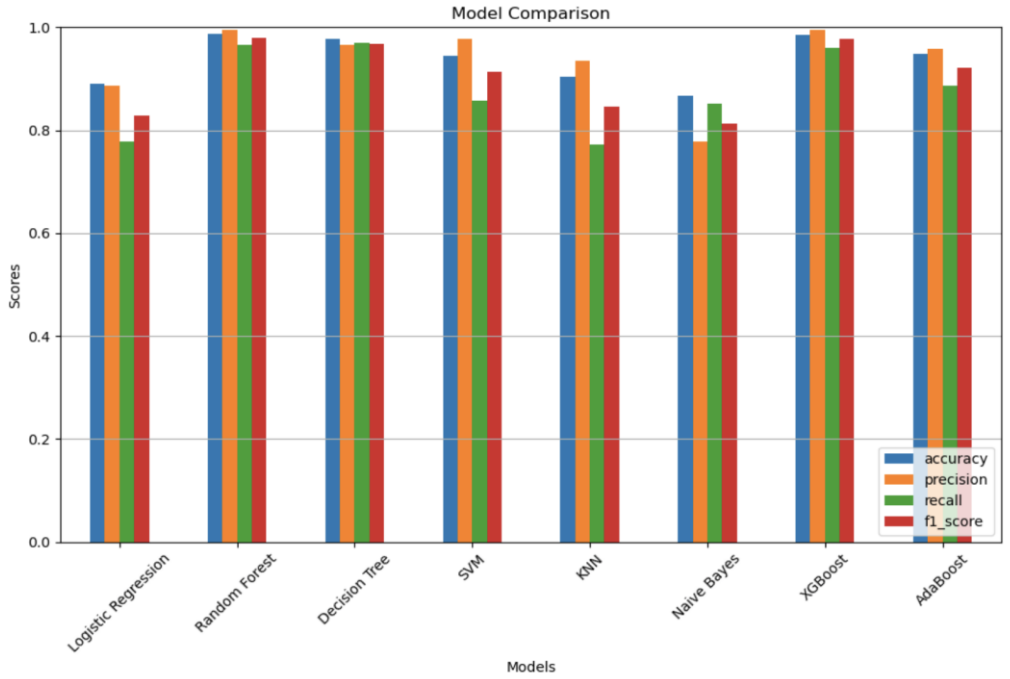


Figure 8: Model Comparison Metrics

For NB, all four metrics are significantly lower compared to other models, especially precision and accuracy. Naive Bayes is not ideal for this application as it struggles to handle complex relationships between features in PCOS datasets. XGB performance is nearly identical to Random Forest, with precision and recall at very high levels. XGBoost's consistency across metrics and its efficiency in handling imbalanced datasets make it a strong choice. AdaBoost shows balanced performance with high precision and recall but falls short of Random Forest and XGBoost. Its metrics are still above average, making it a competitive option, especially when computational efficiency is a concern.

Both Random Forest and XGBoost outperform other models consistently. Their ability to capture complex patterns in data is evident, making them ideal for PCOS prediction. These models are less prone to overfitting, especially when compared to standalone models like Decision Tree or Logistic Regression. Models like SVM prioritize precision but sacrifice recall, which might not be ideal for PCOS diagnosis where missing true cases (low recall) can lead to delayed treatment. Random Forest and XGBoost achieve a better balance, making them preferable for sensitive medical applications. While Logistic Regression and Naive Bayes are simpler to implement and interpret, their lower metrics make them less suited for this task. Models like Decision Tree offer a middle ground, combining interpretability with decent performance. For clinical applications, Random Forest and XGBoost stand out due to their accuracy and robustness. If interpretability is a key requirement, Decision Tree or explainable versions of ensemble models can be considered.

From the evaluation, it was determined that XGBoost was the ideal model for the classification problem. Hyperparameter tuning was then performed on the model using GridSearchCV and the best model was saved to be used in the interface as shown in Figure 9.

```
Best parameters:
learning_rate: 0.2
max_depth: 3
n_estimators: 100

Best Model Metrics:
accuracy: 0.9865996649916248
precision: 0.9948979591836735
recall: 0.9653465346534653
f1_score: 0.9798994974874372
```

Figure 9: Best Parameters and Model Metrics of the XGBoost Model

Figure 9 displays the best parameters and the best metrics achieved by the XGBoost model for PCOS prediction. Dilating the hyperparameters, the learning frequency with a value of 0.2 indicates a moderately fast convergence rate, balancing between speed and accuracy while avoiding overshooting the optimal solution. A relatively shallow depth (3) for the trees suggests the model prioritizes generalization over capturing very complex patterns, particularly to prevent overfitting. With 100 estimators, the model achieves a balance between complexity and computational efficiency, ensuring robust learning while keeping resource usage reasonable. The accuracy indicates that approximately 98.66% of all predictions (both positive and negative cases) were correct. High accuracy is expected since a precision of 99.48%

suggests that the model is highly effective at minimizing false positives (i.e., incorrectly diagnosing someone without PCOS as having it). This is critical in reducing unnecessary anxiety or medical interventions for individuals. A recall of 96.53% indicates that the model successfully identifies the majority of true PCOS cases. An implication that is essential in medical diagnostics is to ensure that most affected individuals are correctly diagnosed and receive appropriate treatment. Reiterating that the F1-score is the harmonic mean of precision and recall, providing a balanced view of the model's performance. Thus, at 97.99%, the high F1-score confirms that the model effectively balances precision and recall, making it reliable for both detection and reducing false positives. From Figure 9, the results and findings support the study's primary choice for PCOS prediction using XGBoost as the ML model. It is highly reliable and can handle the complex, non-linear (categorical) features associated with PCOS datasets.

CONCLUSION

The present study has projected XGBoost as a more suitable model for PCOS prediction and hence developed the PCOS predictive system – "PCOS Predictor" and deployed it with ease for research and clinical purposes. Thus, the deployment of the PCOS predictor on the XGBoost model revealed that the optimal parameters, particularly the shallow max depth, indicate the model is tuned to avoid overfitting while still capturing the key patterns in the data – enhancing its robustness across diverse datasets and populations. Also, the high precision minimizes false positives, avoiding unnecessary treatments or stress for patients misdiagnosed with PCOS – reducing the risk of missed diagnoses. This outcome extends the proposition pointed out by Thakre et al. (2021) and Khanna et al. (2023).

For research or exploratory use, the study results also showed that Decision Tree had the fastest training time and Random Forest had the slowest training time. XGBoost was not too far off from the Decision Tree but had higher accuracy, precision, and recall than the Decision Tree. Consequently, the development of a PCOS predictor that deploys the XGBoost model for prediction establishes a significant contribution to the field as it can aid healthcare professionals and individuals in the early detection and management of PCOS. Thus, the use of the XGBoost algorithm has shown high accuracy and reliability, making the PCOS Predictor a valuable resource in clinical settings. This amounted to another reason why XGBoost is a more suitable model for PCOS prediction, also corroborating the findings by Tiwari et al. (2020) on the impact of categorized features in predictive models and the proposition by Rani et al. (2024) on personalized diagnostics and hyper tuning in line with established criteria. The XGBoost proved to be an excellent choice due to its transparency and relatively high performance. With metrics like these, this XGBoost model is highly deployable in clinical settings.

The results and findings from this study align with the impact of using ML for the prediction of PCOS as pointed out by Khanna et al. (2023). The findings also corroborate the idea of Kumar et al. (2024) on the proposition of building an interface for both clinical and experimental use. The study findings also raised concerns that could be possible limitations for the models on achieving optimal performance. For example, the model only predicts using text data of ultrasound images due to the nature of the dataset needed to test. This means that for prediction to take place, healthcare professionals need to analyze the ultrasound images to find features used in prediction such as endometrium thickness, number of follicles on the left and right ovaries, average follicle size, endometrium thickness – there was no use of advanced imaging results. Also, the current dataset may not encompass the full diversity of the population, which could affect the model's generalizability. This means that there could still be more combinations of features not included in the dataset that led to a PCOS diagnosis.

It is believed that the PCOS Predictor System can be further refined to provide even more reliable and accurate diagnostic support for PCOS. To improve the work, some suggestions were recommended, for example, the RF and XGB should be prioritized for their robust and consistent performance: that is, further improvements can be achieved by optimizing feature selection and ensuring balanced datasets, as PCOS datasets might have imbalances that skew results. The pairing of models like XGB with other tools for interpretability to make predictions more transparent for clinical adoption is also suggested. Future studies should aim to incorporate a larger and more diverse dataset to improve the model's robustness and generalizability. This should include data from different populations to ensure the model can accurately predict PCOS across diverse groups. It is hoped that by addressing these recommendations, the PCOS Predictor System can be further refined to provide even more reliable and accurate diagnostic support for PCOS, improving patient outcomes.

REFERENCES

Bharati, S., Podder, P., and Mondal, M.R.H. (2020). Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms. In the Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP). Dhaka, Bangladesh, June 1486–1489.

Fahs, D., Salloum, D., Nasrallah, M., & Ghazeeri, G. (2023). Polycystic Ovary Syndrome: Pathophysiology and Controversies in Diagnosis. *Diagnostics*, 13(9), 1559. https://doi.org/10.3390/diagnostics13091559

Hatoum, S., Amiri, M., Buyalos, R. P., Sheidaei, A. & Azziz, R. (2024). Prevalence of Polycystic Ovary Syndrome in Adolescent Girls by World Health Organization Region. *Journal of the Endocrine Society*, 8(1), 867-894. https://doi.org/10.1210/jendso/bvae163.1655

Keyif., B., Yavuzcan, A., Kurdoglu, M. (2023). Diagnosing Polycystic Ovary Syndrome Using Triglyceride-Related Indices: Is it Possible Without Rotterdam Criteria? *Crescent Journal of Medical and Biological Sciences*, 10(4), 151–152. https://doi.org/10.34172/cjmb.2023.4007

Khanna, V.V., Chadaga, K., Sampathila, N., Prabhu, S., Bhandage, V. & Hegde, G.K. (2023). A Distinctive Explainable Machine Learning Framework for Detection of Polycystic Ovary Syndrome. *Appl. Syst. Innov.* 6(1), 32-39.

Kumar, A., Singh, J. and Khan, A. A. (2024). A Comprehensive Machine Learning Framework with Particle Swarm Optimization for Improved Polycystic Ovary Syndrome (PCOS) Diagnosis. *Engineering Research Express*, 6(1), 1-23. https://doi.org/10.1088/2631-8695/ad76f9

Lathifah, A., Arham, Z., Hasanati, N., Zulfiandri A. & Nurmiati, E. (2023). Cross-Industry Standard Process for Data Mining (CRISP-DM) for Discovering Association Rules in Graduate Tracer Study Data of Islamic Higher Education Institution. 11th International Conference on Cyber and IT

Service Management (CITSM), Makassar, Indonesia, pp.1-6. https://doi.org/10.1109/CITSM60085.2023.10455691

Lazareva L.M. (2023). The Prevalence of Polycystic Ovarian Syndrome in the Population of Women of Reproductive Age Using the Rotterdam 2003 criteria (Britory Review). *Acta Biomedica Scientifica*, 8 (4), 59-67. https://doi.org/10.29413/ABS.2023-8.4.7

Mohamed, S. M., Nivya, P. S., Devika, C. S., Malayanur, S R., Poornima, N., et al. (2023). A Review of Hyperandrogenism State in Polycystic Ovarian Syndrome. *International journal of Reproduction, Contraception, Obstetrics and Gynecology*. https://doi.org/10.18203/2320-1770.ijrcog20233666

Neven, A. C. H., Forslund, M., Ranashinha, S., et al. (2024). Prevalence and Accurate Diagnosis of Polycystic Ovary Syndrome in Adolescents Across World Regions: A Systematic Review and Meta-Analysis. *European Journal of Endocrinology*, 191(4), S15–S27. https://doi.org/10.1093/ejendo/lvae125

Oliha, F. O and Omobude, E. K. (2023). Aggregating Covid-19 Data on Vaccination and Mortality Progression for Dashboard Visualization: A Recap. *Nigerian Journal of Science and Environment*, 21(2), 638 – 651.

Oliha, F. O. and E. I. Obayagbona, (2023). Dashboarding Vaccine and Routine Immunization Data Services for Selected Healthcare Centers: A User-Centric Evaluation. *Nigerian Journal of Science and Environment*, 21(2), 500-513.

Poojitha G, Talla V. (2024). Diagnosis and Treatment of Polycystic Ovarian Syndrome: An Update. *Pharm Pharmacol Int J*, 12(5):185-189. https://doi.org/10.15406/ppij.2024.12.00450

Rani, R., Priyanka, R. Y., Bajal, N., Mandal, B. S., Kumar, A. P. (2024). Polycystic Ovary Syndrome (PCOS): from Diagnosis to Treatment. *International Journal of Environment Agriculture and Biotechnology*, 9(4), 031-037. https://doi.org/10.22161/ijeab.94.5

Rani, R., Yadav, P., Bajal, N., Mandal, B S., Kumar, A. P. (2024). Polycystic Ovary Syndrome (PCOS): from Diagnosis to Treatment. *International Journal of Environment Agriculture and Biotechnology*, 9(4), 031-037. https://doi.org/10.22161/ijeab.94.5

Siddiqui, S., Mateen, S., Ahmad, R., Moin, S. (2022). A Brief Insight into the Etiology, Genetics, and Immunology of Polycystic Ovarian Syndrome (PCOS). *J. Assist. Reprod. Genet*, 39, 2439–2473.

Taieb, A., & Feryel, A. (2024). Deciphering the Role of Androgen in the Dermatologic Manifestations of Polycystic Ovary Syndrome Patients: A State-of-the-Art Review. *Diagnostics*, 14(22), 2578. https://doi.org/10.3390/diagnostics14222578

Thakre V., Vedpathak, S., Thakre, K., Sonawani, S. (2020). PCOcare: PCOS Detection and Prediction Using Machine Learning Algorithms. *Biosci. Biotechnol. Res. Commun*. 13(2), 240 – 244.

Thorat, S., Ranjan, N., Gangal, V., Jogdand, T., Kulat, S. & Borde, S. P. R. (2024). Identification of Polycystic Ovarian Syndrome through Clinical Metrics and Ultrasonographic Imaging Analysis. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, pp.1-6. https://doi.org/10.1109/InC460750.2024.10649348

Tiwari, S., Kane, L., Koundal, D., Jain, A., Alhudhaif, A., Polat, K., Zaguia, A., Alenezi, F., & Althubiti, S.A. (2022). SPOSDS: A Smart Polycystic Ovary Syndrome Diagnostic System using Machine Learning. *Expert Syst. Appl*. 203, 592-603.

Unfer, V., Kandaraki, E., Pkhaladze, L., Roseff, S., Vazquez-Levin, M. H., Laganà, A.S., et al. (2024). When one size does not fit all: Reconsidering PCOS Etiology, Diagnosis, Clinical Subgroups, and Subgroup-Specific Treatments. *Endocrine and Metabolic Science*, 14, 345-353. https://doi.org/10.1016/j.endmts.2024.100159

Vinothini, S., Vaishnavi S., and Mythili (2024). Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Machine Learning. In the IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, pp.1-9. https://doi.org/10.1109/InC460750.2024.10649123