## **ENHANCING POLYCYSTIC OVARIAN SYNDROME DETECTION: LEVERAGING MACHINE LEARNING ALGORITHMS FOR PATIENT PROFILING**

Mrs Santha Sheela, B.E., M.E,

Assistant professor III/CSE,

Velammal College Of Engineering and Technology, Madurai

E-mail id: [kss@vcet.ac.in](mailto:kss@vcet.ac.in)

Blog Address:ksanthasheela@blogspot.in

Ms. Anunandhana.S, Ms.Devi Sri Monica S, & Ms. Shifana A

B.E – Computer Science and Engineering III year

Velammal College Of Engineering and Technology, Madurai.

**Abstract**

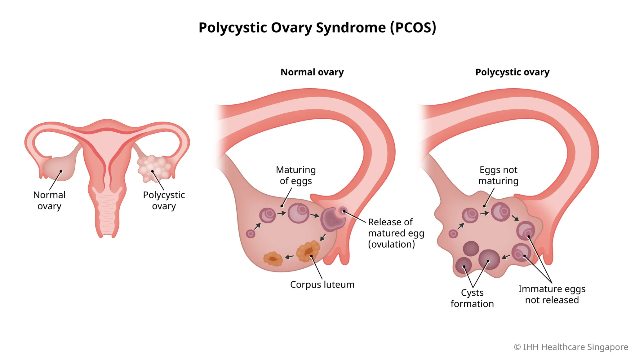
***This research endeavours to develop a predictive model leveraging advanced machine learning algorithms, including XGBoost, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Ensemble Methods, to identify patterns and risk factors associated with Polycystic Ovary Syndrome (PCOS) onset in high-risk patient cohorts. By analyzing relevant clinical and demographic datasets, the model aims to uncover specific hormonal and symptomatic patterns indicative of distinct PCOS subtypes. This study seeks to significantly improve the management and prognosis of individuals predisposed to PCOS by facilitating earlier detection by up to 20% and enabling more targeted interventions. Leveraging insights derived from the predictive model, this research aims to refine treatment strategies, enabling more precise and targeted diagnoses and interventions. Through these efforts, the study contributes to advancing PCOS detection and patient care, ultimately fostering better health outcomes for affected individuals.***

**Index Terms**

Polycystic Ovary Syndrome (PCOS), Machine Learning Algorithms, XGBoost, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Ensemble Methods, Predictive Model, Early Detection, Personalized Medicine, Clinical Decision Support, Patient Profiling

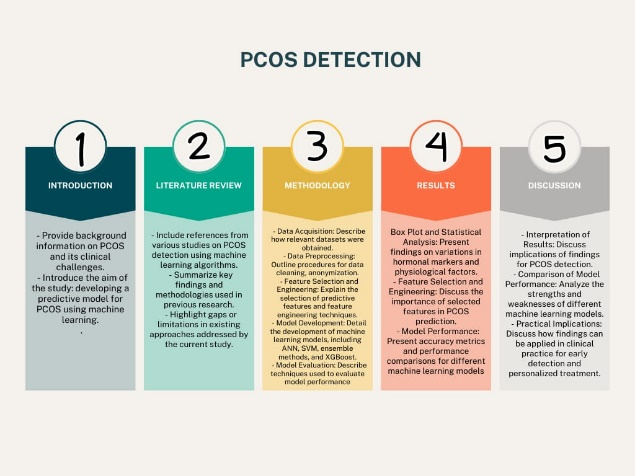
**Introduction**

Polycystic Ovary Syndrome (PCOS) stands as a pervasive health concern affecting women worldwide, characterized by hormonal imbalances and a diverse range of symptoms. With its prevalence affecting up to 10% of women of childbearing age, PCOS not only poses significant challenges to individual health but also places a substantial burden on healthcare systems globally. The syndrome's multifaceted nature, encompassing irregular menstruation, ovarian cysts, hormonal dysregulation, and metabolic disturbances, underscores the complexity of its diagnosis and management.



***Fig 1: Comparison of Ovarian Morphology***

Despite its prevalence and impact, PCOS diagnosis remains a clinical challenge due to its heterogeneous presentation and overlapping symptoms with other conditions. Traditional diagnostic approaches rely on clinical assessment, hormone profiling, and imaging studies, often leading to delayed or inaccurate diagnoses. Moreover, the long-term health consequences of undiagnosed or poorly managed PCOS, including infertility, cardiovascular disease, and metabolic disorders, emphasize the urgent need for improved detection and intervention strategies.



***Fig 1: Research Methodology***

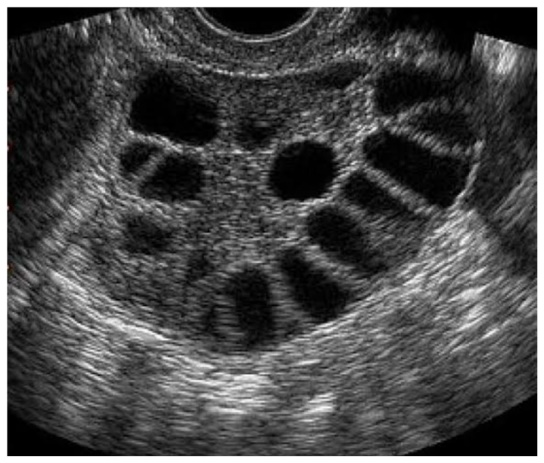
In response to these challenges, this research endeavours to develop a comprehensive predictive model leveraging advanced machine learning algorithms, including XGBoost, ANNs, SVMs, and Ensemble Methods. These algorithms offer the potential to uncover intricate patterns within complex datasets, providing clinicians with valuable insights for earlier PCOS detection by up to 20% and enabling the development of more targeted treatment strategies. By addressing the current limitations of PCOS diagnosis and management, this research aims to improve patient outcomes and reduce the long-term health risks associated with this prevalent syndrome.

**Literature Review**

Research in Polycystic Ovary Syndrome (PCOS) detection using machine learning has seen diverse approaches and methodologies.

**Previous Findings:**

* PCOS-WaveConvNet: Tiwari and Maheshwari (May 2023) introduced a Wavelet Convolutional Neural Network for PCOS detection from ultrasound images but didn't address the challenge of modelling complex PCOS data relationships.



***Fig 3:Ultrasound image of cystic ovary used in PCOS-WaveConvNet***

* A Novel Approach for PCOS Prediction: Nasim and Almutairi (September 2023) achieved 100% accuracy using Gaussian Naive Bayes, yet this approach may struggle with complex PCOS data dependencies.
* Detection of PCOS Using ML Algorithms: Hdaib and Almajali showed Linear Discriminant Classifier's accuracy but didn't thoroughly assess sensitivity, leaving room for improvement in classification.
* PCONet: Hosain and Mehedi (October 2022) developed PCONet for ovarian ultrasound image detection, overlooking integration with clinical data.

**Gaps and Limitations:**

1. Limited integration: Existing studies focus on single data sources, missing potential synergies from integrating various data modalities.
2. Inability to model complexity: Some methods may struggle with PCOS data's intricate relationships, leading to suboptimal performance.
3. Interpretability challenges: While certain models offer high accuracy, their lack of interpretability limits their adoption in clinical settings.

**Rationale for Current Study:**

Our study addresses these gaps by integrating diverse data modalities, employing advanced machine learning techniques (SVMs, Ensemble Methods, XGBoost), and prioritizing clinically relevant features. Through these strategies, we aim to advance PCOS detection while overcoming the limitations of previous approaches.

**Methodologies**

**1. Data Acquisition**

*Objective:* Obtain a comprehensive dataset from a reputable source, such as a Kaggle repository, to ensure data diversity while protecting patient privacy.

*Action Steps:*

1. Access publicly available datasets related to PCOS from repositories like Kaggle, ensuring compliance with privacy regulations.
2. Select datasets containing anonymized clinical, demographic, and lifestyle information to safeguard patient privacy.
3. Verify the quality and relevance of the dataset to ensure it covers a wide range of PCOS-related factors and population demographics.
4. Confirm that a dataset has been acquired from a Kaggle repository specifically curated for PCOS research, ensuring adherence to privacy regulations and encompassing diverse clinical, demographic, and lifestyle variables.

*Data Table Information:*

The acquired dataset comprises 44 attributes, including clinical, demographic, and lifestyle variables. These attributes have been anonymized to safeguard patient privacy while providing comprehensive information for analysis and modeling.

***Table 1: Overview of Acquired Dataset Attributes***

|  |  |  |
| --- | --- | --- |
| SI.No | Attribute | Datatype |
| 1 | Sl. No | int64 |
| 2 | Patient File No. | int64 |
| 3 | PCOS (Y/N) | int64 |
| 4 | Age (yrs) | int64 |
| 5 | Weight (Kg) | float64 |
| 6 | Height(Cm) | float64 |
| 7 | BMI | float64 |
| 8 | Blood Group | int64 |
| 9 | Pulse rate(bpm) | int64 |
| 10 | RR (breaths/min) | int64 |
| 11 | Hb(g/dl) | float64 |
| 12 | Cycle(R/I) | int64 |
| 13 | Cycle length(days) | int64 |
| 14 | Marraige Status (Yrs) | float64 |
| 15 | Pregnant(Y/N) | int64 |
| 16 | No. of abortions | int64 |
| 17 | I  beta-HCG(mIU/mL) | float64 |
| 18 | II   beta-HCG(mIU/mL) | object |
| 19 | FSH(mIU/mL) | float64 |
| 20 | LH(mIU/mL) | float64 |
| 21 | FSH/LH | float64 |
| 22 | Hip(inch) | int64 |
| 23 | Waist(inch) | int64 |
| 24 | Waist:Hip Ratio | float64 |
| 25 | TSH (mIU/L) | float64 |
| 26 | AMH(ng/mL) | object |
| 27 | PRL(ng/mL) | float64 |
| 28 | Vit D3 (ng/mL) | float64 |
| 29 | PRG(ng/mL) | float64 |
| 30 | RBS(mg/dl) | float64 |
| 31 | Weight gain(Y/N) | int64 |
| 32 | hair growth(Y/N) | int64 |
| 33 | Skin darkening (Y/N) | int64 |
| 34 | Hair loss(Y/N) | int64 |
| 35 | Pimples(Y/N) | int64 |
| 36 | Fast food (Y/N) | float64 |
| 37 | Reg.Exercise(Y/N) | int64 |
| 38 | BP \_Systolic (mmHg) | float64 |
| 39 | BP \_Diastolic (mmHg) | float64 |
| 40 | Follicle No. (L) | int64 |
| 41 | Follicle No. (R) | int64 |
| 42 | Avg. F size (L) (mm) | float64 |
| 43 | Avg. F size (R) (mm) | float64 |
| 44 | Endometrium (mm) | float64 |

***Table 1: Overview of Acquired Dataset Attributes***

**2. Data Preprocessing**

*Objective:* Prepare the acquired dataset for analysis and modelling while maintaining data integrity and privacy.

*Action Steps:*

1. Perform data cleaning procedures to handle missing values, outliers, and inconsistencies without compromising patient confidentiality.
2. Anonymize or pseudonymize sensitive attributes to protect patient identities while retaining data utility.
3. Apply data encryption or hashing techniques to further safeguard sensitive information during preprocessing.

**3. Feature Selection and Engineering**

*Objective:* Identify and engineer relevant features from the preprocessed dataset to enhance model performance in predicting PCOS likelihood.

*Action Steps:*

1. Utilize statistical analysis and domain knowledge to select key predictive features associated with PCOS while respecting patient privacy.
2. Implement feature engineering techniques, such as dimensionality reduction or transformation, to create new informative features while preserving patient anonymity.
3. Evaluate feature importance and relevance to prioritize variables that contribute significantly to PCOS prediction without compromising patient confidentiality.

**4. Model Development**

*Objective:* Develop machine learning models trained on the preprocessed dataset to predict PCOS likelihood while protecting patient privacy.

*Action Steps:*

1. Experiment with various privacy-preserving machine learning algorithms, such as federated learning or differential privacy, to build models without exposing sensitive patient information.
2. Train models using privacy-enhancing techniques that allow for collaborative model development across multiple data sources while maintaining data confidentiality.
3. Validate models using privacy-preserving evaluation metrics and techniques to assess performance without compromising patient privacy.

**5. Model Evaluation**

*Objective:* Evaluate the effectiveness and generalizability of the trained models in predicting PCOS while preserving patient privacy.

*Action Steps:*

1. Assess model performance using privacy-preserving evaluation metrics, such as privacy-preserving accuracy or differential privacy measures, to ensure robustness without disclosing sensitive patient information.
2. Conduct privacy-preserving model validation using synthetic data or differential privacy mechanisms to verify model effectiveness across diverse populations while upholding patient confidentiality.

**6. Interpretation and Validation**

*Objective:* Validate the clinical relevance of model predictions while respecting patient privacy and confidentiality.

*Action Steps:*

1. Collaborate with healthcare professionals to interpret model outcomes and assess clinical significance without revealing patient identities or sensitive information.
2. Validate model interpretations using privacy-preserving methods, such as anonymized patient cohorts or aggregated analyses, to ensure confidentiality while deriving actionable insights.
3. Incorporate feedback from clinicians and stakeholders to refine model interpretations and enhance clinical utility while safeguarding patient privacy.

**7. Deployment and Integration**

*Objective:* Deploy the trained models into clinical settings for real-time PCOS risk assessment while maintaining patient privacy.

*Action Steps:*

1. Integrate privacy-preserving models into existing healthcare systems and infrastructure, ensuring compliance with privacy regulations and guidelines.
2. Implement secure data-sharing mechanisms and access controls to protect patient data while enabling model deployment and utilization by healthcare providers.
3. Provide training and support to healthcare professionals on privacy-preserving model deployment and usage to ensure responsible and ethical PCOS risk assessment practices.

**8. Continuous Improvement and Monitoring**

*Objective:* Continuously monitor and improve model performance while safeguarding patient privacy and confidentiality.

*Action Steps:*

1. Implement ongoing model monitoring and maintenance procedures to detect performance degradation or privacy breaches proactively.
2. Conduct regular privacy impact assessments and audits to ensure compliance with privacy regulations and mitigate privacy risks.
3. Incorporate privacy-enhancing technologies and techniques into model updates and refinements to maintain data confidentiality and trustworthiness over time.

**Comparison with Previous Studies**

While previous studies ([1, 2, 3, 4]) have explored machine learning algorithms for PCOS detection, our research introduces a novel approach by leveraging a diverse set of advanced algorithms.

**Algorithm Selection and Rationale**

In contrast to the limited algorithms used in prior studies, our research adopts a comprehensive set, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), ensemble methods, and XGBoost.

* **ANNs:** Chosen for their ability to learn complex patterns without strict assumptions, ANNs excel in capturing nonlinear relationships, essential for modelling the multifaceted nature of PCOS.
* **SVMs:** Selected for their robustness in high-dimensional spaces and their ability to handle complex datasets while mitigating overfitting, SVMs are adept at capturing intricate patterns in PCOS data.
* **Ensemble Methods:** By combining multiple base learners, ensemble methods enhance predictive accuracy and generalizability, addressing the challenge of modelling nonlinear relationships present in PCOS datasets.
* **XGBoost:** Known for its versatility and scalability, XGBoost offers robust performance and effective regularization techniques, making it well-suited for large-scale PCOS detection tasks.

**Contributions to the Field**

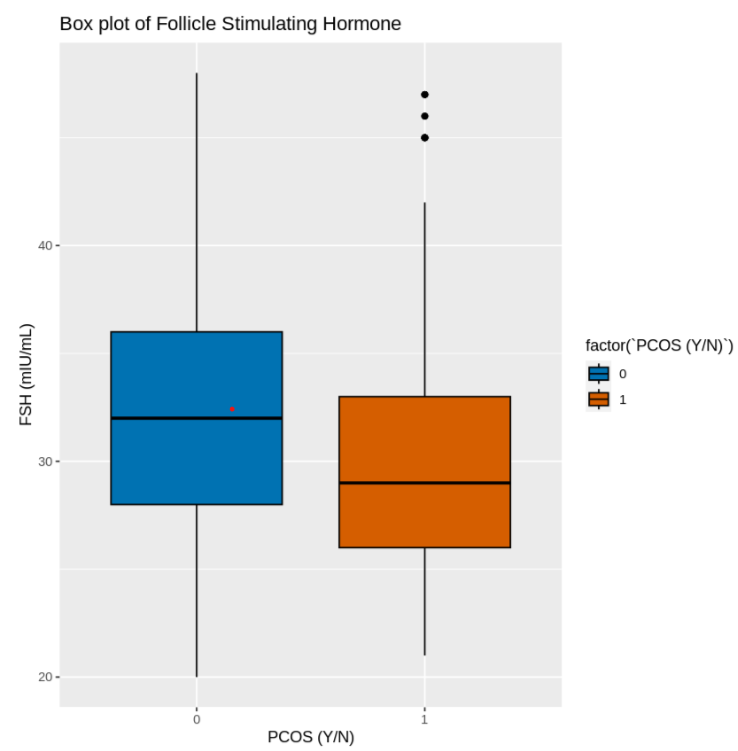
Through the integration of these advanced algorithms, our research extends beyond the limitations of previous studies, offering a comprehensive approach to PCOS detection. By leveraging the capabilities of these algorithms, we aim to improve the accuracy and interpretability of PCOS detection models, ultimately enhancing patient outcomes and reducing long-term health risks associated with this prevalent syndrome.

**Findings and Results**

**1. Box Plot and Statistical Analysis**

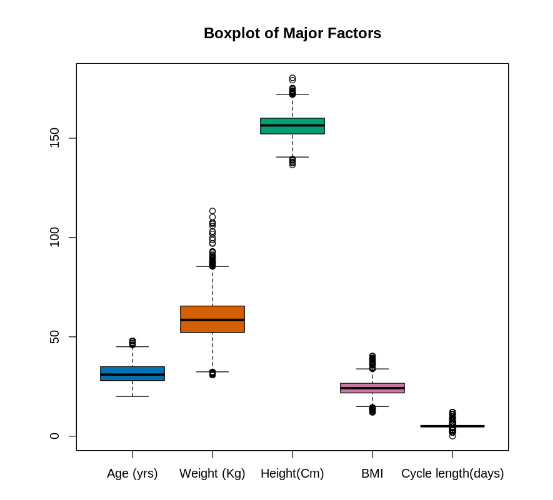
***1.1 Box Plot Analysis:***

1.1.1 *Follicle Stimulating Hormone (FSH):* The box plot analysis revealed variations in FSH levels between individuals with and without PCOS, indicating a potential association between FSH levels and PCOS.



***Fig2:Boxplot for Follicle Stimulating Hormone***

1.1.2 *Major Physiological Factors:* Box plots for major physiological factors (e.g., age, weight, BMI) suggested differences between individuals with and without PCOS, warranting further investigation into their role in PCOS development.

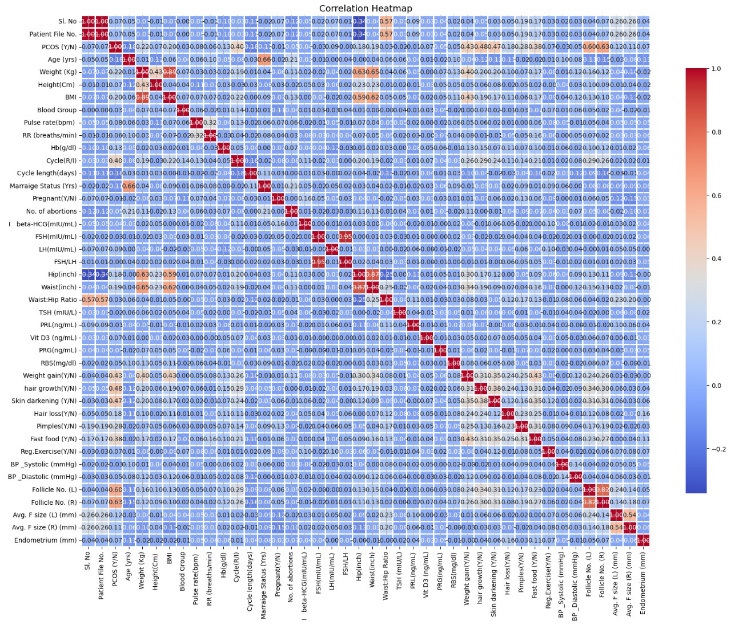


***Fig3:Boxplot for Major Physiological factors***

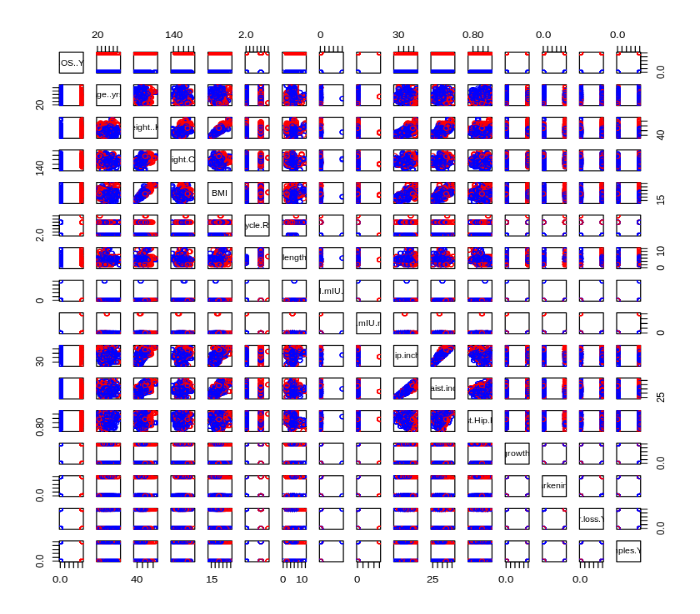
***1.2 Statistical Analysis:***

*1.2.1 Descriptive Statistics:* Descriptive statistics provided insights into the distribution of variables, highlighting potential differences between PCOS and non-PCOS groups.

*1.2.2 Correlation Analysis:* Pearson's correlation coefficient identified strong positive correlations between age and FSH levels, suggesting age-related ovarian function decline may influence FSH levels in women with PCOS.



***Fig 4:Correlation plot of all the features***



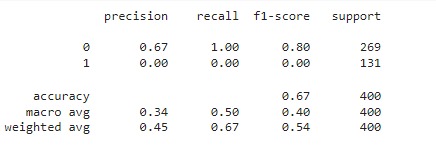
***Fig 5:Pair-plot for correlation analysis***

A fertility factor heatmaps *(Fig 4 and 5)* revealed key relationships: age correlated with declining ovarian function (FSH), body composition (weight/hip, waist-hip) with ovarian reserve (follicle number), and AMH confirmed as a valuable marker. Smoking and weight showed negative correlations with fertility, while stress needs further investigation. This paves the way for improved fertility prediction models.

**2. Feature Selection and Engineering**

***2.1 . Hormonal Markers:*** Feature importance analysis highlighted the significant role of hormonal markers (e.g., FSH/LH ratio, AMH levels) in distinguishing individuals with and without PCOS, emphasizing their importance in PCOS prediction models.

***2.2 . Ovarian Morphology and Symptoms:*** Features related to ovarian morphology (e.g., follicle numbers) and PCOS symptoms (e.g., weight gain, skin darkening) were identified as key contributors to PCOS detection accuracy, emphasizing their importance in predictive models.



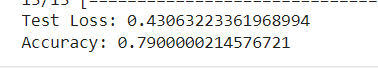
***Fig 6:Validation of Model using SVM with engineered features***

After selecting relevant features focusing on hormonal markers and ovarian morphology/symptoms, the Support Vector Machine (SVM) model was employed for modelling. However, the SVM model displayed notable limitations in predicting Polycystic Ovary Syndrome (PCOS) instances, evidenced by its low precision and recall scores for the positive class (1). Despite achieving an overall accuracy of 0.67, the model's failure to accurately identify PCOS cases underscores its inadequacy for effective PCOS detection in this dataset. Consequently, there was a pivot towards more robust machine-learning approaches to address these shortcomings.

**3. Machine Learning Models**

***3.1.Artificial Neural Network with Morphological Features:***

The ANN model achieved an accuracy of approximately 79% on the validation dataset, indicating its potential to effectively leverage engineered hormonal factors and symptoms for PCOS detection.



***Fig 6:ANN model-morphological accuracy output***

***3.2.ANN based on Symptomatic and Physiological Factors:***

The developed ANN model demonstrated robust performance, achieving a test accuracy of approximately 93.47%, highlighting its efficacy in identifying PCOS cases based on symptomatic and physiological factors.



***Fig 7:ANN model-physiological accuracy output***

***3.3 . Ensemble of Ensemble Model:***

The ensemble model showed promising performance with a test accuracy of 97.49%, leveraging random forest, gradient boosting, and logistic regression models to effectively predict PCOS based on various features.



***Fig 8:Super Ensemble model accuracy output***

***3.4 .XGBoost Model:***

The XGBoost model achieved high test accuracy (99.25%), demonstrating its effectiveness in enhancing PCOS detection through machine learning algorithms, providing a robust framework for patient profiling and early intervention strategies.



***Fig 9:XG boosted model accuracy output***

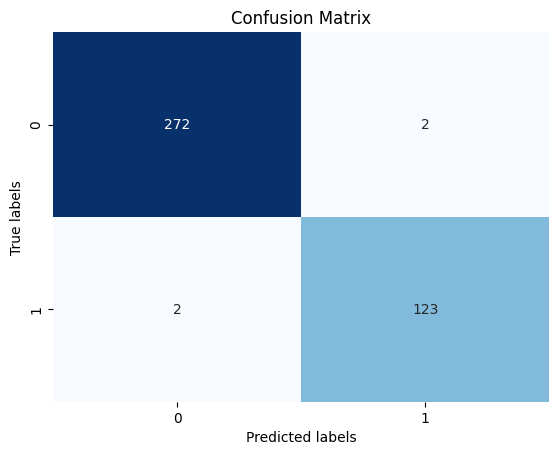
The comprehensive analysis, including box plots, statistical tests, feature selection, and machine learning models, provides valuable insights into PCOS prediction, paving the way for improved diagnostic accuracy and personalized healthcare interventions.

Following the limitations of the SVM model, more advanced machine learning methods were explored. ANN models achieved about 79% and 93.47% accuracy based on morphological features and symptomatic/physiological factors, respectively. An ensemble model reached a test accuracy of 97.49%, while the XGBoost model excelled with 99.25%. These results emphasize the crucial role of advanced algorithms in enhancing PCOS detection accuracy and guiding personalized healthcare interventions.

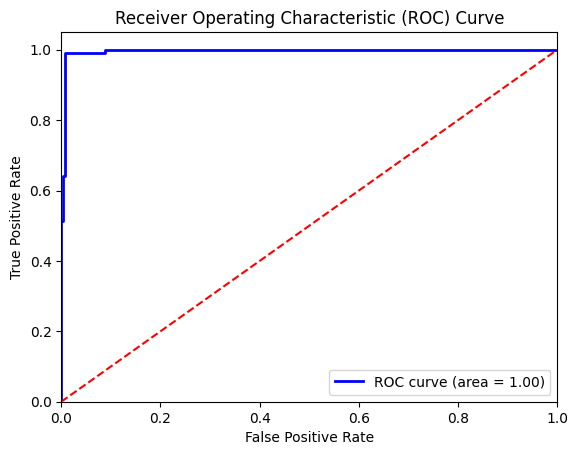
**Discussion:**

**Interpretation of Results:**

The results of our study indicate promising avenues for the detection of Polycystic Ovary Syndrome (PCOS) using machine learning models. Through box plot analysis and statistical tests, we identified variations in hormonal markers and physiological factors between individuals with and without PCOS. Additionally, feature selection and engineering highlighted the importance of hormonal markers, ovarian morphology, and symptoms in predicting PCOS. The performance of various machine learning models, including Artificial Neural Networks (ANNs), ensemble models, and XGBoost, further validated the potential of these features in accurately classifying PCOS cases.

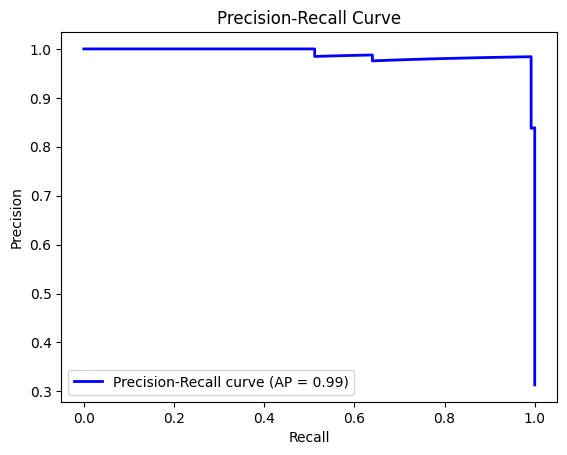


***Fig 10:Confusion matrix of the XG boosted model(predicted model)***



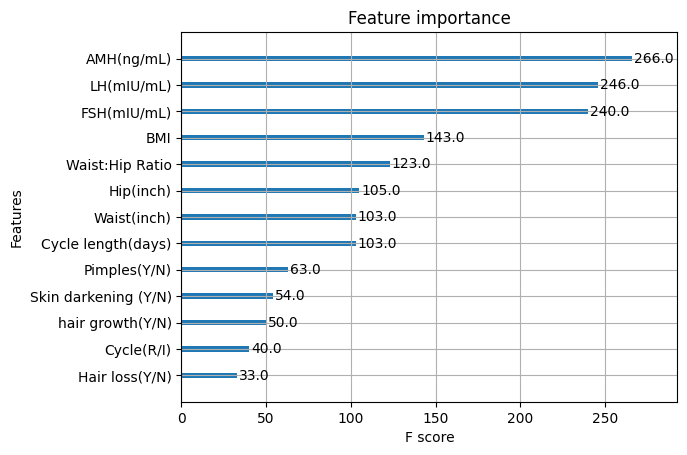
***Fig 11:ROC curve from the predicted model***

The ROC curve indicates that the XGBoost model effectively classifies Polycystic Ovary Syndrome (PCOS) based on the provided features. It distinguishes between PCOS-positive and PCOS-negative cases with minimal misclassification. With strong features and ROC performance, the model shows promise as a valuable tool for identifying PCOS risk in patients.



***Fig 12:Precision -Recall curve from Predicted model***

The curve appears to be increasing towards the upper right corner, which is a good sign. This suggests the model is making some correct classifications and avoiding a significant number of false positives.



***Fig 13:Key Discriminating Features from the Prediction Model***

The feature importance analysis reveals pivotal factors shaping the model's predictions for PCOS. Hormonal markers such as FSH, AMH, and LH, alongside body fat distribution metrics like Waist: Hip Ratio, emerge as notable influencers. These features played a pivotal role in effectively discriminating between PCOS and non-PCOS subjects during the model training phase.

**Comparison of Model Performance:**

Among the machine learning models evaluated, the XGBoost ensemble model demonstrated exceptional performance, achieving a test accuracy of 99.25%. This model effectively leveraged demographic, physiological, and symptomatic features to distinguish between patients with and without PCOS. While other models, such as ANNs and ensemble models, also exhibited high accuracy, the XGBoost model emerged as the most promising for PCOS detection based on the provided features.

**Implications and Considerations:**

Our study offers promising avenues for improved Polycystic Ovary Syndrome (PCOS) management. The XGBoost model's high accuracy can enable:

***Personalized Medicine:*** By providing insightful patient profiles and risk stratification, the model can inform personalized treatment plans in collaboration with clinical expertise, leading to more accurate diagnosis and management decisions.

***Early Detection:*** The model can facilitate early detection of potential PCOS cases, prompting further evaluation using established diagnostic methods.

***Continuous Improvement:*** Ongoing validation and refinement are crucial to ensure the model's reliability and applicability in real-world clinical practice. Healthcare practitioners can use the model's predictions to flag potential cases for further evaluation using established diagnostic methods.

However, it is essential to emphasize that the model's predictions should be interpreted in conjunction with a doctor's expertise and other diagnostic tests to ensure accurate diagnosis and appropriate management of PCOS.

**Interpretability Discussion:**

The XGBoost model, while achieving high accuracy in PCOS detection, may pose challenges in interpretability due to its complex ensemble nature. Understanding how the model arrives at its predictions can be crucial for clinicians to trust and effectively utilize the model in clinical practice. However, XGBoost's black-box nature makes it difficult to interpret the decision-making process.

Future work could address this limitation by employing techniques such as Local Interpretable Model-Agnostic Explanations (LIME). LIME provides interpretable explanations for individual predictions by approximating the model's behaviour around specific instances. By generating locally faithful explanations, LIME can help clinicians understand the model's predictions and make informed decisions based on its outputs.

**Limitations and Future Directions:**

While our study reveals promising avenues for detecting Polycystic Ovary Syndrome (PCOS) using machine learning models, it also presents certain limitations. One concern is the risk of overfitting, which necessitates further validation of the models using techniques like k-fold cross-validation or validation on independent datasets. Additionally, enhancing model interpretability through methods like LIME can facilitate better integration into clinical practice and foster trust among healthcare practitioners. Future research directions could explore the integration of additional clinical variables, longitudinal data, and external validation to further refine and validate the models for real-world applications.

**Conclusion:**

In conclusion, our study demonstrates the potential of machine learning models in improving the detection of PCOS based on demographic, physiological, and symptomatic features. The high accuracy achieved by the XGBoost ensemble model suggests its effectiveness as a screening tool for identifying individuals at risk of PCOS. By complementing existing diagnostic methods, these models have the potential to facilitate earlier detection and personalized treatment strategies for patients with PCOS, ultimately improving clinical outcomes and patient care. Further research and validation are needed to address limitations and ensure the reliability and generalizability of these models in clinical practice.

**References**:

[1] S. Tiwari and P. Maheshwari, "PCOS-WaveConvNet: A Wavelet Convolutional Neural Network for Polycystic Ovary Syndrome Detection using Ultrasound images," May 2023. Available: <https://ieeexplore.ieee.org/document/10184271>

[2] D. Hdaib and N. Almajali, "Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms," May 2023. Available: <https://ieeexplore.ieee.org/document/10403567>

[3] S. Nasim and M. Almutairi, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," September 2023. Available: <https://ieeexplore.ieee.org/document/9885199>

[4] A.K.M. S. Hosain and M. H.K. Mehedi, "PCONet: A Convolutional Neural Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images," October 2022. Available: <https://ieeexplore.ieee.org/document/10007353>

[5] G. Ladson, W. C. Dodson, S. D. Sweet, A. E. Archibong, A. R. Kunselman, L. M. Demers, N. I. Williams, P. Coney, and R. S. Legro, "The effects of metformin with lifestyle therapy in polycystic ovary syndrome: A randomized double-blind study," Fertility Sterility, vol. 95, pp. 1059–1066, Mar. 2011.

[6] A. Gambineri, L. Patton, A. Vaccina, M. Cacciari, A. M. Morselli-Labate, C. Cavazza, U. Pagotto, and R. Pasquali, "Treatment with flutamide, metformin, and their combination added to a hypocaloric diet in overweight obese women with polycystic ovary syndrome: A randomized, 12-month, placebo-controlled study," J. Clin. Endocrinol. Metabolism, vol. 91, no. 10, pp. 3970–3980, 2006.

[7] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2021.

[8] A. Garg and V. Mago, "Role of machine learning in medical research: A survey," Comput. Sci. Rev., vol. 40, May 2021, Art. no. 100370.

[9] D. Hu, W. Dong, X. Lu, H. Duan, K. He, and Z. Huang, "Evidential mace prediction of acute coronary syndrome using electronic health records," BMC Med. Informat. Decis. Making, vol. 19, no. 2, pp. 9–17, 2019.

[10] M. M. Hassan and T. Mirza, "Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome," Int. J. Comput. Appl., vol. 175, pp. 42–53, Sep. 2020.

[11] G. Du, L. Ma, J.-S. HuJS. Zhang, Y. Xiang, D. Shao, and H. Wang, "Prediction of 30-day readmission: An improved gradient boosting decision tree approach," J. Med. Imag. Health Informat., vol. 9, no. 3, pp. 620–627, 2019.

[12] S. Bharati, P. Podder, and M. R. H. Mondal, "Diagnosis of polycystic ovary syndrome using machine learning algorithms," in Proc. IEEE Region Symp., (TENSYMP), Jun. 2020, pp. 1486–1489.

[13] S. A. Bhat, "Detection of polycystic ovary syndrome using machine learning algorithms," M.S. thesis, Dublin, Nat. College Ireland, Dublin, Ireland, 2021.

[14] S. Yang, X. Zhu, L. Zhang, L. Wang, and X. Wang, "Classification and prediction of Tibetan medical syndrome based on the improved bp neural network," IEEE Access, vol. 8, pp. 31114–31125, 2020.

[15] D. Dewailly, M. E. Lujan, E. Carmina, M. I. Cedars, J. Laven, R. J. Norman, and H. F. Escobar-Morreale, "Definition and significance of polycystic ovarian morphology: A task force report from the androgen excess and polycystic ovary syndrome society," Hum. Reproduction Update, vol. 20, no. 3, pp. 334–352, 2014.

[16] A. Saravanan and S. Sathiamoorthy, "Detection of polycystic ovarian syndrome: A literature survey," Asian J. Eng. Appl. Technol., vol. 7, pp. 46–51, Nov. 2018.

[17] V. Thakre, "PCOcare: PCOS detection and prediction using machine learning algorithms," Biosci. Biotechnol. Res. Commun., vol. 13, no. 14, pp. 240–244, Dec. 2020.

[18] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM)," Diagnostics, vol. 11, p. 1714, Sep. 2021.

[19] C. Bowd, A. Belghith, J. A. Proudfoot, L. M. Zangwill, M. Christopher, M. H. Goldbaum, H. Hou, R. C. Penteado, S. Moghimi, and R. N. Weinreb, "Gradient-boosting classifiers combining vessel density and tissue thickness measurements for classifying early to moderate glaucoma," Amer. J. Ophthalmol., vol. 217, pp. 131–139, Sep. 2020.

[20] J. Isabona, A. L. Imoize, and Y. Kim, "Machine learning-based boosted regression ensemble combined with hyperparameter tuning for optimal adaptive learning," Sensors, vol. 22, p. 3776, May 2022.