The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
19(2): S2: 507-512, 2024 www.thebioscan.com

An Analysis of Polycystic Ovarian Syndrome (PCOS) Among College Students in Ahilyanagar, Maharashtra, Based on Different Statistical Models

Anuja Bhalerao[1], Ivan Aranha[1]*, Yogesh Yewale[2], Urmila Shinde[3], and Gautam Aher[3]

[1]Department of Zoology, Ahmednagar College, Ahilyanagar, Savitribai Phule Pune University, Maharashtra 414 001

[2] Department of Statistics, Ahmednagar College, Ahilyanagar, Savitribai Phule Pune University, Maharashtra 414 001

[3] Department of Obstetrics and Gynaecology, Dr. Vithalrao Vikhe Patil Foundation's Medical College, Ahilyanagar, Maharashtra 414001

Corresponding author: ivanaranha@gmail.com

DOI: 10.63001/tbs.2024.v19.i02.S2.pp507-512

KEYWORDS
Polycystic Ovarian Syndrome (PCOS), Hyperandrogenism, Oligomenorrhea, insulin resistance, Random Forest, Naive Bayes, Support vector machine, Logistic regression.

Received on:

26-10-2024

Accepted on:

22-11-2024

Published on:

30-12-2024

ABSTRACT

Polycystic Ovarian Syndrome (PCOS) creates a hormonal imbalance among females, compromising their reproductive health. Apart from disruption in reproductive health, this syndrome causes disturbances in various attributes of an individual, such as psychological, metabolic, and circadian rhythm. It also increases the risk of other diseases such as non-alcoholic fatty liver and cardiovascular disorders. The existence of PCOS tendency for a long time causes Infertility; hence, its early detection will help in managing its aetiology. Detecting PCOS is difficult due to the complexity of PCOS pathogenesis, so the Rotterdam criteria are the gold standard. The present study aims to understand the prevalence of PCOS among college-going girls in Ahilyanagar city (formerly Ahmednagar). The study included 1350 females between the ages of 16 to 35. 111 females were clinically diagnosed with PCOS, and a prevalence of 8.22% in Ahilyanagar city, Maharashtra, was detected. Based on the chi-square analysis, PCOS females were associated with hyperandrogenism, oligomenorrhea, and skin darkening; Acanthosis nigricans, an indicator of insulin resistance, was also positive. A study of lifestyle parameters showed a positive association with PCOS. Four statistical models are proposed for the diagnosis of PCOS: Random Forest, Naïve Bayes, Support Vector Machine, and Logistic regression model. Naïve Bayes is the most effective model for the collected data, with an accuracy of 91.48 and sensitivity of 0.9090, followed by Logistic regression with an accuracy of 91.11 and sensitivity of 0.9005.

INTRODUCTION

Polycystic Ovarian Syndrome (PCOS) is characterised by hormonal imbalance, which is often known to have noticeable decrease in FSH and Progesterone levels and increase in LH, Estrogen, Androgens, and Testosterone levels (Azziz, 2018). The hormonal imbalance further characterises irregularity (oligomenorrhea) or absence of the menstrual cycle, hyperandrogenism, insulin resistance, and obesity. The complex aetiology of PCOS often leads to its misdiagnosis. Despite significant progress in understanding the pathophysiology and diagnosis of the disorder over the past 20 years, the disorder remains underdiagnosed and misunderstood by many practitioners (Dokras et al., 2017). Rotterdam criteria have proven to be a gold standard for the correct clinical diagnosis of PCOS (Bhalerao and Aranha, 2021, 2023). It includes the fulfillment of any two characters from the given three characters, such as Polycystic ovarian Morphology (> 12 antral follicles in one ovary), Hyperandrogenism (Clinical and Biochemical), Menstrual irregularities (< 8 cycles per year or >35 days between cycles) (Christ and Cedars, 2023). Globally, the prevalence of PCOS is 15-22%. Literature has reported differences in the prevalence within different populations and ethnicities (Chen and Fang, 2018; Ganie et al., 2019). Being a lifestyle condition, PCOS shows varied prevalence among populations of different regions in India; Mumbai, 21.05% in 2019, Bengaluru, 9.13% in 2011, Karnataka, 9.1% in 2016, Chennai, 6% in 2017, Lucknow, 3.7% in 2012. (Aggarwal et. al, 2019; Ganie et al., 2019). Various lifestyle parameters have proved to be pioneers in developing PCOS condition among females, such as poor sleep quality, Exposure to Endocrine disruptors via cosmetics, room freshener, car freshener, Perfume usage, and Fast-food consumption (Aranha and Bhalerao, 2023). Disruption of circadian rhythm shows a positive association with PCOS by showing higher serum melatonin and cortisol (Heydari and Ramdass, 2025). Consumption of food rich in Advanced Glycation End Products (AEGs) leads to ovarian dysfunction, hyperandrogenism, insulin resistance, and obesity among PCOS females (Mouanness et al., 2022). Exposure to Endocrine modulators present in cosmetic products such as perfumes, air fresheners, and deodorants causes a hormonal imbalance, often leading to estrogen dominance contributes to the pathogenesis of PCOS (Kshetrimanyum et al., 2019).

This cross-sectional study aimed to find out the prevalence of PCOS among college-going girls. PCOS, being a complicated syndrome, often remains undiagnosed, misdiagnosed, and ignored by many students due to its complex aetiology. An attempt is made in this study to prepare a statistical model for the diagnosis of PCOS, which will help students for early consultation with a doctor regarding their reproductive health. The statistical models chosen for this study are Random Forest, Naïve Bayes, Support Vector Machine, and Logistic regression model.

METHODOLOGY

Sample collection

A survey questionnaire was designed containing questions in concern to various clinical parameters such as duration of menstrual cycle, menstrual phase duration, Hyperandrogenism characterized by acne, facial hair, hair loss, and weight gain, Skin Darkening (Acanthosis nigricans) indicating insulin resistance. The questionnaire also included questions regarding lifestyle parameters contributing to PCOS, such as sleep duration, plastic bottle usage, perfume usage, and fast-food consumption.

Inclusion criteria for the survey

Participants in this survey were defined as females between the ages of 16 and 35 who were informed about the survey's purpose.

Statistical analysis

A chi-square test of goodness fit was performed to observe the association of included clinical and lifestyle parameters with PCOS. For chi-square analysis, R Gui software R- R-4.1.5 Windows was used. For chi-square analysis null hypothesis (H0) and alternative hypothesis (H1).

H0: There is no association of the given clinical/lifestyle parameter and PCOS

H1: There is an association between the given clinical/lifestyle parameter and PCOS

Based on the p-value, the H0 will be accepted or rejected. If the p-value is<0.05, we will reject H0 hypothesis and accept H1.

All the clinical and lifestyle parameters were subjected to Random Forest, Naïve Bayes, Support vector machine, and Logistic regression model to help with the diagnosis of PCOS.

a) ***Random Forest analysis:*** It is an algorithm proposed by Leo Breiman in 2001, which includes a combination of decision trees to make predictions. This procedure is consistent because it depends on strong features and not noise variables (Breiman 2001; Biau 2012). Two features are used to build the model. These are the Clinical Score and Lifestyle Score. Data is split into training data (80 %) and testing data (20%) in a random manner. A decision tree model is built using training data. Steps 2 and 3 are repeated 100 times to construct decision trees (100) for each training data. Voting is done at this stage, and the value that occurs the maximum number of times is taken as the predicted value. (Majority vote). This model is then used to make the predictions on test data. To evaluate the model's performance, a confusion matrix and accuracy are obtained.

b) ***Naïve Bayes analysis:*** This algorithm is widely used for data analysis of large data because of its fast and simple structure (Chen et al., 2021). It is based on Bayes' theorem, and all the predictions are independent of each other (Peretez et al., 2024). The theorem states,

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$$

$P(Y|X)$: Posterior probability of Class Y given feature X

$P(X|Y)$: Likelihood

$P(Y)$: Prior Probability of Class Y

$P(X)$: Marginal likelihood

c) ***Support vector Machine analysis:*** This algorithm uses classification and regression tasks therefore considered a powerful supervised algorithm. It works by separating data points into different categories, maximizing the margin between the classes. It handles complex datasets using the "*kernel trick*" for non-linear classification. (Theodoros and Massimiliano 2001; Cervantes et al., 2020).

d) ***Logistic regression model analysis:*** It is a statistical analysis that uses the probability of a binary outcome based on one or more predictor variables. It is a type of regression analysis where the dependent variable is categorical, predicting the probability of an outcome occurring (Peng et al., 2002; Sperandei, 2014). The predicted value in the logistic regression model is a probability; hence, it lies between 0 and 1. The predicted value greater than 0.5 is considered as 1, otherwise it is treated as 0.

Confusion Matrix

It is a matrix that summarizes the performance of the model on test data. It consists of four entries, *viz.*, true positive (correct prediction of positive outcome), true negative (correct prediction of negative outcome), false positive (incorrect prediction of positive outcome), and false negative (incorrect prediction of negative outcome).

Predictions	Predicted PCOS	Predicted No PCOS
Actual PCOS	TP	FN
Actual No PCOS	FP	TN

Accuracy, sensitivity, and precision were calculated for each model using the following formulas:

Accuracy: It is the ratio of the total correctly classified observations to the total observations

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}}$$

Precision: The proportion of true positive predictions among all positive predictions.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}}$$

Recall (Sensitivity): The proportion of true positive predictions among all actual positive instances.

$$\boldsymbol{Sensitivity = \frac{TP}{TP + FN}}$$

RESULTS

Out of 1350 females recruited in study 111(8.22%) were seen to have a clinical diagnosis of PCOS. 128 (9.48%) females showed irregularity in the menstrual cycle but did not show clinical diagnosis of PCOS. Whereas, 1,111(82.29%) females were exhibiting a healthy and normal menstrual cycle. The prevalence of PCOS was estimated to be 8.22% among college-going girls. The chi-square analysis showed a positive association of clinical and lifestyle parameters with the pathogenesis of PCOS. The H0 proposed Table 1 shows the subscale scores for Irregularity in menstrual cycle for all the female candidates, considering the cycle duration. Among 111 females diagnosed with PCOS, 66 exhibited severe irregularity in menstrual cycle, which showed a varied range of cycle duration, including more than 45 days, 3months gap, 38 days cycle, and 20-22 days cycle. While the remaining females exhibited mild irregularity in their menstrual cycle. Those females who were not diagnosed with PCOS also showed mild to severe irregularity in their cycle. Among non-PCOS females, 57 exhibited severe irregularity in the menstrual cycle. Since the p-value is <0.05, we reject the null hypothesis stating that irregular menstrual cycle is associated with PCOS. Irregularity in the Menstrual bleeding pattern is observed among PCOS and non-PCOS females. The excessive bleeding duration is observed among 8 PCOS females, while 45 non-PCOS females also exhibit the same pattern. By rejecting the H0 hypothesis, we state that Irregularity in menstrual bleeding is associated with PCOS.

Skin darkening is observed among all 111 PCOS subjects, whereas 134 non-PCOS females also exhibited the same trait, as stated in Table 3. Skin darkening around lips, neck, fingers, toes, and ankles was observed in these females. Since p value is 0.0004998, which is less than 0.05, H0 is rejected, concluding Skin darkening (Acanthosis nigricans) shows association with PCOS. Tables 4,5,6, and 7 indicate the presence and absence of Acne pattern, weight gain, hair loss, and facial hair growth, respectively, which are the clinical markers of Hyperandrogenism, a pioneer trait for the diagnosis of PCOS. Acne in excess, indicating clinical hyperandrogenism, was observed among 74 PCOS females and 649 non-PCOS females. The occurrence of acne was observed on cheeks and lower chin, forehead, back, and even on the upper arms region. Weight gain was seen among 101 PCOS females and 304 non-PCOS females. 10 PCOS subjects did not exhibit weight gain. Hair loss was seen among all 111 PCOS subjects, while only 334 females exhibited the same trait. Facial hair growth was significantly observed on the Lower lip, upper lip, sides of cheeks, and breasts among 61 PCOS females, while 50 PCOS females did not exhibit this trait. 200 non-PCOS females have shown facial growth patterns similar to those of PCOS subjects. The null hypothesis for acne pattern, weight gain, hair loss, & facial hair growth is rejected since the p value is less than 0.05, indicating all of these clinical markers of hyperandrogenism show association with PCOS.

Body mass index (BMI) is a measure of obesity among individuals. Table 8 shows the range of BMI values among PCOS and non-PCOS females. Based upon the BMI index among PCOS females, 03 are obese, 06 are overweight, 80 are thin, and 22 are healthy. Among the non-PCOS females, 06 are obese and 20 are overweight. The p-value for BMI is less than 0.05; therefore, H0 is rejected, and BMI shows a positive association with PCOS.

PCOS being a lifestyle disorder, sleep duration, daily exercise, plastic bottle usage, cosmetic product usage, perfume usage, air freshener usage, and fast-food consumption contribute a positive association with PCOS, which has been proved by the p-values of each, which are less than 0.05, thus rejecting the H0. Excessive sleep disruption was observed among 31 PCOS females, while mild disruption in sleep was observed in 46 PCOS females. 260 non-PCOS females also seem to have disrupted sleep patterns. All PCOS females did not participate in daily exercise; lower physical activity is an indicator of a sedentary lifestyle. 626 non-PCOS females also opted for a sedentary lifestyle. All 111 PCOS females showed daily usage of a plastic bottle for drinking water. Moderate to excessive cosmetic product usage was observed among 37 PCOS females and 354 non-PCOS females. All the PCOS females showed the usage of perfume on a rare, occasional, and daily basis. Daily perfume usage was observed among 31 PCOS females and 245 non-PCOS females. Air freshener usage on a daily basis was observed among 14 PCOS and 96 non-PCOS females. The p-values of the Chi-square test for goodness fir given for Clinical parameters in Table 16 and lifestyle parameters in Table 17 state their positive association with PCOS. Therefore, we can say that clinical and lifestyle parameters weigh equally for the pathogenesis of PCOS.

Table 18 shows the accuracy and sensitivity values of all four statistical models used for the diagnosis of PCOS. Since the sensitivity of Random Forest analysis is 0.375, even though its accuracy of 92.22% is highest among all four models, we do not consider it a good diagnostic model. The sensitivity of the Naïve Bayes model is 0.9090, which highest among all the models, with an accuracy of 91.48% proving to be an excellent model for the diagnosis of PCOS. The second preferred model for diagnosis of PCOS will be Logistic linear analysis because it has a sensitivity of 0.9000 and an accuracy of 91.11%.

Table 1: Chi-square analysis of Irregular Menstrual Cycle.

Menstrual cycle Duration	PCOS absent	PCOS present
28 days	697	0
28-32 days	221	27
33 days	149	18
33-45 days	45	07
> 45 days, 3 months, 20 days, 22 days, 38 days	127	59

Table 2: Chi-square analysis of Menstrual bleeding pattern.

Bleeding pattern	PCOS absent	PCOS present
Normal	897	0
4-5 days, 5-6 days	67	75
7 days, 6-7 days	05	01
3 days	225	27
2 days, 10-15 days, 25 days, 9-10 days,7-8 days	45	08

Table 3: Chi-square analysis of Skin darkening.

Skin darkening	PCOS absent	PCOS present
No	1105	0
Yes	134	111

Table 4: Chi-square analysis of Acne pattern.

Acne pattern	PCOS absent	PCOS present
No	480	0
1-3 acne anywhere on face	44	31
3-5 acne cheeks, forehead	66	06
5-10 acne on cheek and lower chin & and forehead	89	06
>10 acne on cheek and lower chin, forehead & back	560	68

Table 5: Chi-square analysis of Weight gain.

Weight gain	PCOS absent	PCOS present
No	934	10
Yes	304	101

Table 6: Chi-square analysis of Hair loss.

Hair loss	PCOS absent	PCOS present
No	905	0
Yes	334	111

Table 7: Chi-square analysis of Facial hair growth.

Facial hair growth	PCOS absent	PCOS present

No	1001	0
Yes, very less	38	50
Yes, upper lip	01	09
Yes, upper &lower lip	178	45
Yes, the Lower lip, the upper lip, sides of the cheek	13	04
Yes, the Lower lip, the upper lip, the sides of the cheek, breast	08	03

Table 8: Chi-square analysis of BMI.

BMI	PCOS absent	PCOS present
Less than 18.5(Thin)	1128	80
18.5 to 24.9 (Healthy)	85	22
25 to 29.9(overweight)	20	06
>30(obese)	06	03

Table 9: Chi-square analysis of daily sleep duration.

Sleep duration	PCOS absent	PCOS present
8-9hrs	89	0
7-8 hrs	386	34
6-7 hrs	504	46
5-6 hrs	203	22
4-5 hrs	57	09

Table 10: Chi-square analysis of Daily exercise.

Daily exercise	PCOS absent	PCOS present
Yes	613	0
No	626	111

Table 11: Chi-square analysis of Plastic bottle usage.

Plastic bottle usage	PCOS absent	PCOS present
No	488	0
Yes	751	111

Table 12: Chi-square analysis of Cosmetic product usage.

Cosmetic product usage	PCOS absent	PCOS present
No	339	0
Minimal Usage	546	74
Moderate usage	152	21
Maximum usage	90	05
Excessive usage	112	11

Table 13: Chi-square analysis of Perfume usage.

Perfume usage	PCOS absent	PCOS present
No	86	0
Rare	485	44
Occasionally	423	36
daily	245	31

Table 14: Chi-square analysis of Air freshener usage.

Air freshener usage	PCOS absent	PCOS present
No	649	0
Rarely	493	97
Daily	96	14

Table 15: Chi-square analysis of Fast-food consumption.

Fast food consumption	PCOS absent	PCOS present
No	63	2
Rarely	30	3
2-3 times a week	232	27
7 times a week	410	36
More than 7 times a week	71	43
Daily	433	0

Table 16: Chi-square analysis values for Clinical parameters.

Name of clinical parameter	Chi-squared value	P value	H0	H1
Irregularity in the menstrual cycle	204.21	0.0004998	Rejected	Accepted
Skin darkening (Acanthosis nigricans)	545.48	0.0004998	Rejected	Accepted
Acne	158.08	0.0004998	Rejected	Accepted
Weight gain	210.86	$<2.2*10^{-16}$	Rejected	Accepted
Hair loss	242.67	$<2.2*10^{-16}$	Rejected	Accepted

Facial Hair Growth	506.52	<2.2*10 $^{-16}$	Rejected	Accepted
Bleeding pattern	460.54	<2.2*10 $^{-16}$	Rejected	Accepted
BMI	40.806	7.187*10 $^{-9}$	Rejected	Accepted

Table 17: Chi-square analysis values for Lifestyle parameters.

Name of Lifestyle parameter	Chi- squared value	P value	H0	H1
Sleep duration	11.282	0.02357	Rejected	Accepted
Daily Exercise	98.61	<2.2*10 $^{-16}$	Rejected	Accepted
Plastic bottle usage	66.774	3.045*10 $^{-16}$	Rejected	Accepted
Cosmetic products usage	46.403	2.031*10 $^{-9}$	Rejected	Accepted
Perfume usage	11.111	0.01114	Rejected	Accepted
Air freshener usage	113.83	<2.2*10 $^{-16}$	Rejected	Accepted
Fast food consumption	174.23	<2.2*10 $^{-16}$	Rejected	Accepted

Table 18: Compilation of Accuracy and sensitivity values for various statistical models.

Statistical Model name	Accuracy value	Sensitivity value
Random Forest analysis	91.48%	0.375
Naïve Bayes analysis	92.22%	0.9090
Support Vector Machine analysis	92.22 %	0.87
Logistic regression analysis	92.96%	0.9000

DISCUSSION

In previous literature, the PCOS prevalence was reported as 21.05% (Aggarwal et al., 2019), the national prevalence according to NIH 1990 criteria was 7.2% and by Rotterdam criteria was 19.6% (Ganie et al., 2019; Bhalerao and Aranha, 2021, 2023). In the present study, following the Rotterdam 2003 criteria, the prevalence estimated is 8.22%. The above literature included the study population of urban cities. Considering the study area of Ahilyanagar city, the population included both rural and urban types. The decrease in incidence rate might be due to less awareness about the disease and misdiagnosis over the complexity of PCOS.

Obesity and overweight nature were observed among 8.1% PCOS females, while the majority of 72.07% PCOS females were lean. Insulin resistance, being an important parameter (Patel 2018) of PCOS etiology, seems to affect 100% PCOS females. This feature was also seen to be occurring among 9.93 of percentage population of non-PCOS females. The occurrence of this clinical feature indicates a high risk of Type 2 Diabetes mellitus among PCOS as well as normal females. The hyperandrogenism markers seem to affect the PCOS population as in 66.66% for acne, 90.99% for weight gain, 100% for Hair loss, and 54.95% for facial hair growth. The normal population also showed the occurrence of acne 52.38%, weight gain 24.53%, hair loss26.96 26.96%, and facial hair 16.14%. The appearance of these characters might indicate the future risk of hormonal imbalance among these females and an indicator to resent the current lifestyle choices.

Irregularity in menstrual cycle has shown a positive association with PCOS among the PCOS females; 59.46% females showed strong irregularity within their menstrual cycle, while 13.88% of the normal female population also exhibited the same trait. Irregularity in the menstrual cycle being a strong clinical parameter for the diagnosis of PCOS, proper assessment is regularly required for early detection of PCOS. The longer the persistence of PCOS tendency and late diagnosis can severely compromise the fertility of females (Azziz, 2018).

Lifestyle parameters, along with metabolic and hormonal imbalance, equally contribute to the development of PCOS etiology (Kshetrimanyum, 2019; Aranha and Bhalerao, 2023). The endocrine modulators in cosmetics, such as Parabens and Phthalates, contribute to increased estrogen levels, which ignites the loop of hormonal imbalance. The same principle applies to the chemicals lodged in perfumes and air fresheners, causing endocrine loop disruption (Kshetrimanyum, 2019). Plastic bottles tend to release Bisphenol A, which is a potent chemical for increasing the risk of Type 2 Diabetes Mellitus among individuals (Siddique et al., 2020). Sleep cycle disruption affects the melatonin levels, in turn spiking up the cortisol levels, inducing stress in our body (Heydari and Ramdass, 2025). Considering the adverse effects of the above-mentioned lifestyle disorders and their association with PCOS gives a chance to reflect on them and rethink our choices. Because dependency on drugs for a lifetime is a difficult way to live. Managing the etiology of PCOS is a more effective way to maintain fertility among reproductive females (Patel, 2018).

The choice of Naïve Bayes' model has proven to be effective for preparing a diagnostic model which can help students to assess their reproductive health and provide them a basic outline to estimate the timings for their gynaecologist's visit. Due to a lack of awareness, hesitation to speak, this syndrome remains undetected. So diagnostic model can always help female students in determining the status of their reproductive health. Earlier detection can help in managing the etiology soon, maintaining the fertility of females. Further research is necessary for the estimation of Biochemical markers of PCOS, leading to clinical diagnosis, which can provide biochemical support for clinical features included in this study. This can add to further refinement in the proposition of the best statistical model for PCOS diagnosis.

CONCLUSION

The present study highlights a lower prevalence of PCOS (8.22%) in Ahilyanagar, as compared to earlier urban-centric studies, potentially due to underdiagnosis and limited awareness, especially in mixed rural-urban populations. A significant proportion of PCOS females were lean (72.07%), emphasizing that obesity is not the sole phenotypic indicator. Insulin resistance was universally present among PCOS participants and also affected nearly 10% of non-PCOS females, indicating a broader metabolic risk across the population. Hyperandrogenic features such as acne, weight gain, hair loss, and facial hair were markedly higher in the PCOS group, though also present in the non-PCOS group, suggesting early signs of hormonal imbalance in the general population. Menstrual irregularities, strongly associated with PCOS, were observed in over half of the affected individuals, reinforcing its diagnostic significance.

Lifestyle and environmental factors, including exposure to endocrine-disrupting chemicals and poor sleep hygiene, emerged as key contributors to PCOS etiology. This underscores the need for preventive strategies centered on lifestyle modification rather than solely relying on long-term pharmacological interventions. The application of the Naïve Bayes diagnostic model has proven effective in the early identification of at-risk individuals, particularly in young, hesitant populations such as students. Early diagnosis is crucial for timely intervention, helping to preserve fertility and improve long-term health outcomes. Future research should focus on integrating biochemical markers to validate and enhance diagnostic accuracy, enabling the development of more refined, evidence-based predictive tools for PCOS.

Acknowledgment

The authors sincerely thank all the participants for their active involvement in the survey and for generously sharing their information. AB and IA extend their heartfelt gratitude to Dr. R. J. Barnabas, Chairman of B.P. Hivale Education Society's Ahmednagar College, Ahilyanagar, for his unwavering support, guidance, and encouragement throughout this study.

REFERENCES

- Aggarwal M, Yadav P, Badhe S, Deolekar P. (2019) A cross sectional study on prevalence of PCOS and risk factors associated with it among medical students. *Indian Journal of Obstetrics and Gynecology Research*;**6(4)**:522–526.
- Aranha, I., & Bhalerao, A. (2023). Treating PCOS: diet, nutrition, medication and lifestyle changes. Advances in, 75.
- Azziz R (2018) Polycystic Ovary Syndrome. *Obstet Gynecol*, **36**: 132:321.
- Bhalerao, A., & Aranha, I. (2021). Polycystic ovarian syndrome (PCOS), a distress of female reproductive health. Shanlax International Journal of Arts Science and Humanities, 8(S1-Feb), 46-53.
- Bhalerao, A., & Aranha, I. (2023). Polycystic Ovarian Syndrome (PCOS): A Lifestyle Upshot and Cluster of Maladies. Quantum Journal of Medical and Health Sciences, 2(1), 50-62.
- Biau, G (2012) Analysis of Random Forest model. The Journal of Machine Learning Research, 13(1), 1063-1095.
- Breiman, L (2001). Random Forests. *Machine Learning* **45**, 5–32.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020) A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189-215.
- Chen Y and Fang S (2018) Potential genetic polymorphisms predicting polycystic ovary syndrome. *Endocrine Connections,* **7**, R187–R195.
- 1Chen, H., Hu, S., Hua, R. *et al.* (2021) Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J. Adv. Signal Process.* , 30.
- Christ, J. P., & Cedars, M. I. (2023). Current Guidelines for Diagnosing PCOS. *Diagnostics (Basel, Switzerland)*, *13*(6), 1113.
- 1Dokras A, Saini S, Gibson-Helm M, Schulkin J, Cooney L,Teede H.(2017) Gaps in knowledge among physicians regarding diagnostic criteria and management of polycystic ovary syndrome. *Fertil Steril*; **107**:1380–1386.e1.
- Ganie M A, Vasudevan V, Wani I A, Baba M S, Arif T and Rashid A (2019) Epidemiology, Pathogenesis, genetics & management of polycystic ovary syndrome in India. *Indian J Med. Res.* **150**, 333-344.
- Heydari, T., & Ramdass, P. V. A. K. (2025) Circadian rhythm disruption and polycystic ovary syndrome: a systematic review and meta-analysis. *AJOG global reports*, *5*(2), 100479.
- Kshetrimayum C, Sharma A, Mishra V.V, Kumar S (2019) Polycystic ovarian syndrome: Environmental/occupational, lifestyle factors; an overview. *Journal of the Turkish-German Gynecological Association*.
- Mouanness, M., Nava, H., Dagher, C., & Merhi, Z. (2022) Contribution of Advanced Glycation End Products to PCOS Key Elements: A Narrative Review. *Nutrients*, *14*(17), 3578.
- Patel S. (2018) Polycystic ovary syndrome (PCOS), an inflammatory, systemic, lifestyle endocrinopathy. J Steroid Biochem Mol Biol., 182:27-36.
- Peng, J., Lee, K., & Ingersoll, G. (2002). An Introduction to Logistic Regression Analysis and Reporting. Journal of Educational Research - J EDUC RES. 96. 3-14.
- Peretz, O., Koren, M., Koren, O (2024) Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. Engineering Applications of Artificial Intelligence, 136(B).
- 2Siddique, S., Zhang, G., Kubwabo, C. (2020) Exposure to bisphenol a and risk of developing type 2 diabetes: A mini review, Emerging Contaminants. 6, 274-282.
- Sperandei S. (2014). Understanding logistic regression analysis. *Biochemia medica*, *24*(1), 12–18.
- Theodoros, E., & Massimiliano, P. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257.