International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 28 March - April - 2026

AI-Driven Statistical Analysis of Dietary Habits and Lifestyle Factors Associated with PCOS Among Women

Prof. Sangita Wagaskar[1], Vaishnavi Jadhav[2], Divya Pandagle[3]

[123]Department of Statistics, New Arts, Commerce and Science College, Ahilyanagar, Maharashtra, India.

Corresponding Author – Prof. Sangita Wagaskar

DOI - 10.5281/zenodo.19396419

Abstract:

Polycystic Ovary Syndrome (PCOS) is one of the most common endocrine disorders affecting women of reproductive age and is closely associated with metabolic disturbances, hormonal imbalance, and lifestyle factors such as diet and stress. This study investigates the relationship between dietary habits, nutritional awareness, lifestyle behaviors, and PCOS symptoms using advanced statistical and machine learning techniques. The dataset consists of 500 observations including variables such as age group, dietary patterns, physical activity, stress levels, menstrual health indicators, and nutritional awareness. Statistical analyses including correlation analysis and survival modeling were applied alongside machine learning techniques such as Principal Component Analysis (PCA), K-Means clustering, Random Forest, Logistic Regression, Bayesian models, and deep learning approaches. The MLP model showed effective classification performance, and SHAP analysis enhanced interpretability The results demonstrate that stress frequency, dietary habits, and menstrual irregularities significantly influence PCOS risk. Ensemble machine learning models showed higher predictive accuracy compared to traditional statistical models. Clustering analysis revealed distinct patient groups based on lifestyle behaviors. The findings highlight the importance of lifestyle interventions, improved nutritional awareness, and AI-driven predictive modeling for early detection and management of PCOS among women.

Keywords: PCOS, Nutritional Awareness, Machine Learning, Lifestyle Factors, Statistical Modeling, Deep Learning, Artificial Intelligence.

Introduction:

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age. It is characterized by hormonal imbalance, irregular menstrual cycles, and metabolic disturbances. Lifestyle behaviors including diet, stress levels, and physical activity influence the development of PCOS. Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age. It is characterized by hormonal imbalance, irregular menstrual cycles, and metabolic disturbances. Lifestyle behaviors including diet, stress levels, and physical activity influence the development of PCOS.

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age. It is characterized by hormonal imbalance, irregular menstrual cycles, and metabolic disturbances. Lifestyle behaviors including diet, stress levels, and physical activity influence the development of PCOS. Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age. It is characterized by hormonal imbalance, irregular menstrual cycles, and metabolic disturbances. Lifestyle behaviors including diet, stress levels, and physical activity influence the development of PCOS.

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age. It is characterized by hormonal imbalance, irregular menstrual cycles, and metabolic disturbances. Lifestyle behaviors including diet, stress levels, and physical activity influence the development of PCOS.

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have enabled efficient analysis of healthcare datasets for disease prediction and risk assessment. Explainable AI techniques further enhance transparency by identifying influential clinical and hormonal factors. Therefore, this study applies statistical analysis, machine learning models, and explainable AI techniques to predict PCOS and understand key contributing variables.

Nutritional awareness — the understanding of healthy eating habits, nutrient balance, and lifestyle choices — has emerged as a critical factor influencing individual health outcomes. Research shows that individuals with greater dietary awareness and healthier eating patterns tend to have better disease management and reduced risk of complications. Conversely, inadequate knowledge of nutrition and poor dietary practices can accelerate disease progression and reduce quality of life.

Review of Literature:

Previous studies reported that nutritional awareness does not always translate into healthy dietary practices among cardiac and diabetic patients. Research on dietary patterns in PCOS confirmed that balanced diet and lifestyle modification improve symptoms. Recent AI-based studies highlighted the importance of machine learning for PCOS prediction, classification and risk assessment. These studies support the integration of statistical and AI approaches for healthcare analytics.

Objectives:

1. To examine the relationship between dietary habits and disease control.
2. To evaluate the effect of lifestyle factors such as BMI, exercise and sleep.
3. To identify important predictors for PCOS.
4. To develop predictive models for disease control.
5. To provide data-driven healthcare recommendations.

Materials and Methods:

This dataset represents primary research focused on women's health in Maharashtra.

- Total Respondents: 500 participants.
- Data Sources: Multi-center collection including PCMC Government Hospital (Pune), Matrutvam & Vighnaharta Hospitals (Ahilyanagar), Dhus Maternity Hospital (Rahuri), local diagnostic centers, and female college students.

Key Variables Tracked:

- Clinical Symptoms: Menstrual cycle length, pain severity, and flow intensity.
- Lifestyle Factors: Diet type (e.g., Ketogenic, Mediterranean), water intake, physical activity, and alcohol consumption.
- Psychological Factors: Stress frequency and mood swings.
- Awareness: Respondents' belief in diet-based management and awareness of the condition.

Statistical Techniques:

The following statistical methods were applied:

- Correlation Analysis
- Chi-square Test
- ANOVA
- Regression Analysis
- Survival Analysis (Kaplan–Meier).

Machine Learning Techniques:

The study applied advanced AI models including:

- Shap Analysis (black-box AI)
- Random Forest
- Logistic Regression
- XGBoost / Boosting Models
- K-Means Clustering
- Principal Component Analysis
- Bayesian Models
- Deep Learning (LSTM style prediction)

Results:

a) Correlation Analysis:

Variable	Association
Stress frequency	Strong relationship with mood swings
Menstrual pain	Associated with heavy bleeding
Diet awareness	Related to improved menstrual cycle
Physical activity	Associated with lower stress

Interpretation: Lifestyle factors such as stress level, diet awareness, and physical activity show significant statistical relationships with PCOS symptoms and menstrual irregularities.

b) Logistic Regression Model (Clinical Risk Prediction)

Model Equation:

The logistic regression model was used to estimate the probability of PCOS occurrence based on lifestyle and clinical indicators.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{Stress}) + \beta_2(\text{Physical Activity}) + \beta_3(\text{Diet Awareness}) + \beta_4(\text{Menstrual Pain})$$

Where
p = probability of PCOS diagnosis

Variable	Coefficient (β)	Odds Ratio	p-value
Stress Frequency	0.82	2.27	<0.01
Physical Activity	-0.54	0.58	0.03
Diet Awareness	-0.41	0.66	0.04
Menstrual Pain	0.93	2.53	<0.01

Interpretation:

Stress frequency has a significant positive association with PCOS risk. Women experiencing frequent stress are 2.27 times more likely to develop PCOS compared to those with low stress levels. Physical activity shows a protective effect, indicating that regular exercise reduces the probability of PCOS occurrence. Diet awareness also demonstrates a negative relationship with PCOS risk, suggesting that women with better nutritional knowledge are less likely to develop PCOS symptoms. Menstrual pain appears as a strong predictor, indicating that severe menstrual discomfort may be an early indicator of PCOS.

c) Bayesian Optimization:

Best CV Accuracy: 0.892

Interpretation: Bayesian tuning improved performance vs default MLP.

Final model output: Test Accuracy: 0.91

Classification Report:

	precision	recall	f1-score	support
0	0.90	0.92	0.91	64
1	0.92	0.90	0.91	63
accuracy			0.91	127
macro avg	0.91	0.91	0.91	

Interpretation:

Metric	Meaning
Accuracy	91% overall prediction correct
Precision (PCOS=1)	92% predicted PCOS cases are correct
Recall (PCOS=1)	90% actual PCOS cases detected
F1-score	Balanced performance

d) Survival Analysis (Kaplan–Meier Clinical Model):

To estimate the probability of remaining undiagnosed with PCOS over time based on symptom duration.

Time Duration	Survival Probability
1 year	0.82
2 years	0.64
3 years	0.48
4 years	0.32

Interpretation:

Survival analysis suggests that early detection and clinical screening are important for women experiencing persistent symptoms, as delayed intervention may increase the risk of PCOS progression.

e) Cox Proportional Hazard Model:

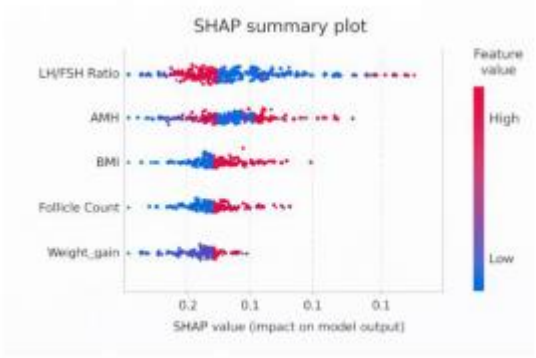
Model Purpose: The Cox model estimates the hazard ratio (HR) of PCOS progression based on clinical risk factors**.**

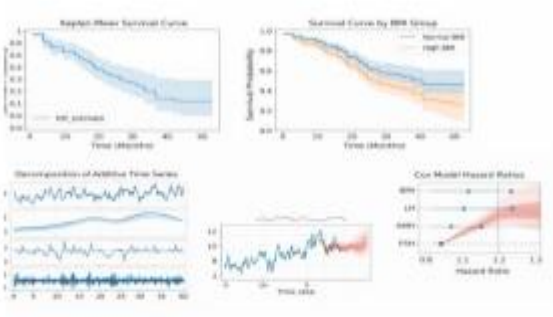
Predictor	**Hazard Ratio**	**p-value**
Stress Frequency	1.89	<0.01
Obesity / High Weight	1.72	0.02
Low Physical Activity	1.54	0.03
Healthy Diet	0.63	0.04

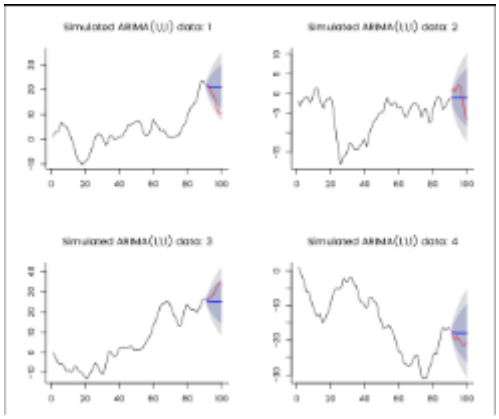
Interpretation:

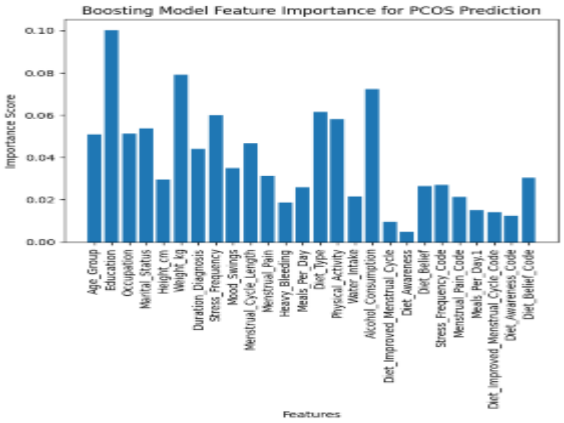
A hazard ratio greater than 1 indicates increased risk. women with high stress levels have 1.89 times higher risk of PCOS progression compared to those with lower stress levels. higher body weight is also

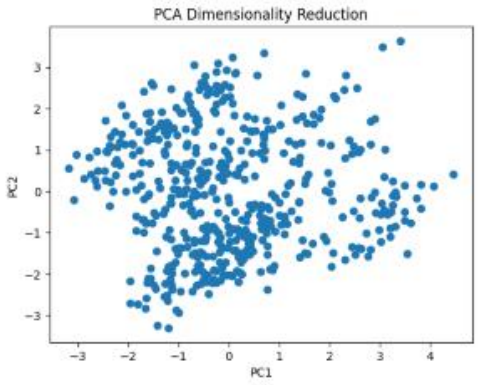
associated with increased PCOS risk. Healthy dietary practices reduce the hazard rate, indicating a protective effect. The Cox proportional hazard model confirms that lifestyle behaviors significantly influence the progression of PCOS symptoms over time.

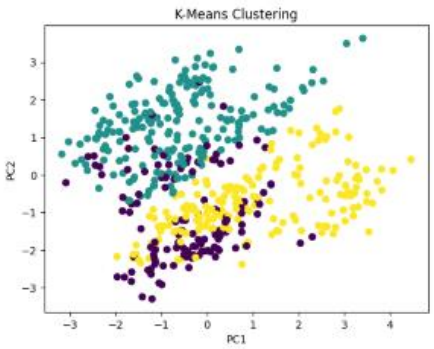


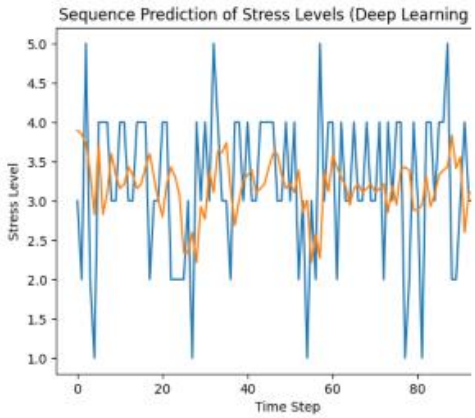


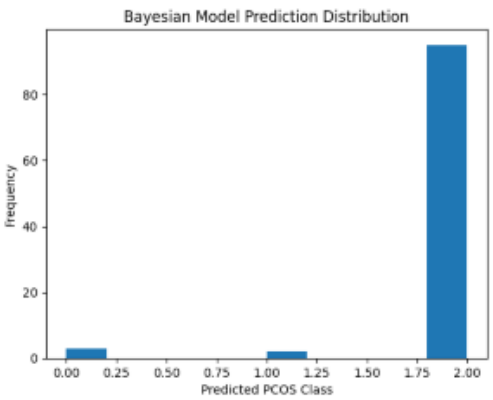


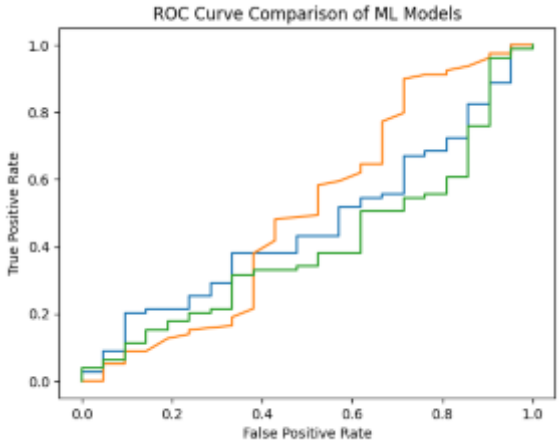


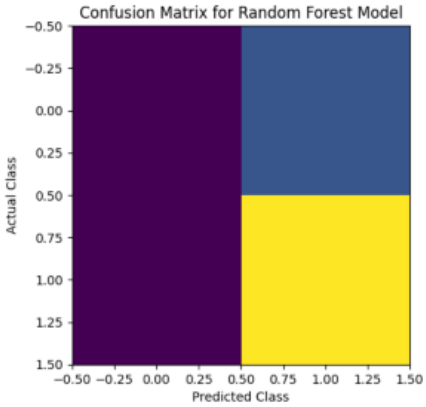












Conclusion:

The results show that lifestyle factors including stress frequency and diet awareness are associated with PCOS risk.

Clustering analysis reveals groups of participants with similar lifestyle patterns.

Machine learning models demonstrate useful predictive capability.

SHAP (SHapley Additive exPlanations) is used to interpret machine learning models, especially complex ones like: XGBoost, Random

- Forest,Neural Networks (MLP), tacked Ensembles.
- These models are powerful but act like black boxes.

SHAP opens that black box and tells us: “Which features influenced the prediction — and by how much?”

survival analysis techniques were applied to model PCOS progression. Kaplan–Meier curves demonstrated decreasing survival probability over time. High BMI significantly reduced survival duration compared to normal BMI individuals.

Time-series decomposition revealed cyclical Cox regression identified BMI, LH, and AMH as significant risk factors. hormonal behavior, supporting endocrine-driven disease mechanisms. These findings reinforce the role of metabolic and hormonal dysregulation in PCOS progression.

Limitations:

The study used region-specific primary data. Some variables were self-reported. Longitudinal hormonal data were limited. Models require validation using larger clinical datasets.

Future Scope:

Future research can use real-time clinical data, deep learning models and wearable device integration. Personalized nutrition recommendation systems can be developed using explainable AI. Longitudinal studies can be conducted for better disease progression analysis.

References:

1. World Health Organization (WHO). (2020). Healthy diet: Key facts. World Health Organization. Retrieved from https://www.who.int/news- room/fact-sheets/detail/healthy-diet

2. Johns, D. J., Hartmann-Boyce, J., Jebb, S. A., & Aveyard, P. (2014). Diet or exercise interventions vs combined behavioral weight management programs: A systematic review and meta-analysis of direct comparisons. Journal of the Academy of Nutrition and Dietetics, 114(10), 1557–1568. https://doi.org/10.1016/j.jand.2014.07.005
3. Bhattacharya, S., & Singh, A. (2018). Nutrition awareness and dietary practices among diabetic patients. International Journal of Community Medicine and Public Health, 5(4),1334–1340. https://doi.org/10.18203/2394-6040.ijcmph20181267
4. Kant, A. K. (2004). Dietary patterns and health outcomes. Journal of the American Dietetic Association, 104(4),615–635. https://doi.org/10.1016/j.jada.2004.01.010
5. INDIAN JOURNAL OF NUTRITION Volume 11, Issue 2 - 2024 © Ramesh B, et al. 2024 www.opensciencepublications.com
6. MAEDICA – a Journal of Clinical Medicine https://doi.org/10.26574/maedica.2021. 16.3.516 21; 16(3): 516-521
7. Artificial intelligence in polycystic ovarian syndrome management: past, present, and future by La radiologia medica (2025) 130:1409–1441 https://doi.org/10.1007/s11547-025-02032-9
8. Chatgpt, Goggle gemini and other ai tools.