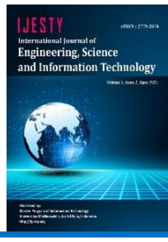
International Journal of Engineering, Science and Information Technology
Volume 5 No. 2 (2025) pp. 491-502
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i2.1364
Research Paper, Short Communication, Review, Technical Paper



A Comparative Study of Data Mining Models using Essential Metrics in the Prediction of the Relation Between Polycystic Ovary Syndrome and Postpartum Depression in Women

Arya P. Pillai*, N.V. Chinnasamy

Karpagam Academy of Higher Education, Coimbatore, India

**Corresponding author Email: aryappillai26@gmail.com*

The manuscript was received on 18 November 2024, revised on 28 December 2024, and accepted on 28 March 2025, date of publication 9 May 2025

Abstract

Around the world today, personal health care is an unavoidable task to be done in human life. The vest emergence of medical science and medical technology is accelerating day by day. In many paradigms, information technology plays a vital role in comparing the past evidence with the present in the medical field and as a result, predictions will be outlined. Data mining and Data mining algorithms in medical care have a major role in improving personal care and in the overall healthcare system. Women with PCOS are more reasonably experience several pregnancy issues, including diabetes mellitus, hypertension, anxiety and mood swings, which may sometimes lead to Postpartum Depression. This paper evaluates a few parameters related to health care and predicts the relationship between PCOS and PPD in women based on data mining approaches.

Keywords: *PCOS, PPD, LR, Naïve Bayes, SVM.*

1. Introduction

Health healthcare Industry is a field that is carried out by using different Data mining techniques. In today's societies, the involvement of women in family care giving and in professional life is outstanding. The WHO states that women's personal and health care has become a top priority nowadays. As a result of this, it is challenging for women to keep a work–life balance; subsequently no time to think about herself.

Polycystic Ovarian Syndrome, on the other hand, is a common hormonal disorder which occurs in women during their reproductive cycle. It is a hormonal disorder which affects the quality of life in women. The earlier identification and diagnosis of PCOS in women is essential, especially during the pregnancy period, as this condition may sometimes lead to Postpartum Depression (PPD) in their postnatal period [15]. Apart from other chemical and clinical changes in women, there are many other external factors which lead a woman to PCOS and finally to the PPD condition. Both conditions (PCOS and PPD) are acute illnesses as they affect the quality of life. Data mining plays an important role in the medical industry as it transforms a vast amount of data into valuable and accurate information [8][10].

According to a study conducted by Metropolis Healthcare recently, one out of five women is found to have PCOS, with the numbers rising every year [1]. Symptoms of PCOS included irregular periods, polycystic ovaries, obesity, Hair growth, Depression and anxietyetc [9][11]. Untreated PCOS in women may lead to serious health outcomes such as high Blood pressure level, Ischemic heart disease, Infertility, endometrial cancer and PPD. Moreover, the women with PCOS are also at risk of other disorders such as mental depression, anxiety, panic disorders, phobias and PPD, which may affect the quality of life [2][4].

Health healthcare Industry is a field in which research is being carried out by using different data mining techniques. Even in the drug industry, Data mining techniques help in identifying those patients where the drug is more effective or where the drug has any side effects. Data mining in health care aims to resolve real-life health problems in the diagnosis and treatment of disease [6][14].

This paper makes an effort to predict the relationship between PCOS & PPD in women based on three models, such as the LR (Logistic Regression) model, NB (Naïve Bayes) model and SVM (Support Vector Machine) model. For the implementation of data, Python, which is a powerful data analysis and manipulation platform, is used in this paper.

II Section includes the related works, and III Section elucidates the materials and models used. In the IV Section, different data mining methods, such as data mining algorithms, are described. V Section focused on Results and Discussions. Eventually, the VI Section deals with the conclusion part of the research article.

2. Literature Review

In 2020, Vaidehi Thakre and Shreyas Vedpathak suggested a methodology for the early identification and forecasting of PCOS treatment. In this paper, 30 significant features were identified using CHI CHI-SQUARE method, and five distinct machine learning models, such as K-Nearest Neighbour, Gaussian NB, LR, SVM and RF models, have been used. The results of the analysis were compared, and it was noticed that the RF model attained the best precision in comparison to other classifiers [3].
In 2021, Preeti Chauhan and Pooja Patil proposed a model for the early detection of PCOS using algorithms. Data dataset is obtained from a survey conducted among women and processed using Google Colab. In this scholarly paper, algorithms such as KNN, Decision Tree, SVM, and NB are applied for the classification of PCOS, and the decision tree classifier was found to be the efficient classifier for the prediction of PCOS. [12]
In 2021, Aggarwal Shivani introduced an article for the study of PCOS prediction using LR, SVM, Decision tree and KNN algorithm. The algorithm works on a data record of 541 records and 41 attributes. After performing feature selection,12 attributes were selected for diagnosing PCOS. [5].
In 2022, a clinical decision tree support system was proposed by S. Sreejith, H. Khanna Nehamiah for diagnosing PCOS. The data set is processed Red Deer algorithm, which provides a decision support system with an accuracy of 89.81 % [13].
In 2021, Pooja Rani, Rajneesh Kumar introduced a Hybrid decision support system for the early finding of cardiac disorders in patients based on the clinical parameters. Recursive feature elimination and the Genetic algorithm are combined for the selection of relevant features from the data set. Adaboost classifier, RF, NB, and SVM models were used, and it has been found that RF has given the most accurate result [7].

2.1. Information Source

To explore and predict if there is a connection between PCOS and PPD in women, a clinical dataset was used. The data in this set were carefully put together at various health centres and Mother-Child hospitals in the Ernakulam district, India. There are health records for a total of 541 female patients, each one containing information about 27 specific clinical and behavioural aspects. We selected the patient features by considering their importance and likely influence on predicting PCOS and PPD. Thanks to the expertise of gynaecologists and psychiatrists, the list includes both physical and environmental factors that may impact hormonal disorders or someone's mental health. Figure 1 covers a broad range of topics by using numerical data and non-numerical categories such as lifestyle habits, health of the heart and uterus, hormone levels and social-psychological factors.
The description and data types of every selected attribute are included in Table 1. A suitable data structure was selected because we wanted to model the data and also explore the data through EDA.

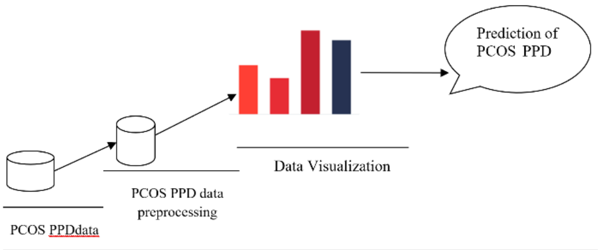


Fig 1. Data modelling of PCOS PPD with relevant steps

Table 1. Attributes with description

Attributes	Description	Information Attribute
Age	Age of Respondent(Majority of respondents are above 20 years)	Numeric
BMI level	Body Mass Index (Maximum level:25)	Numeric
Pulse rate	Pulse rate of respondent(Maximum level should be below 100)	Numeric
HB count	Haemoglobin level (should range between 10 -12.5)	Numeric
Marriage status	Whether the Respondent is married or not	Non-Numeric
Pregnancy	Whether the respondent has completed pregnancy	Non-Numeric
No. of abortions (if any)	Count of abortions of Respondent (If any)	Numeric
FSH	Follicle Stimulating Hormone value of Respondent	Numeric
LH	Luteinizing Hormone value of the Respondent	Numeric
FSH/LH ratio	The ratio of FSH/LH of the Respondent (Maximum level is 2)	Numeric
Waist: Hip Ratio	The Waist: Hip ratio of Respondent (Maximum level is 0.85)	Numeric
TSH	Thyroid Stimulating Hormone value (Maximum level is 5)	Numeric
Weight Gain	Is the respondent's weight increasing?	Non-Numeric
Hair Growth	Whether the Respondent has Hair growth?	Non-Numeric
Skin Darkening	Whether the Respondent's Skin is darkening or not	Non-Numeric
Hair Loss	Whether the Respondent has Hair fall?	Non-Numeric
Pimples	Does Whether respondent have pimples?	Non-Numeric
Fast Food	Is the respondent taking fast food regularly?	Non-Numeric
Regular Exercise	Is the respondent taking regular exercise?	Non-Numeric
BP_Systolic(mm/Hg)	Respondent's blood Pressure (Systolic value, Normal value is 120)	Numeric
BP_Diastolic(mm/Hg)	Respondent's blood Pressure (Diastolic value, Normal value is 80)	Numeric
Follicle no.	Respondent's Follicle size(Maximum size should not be more than 10 mm)	Numeric

Endometrium	Respondent's Endometrium thickness(Normal range is 11)	Numeric
Mood Swings	Whether respondents have mood swings in pre pre-pregnancy or post-pregnancy period?	Non-Numeric
Periods	Does the respondent have regular periods?	Non-Numeric
Postpartum	Whether the Respondent experience postpartum in their delivery period?	Non-Numeric
Duration of Periods	How long do the periods last for the Respondent	Numeric

2.2. Information Preprocessing

Data preprocessing plays a key role in building a machine learning model, and it is crucial for health data since there are often errors in it. The data from the hospitals was cleaned using a multi-step process to confirm the quality and dependability of the data used for training the models.

2.2.1. Data Cleaning

The dataset initially exhibited several inconsistencies, including:

1. Missing values in critical clinical metrics like hormone levels.
2. Redundant features due to overlapping medical terminologies.
3. Inconsistent formats in categorical responses (e.g., "yes", "Yes", "YES" as separate entries).

To resolve these:

1. Missing numerical values were imputed using mean or median imputation strategies, depending on the distribution.
2. Missing categorical values were filled using mode imputation.
3. Inconsistent categorical entries were standardised to binary form ("Yes" = 1, "No" = 0).

2.2.2. Encoding of Categorical Data

To convert all the non-numeric features to integers, ordinal encoding was implemented, as machine learning models require numbers. Ordinal encoding gives each category of a feature a unique integer value. For instance:

Table 2. Categorical Data

Original Value	Encoded Value
Yes	1
No	0

The encoding took place using the sklearn. Preprocessing library in Python. Whether you suffer from PCOS or PMDD, you can use the Ordinal Encoder to effectively and smoothly handle your data on marriage, pregnancy, exercise and mood changes.

2.2.3. Normalisation and Feature Scaling

To prevent domination of features with larger scales (e.g., systolic BP vs. waist-hip ratio), **Min-Max Scaling** was applied to normalise numerical attributes within the range [0,1]. As a result, the model learns faster, and every feature deserves the same consideration.

2.2.4. Feature Selection

Although 27 different features were selected, only a few play a major role in how the models perform their predictions. Because of this, CFS and RFE were used to identify the most important predictors for developing PCOS and PPD. The FSH/LH ratio, BMI, TSH, mood swings, history of pregnancy and exercise patterns were significant when compared with the target variables.

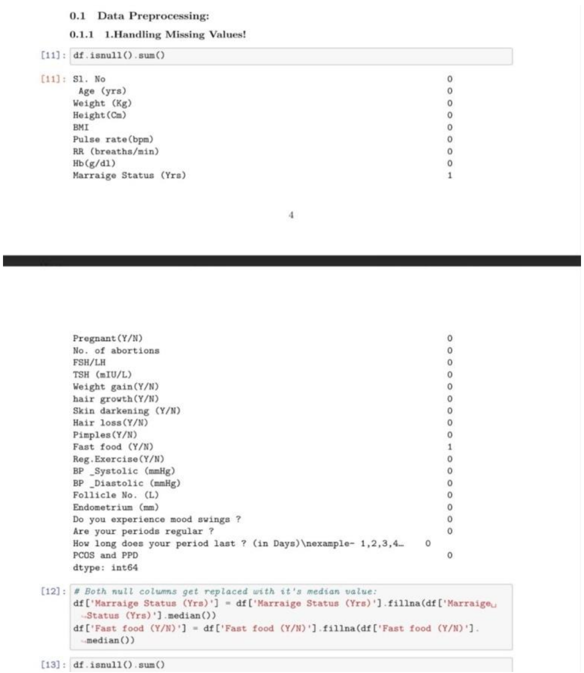


Fig 2. Screen capture of Python code for handling missing values

```
[14]: try:
          df = df.drop(columns=['Sl. No'])
          print("Column 'Sl. No' dropped successfully.")
      except KeyError:
          print("Column 'Sl. No' not found in DataFrame.")

      Column 'Sl. No' dropped successfully.

[15]: print(df.columns)

      Index([' Age (yrs)', 'Weight (Kg)', 'Height(Cm) ', 'BMI', 'Pulse rate(bpm) ',
             'RR (breaths/min)', 'Hb(g/dl)', 'Marraige Status (Yrs)',
             'Pregnant(Y/N)', 'No. of aborptions', 'FSH/LH', 'TSH (mIU/L)',
             'Weight gain(Y/N)', 'hair growth(Y/N)', 'Skin darkening (Y/N)',
             'Hair loss(Y/N)', 'Pimples(Y/N)', 'Fast food (Y/N)',
             'Reg.Exercise(Y/N)', 'BP _Systolic (mmHg)', 'BP _Diastolic (mmHg)',
             'Follicle No. (L)', 'Endometrium (mm)',
             'Do you experience mood swings ?', 'Are your periods regular ?',
             'How long does your period last ? (in Days)\nexample- 1,2,3,4…',
             'PCOS and PPD'],
            dtype='object')

[16]: df.duplicated().sum()

[16]: 0

[17]: df.describe()
```

Fig 3. Screen capture of Python code for data preprocessing

Sampling technique is performed on the dataset, which separates features and target variables as shown in Figure 3.

2.3. Data Visualisation

Data visualisation is a crucial component of exploratory data analysis (EDA), providing a graphical interface to interpret complex datasets and uncover hidden patterns, trends, and correlations. In this study, various ways of displaying data were used to learn about the connections among key factors affecting PCOS and PPD. Drawing various graphs makes it easier for researchers and clinicians to discover the patterns of variables and notice any shared symptoms. Line graphs were created to observe and monitor BMI, haemoglobin levels and hormone fluctuations related to age. Through scatter plots, it was possible to see that FSH/LH might connect to mood swings, while symptoms experienced after childbirth might involve the thyroid. Distribution graphs (for instance, histograms and KDE plots) illustrated the different values of pulse rate, systolic BP and follicle size. Additionally, Pearson correlation matrices were used to make heatmaps that showed how closely related variables are and in which direction the relationship exists. Illustrations similar to those provided in Figure 4 made it possible to select important features and build a useful model for predicting PCOS and PPD among the sample.

Doing Sampling techniques due to imbalanced dataset I have!

```
[38]: # Separate features and target
      X = df.drop(columns=['PCOS and PPD'])
      y = df['PCOS and PPD']

      # Oversample the minority class (1's) to 250 rows
      oversample = RandomOverSampler(sampling_strategy={1: 250})
      X_over, y_over = oversample.fit_resample(X, y)

      # Combine the oversampled minority class with the majority class
      df_over = pd.concat([X_over, y_over], axis=1)

      # Now undersample the majority class (0's) to 275 rows
      undersample = RandomUnderSampler(sampling_strategy={0: 275, 1: 250})
      X_final, y_final = undersample.fit_resample(df_over.drop(columns=['PCOS and
       PPD']), df_over['PCOS and PPD'])

      # Combine the undersampled features and target
      df_final = pd.concat([X_final, y_final], axis=1)
      df_final.head()
```

Fig 4. Screenshot of Python Code ff Sampling Technique

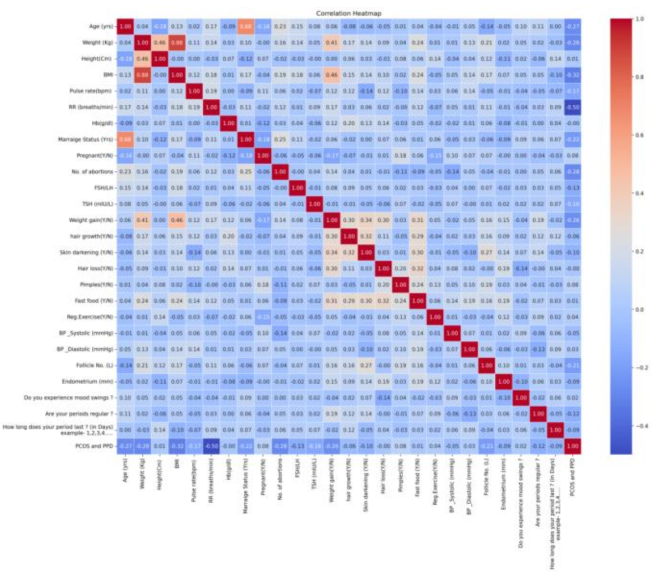


Fig 5. Correlation Heat Map of Variables.

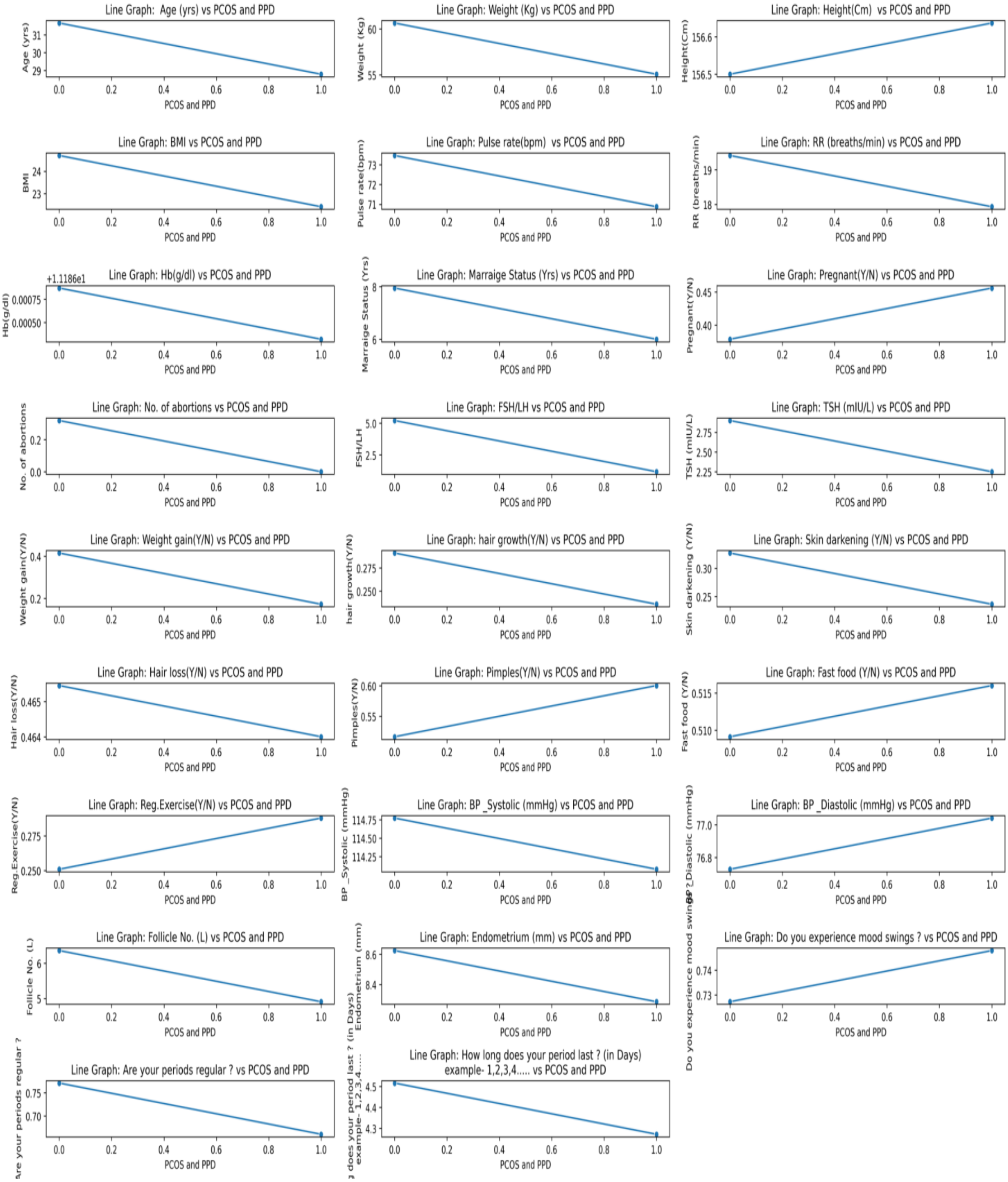


Fig 6. Line Graph of Variables

Fig 7. Scatter Plot of Variables

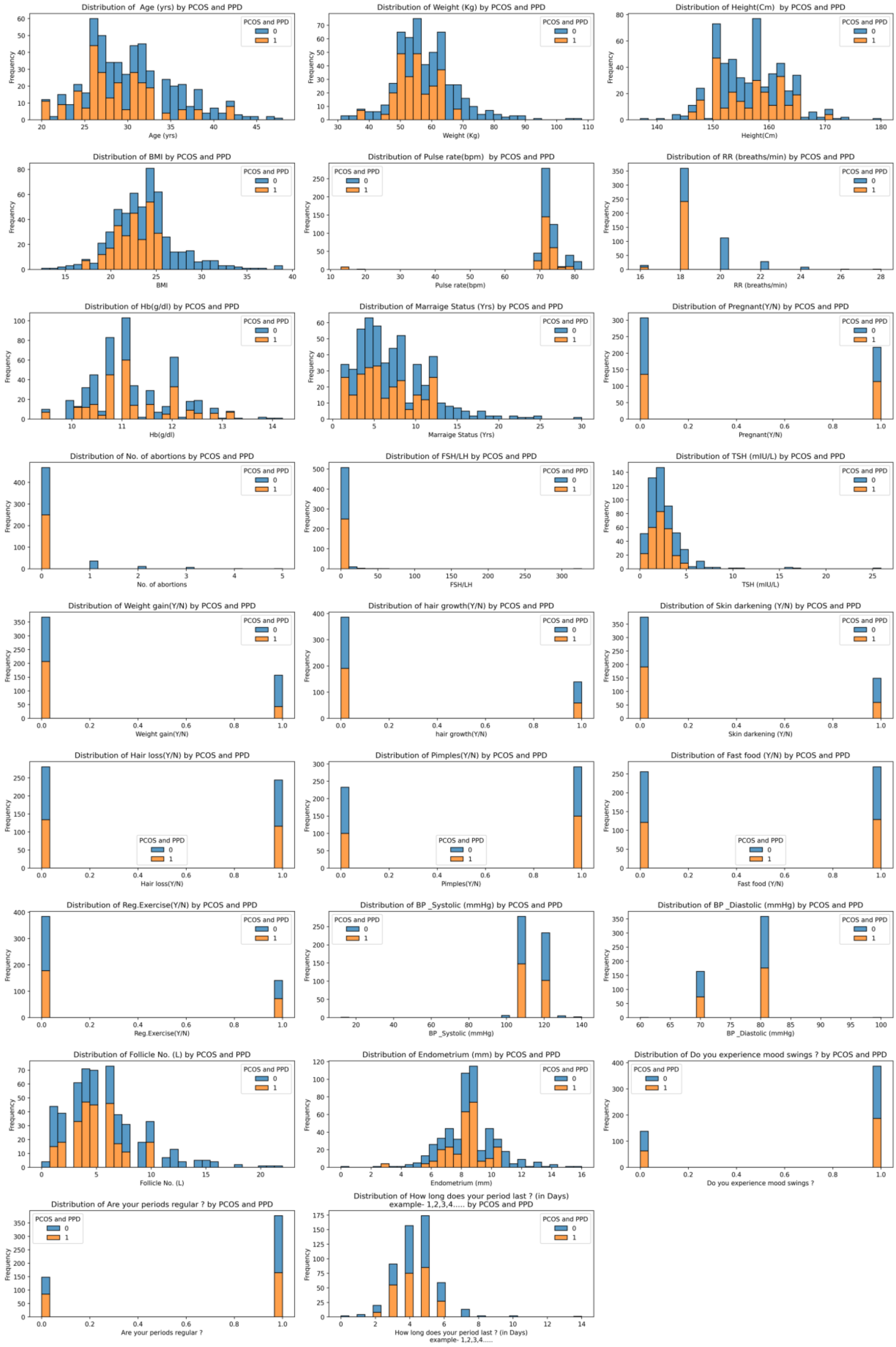


Fig 8. Distribution Graph of Variables

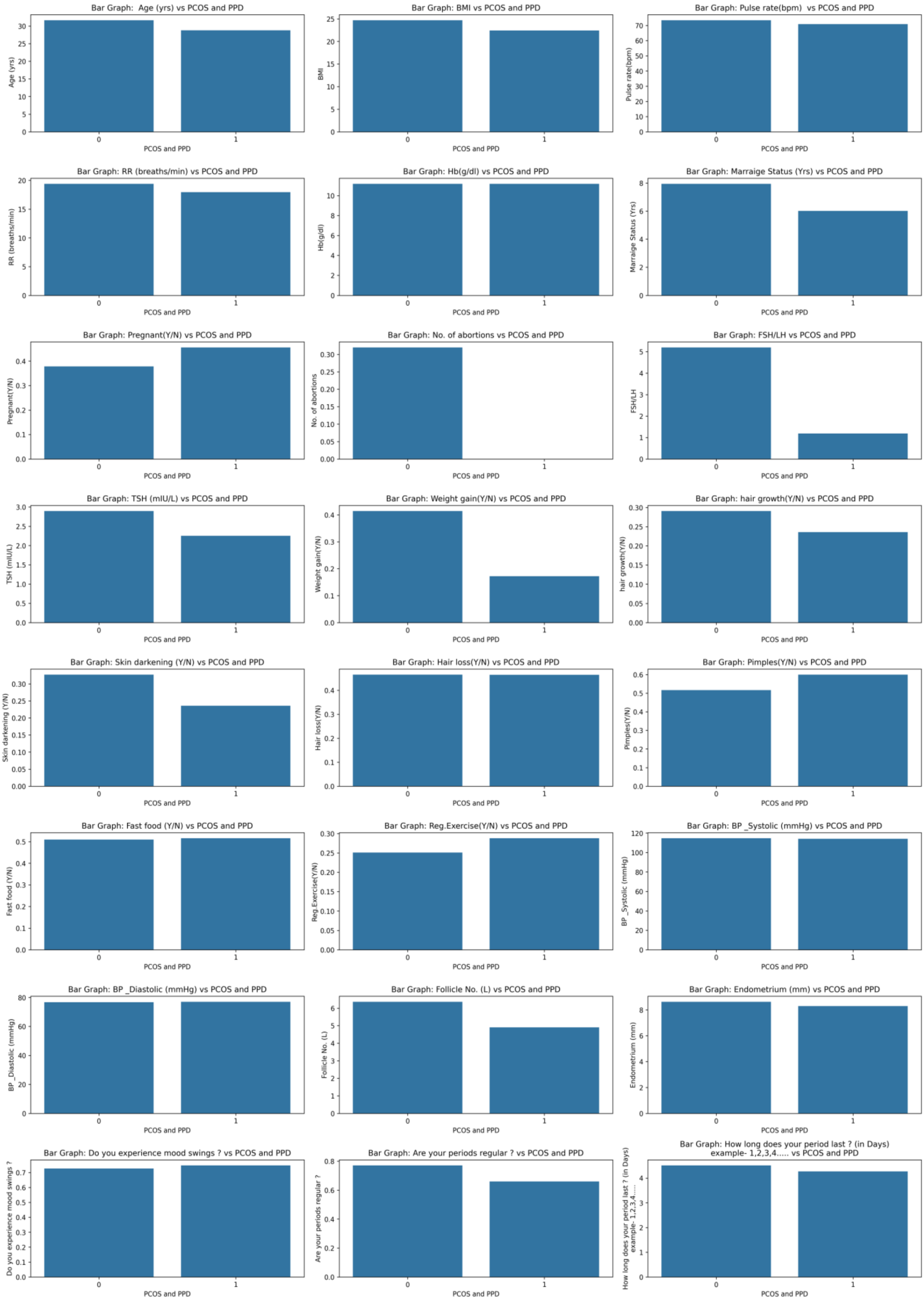


Fig 9. Barplot of Variables

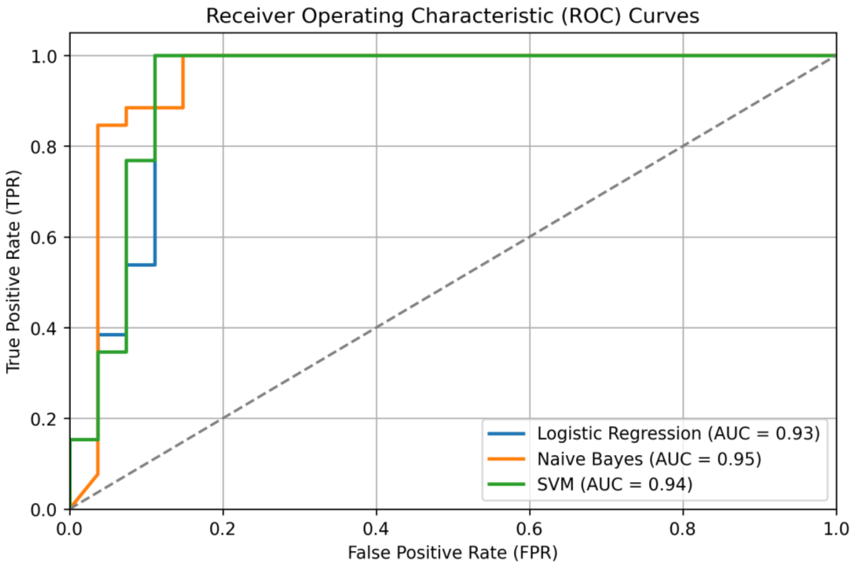


Fig 10. AUC ROC curve of different models

3. Methods

To investigate the association between Polycystic Ovary Syndrome (PCOS) and Postpartum Depression (PPD) among women, we employed three supervised machine learning models**:** Logistic Regression (LR), Naïve Bayes (NB), and Support Vector Machine (SVM). Each model was constructed using Python and aided by scikit-learn, pandas and matplotlib libraries. The architecture used for analysing PCOS-PPD data is depicted in Figure 10. Below, we outline each model and discuss its mathematics, as well as how to implement it.

3.1. Logistic Regression (LR) Model

Logistic Regression is a widely used classification algorithm that predicts the probability of a binary outcome based on one or more independent variables. In this study, we applied Binary Logistic Regression to predict whether PCOS and PPD are *RELATED* (class = 1) or *NOT RELATED* (class = 0).

The Logistic Regression model uses the logistic function, also called the sigmoid function, to map any real-valued number to a value between 0 and 1. The sigmoid function is defined as:

$$\pi(x) = \frac{1}{1+e^{-z}} \quad \text{.........(1)}$$

where:

- $\pi(x)$ is the predicted probability that the output belongs to the "RELATED" class
- e is Euler's number (natural log base)
- $z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$, a linear combination of input features

The probability of the binary outcome can be expressed as:

$$P(Y = 1 \mid x) = \pi(x) = \frac{1}{1+e^{-z}} \quad \text{.........(2)}$$

The model is trained by setting the weights w_i Using maximum likelihood estimation to achieve a minimum loss from the binary cross-entropy function. Logistic regression is the best option here because it is easy to interpret and use.

3.2. Naïve Bayes (NB) Model

Naïve Bayes operates according to Bayes' Theorem, under the idea that different features are unrelated when the class is identified. Although NB is straightforward, it performs well on big datasets and is suitable for both kinds of data.

The posterior probability in Naïve Bayes is given by:

$$P(Y \mid X) = \frac{P(X \mid Y).P(Y)}{P(X)} \quad \text{.........(3)}$$

where:

- $P(Y \mid X)$ is the posterior probability of class Y given predictors X.
- $P(X \mid Y)$ is the likelihood
- $P(Y)$ is the prior probability of class
- $P(X)$ is the evidence (acts as a normalising constant)

In the context of PCOS and PPD, this model computes:

$$P(RELATED \mid features) = \frac{P(features \mid RELATED).P(RELATED)}{P(features)} \quad \text{.........(4)}$$

Let the classifier be defined as:

$$P(PCOS\ and\ PPD \mid features) = \frac{A.B}{C+D} \quad \text{.........(5)}$$

Where:

- $A = P(+features \mid RELATED)$
- $B = P(RELATED)$
- $C = P(+features \mid RELATED) \cdot P(RELATED)$
- $D = P(+features \mid NOT\ RELATED) \cdot P(NOT\ RELATED)$

This model classifies each instance based on the maximum a posteriori (MAP) estimate, assigning the class with the highest probability. Using the NB model, we can successfully process big data that includes many different features.

3.3. Support Vector Machine (SVM) Model

Support Vector Machine is effective for performing classification and regression on supervised data. The purpose of this study is to use SVM to separate the classes of PCOS and PPD by constructing a hyperplane that guarantees maximum accuracy.
A hyperplane in an n-dimensional feature space is represented as:

$$V \cdot xi + c = 0 \quad \text{(6)}$$

Where:

- $V = \{v1, v2, \ldots, vn\}$ *is the weight vector*
- x_i *is the feature vector of the ith instance*
- c *is the bias term*

The classification constraints for the two classes are defined as:

$$V \cdot x_i + c \geq 1, for\ y_i = +1(RELATED) \quad \text{(7)}$$

$$V \cdot x_i + c \leq -1, for\ y_i = -1(NOT\ RELATED) \quad \text{(8)}$$

The hyperplane that provides the greatest margin between the classes is called the optimal hyperplane. The margin is measured as the perpendicular distance from the hyperplane to the closest points in each class, which are called support vectors. The problem is solved by using Quadratic Programming (QP).
SVM works best in high dimensions and can prevent overfitting, except when the data cannot be separated by a line (in which case a kernel trick is applied).

3.4. Model Evaluation Metrics

Each model is evaluated using standard classification metrics such as:

- Accuracy: Percentage of correctly classified instances
- Precision: True positives / (True positives + False positives)
- Recall (Sensitivity): True positives / (True positives + False negatives)
- F1-Score: Harmonic mean of precision and recall
- ROC-AUC Score: Area under the Receiver Operating Characteristic curve

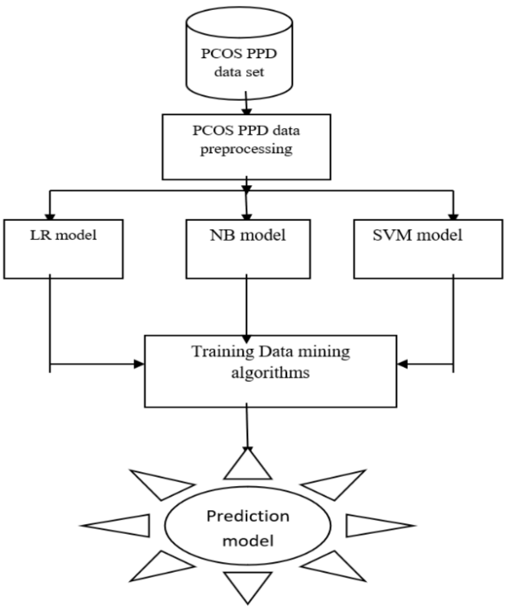


Fig 11. PCOS PPD architecture diagram

4. Results and Discussion

The relationship between PCOS and PPD was studied using health records of 541 women who attended health centres in Kerala, India. It is important to find such relationships early when patients are being treated. To find out, Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM) were used to determine if a subject with PCOS will develop PPD (RELATED) or not (NOT RELATED). Evaluating the model's effectiveness involved measuring Accuracy, F1 Score and AUC-ROC Curve.

Table 2. Terms and description

Terms	Description
R	PCOS and PPD are RELATED
NR	PCOS and PPD are NOT RELATED
ANR	ACTUALLY, NOT RELATED but classified as RELATED
AR	RELATED but classified as NOT RELATED

4.1. Performance metrics: F1 Score and Accuracy of models

Classification algorithms in predictive modelling should be tested to measure their performance and see if they are effective and reliable. In this research, two main evaluation techniques, Accuracy and F1 Score, were applied to test if the predictive models (Logistic Regression, Naïve Bayes and SVM) were able to connect Polycystic Ovary Syndrome (PCOS) and Postpartum Depression (PPD). The idea of accuracy is to check if more negative and positive results are predicted instead of the total number of observations in the data set. Here is the formula for computing accuracy:

$$Accuracy = \frac{R + NR}{R + NR + AR + ANR} \qquad \text{(9)}$$

Where:

- **R:** Instances correctly predicted as RELATED
- **NR:** Instances correctly predicted as NOT RELATED
- **AR:** RELATED, but predicted as NOT RELATED (false negative)
- **ANR:** NOT RELATED, but predicted as RELATED (false positive)

However, Accuracy alone can be misleading, especially in datasets with class imbalance. Therefore, the F1 Score is also used as a complementary metric. The F1 Score measures a model’s performance by equally weighing false positives and false negatives and blending Precision and Recall.

The evaluation results for Logistic Regression (LR), Naïve Bayes (NB), and Support Vector Machine (SVM) models were tabulated, as shown in Table 3. Among these, the LR model achieved the highest F1 Score (0.945) and accuracy (0.943), indicating superior performance in predicting true associations between PCOS and PPD. In addition, Figures 5 to 9 illustrate how differences in certain features affect the predicted outcome.

4.2. AUC ROC curve between different models

The AUC-ROC (Area Under the Receiver Operating Characteristic) curve is a widely adopted metric used to evaluate the performance of binary classification models. It works best for medical predictions, for example, finding out whether PCOS and PPD are related, since this kind of prediction can have vital consequences. A ROC curve uses graphs to represent how useful a classifier is for discrimination, as the threshold is increased. The PTR is displayed on the Y-axis, and the PFR is displayed on the X-axis. According to the definition, these rates consist of the following values:

PTR $=\frac{R}{R+AR}$

PFR $=\frac{ANR}{ANR+NR}$

Figure 10 describes the AUC-ROC curve of the LR, N-band SVM models.

Table 3. Models with F1 Score, Accuracy and R2_Score

Model	**F1 Score**	**Accuracy**
LR	0.945	0.943
NB	0.892	0.886
SVM	0.641	0.642

4.3. Comparative Analysis of Predictive Models for PCOS–PPD Association

Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder that affects approximately 18% of women during their reproductive years. The signs include too much or too little of certain hormones, inconsistent menstruation and problems with metabolism, which often influence a woman’s well-being. The connection between PCOS and Postpartum Depression (PPD) has been suggested by recent studies. Knowing about the link is necessary, mainly for married women, as it allows early diagnosis and quick action by medical workers. To complete this study, both data mining and machine learning were used to classify information on the link between PCOS and PPD using records from 541 patients in Kerala hospitals. This research made use of Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM) algorithms to look into the relationship. Evaluation of the performance was done by considering these two important parameters: Accuracy and the AUC-ROC Curve. According to Table 4, Logistic Regression displayed the highest results with an accuracy of 0.94 and an AUC-ROC of 0.93. Even if Naïve Bayes performed very well on discrimination, with an AUC-ROC of 0.95, it had a slightly lower accuracy of 0.89 compared to Gradient Boosting. Even though the SVM model had a great AUC, it did not perform accurately, as it recorded only 0.64. These findings suggest that Logistic Regression is the most reliable model among the three when considering both accuracy and discriminative capability. Moving forward, the integration of more advanced machine learning techniques and hybrid ensemble models may further enhance predictive capabilities in healthcare diagnostics.

Table 4. Comparative Performance of Models

Model	Accuracy	AUC-ROC
LR	0.94	0.93
NB	0.89	0.95
SVM	0.64	0.94

5. Conclusion

The study investigated whether there is a correlation between PCOS and PPD in women using LR, NB and SVM. For evaluation, we considered the metrics Accuracy, F1 Score and AUC-ROC and applied them to data sets collected in Kerala. Logistic Regression performed better than the others, reaching an accuracy level of 94% while its AUC-ROC was 0.93, meaning the model is reliable for making predictions. On one hand, Naïve Bayes had an AUC-ROC of 0.95. With an accuracy of only 64%, SVM was not the most effective model. One important achievement of this research is that LR can spot women at risk for PPD relying on factors shown in PCOS. With this, diseases can be detected early, and proper care given promptly to patients. There is potential to develop hybrid or ensemble models, including information about patient habits and mental state and apply deep learning technology to make the models more durable. They can also assist in developing strategies for women to protect their mental well-being.

References

[1] Nithin, S. (2022). Perceptions of PCOS between women in urban and rural areas in Kerala, India [Master's thesis, Boston University]. *Boston University Institutional Repository.* https://open.bu.edu/handle/2144/45828BU Open Learning Initiative

[2] Rao, I., & Saxena, M. (2025). Exploring the Connections of the Mental Health and Sustainability. *International Journal of SDG's Prospects and Breakthroughs, 3*(1), 8-14.

[3] Thakre, V., Vedpathak, S., & Thakre, K. (2020). PCOcare: PCOS detection and prediction using machine learning algorithms. *Bioscience Biotechnology Research Communications, 13*(14), 240–244. https://doi.org/10.21786/bbrc/13.14/56ResearchGate+2JEET Journal+2MIT WPU Research Platform+2

[4] Nakamura, H., & O'Donnell, S. (2025). The Effects of Urbanization on Mental Health: A Comparative Study of Rural and Urban Populations. *Progression Journal of Human Demography and Anthropology, 3*(1), 27-32.

[5] Aggarwal, S., & Pandey, K. (2021). An analysis of PCOS disease prediction model using machine learning classification algorithms. *Recent Patents on Engineering, 15*(6), 53–63. https://doi.org/10.2174/1872212115999201224130204Ingenta Connect

[6] Rahim, R. (2024). Quantum computing in communication engineering: Potential and practical implementation. Progress in Electronics and Communication Engineering, 1(1), 26–31. https://doi.org/10.31838/PECE/01.01.05

[7] Sharma, A., & Iyer, R. (2023). AI-powered Medical Coding: Improving Accuracy and Efficiency in Health Data Classification. *Global Journal of Medical Terminology Research and Informatics, 1*(1), 1-4.

[8] Rani, P., Kumar, R., Sid Ahmed, N. M. O., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments, 7*(3), 263–275. https://doi.org/10.1007/s40860-021-00133-6SCIRP

[9] Ziwei, M., Han, L. L., & Hua, Z. L. (2023). Herbal Blends: Uncovering Their Therapeutic Potential for Modern Medicine. *Clinical Journal for Medicine, Health and Pharmacy, 1*(1), 32-47.

[10] Fathima, I., & Abbasi, B. U. D. (2019). Prediction of postpartum depression using machine learning techniques from social media text. *Expert Systems, 36*(4), e12409. https://doi.org/10.1111/exsy.12409ResearchGate

[11] Mohandas, R., Veena, S., Kirubasri, G., Thusnavis Bella Mary, I.., & Udayakumar, R. (2024). Federated Learning with Homomorphic Encryption for Ensuring Privacy in Medical Data. *Indian Journal of Information Sources and Services, 14*(2), 17–23. https://doi.org/10.51983/ijiss-2024.14.2.03

[12] Rahim, R. (2024). Adaptive algorithms for power management in battery-powered embedded systems. SCCTS Journal of Embedded Systems Design and Applications, 1(1), 25-30. https://doi.org/10.31838/ESA/01.01.05

[13] Dokras, A. (2012). Mood and anxiety disorders in women with PCOS. *Obstetrics and Gynecology Clinics of North America, 39*(1), 83–93. https://doi.org/10.1016/j.ogc.2011.10.006

[14] Chauhan, P., & Patel, P. (2021). Comparative analysis of machine learning algorithms for prediction of PCOS. In 2021 International Conference on Communication Information and Computing Technology (ICCICT) (pp. 1–7). *IEEE.* https://doi.org/10.1109/ICCICT50803.2021.9510128

[15] Sreejith, S., & Nehemiah, H. K. (2022). A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier. Healthcare Analytics, 2, 100102. https://doi.org/10.1016/j.health.2022.100102

[16] Kumari, M., & Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction. International Journal of Computer Science and Technology, 2(2), 304–308.SCIRP

[17] Xiao, M., & Yan, C. (2020). Risk prediction for postpartum depression based on random forest. Chinese Journal of Health Statistics, 45(10), 1215–1222. https://pubmed.ncbi.nlm.nih.gov/33268583

[18] Rahim, R. (2024). Energy-efficient modulation schemes for low-latency wireless sensor networks in industrial environments. National Journal of RF Circuits and Wireless Systems, 1(1), 21–27.

[19] Hasab, M. A. H., Khalil, S., & Al-Musawi, A. F. M. (2024). An induced neutrosophic using soft set (J,L) over M with parameter H. *Results in Nonlinear Analysis, 7*(4), 104–114.