Poly Cystic Ovarian Syndrome : A Machine Learning Approach

|  |  |
| --- | --- |
| Aayushi Sinha | Lekha A |
| Department. of Computer Applications | Department. of Computer Applications |
| PES University | PES University |
| Bengaluru, India | Bengaluru, India |
| aayushisinha892@gmail.com | lekha@pes.edu |

Abstract—PCOS is characterized by the ovaries production of an abnormal number of androgens. These androgens can cause more problems with the women’s menstrual cycle.It explores the application of ML techniques to predict PCOD/PCOS, a binary classification problem, and presents a comprehensive literature survey on existing research in this domain. It has been identified as one of the major health issues among women that directly affects fertility during reproductive ages , affecting 1 in every 10 women.. The objective of this study is to statistically evaluate the

metabolic and clinical features, using dimensionality reduction and principal component analysis, and compare the accuracy and F1-Score of different machine learning models.

Index Terms—Poly Cystic Ovarian Syndrome : A Machine Learning Approach

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a condition where a woman’s ovaries produce too many male hormones (andro gens), leading to various problems with her menstrual cycle. PCOS comes with a bunch of symptoms like irregular periods, acne, heavy bleeding during periods, excessive hair growth, dark patches on the skin, weight gain, pelvic pain, oily skin, and trouble getting pregnant. This condition is marked by hy perandrogenism (too much male hormones), insulin resistance (difficulty in using insulin effectively), anovulation (when the ovary doesn’t release eggs properly during the menstrual cycle), and neuroendocrine disruption (confusion in hormonal signals). By using machine learning algorithms, doctors can diagnose PCOS earlier and provide better treatment. LATEX.

II. LITERATURE SURVEY

Due to their significance in today’s society, among the countless issues that face us, those that pertain to women’s health were chosen as the focus of our attention. This section shines a light on the literature work on PCOS. It discusses all similar works done to detect PCOS using machine learning algorithms. Table I summarizes all the methods used for similar research objectives and their respective results in terms of accuracies of the model used to detect PCOS successfully. In the literature on PCOS, there has been limited research on the early detection of PCOS. A few of the best works have been discussed and are compared here. In [1], a novel algorithm RFLR (hybrid random forest and logistic regression) has been applied to metabolical and clinical features to detect

PCOS and provides an accuracy of 89.27learning models help determine the presence of PCOS and develop smart systems to detect PCOS on similar metabolic and clinical features, namely, Logistic Regression, Bayesian Classifier,

This paper mainly focuses on comparing the different techniques and algorithms to detect PCOS sooner and more accurately. Dharshini and Vimala (2020) looked into how machine learning techniques may be used to categorize PCOS in their work. They assessed the effectiveness of several techniques, including support vector machines (SVM), logistic regression,Random Forest Classifier (RFC), which consisted of 268 cases in the dataset. PCOS-positive individuals out of a total of 768 cases, was used by the authors. They beat the other algorithms using SVM achieving an accuracy of 82.16 percent. The authors concluded that SVM is a promising method for this use and that machine learning techniques can be useful in the categorization of PCOS. [5] In their work ”Performance Evaluation of Machine Learning Tech-niques for PCOS Diagnosis,” Soman et al. (2020) examined how well several machine learning approaches performed in diagnosing PCOS. The authors trained and tested a number of classification models, such as logistic regression, decision tree, random forest, SVM, using a dataset of 500 patients with and without PCOS. SVM had the greatest rates of PCOS diagnostic accuracy, according to the findings. The study supports the use of SVM and KNN models for this purpose and emphasizes the potential of machine learning approaches in enhancing the precision and effectiveness of PCOS diagnosis.

Clinical Evaluation Traditionally, PCOS diagnosis involves a combination of clinical evaluation, medical history, physical examinations, and laboratory tests (hormonal assays, ultra-sound imaging, etc.). Blood tests can be used to identify char-acteristic changes in hormone levels, although these changes are not universal. Women with polycystic ovary syndrome may have elevated levels of Testosterone (an ovarian an-drogen hormone that influences hair growth), Oestrogen (an ovarian hormone that stimulates growth of the womb lining (endometrium), Luteinising hormone (LH, a pituitary hormone which influences hormone production by the ovaries and is important for normal ovulation);Insulin (a hormone that is principally involved in utilization of energy from food); andanti-mullerian¨ hormone (which is measures the fertility

level of the ovaries).

Data-Driven Approaches Researchers have started employ-ing data-driven approaches, utilizing machine learning al-gorithms to analyze and interpret large datasets associated with PCOS. These datasets may include hormonal profiles, ultrasound results, lifestyle information, and other relevant patient data.

Ultrasound image analysis

Image processing techniques are used to analyze ultrasound images of ovaries. Enlarged ovaries with multiple small cysts are indicative of PCOS.Computer-aided diagnosis (CAD) sys-tems have been explored to assist healthcare professionals in interpreting ultrasound images.Multiple ablation studies have been conducted to analyze the impact of different stages of machine learning such as image pre-processing in case of performing conventional machine learning technique; transfer learning in case of executing deep learning technique.

* 1. METHODOLOGY

A. Data Collection and Preprocessing

Information Collection and Preprocessing a dataset com-prising of 542 tests reports each having 41 attributes was taken from Kaggle’s store. A comprehen- sive dataset has been compiled, which envelops a range of physical and clinical parameters pertinent to the diagnosis of PCOS and infertility-related conditions

B. Dataset Splitting

Include Selection Out of all the highlights, we will take a chosen few features and prepare our demonstrate. There are 14 qualities out of which are considering

C. Feature Selection

Show Training Train the dataset on a preparing program utilizing an optimiza- tion calculation (e.g. SVM,LR,RFC). Procedures such as early dismissal can be utilized to anticipate overwork.

D. Model Training

Train the dataset on a training program using an optimiza-tion algorithms (e.g. SVM,LR,RFC).

E. Evaluation

* + Evaluation Evaluate the preparing show in the testing setup to evaluate its execution. Common assessment measurements for classification function incorporate precision, exactness, review to check i, F1 score,ts exactness ormulasusing the and ROC-AUfC curve. Visualize the forecast demonstrate and any classification errors to pick up understanding into its qualities and weaknesses
1. Testing

Testing For testing we have utilized the manual test cases which checks the essential highlights of the UI. Separated from this we have also added the assessment measurements for the show testing accracy. To check the model’s precision we have to check the F1 score ,draw a disarray framework, AUC RUC bend, exactness and recall utilizing the formula.

Fig. 1. Process Flow Diagram



G. Deployment

Deployment Once you are fulfilled with the execution of the model, deploy it in a genuine environment. This may include integrating the demonstrate into a computer program or healthcare application

IV. RESULTS

Utilizing administered machine learning calculations, we pre- dicted polycystic ovarian disorder (PCOS) based on clinical parameters. Calculated relapse, bolster vector ma-chines, ran- dom timberlands. Highlight significance exami-nation highlighted testos- terone and luteinizing hormone as noteworthy indicators. This study grandstands the viability of directed ML in PCOS prediction, supporting early location and mediation. With an accuracy score of 97percent, our dis-coveries emphasize the potential for ML-driven approaches to revolutionize PCOS diagnosis and administration, advertising personalized bits of knowledge for improved understanding care. These equations are crucial measurements utilized in binary classification assignments to assess the execution of a classifier. 1. Accuracy: Accuracy = TP + TN TP + TN + FP

* FN 2. Specificity (Genuine Negative Rate: Specificity = TN TN + FP 3. Affectability (Genuine Positive Rate or Recall): Sensitivity = TP TP + FN 4. Exactness (Positive Prescient Value): Precision = TP TP + FP Where TP, TN, FP, FN are truncation for genuine positive, true negative, wrong positive, and untrue negative respectfully. The values of TP, TN, FP, and FN were gotten from the confusion matrices

These formulas are fundamental metrics used in binary classification tasks to evaluate the performance of a classifier.

* 1. Accuracy:

TP +TN

Accuracy = T P + T N + F P + F N

2. Specificity (True Negative Rate:

T N

Specificity = T N + F P

3. Sensitivity (True Positive Rate or Recall):

T P

Sensitivity = T P + F N

4. Precision (Positive Predictive Value):

T P

P recision = T P + F P

Where TP, TN, FP, FN are abbreviation for true positive, true negative, false positive, and false negative respectfully. The values of TP, TN, FP, and FN were obtained from the confusion matrices



Fig. 2. Age Group

V. CONCLUSION

The presentation of these machine learning calculations for prediction of polycystic ovarian disorder speaks to an enor-mous headway in healthcare. We rethink the diagnosis and administration of PCOS by utilizing data-driven methodolo-gies. Clients can bolster the significant information into the Parallel clas- sification models which would be accommodat-ing to anticipate the most accurate or personalized forecast and experiences. Incorporating these sorts of calculations into clinical hone can be instru- mental in reducing the burden of heart disappointment, especially if it is extricated from the information proficiently, and they have an increased the chance of determination at an early arrange of heart failure profiting the quiet at expansive. By and large, it speaks to a paradigm move in care, advertising the chance for way better diagnoses, treatments custom-made particularly to the person, and overall better understanding health.Utilizing administered machine learning algorithms, we anticipated polycystic ovarian disorder (PCOS) based on clinical parameters. Calculated relapse, back vec- tor machines, arbitrary timberlands, and angle boosting machines were utilized, with angle boosting machines yielding the highest exactness of 85 percent

REFERENCES

* 1. Palak Mehrotra, Jyotirmoy Chatterjee and Chandan Chakraborty, Computerized Screening of Polycystic Ovary Syn- drome utilizing Machine Learning Methods, IEEE, 2020.
1. Bedy Purnama, Untari Novia Wisesti, Adiwijaya, Fhira Nhita, Andini Gayatri and Titik Mutiah, ”A Classification of Polycystic Ovary Disorder Based on Follicle Detec-tion of Ul-trasound Images”, 3rd Universal Conference on Informa- tion and Communication Innovation (ICoICT), 2021 3. Priyanka R. Lele and Anuradha D. Thakare, ”Compara- tive Examination of Classifiers for Polycystic Ovary Syndrome Detection utiliz-ing Different Measurable Measures”, International Journal of Building Investigate Innovation (IJERT), vol. 9, no. 03, Walk 2020, ISSN 2278-0181,2015 4. Namrata Tanwani, ”Detecting PCOD utilizing Machine Learning”, Diary of Cutting edge Patterns in Designing and Science, vol. 07, no. 01, 2020, ISSN 2348-3121. 5. J. Madhumitha, M. Kalaiyarasi and S. Sakthiya Ram, ”Automated Polycystic Ovarian Disorder Recognizable proof with Follicle Recognition”, 2017 3rd Worldwide Con-ference on Signal Handling and Communication. 6. Pijush Dutta, Shobhandeb Paul and Madhurima Ma- jumder, ”An Productive Destroyed Based Machine Learning classi- fication for Forecast Discovery of PCOD”, Investigate Square, Novem-ber 2021. 7. A. S. Kumar, S. Annamalai, M. Kumaresan, P.

Manikan- dan, R. Sekaran and H. A. Pai, ”CNN-Based Exami-nation of Ul- trasound Pictures for PCOS Diagnosis,” 2023 3rd International Conference on Innovative Headways in Compu-tational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 347-350, doi: 10.1109/ICTACS59847.2023.10390451. 8. N. Kaur, G. Gupta and P. Kaur, ”Transfer-Based Deep Learning Procedure for PCOS Location Utilizing Ul- trasound Images,” 2023 Worldwide Conference on Net- work, Interactive media and Data Innovation (NMIT- CON), Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMIT- CON58196.2023.10276245 ,