International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 08, Volume 09 (August 2022) https://www.irjcs.com/archives

# POLYCYSTIC OVARY SYNDROME USING MACHINE LEARNING

**V Babitha**
Student, Information Science & Engineering,
Vemana Institute of Technology, Bengaluru, India
**Varun B V**
Student, Information Science & Engineering,
Vemana Institute of Technology, Bengaluru, India
**Upendra K**
Student, Information Science & Engineering,
Vemana Institute of Technology, Bengaluru, India
**Supreeth A**
Student, Information Science & Engineering,
Vemana Institute of Technology, Bengaluru, India

**Publication History**
Manuscript Reference No: IRJCS/RS/Vol.09/Issue08/SPAUCS10095
Research Article | Open Access
Peer-review: Double-blind Peer-reviewed
Article ID: IRJCS/RS/Vol.09/Issue08/SPAUCS10095
Volume 2022 | Article ID SPAUCS10095 http://www.irjcs.com/volumes/Vol09/iss-08/15.SPAUCS10095.pdf
Received: 25, July 2022 | Accepted: 04, August 2022 | Published: 12, August 2022
**doi:** https://doi.org/10.26562/irjcs.2022.v0908.15
**Citation:**Babitha,Varun,Upendra,Supreeth (2022).Ploycystic Ovary Syndrome using Machine Learning. International Research Journal of Computer Science (IRJCS), Volume IX, Issue VIII, 2022, Pgs228-234
**BibTeX key:**
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar


**Abstract:** One of the most common hormonal conditions in women of reproductive age is polycystic ovarian syndrome (PCOS). Since pcos is frequently linked to obesity, type 2 diabetes mellitus, and high cholesterol levels, early detection and treatment are crucial. By counting the number of follicles in an ovarian ultrasound image and adding clinical, biochemical, and imaging characteristics, pcos can be automatically detected and patients divided into normal and pcos-affected groups. The body mass index (BMI), hormone levels, and other factors are all taken into consideration by the Support Vector Machine method, which is utilized for categorization.
**Keywords:** Random Forest Regression, K nearest Neighbors, Gaussian Naive Bayes, and Logistic Regression.

## I. INTRODUCTION

Preterm abortions, infertility, ovulation, and other issues have a significant impact on the current female population worldwide. It has been noted that PCOS, a disorder common in women of reproductive age, plays a significant role in the causes of infertility. PCOS affects more than five million women globally who are of reproductive age. It is an endocrine condition defined by alterations in the levels of female hormones and aberrant male hormone synthesis. Ovarian dysfunction brought on by this illness increases the risk of infertility and miscarriage. Obesity, an irregular menstrual cycle, an excess production of the male hormone, acne, and hirsutism are all signs of PCOS. Due to the wide range of symptoms and the existence of numerous risk factors, PCOS is very challenging to diagnose. Patients with PCOS are now burdened by the time and expense associated with multiple clinical testing and ovary scans. This research suggests a system for the early identification and prediction of PCOS using an ideal and minimal but effective clinical and metabolic parameter, which serves as an early marker for this condition, in order to solve this issue. The 541 women who participated in a patient survey conducted during medical consultations and clinical examinations provided the data sets needed for the development of this system. Using SPSS V 22.0, 8 probable features are found out of the 23 characteristics from clinical and metabolic test data based on their importance. Several machine learning algorithms, including the Naive Bayes classifier method, logistic regression, K-Nearest neighbor (KNN), Classification and Regression Trees (CART), Random Forest Classifier, and Support Vector Machine (SVM) in Spyder Python IDE, are used to classify PCOS using the feature set modified with Principal Component Analysis (PCA). Results showed that RFC, which has an accuracy of 89.02 percent, is the most suited and accurate method for the PCOS prediction. Together, technology and humanity can pave the road for better health care and services. Machine learning is a subtype of artificial intelligence that gives the system the capacity to autonomously learn and develop without being explicitly programmed.

It mainly focuses on creating algorithms that can access the available datasets and utilize data for the network's learning needs. Applications of machine learning, such as detection, data prediction, image identification, etc., have a significant impact on the health industry.

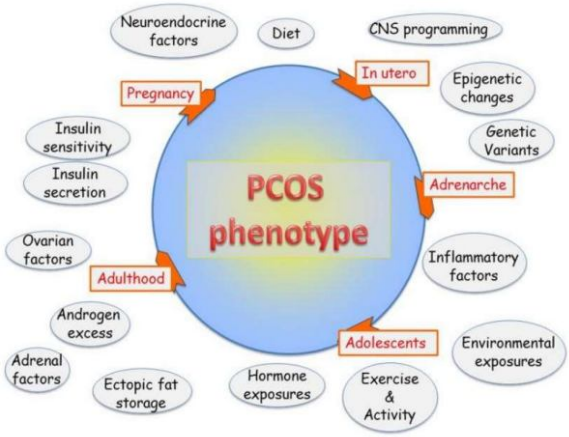


Fig 1: Factors contributing to PCOS

According to Fig. 1, PCOS phenotypes include oligo ovulation, polycystic ovaries, and clinical hyper and rogenism. Most people with obesity and/or PCOS also have insulin resistance, which is more common and more severe in obese PCOS patients than in non-obese patients.

## II. SCOPE

A frequent hormonal illness that affects females in the reproductive age group is polycystic ovarian syndrome (PCOS). Given that PCOS is frequently linked to obesity, type 2 diabetes mellitus, and high cholesterol levels, early diagnosis and treatment of the condition are crucial. In this study, PCOS is automatically detected by counting the number of follicles in an ultrasound scan of the ovaries and then classifying patients into normal and PCOS-affected groups using clinical, biochemical, and imaging criteria. In females, menstruation issues, hirsutism, endocrine disorders, acne, obesity, etc. are the primary causes of this illness. Using ovaries' ultrasound pictures, ovarian follicles are identified. It is difficult to identify objects in ultrasound images, including follicles in the ovary, the development of the fetus, tumors existence, and the growth of the fetus. A physical count of the follicular cysts in the ovary is now utilized by doctors to diagnose PCOS and determine whether it is present or not.

## III. OBJECTIVE

One of the significant, most common hormonal problems experienced by women of childbearing age is polycystic ovarian syndrome (PCOS). This is a diverse endocrine problem that greatly increases the risk of infertility, ovulation, cardiovascular disease, type 2 diabetes, obesity, and other conditions. Nearly 12 to 21 percent of women of reproductive age have PCOS, and among them, 70 percent do not have a diagnosis. PCOS can be treated to a certain extent with regulated medications and lifestyle changes.

## IV. LITERATURE SURVEY

The issues relating to women's reproductive health were chosen as a focus of our interest among the myriad issues that surround us because of their significance to the modern society. A thorough review of PCOS studies and tools for diagnosing the condition was conducted. The complicated endocrine condition known as Polycystic Ovary Syndrome (PCOS) affects between 5 and 10 percent of Indian women of reproductive age, according to the literature [16]. It raises the risk of insulin resistance, obesity, cardiovascular disease, and psychological disorders [17] and is a significant contributor to anovulatory infertility. The signs and symptoms of PCOS may differ from patient to patient. Some of these include irregular menstruation cycles, acne, being overweight, a higher risk of infertility, significant hair loss, balding on the top of the head, and an increase in facial hair development [1]. When there are more than 12 follicles per unit area and they are evident on a radiological scan, PCOS is often suspected [15].

Some authors have suggested increasing the cutoff value from 12 to 20 follicles or doing away with ultrasound entirely in favors of alternative indicators, like serum anti-Mullerian hormone (AMH) [1, 10]. The accurate application of a few well-established, standardized diagnostic techniques is all that is necessary for the diagnosis of PCOS. Delays in PCOS diagnosis have a significant impact on patients' wellbeing [21]. According to Escobar et al. [11], treatment should be symptom-focused, ongoing, dynamic, and tailored to the patient's unique requirements, expectations, and changing circumstances. In [6], Joham et al. looked explored the relationship between PCOS and the infertility rate among women in this population. They found that PCOS-positive women were considerably more likely to take fertility hormone therapy. Strategies to improve PCOS diagnosis and the factors influencing fertility are crucial given the frequency of the condition and the financial and health toll that infertility takes. This is due to the fact that, regardless of BMI, infertility is said to be 15 times more common among women who report having PCOS [18].

 

## V DESIGN

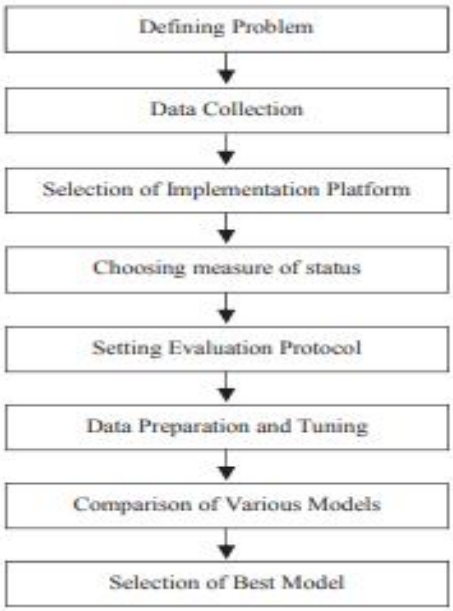


**Fig 2:** Data flow diagram

### 1 Defining Problem

The most crucial stage is to define the problem clearly, taking into account both the model's inputs and the results anticipated from them. It is predicated on the idea that given inputs, results may be predicted.

### 2 Data Collection

The crucial stage that will determine how effective the model will be. Additionally, as the amount of data collected rises, the model's accuracy rises as well, improving performance. There are several ways to obtain data, including in real time or via platforms with repositories like Kaggle and the UCI machine learning repository, which is one of the most popular ones.

### 3 Selection Of Implementation Platform

The right platforms for statistical analysis should be chosen, as well as tools for the effective operation of machine learning approaches. The patient data may be input into the system and PCOS status can be acquired as an output. HTML with SQL is utilized to develop a suitable user interface for this research. SPSS V22.0 is used to determine the importance of features.

### 4 Choosing Measure Of Status

Success in classification issues is evaluated by calculating the model's accuracy and precision. We took into account the following evaluation metrics for this study:

- Accuracy : $\frac{(TP+TN)}{(TP+TN+FP+FN)}$
- Precision (P) : $\frac{(TP)}{(TP+FP)}$
- Sensitivity / Recall (R) : $\frac{(TP)}{(TP+FN)}$
- Specificity : $\frac{(TN)}{(TN+FP)}$
- F1 score : $2 * \left(\frac{P*R}{P+R}\right)$

### 5 Setting Validation Protocol

Keeping a hold out validation set, or, in this method, separating out a portion of the data for testing as test data and leaving the rest as train data

### 6 Data Preparation

This taxing procedure entails processing categorical data, handling missing data, scaling features, and selecting significant characteristics. 'Nan' is used to represent missing values in the dataset. Because a model cannot read a missing value, samples with missing values will either be eliminated or substituted with pre-built estimators before the model is confronted. Likewise, ordinal and nominal variables need to be taken into account appropriately before being fed into the model. The dataset has to be normalized and standardized if the nature of the data requires it. Reduced data dimensionality is the final way to prevent over fitting. This is accomplished by lowering the quantity of feature sets in the dataset. Principal component analysis (PCA), which finds patterns in datasets and correlations between characteristics, is used to execute it in the Spyder Python IDE. After that, the directly eliminating such attributes eliminates the associated data.

 

Using SPSS V22.0, the best features obtained by the PCA method are tested for their ability to distinguish between PCOS status. The level of statistical significance between the two classes of PCOS and Non PCOS patients is determined by a process known as the independent sample t-test. Potential characteristics are those with significance less than 0.01.

### 7 Comparison of Various Models

A baseline will be established with this phase. A variety of classifier algorithms are used in the study, which uses a chosen collection of features. The following list includes some of the several machine learning algorithms that have been tested and found to perform well in the detection of PCOS and Non-PCOS conditions:

- Logistic Regression (LR)
- K-nearest Neighbors (KNN)
- Classification and Regression Tress (CART)
- Random Forest Classifier
- Naive Bayes Classifier
- Support Vector Machine

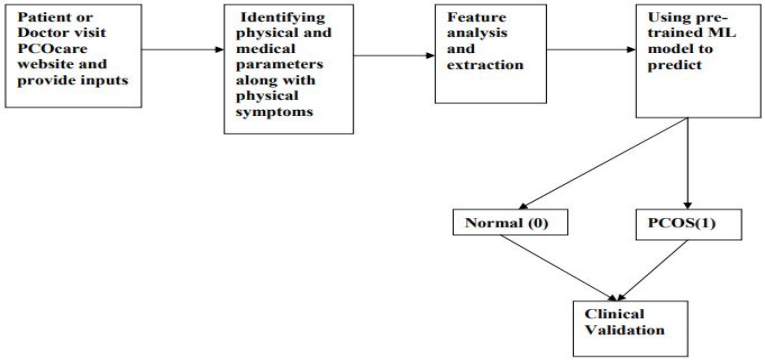


Fig 3: Architecture Diagram

Fig 3 above depicts the system's overall view. It displays the various components that go into creating the system, such as Patient or Doctor visits PCOcare website and provides inputs Identification of physical, medical, and symptom parameters. Feature analysis and extraction. Using a trained machine learning model, predict Normal (0), PCOS (1), and Clinical Validation.

## VI. IMPLEMENTATION

**1 Data Collection:** The more characteristics and samples there are the more distinct and precise the model will be. In this study, we took into account a total of 541 samples that were gathered from various hospitals and clinics in and around the Thrissur district. Each patient was asked for their informed consent, and it was guaranteed that the information gathered would be anonymous.

- **Patient History Collection:** Each patient is given their own proforma, allowing the proforma to hold all of the information pertaining to that patient. These Proforma were brought along for patient surveys and clinic data collecting. At a later time, a single data sheet including all of the Proforma data obtained is created. The characteristics that were incorporated into the proforma were deduced from several literature reviews and clinical surveys that were done; all participants' identities were kept confidential throughout the whole procedure.
- **Real-time data collection:** A patient interface is necessary for the patient data input necessary for the real-time identification of PCOS. The patient's personal information for future references as well as the factors that rule out PCOS are examples of the data that may need to be input. The entered data will be used to feed the data processing algorithm that has been developed. A front end is built using HTML, and the data entered in the front end is written into an excel sheet using SQL, creating the interface.

**2 Parameter Selection:** Data gathered from the survey's results were incorporated into the proposed model, i-Hope. Finalizing the parameters was done with the assistance of professional opinions and taking current research that in some way or another affected PCOS into consideration. The features are altered to eliminate any correlation between them that can negatively impact the categorization outcomes. Table I is a list of the finalized parameters used before transformation. These metrics cover physiologic, metabolic, and biochemical characteristics. Also taken into account were the results of the ultra sound scan, which contained data regarding the cysts that had formed, including the number of cysts present in the right and left ovary and their size.

 

**Table 1:** FINALIZED PARAMETERS

| Sl No | Parameters | Value |
|---|---|---|
| 1 | Age | 15-35 |
| 2 | BMI | <24(normal),<br>>24(abnormal) |
| 3 | Cycle Length and regularity | Long, normal or short<br>Regular/Irregular |
| 4 | LH : FSH Ratio | Normal/abnormal |
| 5 | Waist : Hip Ratio | Normal/abnormal |
| 6 | Weight gain | Yes(y)/No(n) |
| 7 | Excess facial or body hair | Yes(y)/No(n) |
| 8 | Dark areas on skin | Yes(y)/No(n) |
| 9 | Pimples | Yes(y)/No(n) |
| 10 | Blood Pressure | Normal/abnormal |
| 11 | Diabetes<br>(before and after food) | Normal/abnormal |
| 12 | Fast food intake | Yes(y)/No(n) |
| 13 | Regular exercise | Yes(y)/No(n) |
| 14 | Loss of hair | Yes(y)/No(n) |
| 15 | No. of follicles<br>($L^a$ and $R^b$ ) | High ,medium, low |
| 16 | Size of follicles<br>($L^a$ and $R^b$ ) (mm) | >10 |
| 17 | TSH (mIU/L) | 0.4-4(Normal) |
| 18 | AMH | 1-4(Normal)<br>>1(Abnormal) |
| 19 | PRL | 2-29 (Non-Pregnant Females)<br>10-209 (Pregnant Females) |
| 20 | Vit D3 | 20-50 (Normal)<br>>12 (Abnormal) |
| 21 | PRG | 1.5-12.4 (Normal) |

[a] Left ovary, [b] Right ovary

## VII. RESULTS

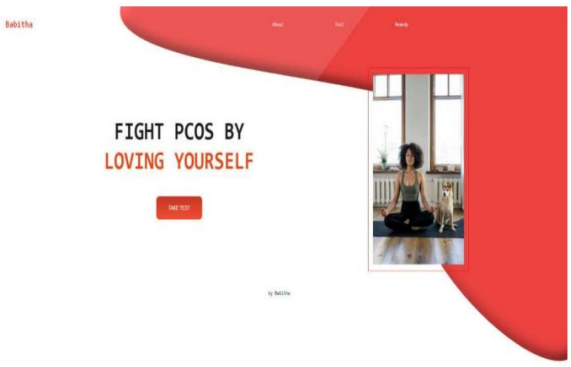


**Fig 4:** Homepage for attending test

Fig 4 shows the Homepage of the application where we can upload the data and take a test.

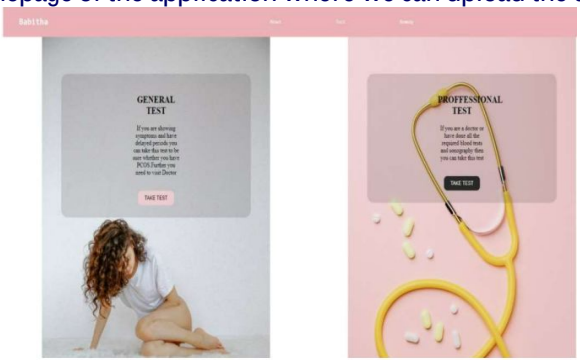


**Fig 5:** Selection of tests

Fig 5 shows the browsing here we can take general test and professional test.

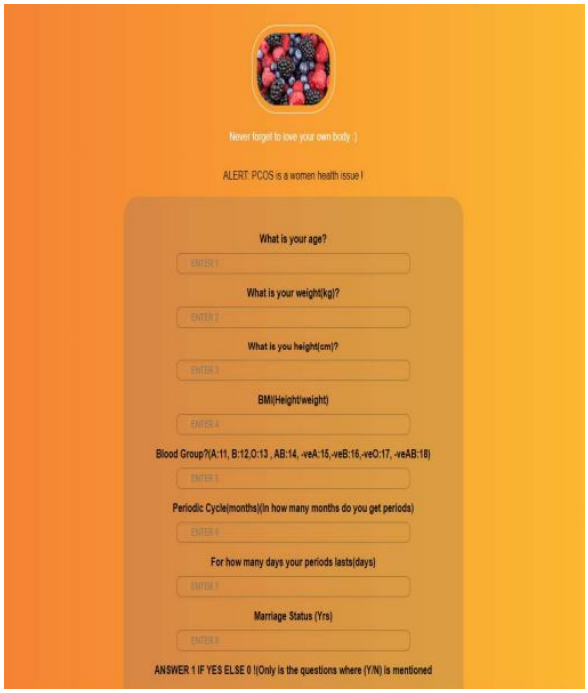


Fig 6: Data Entry form

Fig 6 shows that we can enter the data to get the result.

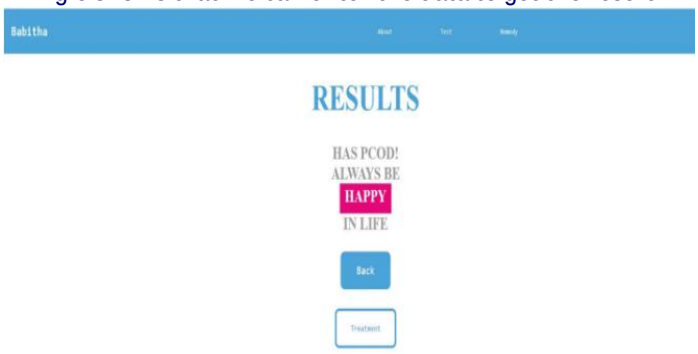


Fig 7(a): Results Show PCOS is Present

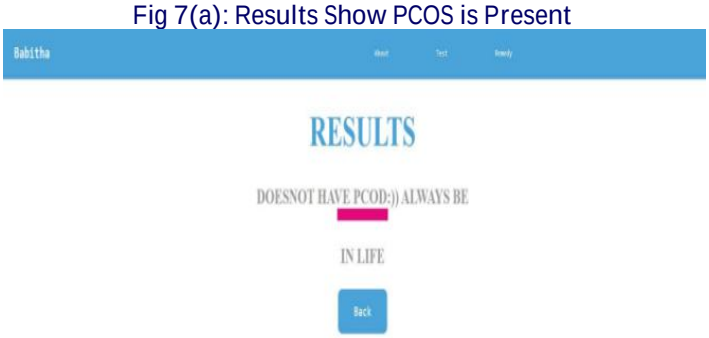


Fig 7(b): Results Show PCOS is Not Present

Fig 7 (a) (b) shows the result of the project Shows PCOS is present or not.

## VIII. CONCLUSION

An ovulation and infertility may occur from PCOS. The clinical and metabolic indicators that serve as biomarkers for the disease are part of the diagnostic criteria. A technique that automates the detection of PCOS based on a small number of possible markers has been developed. Our strategy entails creating a feature vector from real-time patient data collected during clinical and radiological examinations while they are in a medical facility. The PCA feature transformation and statistical significance were used to identify eight metabolic and ultrasound imaging variables that have potential for differentiating between PCOS patients and healthy individuals. The RF algorithm is found to perform better than the other methods.

 

## REFERENCE

1. J.J. Cheng and S. Mahalingaiah, 2018. Analyzing and categorizing results of pelvic ultrasounds to identify polycystic ovaries p.254870 on bioRxiv.
2. U.N. Wisesty and R.M. Dewi, March 2018. Using a competitive neural network, polycystic ovary classification from ultrasound pictures. In Conference Series in Journal of Physics (Vol. 971, No. 1, p. 012005). IOP Publications
3. i-HOPE: Detection and Prediction System for Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques Biomedical Engineering Amsy Denny Department Sahrdaya College of Engineering and Technology Sahrdaya College of Engineering and Technology, Anita Raj Department of Biomedical Engineering 978-1-7281-1895-6/19/$31.00 © 2019 IEEE
4. Automated Polycystic Ovarian Syndrome Detection Using Follicle Recognition Department of Instrumentation and Control, Cummins College of Engineering for Women, Pune, India, Sharvari S. Deshpandei1,AsmitaWakankar2, sharvari.777@gmail.com,asmitawakankar@yahoo.com, ISBN No. 978-1-4799-3914-5/14/$3l.00 ©2014 IEEE
5. PCOcare: Machine Learning Algorithms for PCOS Detection and Prediction Dr. Shreyas Vedpathak World Peace University at MIT Sunil Vaidehi Thakre Pune, India's MIT World Peace University Special Issue of Biosc. Biotech. Res.Comm. Vol. 13 No (2020) Pp-240-244
6. CNN is being used to research and find disorders associated to PCOS. Srilatha K. Sathyabama Institute of Science and Technology, Chennai, M Sumathi, P Chitra, R Sakthi Prabha Chitraperumal@gmail.com and sumathi.ece@sathyabama.ac.in Materials Science and Engineering IOP Conf. Series 1070 (2021) 012062 ICRIET 2020 IOP Publishing doi:10.1088/1757-899X/1070/1/012062
7. Structural Normalized Square Similarity Detection Approach for Polycystic Ovary Syndrome Detection and Prediction 1 Dr. L. Thara, two Divya T.M. 1Associate Professor, Department of Computer Science, PSG College of Arts & Sciences, Coimbatore, India 2Researcher at the computer science department of the PSG College of Arts & Sciences in Coimbatore, India 2021; 8(6): 2834–2842; Nat. Volatiles & Essent. Oils