

Research Article

Classification of Polycystic Ovary Syndrome (PCOS) data using machine learning algorithms

Priyanka R. Lele and Anuradha. D. Thakare

Department of Computer Engineering Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

in detected in women or girls in their reproductive age. PCOS is caused due to imbalance in the hormone level. PCOS lead to irregularity in menstrual periods. The irregular periods further cause formation of cysts (follicles) in either or both ovaries as well as infertility. PCOS symptoms include abnormal periods, disturbance in androgen levels, ovaries having cysts, increased BMI and some other abnormalities in hormone level of LH, FSH, DHEAS, Fasting insulin and Fasting blood sugar. This disease should be detected and prevented as early as possible. An intelligent software will help the health care providers to examine the patients and diagnose the risk of PCOS. This paper is work in progress of our proposed research to detect PCOS. Various machine learning algorithms are implemented on the dataset.

Keywords: Polycystic ovary syndrome, sonography, machine learning, classification.

Introduction

Polycystic ovary syndrome (PCOS) is a heterogeneous endocrine disorder affecting most women of reproductive age. According to the previous research, nearly 5-10 % of reproductive age female are affected with this abnormality. Women with this ovarian dysfunction can increase risk the of cardiovascular disease, type 2 diabetes mellitus, obesity, hypertension, gynecological cancer.

Moreover, recent studies have shown greater risk of first Trimester miscarriage[1]. On ultrasonography the following changes may include such as showing of increased levels of androgen hormone, hirsutism, acne, amenorrhoea or oligo, morphological change and an ovulation, these symptoms are found with the patients having PCOS[2].

PCOS is a plight that leads to small, cysts which are harmless will develop on ovaries. These cysts caused with the hormonal imbalance. The typically affected hormones are, Androgens Estrogen, Progesterone, Testosterone, Luteinizing Hormone. The over production of Androgen hormone, a male hormone, in female body that lead to hyperandrogenism. So, increased testosterone in a women's body can abolish normal menstruation and ovulation[3]. The Fig.1 shows the polycystic ovary in a women.

The symptoms of PCOS are; irregular menstrual periods (they can be missing or can be scanty), infertility, excessive hair growth (unwanted or excessive hair growth on face and/or body), partial baldness, increased Body Mass Index, huge amount of acne, swollen breasts before menstruation cycle, heavy periods, high blood pressure, cysts formation in ovaries [5].

The exact cause for Polycystic Ovary Syndrome is not known. But according to the previous research causes can be insulin resistance, genes, obesity, inflammation, higher androgen levels, etc.

There are some risks that are related to PCOS. The women having PCOS may lead to risks like infertility, type 2 diabetes, abnormal uterine bleeding, high

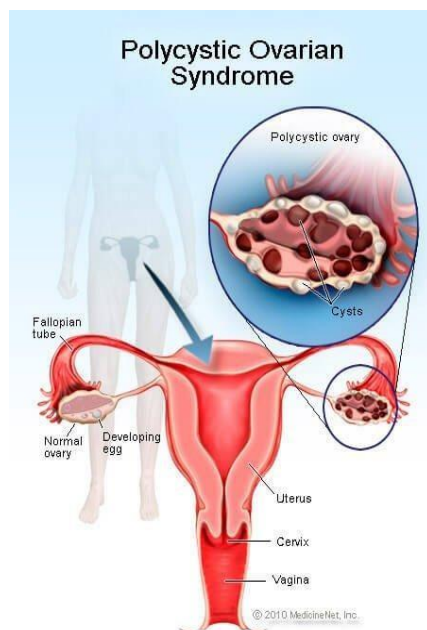


Fig. 1 Polycystic Ovary [4]

cholesterol, fatty liver, liver diseases, depression, anxiety, high blood pressure, obesity, metabolic syndrome, miscarriages, cardiac risks[6]. Multiple tests need to be conducted to diagnose PCOS. To diagnose PCOS and find out other causes of your symptoms, a doctor may consult to you about medical history and do a physical exam and different tests like Pelvic exam, Pelvic ultrasound and some blood tests.

Literature Survey

“Community Screening For PCOS Amongst Adolescent ladies during a Semi geographic region In West Bengal”[7], which proposed 2015 adolescent ladies of West Bengal were asked concerning their initiation of periods. They were conjointly examined for clinical options like excess in steroid hormone. per metropolis criteria, the cases were ended by having PCOS or not having PCOS.

“Early Endocrine, Metabolic, and Sonographic Characteristics of Polycystic Ovary Syndrome (PCOS): Comparison between Non obese and Obese Adolescents” [8], which proposed 11 non-obese and 22 obese adolescents with PCOS and 15 obese controls were chosen. The objective was to characterize early endocrine and metabolic changes in mid-aged women with PCOS and to determine whether the differences between non-obese and obese women are detected early. Comparison between obese PCOS and non obese PCOS done with F test.

“Ultrasound features of polycystic ovaries relate to degree of reproductive and metabolic disturbance in polycystic ovary syndrome”[9], which proposed 49 women (aged between 19 to 36) diagnosed with PCOS were chosen. Evaluation of menstrual cycle and also physical exam assess various parameters(height, weight , BMI, blood pressure, etc) was performed. Study of Antral follicle count(AFC), number of follicles per follicle size, ovarian volume(OV), stromal area(SA), ovarian area(OA), stromal to ovarian area(S/A), stromal index(SI) is performed. Spearman rank was used for correlation between different parameters.

“Data mining to elicit predominant factors causing infertility in women”[10], which proposed a sample population of 575 patients who were getting treatment in endometrial research center in Trichy were selected. Physical and hormonal factors were taken into account. Based on these factors, patients were categorized into 154 fertile women and 421 infertile women. Questionnaire was created which consist of various parameters regarding the factors that influenced the infertility. Statistical analysis and CFS subset evaluator was deployed here. Two methods of classification were used J48 algorithm and Random tree algorithm. K-means clustering and Association rule mining was performed.

“Temperament and Character Differences of Patients with Polycystic Ovary Syndrome”[11], which proposed a total of 91 subjects were included in the study. Clinical and/or biochemical signs of hyperandrogenism; oligo/amenorrhea and polycystic ovaries in ultrasound were considered. Modified Ferriman– Gallwey score (FGS). The relationship between test scores were tested with Mann–Whitney U and independent samples t-test.

Proposed System

A. Proposed Approach

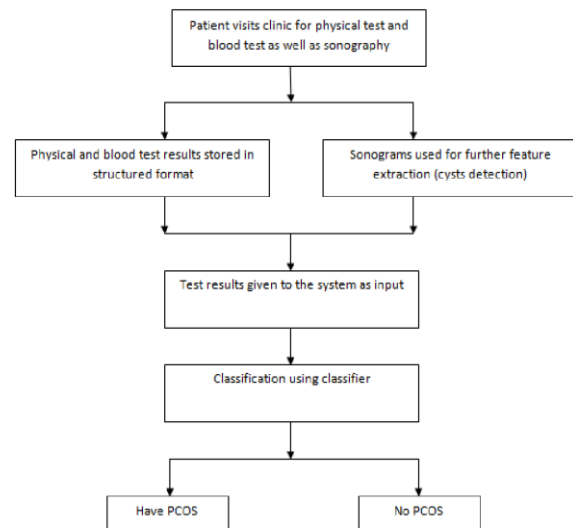


Fig 2. System approach for proposed system

The patients visits the clinic for the required test that needs to be conducted. Physical symptoms like acne, facial hair growth and irregular periods as well as hormone tests for LH(Luteinizing hormone), FSH(Follicle-stimulating hormone), androgen levels, DHEAS, fasting insulin, fasting blood sugar should be examined.

Applied mathematics analysis is performed.

B. Mathematical Background:

Let s be the system with set of parameters P

$S = \{ P_i, O_j, C, \dots \}$

Where,

$P_i = \{ P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}, P_{12}, P_{13} \}$

P_1 - age

P_2 - height (ft)

P_3 - weight (kg)

P_4 - irregular periods

P_5 - hirsutism

P_6 - acne

P_7 - LH(Luteinizing hormone)

P_8 - FSH(Follicle stimulating hormone)

P_9 – androgen level
 P_{10} – DHEAS (\emptyset)
 P_{11} – fasting insulin
 P_{12} – fasting blood sugar
 P_{13} - sonography
 $PBAP A(|) ()$
 $PAB(|) \square$ _____ (2)
 $PB()$

Using Bayes theorem, the probability of event A happening, given that event B has occurred can be found. Here, B is the evidence(used to predict) and A is the hypothesis(assumed). The assumption made here is the variables are independent. That is, the presence of a one particular feature does not affect the other feature.

$O = \{ Y, N \}$
 O_Y - Yes(have PCOS)
 O_N - No (do not have PCOS)

C. Working of Gaussian naïve bayes:

Naïve Bayes is a classifier that is based on probabilities. Training is fast because only the probability of each class and the probability of each class when different input (x) values given need to be calculated. No coefficients need to be fitted by optimization procedures. There two probabilities that need to be considered in this algorithm.

They are:

- Class probabilities: Probabilities of each class in training dataset.
 - Conditional probabilities: Probabilities of each input value given each class value.
- Calculating Class Probabilities

The class probabilities are simply the count(frequency) of instances that belong to each class divided by the total number of instances.

Calculating Conditional Probabilities

The conditional probabilities are the count(frequency) of each attribute value for a given class value divided by the count(frequency) of instances with that class value.

Results & Discussion

The experimentation is carried out on the dataset created and well known machine learning algorithms. The objective of using various algorithms is to identify the most suitable algorithms for classification of the dataset created. The Machine Learning algorithms like Support Vector Machine, Linear Regression and Gaussian Naïve Bayes are used for classification and performance is analyzed statistically. Statistical result like Accuracy is calculated and compared using functions available in Python. Table 1 depicts the statistical results.

Table 1. Classification Results

Classification Algorithms	Classification Accuracy (%)
Support Vector Machine	75
Linear Regression	75
Gaussian Naïve Bayes	89.18

From the results, it is observed that, Gaussian Naïve Bayes has the highest accuracy of 89.18% and is super sitting all other classification algorithms. Whereas; other algorithms have accuracy score of 75% which is comparatively 14.18% less.

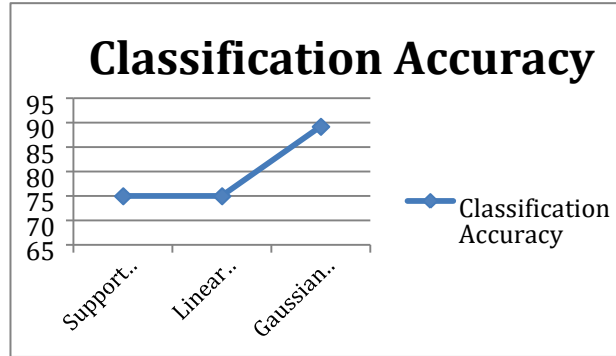


Fig. 5 Line Graph: (x axis: Algorithms / y axis: Accuracy)

The above graph has classification algorithms on its X-axis and the classification accuracy on its Y-axis. It is observed that Gaussian Naïve Bayes algorithm has highest accuracy. The other algorithms may perform well with increased dataset size. This analysis will be carried out further for the proposed research.

Proposed method for classification of PCOS Sonography images using Gaussian Naïve Bayes:

The Gaussian Naïve Bayes produce very accurate classification results with minimum training time compared to conventional supervised and unsupervised learning. The sonography images will be given as input to the classifier. The images are categorized in two named: PCOS images and normal images. PCOS images are the result of the ultrasound/sonography test taken of the both or either ovaries. Using feature extraction algorithm, features like the cysts formation on the outer edge of the ovary in the sonography image will be detected and based on this, the classification will be performed.

Conclusion

In this paper, some of the main Classification algorithms were used to evaluate the classification technique, like Support Vector Machine, Linear Regression and Gaussian Naïve Bayes algorithm to detect whether the patient may have PCOS or not. Classification techniques were considered in the study as it helps us to predict if

the patient has Polycystic Ovarian Syndrome or not based on the JOG0H 1527, 22 January 2019 symptoms. The dataset used in this study has been collected via reference from a doctor who helped to construct the dataset used as input for the machine learning model. It has been noticed that the accuracy of the Gaussian Naïve Bayes has the highest accuracy of 89.18. These models can provide help to the doctors to recognize the disease much faster, thus the treatment can be started as soon as possible.

References

- [1] "Automated screening of Polycystic Ovary Syndrome using machine learning techniques", Palak Mehrotra, Jyotirmoy Chatterjee, Chandan Chakraborty, Biswanath Ghoshdastidar, Sudarshan Ghoshdastidar, 2011 Annual IEEE India Conference DOI: 10.1109/INDCON.2011.6139331
- [2] "A Novel Framework for Filtering the PCOS Attributes using Data Mining Techniques", Dr. K. Meena, Dr. M. Manimekalai, S. Rethinavalli, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS010787 Vol. 4 Issue 01, January-2015
- [3] "A Literature Inspection on Polycystic Ovarian Morphology in Women using Data Mining Methodologies", Neetha Thomas, Dr. A. Kavitha, International Journal of Advanced Research in Computer Science Volume 9, No. 1, January-February 2018 ISSN No. 0976-5697
- [4] https://www.emedicinehealth.com/polycystic_ovarian_syndrome_pcose/article_em.htm
- [5] "Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques", Palvi Soni, Sheveta Vashisht, Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018) IEEE Xplore Part Number:CFP18AWOART; ISBN:978-1-5386-4765-3
- [6] <https://www.healthline.com/health/polycystic-ovary-disease>
- [7] "Community Screening For PCOS Amongst Adolescent Girls In A Semiurban Area In West Bengal", Jayanta Pal, Barindra Nath Mallick. J. Evid. Based Med. Healthc., pISSN 2349-2562, eISSN- 2349-2570/ Vol. 3/Issue 100/Dec. 15, 2016
- [8] "Early Endocrine, Metabolic, and Sonographic Characteristics of Polycystic Ovary Syndrome (PCOS): Comparison between Nonobese and Obese Adolescents", Miriam e. Silfen, michelle r. Denburg, alexandra m. Manibo, rogerio a. Lobo, Richard jaffe, michel ferin, lenore s. Levine, and sharon e. Oberfield. JCEM88(10):4682-4688 Copyright © 2003 by The Endocrine Society. doi: 10.1210/jc.2003-030617
- [9] "Ultrasound features of polycystic ovaries relate to degree of reproductive and metabolic disturbance in polycystic ovary syndrome", Jacob P. Christ, Heidi Vanden Brink, Eric D. Brooks, Roger A. Pierson, Donna R. Chizen and Marla E. , Lujan Fertility and Sterility. Vol. No. 2015 0015-0282/\$36.00
- [10] mining to elicit predominant factors causing infertility in women ", Ms. N. Vijayalakshmi, Ms. M. UmaMaheswari, IJCSMC, Vol. 5, Issue. 8, August 2016
- [11] "Temperament and Character Differences of Patients with Polycystic Ovary Syndrome", Ahmet O' ztu' rk, Suna Kabil Kucur, Ali Seven, Erdem Deveci, Halime S, encan, Alperen Kilic, Onur Yilmaz, Alperen Kilic, PII: S2468-7847(18)30314-3 DOI: <https://doi.org/10.1016/j.jogoh.2019.01.006> Reference: