

PREDICTION OF PCOD AMONG WORKING WOMEN IN IT SECTOR USING HEALTHCARE DATA SET

Nayana Habeeb¹ Dr. S. Manju Priya²

ABSTRACT

Infertility and abortions are becoming common among women day by day. PCOS (Poly Cystic Ovarian Syndrome) or PCOD (Poly Cystic Ovary Disorder) is one of the major endocrine disorders among women of reproductive age. It is characterized by hormone imbalance. The affected women have a higher concentration of male hormone than female hormone. The warning signs of this disease include irregular menses, oily skin, acne, hypertension and a high risk of mood and anxiety disorders. A recent survey has revealed that PCOD is more likely to occur among working women, especially employed in IT sector. In general, there is no prescribed medicine or treatment for this lifestyle disease. Early detection and its prevention are inevitable. So, adoption of a healthy lifestyle could mitigate the ill effects. The time and the amount spent on conducting clinical tests and scanning the ovaries create a lot of burden for these women. To bring an efficient solution to this issue, this paper suggests a prediction model with an optimum accuracy for early identification and prediction of PCOS from a set of promising clinical, physical and metabolic parameters, which act as early markers for this syndrome. Data-mining tools and techniques, along with machine learning [1], can be used among different sections for an effective data prediction.

Key words: PCOS/PCOD, Data Mining, IT sector, Healthy lifestyles

1. INTRODUCTION

Polycystic Ovarian Syndrome, also known as PCOS or PCOD [Polycystic Ovarian Disorder], is creating a lot of disturbance in the normal life style of women and has become a leading cause of infertility among them across the world. It's a condition in which lot of cysts get developed in the side and edge of ovaries. These cysts are actually egg-containing follicles that do not develop properly because of hormonal problems. They can be brought under control to some extent with proper medication and adoption of some healthy life style. When normal methods fail to effect noticeable changes, aggressive treatment procedures like surgical drilling of ovaries are to be opted for improving the ovulation ability thereby reducing the male hormone level. It may further lead to breast, uterine or endometrial cancer, if left untreated for long. Women with PCOS display a variety of symptoms which include

- (A) Gynecological disorders such as failure in ovulation, late menopause, endometrial cancer and infertility,
- (B) Metabolic disorders that include insulin resistance, type 2 diabetes and dyslipidemia,
- (C) Cardiac disorders that include hypertension and cardiovascular diseases,
- (D) Physical disorders like obesity, acne, hirsutism, hair loss and baldness and
- (E) Psychological disorders that include depression, stress and anxiety. The diagnosis of PCOD is based on clinical examination, radiological scanning and so on. As life style is the main factor that causes this condition, early detection of the same would be able to bring a drastic change in their fast and furious life.

¹Research Scholar, Department of Computer Science
Karpagam Academy of Higher Education, Coimbatore.

²Professor, Department of CS, CA & IT
Karpagam Academy of Higher Education, Coimbatore.

So, a drastic change in life style may minimize the number of tests and clinical examination that would be required otherwise.

Causes of PCOS

- Unhealthy lifestyles and food habits
- Hormonal variations
- Laziness and obesity
- High stress level
- Usage of contraceptive pills
- Heredity
- Variations in the level of insulin

2. LITERATURE SURVEY

In [1], Akash C. Jamgade proposed that higher accuracy could be achieved using machine learning. He implemented K-means algorithm on structured and unstructured data. He obtained a high accuracy of 95% through his works.

In [2], Dr.V.Krishnaveni, proposed a prediction model for PCOS that was capable of helping healthcare providers in taking ultimate decisions on their patients. This model helped to identify susceptible cases by applying computational intelligence and data mining techniques.

In [3], Palvi Soni and Sheveta Vashisht proposed a methodology that used segmentation along with CNN to avoid redundant and useless data thereby acquiring a higher accuracy and precision. They tried to explore image segmentation based on region by appending CNN along with it. This method claimed to produce accurate results in shorter time. Classification done after segmentation was able to classify the results based on severity like high, medium, low.

In [4], C. Gopala Krishnan and M. Iyapparaja proposed a method for identifying factors and techniques used in the prediction of PCOS with the help of ultrasound images. They also compared the outcomes with those of the existing works. Here the histology and appearance of ovary were taken into consideration.

In [5] Ms. Palvi tried to review the PCOS in terms of its early symptoms and causes. They also discussed the tasks and techniques in data-mining for predicting the disease. They were able to attain 86% accuracy in the proposed system.

In [6] Vikas B et al. tried to point out accuracies and other performance measures of Naïve Bayes, Decision Tree and Neural Networks to predict whether a person was likely to have PCOS or not. The data set for this study was collected by conducting a survey among women, and the outcome showed that the algorithms had near accuracy value permitting the user to choose any one of them.

In [7] Madhu Kumari et al. found prevalence of Polycystic Ovarian Syndrome to be 11.97% among women of reproductive age in the city of Mumbai (India). The category which included the highest number of sufferers was the student's category, with 21.27% of them suffering from the disorder.

In [8] Amsy Denny et al. found a machine-learning-model-based diagnostic aid for PCOS. He chose metabolic Parameters for his study, which were biomarkers of the disease. They tried to use a minimal set of potential markers. The methodology involved formulation of a feature vector based on real time data collected from patients during clinical and radiological investigations during hospital visits.

In [9] Syeda Sidra et al. made use of evidence-based clinical checklist and statistical analysis. They concluded that patients with PCOS lived a poor quality of life mainly due to depression, presence of acne etc. Therefore, guidelines for management of PCOS should review the recommendations regarding the use of pharmacological agents for these conditions.

In [10] Wendy M. Wolf et al. found that there was neither racial nor ethnic influence on the occurrence of PCOS. The concept of sub-populations was implemented for a better understanding of the complexity and prevalence of PCOS.

In [11] Palvi Soni, Sheveta Vashisht presented various PCOS symptoms, its causes and treatments. They also proposed several data-mining tasks and techniques that were desirable for predicting PCOS.

In [12] Dhiraj Dahiwade et al. proposed general disease-prediction system using Convolutional Neural Network (CNN) and KNN (K-Nearest Neighbor) approach. They produced the output in minimal time and at a low cost. They maintained that CNN was better than KNN in terms of accuracy and time.

In [13] Pahulpreet Singh Kohli, Shriya Arora, proposed a model that used three different classification algorithms having the advantage of classifying three separate databases of the disease. They used backward modelling with p-value test for selecting the apt feature.

In [14] Mrunmayi Patil et al. proposed a model that provided a detailed report on alterations in individual life style. They used Machine learning and Support Vector Machine algorithms for better accuracy. The model also took into account climatic conditions and pollution level making it more flexible for customization.

In [15] the authors stated that data visualization was a more representable class and that humans would extract and predict the meaning of data from visualization.

3. PREDICTION OF PCOD AMONG WORKING WOMEN IN IT SECTOR USING HEALTHCARE DATASET

Busy lifestyle and fast food habits make working techies more prone to PCOS. Doctors state that there has been a 20 % increase in PCOS patients. However, symptoms vary from patient to patient. So, it is important to make a study in this field and develop a prediction model that could predict the PCOS at an early stage.

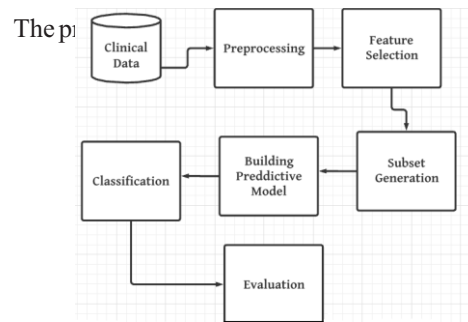


Figure 1: Process Model

The clinical data contain data collected from different categories like symptoms found in the PCOS victims, medical history, clinical signs, scan findings, blood test results and surveys or questionnaires. Feature-selection involves selection of productive features that are capable of providing more accurate results from the available data set. Predictive model construction involves choosing an appropriate method from the available data-mining techniques. Different testing methods can be analyzed to check the accuracy of the prediction model.

4. CONCLUSION

The proposed model can act as a decision-support system for the physicians to treat the victims diagnosed with PCOS, and eventually help them maintain a normal life style. It has an added advantage of minimal execution time. By passing through various stages in the workflow, along with the application of appropriate tasks and techniques in data mining, optimum accuracy can be achieved, which is absent in the existing defined models.

References

[1] Akash C. Jamgade” Disease Prediction Using Machine Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 05 | May 2019

[2] V Krisnaveni “A roadmap to a clinical prediction model

- with computational intelligence for PCOS ”International Journal of Management, Technology and Engineering , Volume IX, Issue II, FEBRUARY/2019.
- [3] Palvi Soni, Sheveta Vashisht. “Image Segmentation for detecting Polycystic Ovarian Disease using Deep Neural Networks” International Journal of Computer Sciences and Engineering. Vol.-7, Issue-3, March 2019
- [4] C. Gopalakrishnan, M. Iyapparaja. “A Detailed Research on Detection of Polycystic Ovary Syndrome from Ultrasound Images of Ovaries”. International Journal of Recent Technology and Engineering, Volume-8, Issue-2S11, September 2019
- [5] Ms. Palvi. “Review of Polycystic Ovary Syndrome Along with Data Mining”. IJSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 04, 2018
- [6] Vikas B et.al “A Critical Study of Polycystic Ovarian Syndrome (PCOS) Classification Techniques” IJCEM International Journal of Computational Engineering & Management, Vol. 21 Issue 4, July 2018
- [7] MadhuKumari et.al “Prevalence of Polycystic Ovary Syndrome among Women in Mumbai and Association of its Symptoms with Work Hours” International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 7, July 2017
- [8] Amsy Denny et.al “I-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques” IEEE Region 10 Conference (TENCON 2019)
- [9] Syeda Sidra et.al” Evaluation of clinical manifestations, health risks, and quality of life among women with polycystic ovary syndrome”, PLOS ONE|<https://doi.org/10.1371/journal.pone.0223329>
- October 11, 2019
- [10] Wendy M. Wolf et.al,”Geographical Prevalence of Polycystic Ovary Syndrome as Determined by Region and Race/Ethnicity ” International Journal of Environmental Research and Public Health, 15, 2589, 2018
- [11] Palvi Soni, Sheveta Vashisht,” Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques”, Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018), IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-4765-3, 2018
- [12] Dhiraj Dahiwade et.al “Designing Disease Prediction Model Using Machine Learning Approach”, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, 2019
- [13] Pahulpreet Singh Kohli, Shriya Arora, “Application of Machine Learning in Disease Prediction”, IEEE, 4th International Conference on Computing Communication and Automation (ICCCA), 2018
- [14] Mrunmayi Patil et.al ,” A Proposed Model for Lifestyle Disease Prediction Using Support Vector Machine”, IEEE – 43488, 9th ICCCNT 2018 July 10-12, 2018, IISC, Bengaluru
- [15] K. Meena, et.al, “A Literature Review on Polycystic Ovarian Syndrome and Data Mining Techniques,” vol. 4, no. 12, pp. 780–785, 2014.