

Review of Polycystic Ovary Syndrome Along with Data Mining

Ms. Palvi

Department of Computer Engineering
Lovely Professional University, India

Abstract— One of the prevalent disease that has been faced by women from last two decades is PCOS (polycystic ovary syndrome). It is also called PCOD (Polycystic ovary disease). It can become one of the biggest cause of infertility in women. Today we live in a modern world so we are available with tremendous amount of data. A recent study has revealed that about 18% of the women in India, mostly from the East, suffer from this syndrome. PCOS is characterized by overweight, irregularity in menstrual cycle, oily skin and darkened acne marks, hypertension and metabolic abnormalities/dysfunction[1]. This can become cause of type 2 diabetes. Data Mining techniques are used for predicting the same disease. In this paper we have reviewed about PCOS, early symptoms, causes, data mining tasks, techniques of data mining, data mining and medical field and the conclusion.

Key words: PCOS/PCOD, apnoea, Clustering and Classification

I. INTRODUCTION

Polycystic Ovarian Syndrome also known as PCOS or PCOD [Polycystic Ovarian Disorder] is a very common hormonal disorder and a leading cause of female infertility worldwide. PCOS is also called as Stein-leventhal Syndrome after two doctors who first described it in 1935[15]. Polycystic Ovarian Syndrome (PCOS) is a condition in which women typically have many number of small cysts around the edge of their ovaries. Polycystic ovaries mean the ovaries consists of large number of cysts that are not bigger than 8mm and develop more follicles than normal every month. Polycystic ovary start maturing at least twice as many follicles compared normal most of which enlarge[2] and mature but do not release an egg. The cysts are the egg containing follicles that do not develop properly because of hormone disturbance.

II. SYMPTOMS OF PCOS

Symptoms of polycystic ovary include increased level of androgens, irregular menses, scanty mensus, acne, excess body and facial hair, acne, pelvic pain, difficulty getting pregnant, and patches of thick, darker, velvety skin. Associated conditions include type 2 diabetes, obesity, obstructive sleep apnea, heart disease, mood disorders, and endometrial cancer. PCOS is due to a combination of genetic and environmental factors. Risk factors include obesity, not enough physical exercise, and a family history of someone with the condition. Diagnosis is based on two of the following three findings: no ovulation, high androgen levels, and ovarian cysts. Cysts may be detectable by ultrasound. Other conditions that produce similar symptoms include adrenal hyperplasia, hypothyroidism and hyperprolactinemia.

Actually PCOS has no cure. Its treatment may involve lifestyle changes such as weight loss and exercise. Birth also help with improving the regularity of periods,

excessive growth of hair and acne. Metformin medicines and anti-androgens may also help in some extent. Some duplicate treatments are used but these are temporary. Other treatments of acne and hair removal techniques may be used. Efforts to improve fertility include weight loss, clomiphene, or metformin. PCOS is the most common endocrine disorder among girls between the ages of 18 and 44. It affects estimatedly 2% to 20% of this age group depending on how it is defined. It is one of the leading causes of poor fertility. The earliest known description of what is now recognized as PCOS dates from 1721 in Italy.

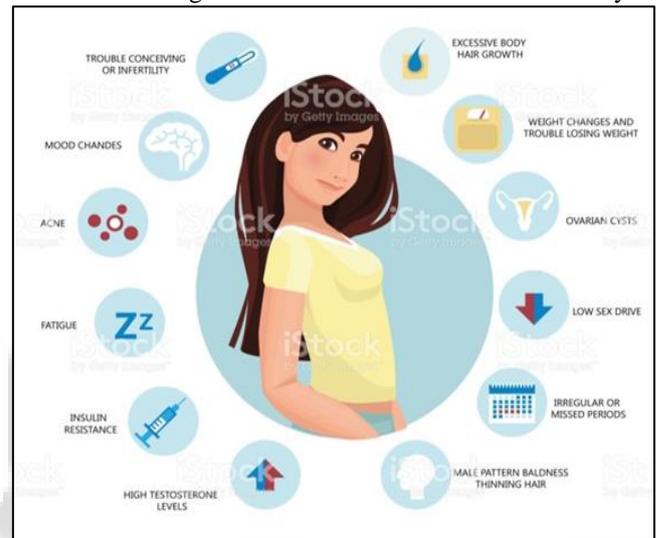


Fig.; 1: Symptoms of PCOS[14]

A. Signs and symptoms of PCOS

1) PCOS signs include three major categories:

- 1) Menstrual disorders: The basic menstrual disorders include oligomenorrhea and amenorrhea. In oligomenorrhea there are fewer than nine menstrual cycles in a year. In amenorrhea there are no menses for three consecutive months..
- 2) 2. Infertility: Chronic anovulation can become major cause of infertility. Anovulation is a state when the ovaries unable to release an oocyte during a menstrual cycle. Therefore, ovulation may hinder. However, a woman who does not ovulate at each menstrual cycle is not necessarily going through menopause.
- 3) 3. Metabolic Syndrome: It includes central obesity and other symptoms that are basically associated with insulin resistance. The risk of cardiovascular disease can be high in this. Serum insulin, insulin resistance, and homocysteine levels are higher in women with PCOS.
- 4) 4. Excessive male hormones: It is also Known as hyperandrogenism, common signs are acne and hirsutism (male pattern of hair growth, such as on the face, chin or chest), but it may produce hypermenorrhea (heavy and prolonged menstrual periods), androgenic alopecia (increased hair thinning or diffuse hair loss), or other symptoms.

III. CAUSES OF PCOS

A. Unhealthy lifestyle and diet:

PCOD/PCOS can also arise due to unhealthy food items such as white sugar, fast food, more carbs, excessive use of white flour in diet. To avoid this disease healthy diet should be followed and regular exercise should be done so as to get rid of it.

B. Hormonal imbalance:

Hormones are those substances that help the body systems to perform its function in a proper channel. Any disturbance or imbalance of the hormones can affect the body drastically. Hormonal imbalance causes problems not only with fertility and ovulation, but also with hair, skin, weight and more.

Some doctors will tell you to go for surgeries to remove the ovaries, but this will not help in the long run. Good nutrition can be boon for curbing this.

C. Laziness:

Laziness can also arise PCOS/PCOD.

D. Stress:

Overthinking and stress can lead to PCOD. A stressful life can arise problems.

E. Contraceptive pills:

If women consume contraceptive pills to curb pregnancy that can also affect reproductive system.

F. Hereditary:

PCOS/PCOD can move from generations to generations if one of the female in family is suffering from this disease, the coming generation would also be affected.

G. Obesity:

Weight can seriously affect body in many ways. If the body mass index of women is more than 24 she can most probably be the sufferer of this disease.

H. Raised level of insulin:

If insulin in the body is more than needed then PCOD/PCOS can arise.

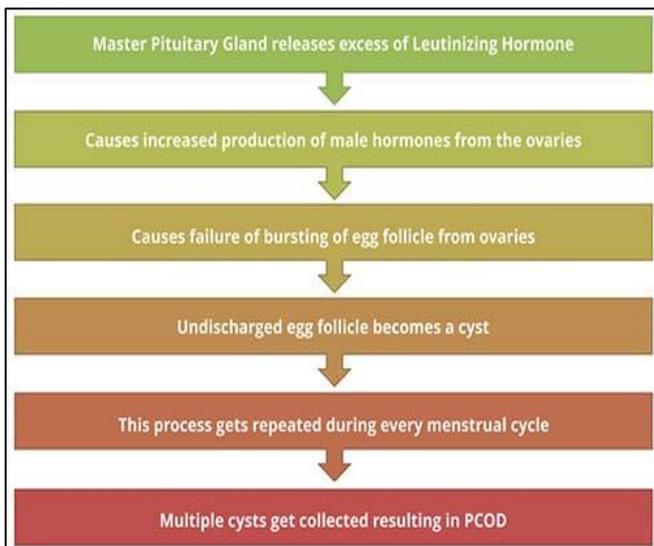


Fig. 1: Causes of PCOS[16]

IV. CAUSES OF PCOS

Data mining is the process of getting relevant data according to one's business values and interests from the large collection of data using various techniques and algorithms such as Association, Clustering and Classification[3].

The steps involved in knowledge extraction are as follows:

A. Data Cleaning:

The information obtained may contain some bugs which is preprocessed here to obtain relevant data.

B. Data Integration

Integration of the data is done here. Various steps are used to join the data from different data sets.

C. Data Selection

The data which is required is selected for user application and further steps.

D. Removal of noisy data

Since they are large amount of data it captures more space, so using this method we mitigate the space but it achieves the same results.

E. Data Mining

A knowledge discovery process to extract the essential data.

F. Pattern Evaluation:

It is the process in which a pattern is identified. Clustering can also be deployed here.

G. Knowledge Representation:

This is the final stage in which the knowledge is represented using varied visualization techniques.

V. DATA MINING TASKS

Data mining is the process of extracting hidden knowledge from various sources of data via data warehouse[4]. That knowledge can be symmetric or asymmetric depending upon the type of the data that data warehouse possesses. In today's era data mining can be considered the most valid source for solving the ambiguity of the data[5]. Data mining tasks are divided into two broad categories- Predictive and Descriptive. Data Mining based forecasts gave the capacity to recognize the rising of Polycystic Ovarian Syndrome and the discoveries can push the need to control the different parts bringing about the infection[6]. Following are the different data mining tasks that can be used in this problem prediction.

Tasks are as follows:

- Classification
- Estimation[3]
- Prediction
- Association rule mining
- Clustering
- Outlier detection[1]
- Visualization[3]

A. Classification:

The classification is one of the most vital tasks in data mining. The idea behind this is to classify the given data records into one of the many probable cases which are

known already. Classification tasks can make use of any one strategy[4]. If the data are classified without looking at the training data, this kind of classification is known as priori classification. But in converse if the data were classified with the help of training data this is known as posteriori classification

B. Estimation:

The valued outcomes of continuous variables are accomplished by the task is called as estimation [3]. To get the value for some not known continuous variables, when the input raw data is given then we can use the estimation.

C. Prediction:

The result that is conceived from the estimation or classification task is known as prediction. It differs in one of the accent. In the classification, we cannot go back later to check whether the classification was correct or not, when a phone line is used for transaction of credit card or internet access as fraudulent there the data mining is used [3]. Due to the incomplete knowledge, the uncertainty may occur out in the real world, but the relevant actions should take place earlier whether the classification may be incorrect or correct.

D. Association Rule:

In the given data set or database, the certain relationships for association connotes among the objects set and the rule used is called as an association rule. To find interesting and unique relations and patterns between variables in databases, we use association rule mining. Apriori Algorithm is one of the very commonly used algorithms for finding the different patterns[1] The set of items are denoted by P and Q, the expression in the form $P \rightarrow Q$ is used to represent the association rule which contains the literals set (called items) for each transaction in the given transactions set [3].

E. Clustering:

The number of clusters or small subgroups is formed by means of sectioning the various groups in the task known as clustering. The clustering or association of files or target groups which have similarities. Like the classification, the clustering should not swear on the predefined groups and this isolates the classification and clustering and so there will be no predefined classes for clustering [3]. Depending on the self-similarity, the grouping of records takes place.

F. Outlier Detection:

When any data entity or data input does not belong to a particular group or cluster and deviates from the usual data inputs, then such a process is called as outlier analysis. This technique is usually used to detect fraudulent transactions or noisy data in the database[1].

G. Visualization:

In the descriptive data mining, visualization of data is the most representable class. From the visual scenes, the meaning is extracted and predicted by the human being and it is more worthful than association rule only when the picture is really worth and difficult to get meaningful visualization [3].

VI. TECHNIQUES IN DATA MINING TASKS

There are following techniques in data mining that can be used.

A. Fuzzy Logic:

An environment which consists of imprecision and risk, where we actually dealing with the fuzzified values whose range lies in between 0 and 1. There are only range criteria for the getting values. The following are the essential characteristic for the fuzzy logic:

- The limiting case of the approximate logical reasoning is taken for the exact reasoning in the fuzzy logic[3].
- It is the matter of degree for everything in the fuzzy logic.
- It can be fuzzified the logical system using ranges[3].
- The collection of equivalently or elastic, constraint of fuzzy on the variables collection is described as knowledge in the fuzzy logic.

B. Rough Set Techniques:

By using rough set techniques, the main goal is to obtain learning from the approximation of concepts. It gives us a scope to find out latent patterns within the data provided. It can be used for getting features and patterns and also for generation of rules used in decision tree[1].

C. Machine Learning:

In the models of machine learning, there will be information of non-quantifiable. The independence of variables and parameters is divided and preassumed by the methods or models. It is an important part of Artificial intelligence in which the computers achieved the ability to learn without being programmed. These can be supervised learning or unsupervised learning.

1) Neural Network:

Neural Network or an artificial neural network is a biological system that detects various patterns and makes predictions according to defined pattern. The greatest breakthroughs in neural network in recent years are in their application to real world problems like customer response prediction, fraud detection etc[7]. Data mining techniques such as neural networks used for increasing business intelligence across a variety of business applications. These days Neural Networks are used in a various applications. Artificial neural network have become very powerful tool in tasks like speech recognition, pattern recognition, decision problem or predication applications.

2) Decision Trees:

A decision tree is a flow chart like structure where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes or class distribution[7].

VII. DATA MINING AND MEDICAL FIELD

Very general ideas about various kinds of classification algorithms that are basically used in data mining. For predicting precise and true results classification techniques are used[8]. Criteria sensitivity is deployed for predicting performances. K-Means clustering is increasing the proficiency of the output. This is the most efficient technique to predict the dengue patients with serotypes and dengue dataset was fully clustered.

Author predict infertility problem in women[9]. Predictive Modelling techniques are used to diagnose various infertility issues in women. J48 Algorithm was also deployed here after applying both random forest and J48 algorithm, random forest considered to be the best for detecting this particular problem.

Author defined various categories of heart problem. Here prediction of heart attack problem is resolved and techniques for predicting heart problem is described by using two important algorithms that are Naïve Bayesian and Decision tree and accuracy of two algorithm is calculated[10]. Decision tree is considered as the best among two.

In this paper the author proposed various types of techniques for predicting diseases. Here decision tree, rule induction and naïve Bayesian classifier has been used so that to predict efficiency of particular algorithm[11]. In this author considered Naïve Bayesian classifier to be the best among three. Decision tree according to author predict least accuracy.

Author proposed methodology for both fertile and infertile patients. WEKA was used for applying various data mining techniques like Statistical analysis, Associative rule mining, Clustering, Classification and Subset evaluation. This proved helpful in extracting key information regarding infertility[3]. Two classification methods were used on the same record set to produce almost same kind of results at different levels of accuracy. Among them, J48 pruned tree has been found to be more accurate[12]. Clustering is also carried out to verify the output of previous methods. From the information gained an attempt is also made to build a decision tree model for prediction of infertility. The accuracy of prediction is 86%.

In this survey paper, authors reviewed the disease called PCOS (Polycystic Ovary Syndrome) and their symptoms, treatment for the PCOS also overviewed the data mining tasks and techniques that are present. In future research, a new algorithm using Rough set Theory and Data Mining techniques is to be proposed for effective classification, accuracy of the positive factors causing Polycystic Ovarian Syndrome[3]. By proposing a hybrid approach of combining use of rough set theory and artificial neural network for prediction of PCOD, the Rough set theory will be used as preprocessing tools for the input data and Artificial Neural Network will be used as predictor.

VIII. CONCLUSIONS

In this survey paper, we reviewed the disease called PCOS (Polycystic Ovary Syndrome) and their symptoms, causes for the PCOS also overviewed the data mining tasks and techniques that are present. In future research, a new algorithm using Data Mining techniques is to be proposed for effective classification, accuracy of the positive factors causing Polycystic Ovarian Syndrome.

ACKNOWLEDGEMENTS

The author would like to devote her sincere thanks to Almighty God and her family.

REFERENCES

- [1] "A Literature Review on The R Ising Phenomenon," vol. 10, no. 2, p. 22311963, 2017.
- [2] A. A. Sawant, S. Patil, and S. Shah, "Review Article Review on PCOD / PCOS & its Treatment in different Medicinal Systems – Allopathy, Ayurveda, Homeopathy," vol. 1, no. 1, pp. 1–16, 2017.
- [3] K. Meena, M. Manimekalai, and S. Rethinavalli, "A Literature Review on Polycystic Ovarian Syndrome and Data Mining Techniques," vol. 4, no. 12, pp. 780–785, 2014.
- [4] S. K. Arts, "A Survey on Data Mining Classification Algorithms," no. July, pp. 264–268, 2017.
- [5] K. Meena, "A Novel Framework for Filtering the PCOS Attributes using Data Mining Techniques," vol. 4, no. 1, pp. 702–706, 2015.
- [6] S. Rethinavalli and M. Manimekalai, "A Hypothesis Analysis on the Proposed Methodology for Prediction of Polycystic Ovarian Syndrome," vol. 6, no. 11, p. 2016, 2016.
- [7] N. Jain and V. Srivastava, "DATA MINING TECHNIQUES: A SURVEY PAPER," pp. 116–119, 2013.
- [8] P. Manivannan, "Dengue Fever Prediction Using K-Means Clustering Algorithm," pp. 1–5, 2017.
- [9] M. S. Simi and S. Nayaki, "Data Analytics in Medical Data: A Review," pp. 2–5, 2017.
- [10] P. N. and P. R. Dr Ravi Kumar, "Usage of Data mining techniques in predicting the Heart diseases – Naïve Bayes & Decision tree," Int. Conf. circuits Power Comput. Technol. [ICCPCT], 2017.
- [11] A. H. Shurrab and A. Y. A. Maghari, "Data Mining Techniques," pp. 625–631, 2017.
- [12] M. Computing, "Data Mining To Elicit Predominant Factors Causing," vol. 5, no. 8, pp. 5–9, 2016.
- [13] <https://www.health-total.com/pcod/causes/>
- [14] <https://www.istockphoto.com/in/vector/pcos-symptoms-infographic-gm925882154-254069364>
- [15] https://en.wikipedia.org/wiki/Polycystic_ovary_syndrome
- [16] https://www.welcomecure.com/assets/images/disease_conditions/Polycystic-Ovary-Syndrome-PCOS/PCOS-%20Causes.jpg