Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Literature Review on Polycystic Ovarian Syndrome and Data Mining Techniques

[1]Dr. K. Meena, [2]Dr. M. Manimekalai, [3]S. Rethinavalli
[1]Former Vice-President, Bharathidasan University, Trichy, India
[2]Director and Head, Department of Computer Applications, Shrimati Indira Gandhi College, Trichy, India
[3]Assistant Professor, Department of Computer Applications, Shrimati Indira Gandhi College, Trichy, India

Abstract— The women's in the age of reproduction, the heterogeneous, complex and multifaceted disease called as PCOS (Polycystic ovary syndrome) affects about 5 to 10 % of the women. Metabolic syndrome (frequent metabolic traits) called as hypertension, hyperinsulinemia, abdominal obesity and dyslipidemia and these are characterized along with the resistance of insulin like chronic anovulation, hyperandrogenism and polycystic ovaries which ends with the serious diseases like coronary disease, endometrial hyperplasia and type 2 diabetes mellitus these are the long term consequences. Anovulatory infertility is commonly most caused by means of the one of the above. Depending on the ethnic background, the PCOS is variable with the phenotype of women and it is mostly determined by metabolic alterations and obesity and it goes on to the postmenopausal age starting from the adolescence which change during the life span due to the PCOS heterogeneous clinical features. For decision making purpose in PCOS phenotype (i.e to discover the main cause of PCOS with many number of attributes (tests)), an approach that used to store and manipulating the valued data for predicting the causes earlier and this process in information technology is called as data mining. In this paper, we have reviewed the causes, syndrome of the PCOS, treatment for PCOS and the data mining techniques that used to predict the factors which leads to severe disease in future by means of analyzing the important factors and classify them for better result.

Keywords— PCOS, Artificial Neural Network, GA, Fuzzy, Rough Set, SVM.

I. INTRODUCTION

The most common endocrinological issue which pretending the women is PCOS (Polycystic ovary syndrome). On ultrasonography the evidence in the ovary, that is the following changes may include such as showing of increased levels of serum androgen, hirsuitism acne, amenorrhoea or oligo, morphological change and anovulation, these characteristic are demonstrated with the patients with PCOS. The agreed criteria in Rotterdam 2003 and this is current practice uses diagnostically [1]. The higher prevalence for the patients with PCOS are obese than the general population about 50%. The long-term morbidity is resulted by means of insulin resistance due to the condition of the metabolic element.

In the ovaries, the clusters of pearl-sized and small cysts frequently induces PCOS and because there a many number of cysts it is referred as Polycystic. The immature eggs are contained by the cysts, which were in the form of fluid-filled. Androgen produces a large number of male hormones, and which is contributed in PCOS and this is referred as symptoms for women [2]. In the adult age, the PCOS phenotype development brings a major role in the programming of utero fetal and it is conceived that it might be the cause for PCOS and it should be clarified. The phenotype of PCOS last expressions and the result of the menstrual disturbances and characteristic metabolic are due to the interaction of the genetic factors with the obesity [3] (environmental factors) which leads the women for developing of PCOS and it is genetically inclined.

For proposing conclusions, supporting the decision making and the useful information is highlighted by means of the process of auditing, cleaning, metamorphosing and patterning the data is called as the data analysis [4]. In the area of science, social science and the different business, the broad various techniques below the layout of names which has multiple approaches and facts is known as data analysis [5]. The unknown information is disclosed in a database which is detected by the data mining. Associations, classification, clustering, prediction and these functions are included in the data mining [6]. From the many fields like pattern recognition [7], information retrieval [8], neural networks [9], analysis of spatial data [10], database technology [11] and statistics [12], and these techniques have to be integrated for the processes of data mining. A repetitious process with various steps is called as data mining.

II. SYMPTOMS OF POLYCYSTIC OVARY SYNDROME (PCOS)

With the collection of different symptoms and signs, a syndrome disease called PCOS is defined. The symptoms for PCOS might be varying from patient to patient [13].

- The menstrual periods may be missing or it may be irregular.
- The women with PCOS may experience infertility.
- The unwanted or excess growth of hairs in the body or face. On the scalp, the hair might be thinning.

- Around the waist, the weight may be increased or there will be problem of weight losing.
- The skin and acne may be darkened which may include the tags in the skin.
- The women with PCOS may lead to the severe diseases such as problems in the blood vessels, heart, diabetes, cancer in uterine and sleep apnea.

III. TREATMENT FOR PCOS

It is required to prevent the problems of women with PCOS, since no heal for PCOS. Depending on the symptoms, the treatment for PCOS may be vary, if the women want to get pregnant or not, and chances for diabetes and heart disease with lower risk [14]. The women with PCOS might undergo the following treatments. The treatment for PCOS may include:

A. Alteration in the Lifestyle

The women with PCOS may experience the health problems; it is due to obese or overweight. The PCOS can be handled at the healthy level by means to keep your weight with exercising and eating the healthy foods [15]. The following are the tips to include for eating the healthy foods:

• The processed foods and the sugars added foods can be limited.

• For diet, the meats with thin foods, vegetables, fruits and whole-grain products should e added [17].

• For normalize the levels of hormone and use of insulin in the body can be improved, the above can be assisted the sugar levels (lower blood glucose) should be reduced [16].

B. Using Pills for Birth Control

The birth control pills can be suggested for the women for not getting pregnant [18]. The pills can do the following:

• The cycles for menstruation can be controlled.

• The levels for male hormone can be reduced.

• The acne can be cleared.

If the taking of pills is stopped, there should be the abnormal gain the menstruation cycle [19]. The progesterone is suggested for reducing the endometrial cancer risk and menstruation cycle is controlled, but it should not help to growth hair and reduce acne for the women with PCOS.

C. Medicine for Diabetes

For treating the diabetes of type 2 [20] [22], the medicine called Glucophage (Metformin) can be used. The production of testosterone can be lowered and sugar (blood glucose) is controlled by Metformin [21] which affects the insulin. The ovulation can be bring back after the few months of use and the growth of abnormal hair is dumbed. The levels of cholesterol can be improved and the body mass is decreased is the positive effects of the metformin and it has been showed in the recent research.

D. Medicine for Fertility

The women with PCOS might experiences the problem of fertility due to the ovulation lacking is the main reason [22]. The ovulation can be stimulated by means of the medicine for the women to become pregnant. The risk can be increased for the multiple births like triplet, twins in some medicine for fertility [23]. The following are the options of treatment:

• For most patients, the ovulation is stimulated by Clomiphene and it is the therapy of first choice.

• With the lower doses of medicines like the combination of Clomiphene with Metformin for the ovulation of women with PCOS [24] if only when the failure of clomiphene alone.

• When compared to clomiphene, the multiple births risk can be evoked in this treat ent which is called as Gonadotropins but it is more expensive when it is given as shots.

E. Drilling of Ovarian Surgery

When there is no respond for above medicine for fertility, there may be chance for ovulation is increased by a surgery called "Drilling of Ovarian" [25]. A small tool like a telescope can be inserted into the stomach (abdomen) by means of making a small cut in the upper and lower part of belly button (navel) and this technique is called as. The small portion of the ovary can be destroyed by a small needle which holding an electric current. On the ovary, there may be chance for developing a scar tissue by using this technique. It helps the ovulation by reducing the male hormone levels [26].

F. Preventing Extra Male Hormones or Excess Growth of Hair Medicine

For clear acne and the hair growth can be reduced by the Anti-androgen medicines. On hair growth, the impact of male hormones can be reduced by Spironolactone which is used for treating the high blood pressure. With the pills of birth control the anti-androgen is combined. If the women attempting to get pregnant, then the medicine should not be taken [27].

IV. DATA MINING TASKS

For the following specific set of six activities or tasks, the data mining is to be used. i.e

- Classification
- Estimation

- Prediction
- Association Rules or Affinity Grouping
- Clustering
- Visualization and Description

A. Classification

Assigning a new presently object to a class of predefined which is examined and it is contained by the technique called Classification [28]. The set of predefined examples are contained in the training set and the well-defined classes are used to characterize the task of classification. In order to classify the unclassified data, the model that can be utilized is build by using the above task.

B. Estimation

The valued outcomes of continuous variables are dealt by the task is called as estimation [29]. To get the value for some unknown continuous variables, when the input 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 [30]. 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. The set of items are denoted by X and Y, the expression in the form X Y is used to represent the association rule which contains the literals set (called items) for each transaction in the given transactions set [31].

E. Clustering

The number of clusters or subgroups are formed by means of sectioning the various group in the task called as clustering. Like the classification, the clustering should not swear on the predefined classes and this separates the classification and clustering and so there will be no predefined classes for clustering [32]. Depending on the self similarity, the grouping of records takes place.

F. Visualization

In the descriptive data mining, visualization of data is the most powerful class. From the visual scenes, the meaning is extracted and predicted by the human being and it is thousand times worth than association rule only when the picture is really good, and tuff to get meaningful visualization [33].

V. TECHNIQUES IN DATA MINING

From the data collection, the useful information is derived for the each of the above problems. The methods that are used to solve these problems is developed in the field like fuzzy sets, hybrid or rough sets, machine learning, statistics and the data mining is relevant for these fields, especially for the exploration of the conceptual data [34].

A. Fuzzy Logic

An environment of imprecision and uncertainty, the problem for representation of knowledge is dealt with an effective framework for conceptation is provided with the extension of the classical logic system and this extension is viewed as the fuzzy logic [35]. The following are the essential characteristic for the fuzzy logic:

- The limiting case of the approximate reasoning is considered for the exact reasoning in the fuzzy logic.
- It is the matter of degree for everything in the fuzzy logic.
- It can be fuzzified the logical system.
- The collection of equivalently or elastic, constraint of fuzzy on the variables collection is interpreted as knowledge in the fuzzy logic.

For the classification, in many hybrid techniques, fuzzy logic is the most practicable concept, where it is in pure form will not be used for classification.

B. Rough Set Techniques

On the impression on the inability and indiscernbility, the distinguishing of objects takes place using empirical data and it is used for the construction of the binary relations with the estimation of concepts or sets which is dealt by a theory called as Rough Sets [36]. To construct the model, bottom-up approach is used for such estimation for the making of our targets. This rough set have many applications in the field of data mining, they are as follows::

- For attribute value table it is used as the induction of decision rule.

- Generation of template for filtration of data – Basis on the relation of equivalence, the elementary blocks are evoked from the data. In this stage for searching, sometimes the generic algorithms are used.

C. Machine Learning

In the models of machine learning, there will be information of non-quantifiable, incorporating subjective difficulty is meant by statistical methods. The independence of variables and parameters is distributed and assumed by the methods or models [37]. Accuracy for predicting is comparable (better than others) for the machine learning is concluded by the various studies. Like statistical methods, it is also free from structural assumptions and parameters, when compared to statistical methods it results in good performance in the attributed fact [38]. The interpreting of result in the data analysis is the problem for statistical approach is another weakness. The techniques for machine learning are as follows:

1) Artificial Neural Network

Like the biological networks of neuron, the patterns are arranged by the non-linear processing elements composition is called as computational models of artificial neural networks [39]. Each node with an activation value whereas each connection has a weight value of neural network. In the massive parallelism, through the connection of network the data propagation takes place and evoking of the nodes by the activation function. With the adjustments of connection weight, the training of the network takes place [40].

2) Genetic Algorithm

The mechanics of natural genetics and natural selection basis, the search algorithm called genetic algorithm works. With the modern genius of human search to form a search algorithm for randomized exchange of information and this is yet to be designed and in the genetic algorithm with the structured among the fittest of string structures is combined for the survival [39]. Using the fittest of the old pieces and bits, a set of strings is created in the each generation and it is measured for good. It is not a simple random walk when the genetic algorithm is randomized. When the expected performance improvement is needed, on the points of new search, the historical information exploitation works efficiently. Reproduction, crossover and mutation are three important operators that give the good result in genetic algorithm [35]. The following are procedures for searching and the difference of GA's than other normal optimization.

- The working of GA's is not on the parameters itself but with the set of parameter coding.
- Searching of GA's is not a single point but population of points.
- GA's not using any auxiliary knowledge or derivatives but it is uses information of objective function.
- It is not the rules of deterministic but uses the rules of probabilistic transition in GAs.

3) Support Vector Machine

The tasks of regression estimation and binary classification are performed by the learning machines called Support Vector Machines (SVMs). Due to the two important components, it becomes familiar like learning and classification [41]. In the first components, it minimizes the expected error than the classification error minimization like the other techniques of classification. SVM admits the computational methods in efficient by using the mathematical programming theory of duality for the dual problem in the second component.

4) Decision Tree

From the given dataset, for different classes a model is represented by generating the tree and a set of rules using a classification scheme called a DT (Decision Tree). A tree structure like flow chart with internal node, branch and leaf node, where test on an attribute is represent by an internal node, and outcome of the test by the branch and class distributions or classes are represented by the leaf node is called a DT [42]. The root node is the top most nodes of the tree in DT.

D. Statistics

Knowledge abstraction from the data is the problem, which is cope up by the statisticians and the papers for first artificial intelligence were also published. The correlation between the two or more variables is analyzed by the statistical tools in the analysis of correlation is an example. Vector of values are used to describe the clusters which is in the large set of objects by the methods called as analysis of cluster [43]. The important variables which describe the clusters by the point called analysis of factor. Logistic regression, CART, Naïve Bayes, Linear Discriminant, K-nearest Neighbour and Linear Discriminants are the popular techniques for the classification tasks of supervised learning.

VI. CONCLUSION

In this survey paper, we 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. 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 pre--processing tools for the input data and Artificial Neural Network will be used as predictor. The prediction accuracy of conventional statistical methods and proposed hybrid approach of rough neural network method will be compared and evaluated. By comparative study it is suspected that computer techniques handled through the newly proposed ANN algorithm Data Mining promise and address real factors of the Polycystic Ovarian Syndrome problem.

REFERENCES

[1] Bart C.J.M.Fauser, "Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS)", pp.1-17.

[2] Ricardo Azziz, Enrico Carmina, Didier Dewailly, Evanthia Diamanti-Kandarakis, Hector F. Escobar-Morreale, Walter Futterweit, Onno E. Janssen, Richard S. Legro, Robert J. Norman, Ann E. Taylor and Selma F. Witchel, "The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report", Fertility and Sterility_ Vol. 91, No. 2, February 2009 0015-0282, American Society for Reproductive Medicine, Published by Elsevier Inc.

[3] T. M. Barber, M. I. McCarthy, J. A. H. Wass and S. Franks, "Obesity and polycystic ovary syndrome", Clinical Endocrinology (2006) 65, 137–145.

[4] Andrew Kusiak, "Data Mining and Decision Making", Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV, Vol. 4730, SPIE, Orlando, FL, April 2002, pp. 155-165.

[5] Mohammed J. Zaki, Wagner Meira, "DATA MINING AND ANALYSIS- Fundamental Concepts and Algorithms", pp. 1-25.

[6] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare", Journal of Healthcare Information Management — Vol. 19, No. 2, pp. 64-72.

[7] Noriyasu Homma, "Pattern Recognition in Medical Image Diagnosis", pp. 319-336.

[8] Ammar Yassir and Smitha Nayak, "Issues in Data Mining and Information Retrieval", International Journal of Computer Science & Communication Networks,Vol 2(1), 93-98.

[9] Dr. K. Usha Rani, "Analysis Of Heart Diseases Dataset Using Neural Network Approach", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.5, September 2011.

[10] Diansheng Guo, Jeremy Mennis, "Spatial data mining and geographic knowledge discovery—An introduction", Computers, Environment and Urban Systems 33 (2009) 403–408.

[11] Amir Netz, Surajit Chaudhuri, Jeff Bernhardt, Usama Fayyad, "Integration of Data Mining and Relational Databases", Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt, 2000, pp. 719-722.

[12] David J. Hand, "Statistics and Data Mining: Intersecting Disciplines", SIGKDD Explorations, ACM SIGKDD, June 1999 Volume 1, Issue 1, pp. 16-19.

[13] "Polycystic Ovary Syndrome", The American College of Obstetricians and Gynecologistics.

[14] Richard S. Legro, Silva A. Arslanian, David A. Ehrmann, Kathleen M. Hoeger, M. Hassan Murad, Renato Pasquali, and Corrine K. Welt, "Diagnosis and Treatment of Polycystic Ovary Syndrome: An Endocrine Society Clinical Practice Guideline", Journal of Clinical Endocrinology & Metabolism, December 2013, JCEM jc.2013–2350.

[15] Marja Ojaniemi and Michel Pugeat, "An adolescent with polycystic ovary syndrome", European Journal of Endocrinology (2006) 155 S149–S152.

[16] "Polycystic ovary syndrome: what it means for your long-term health Information for you", The American College of Obstetricians and Gynecologistics.

[17] "Polycystic Ovary Syndrome (PCOS)", The aAssociation of UK Dietitians.

[18] Frederick R. Jelovsek, "Which Oral Contraceptive Pill is Best for Me?", pp.1-4.

[19] "For Women with Irregular Menstrual Cycles", Women's Health At Mass General.

[20] Daniela Jakubowicz, Julio Wainstein and Roy Homburg, "The Link between Polycystic Ovarian Syndrome and Type 2 Diabetes: Preventive and Therapeutic Approach in Israel", IMAJ, VOL 14 July 2012, pp.442-447.

[21] "Metformin and Pregnancy", Organization of Teratology and Information Specialists.

[22] Shirley S. Wan G, "Infertility, Diabetes, Obesity and the Mystery of PCOS",pp.1-4.

[23] Seddigheh Esmaeilzadeh, Mouloud Agajani Delavar, Zahra Basirat, Hamid Shafi, "Physical activity and body mass index among women who have experienced infertility", Fatemezahra Infertility and Reproductive Health Research Center, Department of Obstetrics and Gynecology, Babol University of Medical Sciences, Babol, Iran, pp. 499-505.

[24] Stephen Franks, "Polycystic Ovary Syndrome", The New England Journal Of Medicine Sept. 28, 1995 Vol. 333 No. 13, pp. 853-861.

[25] Dimitrios Panidisa, Konstantinos Tziomalosb,Efstathios Papadakisa, Ilias Katsikisa, "Infertility Treatment in Polycystic Ovary Syndrome: Lifestyle Interventions, Medications and Surgery", Front Horm Res. Basel, Karger, 2013, vol 40, pp 128–141.

[26] Shaveta M Malik, Michael L Traub, "Defining the role of bariatric surgery in polycystic ovarian syndrome patients", World J Diabetes 2012 April 15; 3(4): 71-79.

[27] Lyndal Harborne, Richard Fleming, Helen Lyall, Naveed Sattar and Jane Norman,"Metformin or Antiandrogen in the Treatment of Hirsutism in Polycystic Ovary Syndrome", Journal of Clinical Endocrinol Metab, September 2003, 88(9):4116–4123.

[28] Divya Tomar and Sonali Agarwal, "A survey on Data Mining approaches for Healthcare" International Journal of Bio-Science and Bio-Technology Vol.5, No.5 (2013), pp. 241-266.

[29] Parvathi I, Siddharth Rautaray, "Survey on Data Mining Techniques for the Diagnosis of Diseases in Medical Domain", International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 838-846.

[30] V. Manikantan and S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering (IJACTE), Volume-2, Issue-2, 2013 2319 – 2526.

[31] Mahmood A. Rashid, Md Tamjidul Hoque, Abdul Sattar, "Association Rules Mining Based Clinical Observations", pp.1-5.

[32] Anoop Jain, Aruna Bajpai, Manish Kumar Rohila, "Efficient Clustering Technique for Information Retrieval in Data Mining", International Journal of Emerging Technology and Advanced Engineering, pp.12-20.

[33] Li Wei, Nitin Kumar, Venkata Lolla, Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana and Helga Van Herle, "A Practical Tool for Visualizing and Data Mining Medical Time Series", pp.1-6.

[34] Chris Rygielski, Jyun-Cheng Wang, David C. Yen, "Data mining techniques for customer relationship management", Technology in Society 24 (2002) 483–502.

[35] K. Rajeswari, Mahadev Shindalkar, Nikhil Thorawade, Pranay Bhandari, "DSS Using Apriori Algorithm, Genetic Algorithm And Fuzzy Logic", Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 4, Jul-Aug 2013, pp.132-136.

[36] R.Saravana Kumar and G.Tholkappia Arasu, "Rough Set Theory And Fuzzy Logic Based Warehousing Of Heterogeneous Clinical Databases", pp.1-22.

[37] Tapas Ranjan Baitharu, Subhendu Kumar Pani, "A Survey on Application of Machine Learning Algorithms on Data Mining", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-3, Issue-7, December 2013,pp. 17-20.

[38] George D. Magoulas and Andriana Prentza, "Machine Learning In Medical Applications", pp.1-7.

[39] Markus Brameier and Wolfgang Banzhaf, "A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining", IEEE Transactions On Evolutionary Computation, Vol. 5, No. 1, February 2001, 17-26.

[40] Dr. Yashpal Singh, Alok Singh Chauhan, "Neural Networks In Data Mining", Journal of Theoretical and Applied Information Technology, pp. 37-42.

[41] Vanitha.L and Venmathi.A, "Classification of Medical Images Using Support Vector Machine" 2011 International Conference on Information and Network Technology IPCSIT vol.4 (2011), pp. 63-67.

[42] D.Lavanya and Dr. K.Usha Rani, "Performance Evaluation of Decision Tree Classifiers on Medical Datasets", International Journal of Computer Applications (0975 – 8887) Volume 26– No.4, July 2011, pp. 1-4.

[43] Prof. Dr. P. K. Srimani, Manjula Sanjay Koti, "Outlier Mining In Medical Databases By Using Statistical Methods", International Journal of Engineering Science and Technology (IJEST), Vol. 4 No.01 January 2012, pp. 239-246.

