DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5393

Volume 9, No. 1, January-February 2018

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A LITERATURE INSPECTION ON POLYCYSTIC OVARIAN MORPHOLOGY IN WOMEN USING DATA MINING METHODOLOGIES

Neetha Thomas
Research Scholar
Department of Computer Science
Kongunadu Arts and Science College
Coimbatore, Tamil Nadu, India

Dr. A.Kavitha, Ph.D
Associate Professor
Department of Computer Science
Kongunadu Arts and Science College
Coimbatore, Tamil Nadu, India

Abstract: Polycystic Ovarian Morphology (PCOM) or Polycystic ovary syndrome (PCOS), is an endocrine disorder, which can be seen in most of the women nowadays .It affects 5–20% of women of child bearing age. Polycystic means multiple cyst, which are fluid sacs inside overies.It is a condition in which the egg never gets matured for ovulation. This is a hormonal imbalance in women during their puberty, i.e., the age between 15 – 44.Inspected through various research papers including medical papers, as a result ,this paper survey about causes of syndrome, women hormonal level, various treatments, and associated diseases of PCOS etc.The study also incorporates different methodologies used in various areas in medical mining.

Keywords: PCOS, Type 2 diabetics, CVD, KDD, Data Mining.

1. INTRODUCTION

PCOS is characterized as hyperandrogenism, excess production of androgen by ovaries; leads to ovulatory dysfunction and polycystic ovarian morphology (PCOM).PCOS can be lean or obese. Obese women with ovarian syndrome have a preeminent chance of menstrual aberration and hirsute than lean women. They tend to be insulin resistant (IR).It is a condition in which your body's cells become resistant to the effects of insulin, which can increase the risk of type-2 diabetes and cardiovascular disease, metabolic dysfunction etc.It is a characteristic which is seen in both obese and lean PCOS women regardless of weight or BMI [1].But to conform can say that long-term insulin resistance eventually leads to diabetes. The risk of type-2 diabetes mellitus (T2D) among PCOS patients is 5 to 10 fold higher than normal. Reproductive age women [2] with metabolic syndrome or type 2 diabetes mellitus are at high risk for Cardio Vascular Diseases (CVD).

2. HORMONAL LEVEL IN A WOMEN

PCOS is a plight that leads to small, cysts which are harmless will develop on ovaries. These cysts caused with the hormonal imbalance [3].The typically affected hormones are, Androgens Estrogen, Progesterone, Testosterone, (LH) Luteinizing Hormone.

Where pg/mL is picograms per millilitre, ng/mL is nanograms Per Millilitre and IU/L stands for International units/litre. Testosterone, produced in small amounts in women's bodies mostly by the ovaries and adrenal glands. Total testosterone is bound with proteins in the blood and free Testosterone is a free hormone [4].

The over production of Andergon hormone,a male hormone,in female body that lead to hyperandrogenism,So increased testosterone in a woman's body can abolish normal menstruation and ovulation

Table 1.Normal level of hormones in women

Estrogen	Follicular Phase	30-120 pg/ml
	Ovulatory Peak	130-370 pg/ml
	Luteal Phase	70-250 pg/ml
Progestrone	Follicular phase	0.2-1.4 ng/ml
	Luteal phase	4 – 25 ng/ml
	Post-Menopausal	0.1 – 1 ng/ml
Testosterone	Total Teststerone	0.1 - 1.2 ng/ ml
	Free Teststerone	0.3-1.9 ng /ml
Luteinizing Hormone	Luteal phase	0.5 to 16.9 IU/L
	Follicular phase	1.9 to 12.5 IU/L
	Ovulatory Peak	8.7 to 76.3 IU/L

3. SYMPTOMS AND ASSOCIATED DISEASES OF POLYCYSTIC OVARIAN SYNDROME

According to the studies [5], diagnosis of PCOS varies,39% have elevated testosterone level, 26% have the sign of hyperandrogenisam and elevated testrone,59% are diagnosed with obesity,15% have irregular menstruation cycle as well metabolic risk are also elevated on PCOS patients. Variability is vivid in diagnosing PCOS and associated co-morbidities.

The major symptoms of PCOS including:-

1. Hirutisam (over growth of hair, on the face, chest, buttocks)
2. Acne on face ,dark patches on facial skin
3. Alopecia, thinning hair on scalp or male pattern baldness
4. Headaches

5. Sleep Apnea.
6. Infertility, missing of menstruation (oligomenorrhea) [6] for more than 35 days.
7. Anxiety or higher risk of depression in women etc
8. Menstrual cycle fluctuations, like absent periods, irregular menstrual cycles,
9. Light or heavy bleeding that varies monthly.
10. Deepening of the voice.
11. Decreased breast size.
12. Preeclampsia, It is marked by high blood pressure in women
13. Cancer [7]

a) Endometrial cancer
b) Ovarian cancer
c) Breast Cancer

14. For pregnant women with PCOS : it leads to[7],

a) Abortion
b) Gestational Diabetes Mellitus
c) Disorders in Pregnancy
d) Preterm child birth,etc

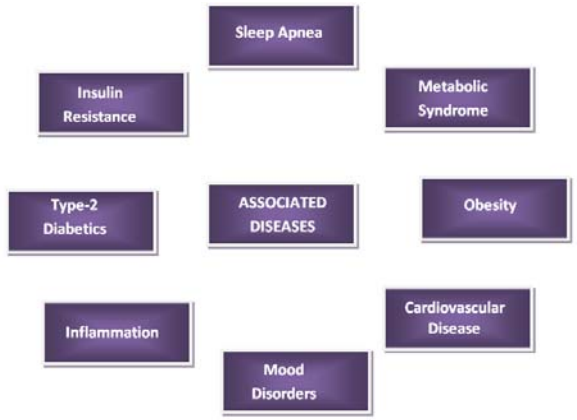

Figure 1 Associated Disease of PCOS

Figure 1 includes a complex number of other symptoms in addition to the symptoms related to the reproductive system are mentioned above [8][9].

4. MANAGEMENT OF WOMEN WITH PCOS

The polycystic ovary Syndrome (PCOS) can be ingrained as 12 or more ovarian follicles, which are 2–9mm in size, and/or a total ovarian volume of more than 10cm [10].PCOS treatment varies according to the patient's physical health and symptoms here listed a few treatment methods or medicine for PCOS.

1. Weight reduction
2. Metformine: Studies are demonstrated that metformine improved ovulatory function in adolescents as well as hyperandrogenisam [11]
3. Clomiphene citrate (CC): To encourage ovulation
4. Laparoscopic Ovarian Drilling (LOD): A safe and effective surgical treatment [12].it is also named as Laparoscopic ovarian diathermy
5. Life style Management: A healthy diet and regular exercise can help lessen the symptoms of PCOS.
6. Gonadotropins: induce ovulation, help to achieve capable follicle for fertilization [7].
7. Treatment of menstrual dysfunction
8. Spironolactone, blocks androgen receptors [7].
9. Cyproterone acetate (CPA), progestin agent which suppresses androgen action [7].
10. Myo-inositol: Treating women with myo-inositol is to reduce androgen levels.
11. Clomid: which stimulates ovulation [13]

5. LITERATURE SURVEY

Medical data mining plays a major role in health care industry. It enables health systems, by analyzing patterns embedded in huge data and converted into meaning full information. Different types of methodologies are incorporated that help to understand medical data by differentiating pathological data and normal data. The following papers are surveyed.

M. Alotaibi,A. Alsinan,[14]did a case study on PCOS in Gulf Countries. They presented the system using mobile health care technologies. Analysis has been done using statistical software SPSS.Sandy Rihana et al,[15] proposed an automated cysts detection and classification in the ovary using multiscale morphological and SVM.

Sharvari S. et al, [16] proposed an automated detection of PCOS by calculating no of follicles in ultrasound ovarian image and assimilate both clinical and biochemical imaging parameters so as to classify patients into normal and PCOS affected. Features are extracted using Multiscale morphological approach and Classification of all the data is done using Support Vector (SVM) algorithm.Yinhui Deng et al [17],introduced filtering method named adaptive morphological filter and watershed algorithm to extract outlines of targets, a clustering method is applied to identify expected follicular cysts.Palak Mehrotra et al, [18] written a paper in Automated Ovarian Follicle Recognition for Polycystic Ovary Syndrome used the techniques like multiscale morphological approach for contrast Enhancement and scanline thresholding is used to extract the contours of the follicles.

S. Sheela et al[19], mentioned about theoretical investigations on PCOS.They used Wiener filters for the removal of speckle noise, extraction of region of interest using segmentation, classify images in maximum accuracy to detect ovarian cyst in short time.Bedy Purnama et al [20], did a classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images by using the methodology feature extraction using Gabor wavelet. The images of follicle are categorized into two groups of texture features as dataset A and dataset B [20].

The above discussed paper fall under PCOS medical image processing.

S. Rethinavalli et al[21] proposed a paper on PCOS prediction using the type of menstruation,Orginal dataset containing 32 attributes ,later reduced the attributes to 7 using feature selection in the pre-processing step.NFRS (Neural Fuzzy Rough Set)and ANN algorithm are used to compare PCO syndrome and without PCO Syndrome. Chi-square test proved the correlation between types of menstruation and PCOS disease.

Various other medical mining areas includes the prediction of coronary heart disease,diabetics,MetS(Metabolic Syndrome),kidney disease, breast cancer etc.Some of the related medical mining works are also added.

Uma Ojha et al[22],had a study on the prediction of breast cancer,they used Eight data mining methods Kmeans,EM,

PAM and Fuzzy c-means,SVM, C5.0, KNN and Naive Bayes.Among these the classification algorithms, C5.0 and SVM have shown 81% accuracy in the prediction of disease.

Vrushali R. Balpande et al[23], proposed a study on the prediction of diabetics using ECLAT methodology. It is an Equivalence Class Clustering and bottoms up lattice traversal Algorithm, which calculates the patient has Un-Control Condition of diabetes.Messan Komi et al[24] proposed a research paper on diabetes prediction system using five data mining classification modelling techniques like Gaussian mixture model (GMM), Support vector machine(SVM), Logistic regression, Extreme learning machines (ELM), Artificial Neural networks (ANN). ANN found to be more effective with 89% accuracy using MATHLAB Tool followed by ELM and GMM.

Yao Xiao et al[25] had a study on preventive cardiovascular health using RFMiner,in that they used 8 base classifiers in which, Naive Bayes and Random forest are best performers.A cascaded classifier is used by combine the advantages of both Naive Bayes and Random Forest. Theresa Princy. R et al [26], suggested a survey on the different classification techniques used in predicting heart disease, they used classification techniques like Naïve Bayes, KNN, Decision Tree Algorithm, Neural Network. Whereas KNN and ID3 found to be more efficient as the risk rate of heart disease was detected.S.Radhimeenakshi [27] proposed a paper for the prediction of heart disease using ANN and SVM Classifiers.SVM is found to more effective in terms of accuracy.

Narander Kumar et al[28],proposed a paper in the prediction of kidney disease using WEKA Tool, in which they used five different classifiers like J48, Naïve Bayes, Random Forest, SVM and k-NN classifiers are compared with performance measures like ROC, kappa statistics, RMSE and MAE.Where Random forest classification holds great accuracy and prediction.

Nahla Barakat [29] proposed a paper on prediction of MetS (metabolic syndrome),before it may lead to CVD and type-2 diabetics. He introduced a hybrid technique in which SVM is the main classification algorithm used.The model is based on cascade generalization concept (CGen-SVM), because it consists of more number of classifiers.

6. KNOWLEDGE DISCOVERY DATABASE PROCESS.

Data mining process of discovering patterns from large data sets [30]. It is the core part of the knowledge discovery process.

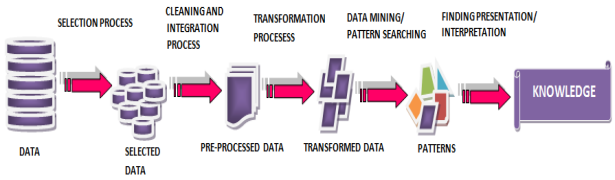

Figure 2

KDD or Knowledge discovery is the selection of relevant data. Figure 2 shows the transformation of the raw information into valid information..KDD includes [31],

1. Data selection
2. Data cleaning
3. Data transformation
4. Pattern searching (data mining)
5. Finding or Presentation/ Interpretation/Evaluation

7. TASKS IN DATA MINING

Data mining tasks are classified as classification, prediction, trend analysis or time series analysis, which is predictive category. Summarization, clustering, association fall in descriptive.

1. Classification: it accurately predicts the target class for each item in the data [32].
2. Prediction: The classification and prediction in data mining snatch the basic trend of the evolution of data which is unknown [33].Prediction attempts forming of patterns that helps to predict the next event in the given available input data. It is the relationship between independent variables dependent and independent variables.
3. Trend analysis: it reveals knowledge about what has taken place in the past and in future. Trend analysis is habitude of collecting and attempting to spot patterns [34].Techniques such as clustering, classification, regression, etc. can use in trend analysis of time series data.
4. Summarization: it is generalized data, which is the main concept in data mining, contains methods to find description of a dataset. Some of the summarization methods are tabulating the mean; standard deviations are often applied for data analysis, data visualization and automated report generation [35]. It includes both primitive and derived data.
5. Clustering: it is technique for the discovery of data distribution and patterns. Clustering is an unsupervised learning [36].Two clustering paradigms are hierarchical and partitioning. The hierarchical clustering includes sequence of partitions, which is nested into sequence of next partitions. In Partitioning clustering, database is partitioned into clusters, in which k partitions optimize a certain criterion [37].
6. Association: It is the relationships between items. They are if/then statements used to understand relationships or the probability of the occurrence of items in a in a relational database or in an information repository .i.e., In a set A= (t1, t2, t3 ...tm} are set of items. Consider D as transactional database; perform set of transactions, in which each transaction D contains set of items. Thus D is the subset of A [36].

8. DATA MINING APPROACHES

Data Mining Techniques are used in many Diagnosis Systems [38]. Data mining approaches includes statistics, machine learning, database systems, neural networks, rough sets and visualization etc [39].

1. Statistical Approaches: the tool included in statistical approach includes Bayesian network, regression analysis, correlation analysis and cluster analysis.
2. Machine Learning Approaches: The two types of machine learning algorithms includes, Supervised learning and unsupervised learning algorithms. Some of

the machine learning approaches are,Linear regression, Random forest ,Support vector machines ,Neural Networks,Naïve Bayes Classifier, Decision trees, Neural Networks, Multi Layer Perception (MLP),k-means, decision tree, Apriori, Self Organizing Maps (SOM) etc [40].

3. Data base oriented approaches: this kind of approaches help to understand about the characteristics of the data. It can be done with the help of attributes; it is termed as attribute oriented induction, where as it focuses on the identification of frequent items sets [39].
4. Rough sets: It is a mathematical tool to discover patterns hidden in data, which is used for data reduction, feature selection, decision rule generation, feature extraction and pattern etc.
5. Visualization: Visual Data Mining helps to deal with large set of information; It gives a yield vision to the user about the data, drawing conclusions and directly interacting with the data [41].

9. CONCLUSION

Data mining is the searching of information from a huge volume of data, medical mining is the searching of medical information from patient's history and pathological data, incorporated with appropriate methodologies. Numerous research are undergoing in medical filed. In this paper we discussed about the basic data mining methods and task for the exploration of information. Also we discussed about the various methodologies used in medical research area for the disease presage. In future research, the work will enhance and expanded for the prognosis of PCOS disease.

10. REFERENCES

[1] Maria Kravariti Katerina K. Naka Sophia N. Kalantaridou Nikolaos KazakosChristos S. Katsouras Antonis Makrigiannakis Evangelos A. ParaskevaidisGeorge P. Chrousos Agathocles Tsatsoulis Lampros K. Michalis," Predictors of Endothelial Dysfunction in Young Women with Polycystic Ovary Syndrome", The Journal of Clinical Endocrinology & Metabolism, Volume 90, Issue 9, 1 September 2005, Pp- 5088–5095.

[2] AnujaDokras,"Cardiovascular disease risk in women with PCOS", STEROIDS**,**Volume 78, Issue 8, August 2013, Pp-773-776.

[3] Ranjitha Sitheswaran, S. Malarkhodi **"**An effective automated system in follicle identification for Polycystic Ovary Syndrome using ultrasound images", Date Added to IEEE *Xplore*: 08 September 2014**.**

[4] E.Sterling, "Hormone Levels and PCOS, Polycystic Ovary Syndrome (PCOS), ObGyn Nurses, Pregnancy and Birth, Infertility

[5] Orquidia Torres, MD,Selma Witchel,DanaRofey,Elizabeth Miller, "Reducing Provider Variation in the Diagnosis of Polycystic Ovarian Syndrome".

[6] Jacqueline A. Boyle PhD, Helena J. Teede PhD, " Irregular menstrual cycles in a young woman", CMAJ, August 5, 2014, 186(11), Pp-852

[7] Fahimeh Ramezani Tehrani and Samira Behboudi-Gandevan,"Contemporary Gynaecologic Practice-chapter 4", Book edited by Atef Darwish.ISBN 978-953-51-1736-0, February 4,2015.

[8] GiovannaMuscogiuri,CinarNese·Stefano Palomba,SilviaSavastano[f]DomenicoTafurig,GiorgioColarieti, GiovanbattistaLa Sala AnnamariaColao, Bulent O.Yildiz ,"Obesity, type 2 diabetes mellitus and cardiovascular disease risk: an up-to-date in the management of polycystic ovary syndrome" .

[9] **"**Endocrinology and Metabolism Clinics of North America". Volume 28, Issue 2, 1 June 1999, Pp- 247-263

[10] H. Prasanna Kumar S, Srinivasan, **"**Segmentation of polycystic ovary in ultrasound images" Published in: Current Trends in Engineering and Technology (ICCTET), IEEE Xplore**:** 01,Dec-14, 2014.

[11] Mariagrazia Stracquadanio, Lilliana Ciotta, "Metabolic Aspects of PCOS: Treatment with Insulin Sensitizers", Pp-111

[12] Subarna Mitra, Prasanta Kumar Nayak, and Sarita Agrawal," Laparoscopic ovarian drilling: An alternative but not the ultimate in the management of polycystic ovary syndrome",Journel of Natural Science and Biology. 2015 Jan-Jun; 6(1), Pp- 40–48

[13] Hotaka Saito,Hiromistu Ohmori" Control of an abnormal human menstrual cycle in PCOS by speed gradient algorithm", Published in**:** SICE Annual Conference (SICE), IEEE Xplore: 27 October 2011.

[14] M. Alotaibi, A. Alsinan." Mobile Polycystic Ovarian Syndrome Management and Awareness System for Gulf Countries ", IEEE, SAI Computing Conference 2016 July 13-15, 2016 | London, UK

[15] Sandy RIHANA, Hares Moussallem, Chiraz Skaf, Charles Yaacoub," Automated Algorithm for Ovarian Cysts Detection in Ultrasonogram", IEEE 2013 2nd International Conference on Advances in Biomedical Engineering.

[16] Sharvari S., Deshpandei, Asmita Wakankar "Automated Detection of Polycystic Ovarian Syndrome Using Follicle Recognition", 2014 IEEE International Conference- (lCACCCT)

[17] Yinhui Deng, Yuanyuan Wang, Ping Chen, "Automated Detection of Polycystic Ovary Syndrome from Ultrasound Images", IEEE- 30th Annual International, British Columbia, Canada, August 20-24, 2008

[18] Palak Mehrotra, Chandan Chakraborty, Biswanath Ghoshdastidar, Sudarshan Ghoshdastida" Automated Ovarian Follicle Recognition for Polycystic Ovary Syndrome" IEEE-2011(ICIIP 2011)

[19] S. Sheela, M. Sumathi" Study and theoretical investigations on PCOS" Published in Computational Intelligence and Computing Research (ICCIC),IEEE Xplore**:** 07 September 2015

[20] Bedy Purnama,Untari Novia Wisesti,Adiwijaya,Fhira Nhita,Andini Gayatri,Titik Mutiah" A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images" Published in ICoICT,IEEE Xplore: 03 September 2015

[21] S. Rethinavalli, Dr. M. Manimekalai," A Hypothesis Analysis on the Proposed Methodology for Prediction of Polycystic Ovarian Syndrome", S., November 2016 | Vol 6, Issue 11, Pp-396-400

[22] Uma Ojha, Dr. Savita Goel,"A study on prediction of Breast cancer Recurrence using Data mining techniques", 2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence, Pp-527-530.

[23] Vrushali R. Balpande, Rakhi D. Wajgi," Prediction and Severity Estimation of Diabetes Using Data Mining Technique", (ICIMIA 2017),Pp-576-580

[24] Messan Komi, J un Li, Y ongxin Zhai, Xianguo Zhang," Application of Data Mining Methods in Diabetes Prediction", 2017 2nd International Conference on Image, Vision and Computing.

[25] Yao Xiao, Ruogu Fang," RFMiner: Risk Factors Discovery and Mining for Preventive Cardiovascular Health", 2017 IEEE/ACM, (CHASE), Pp- 278-279

[26] Theresa Princy. R, J. Thomas," Human Heart Disease Prediction System using Data Mining Techniques, 2016, [ICCPCT].

[27] Mrs.S.Radhimeenakshi,""classification and prediction using heart Disease Risk using Data Mining Techniques of Support Vector Machine and Artificial Neural Network",2016 (INDIACom),Pp-3107-311

[28] Narander Kumar, Sabita Khatri," Implementing WEKA for medical data classification and early disease prediction," Computational Intelligence and Communication Technology"(IEEE-CICT 2017)".

[29] Nahla Barakat," Diagnosis of Metabolic Syndrome: A Diversity Based Hybrid Model", Springer International Publishing Switzerland 2016, pp. 185–198, 2016.DOI: 10.1007/978-3-319-41920-6_14

[30] Lucio Soibelman, M.ASCE and Hyunjoo Kim ,"Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases".

[31] Priyadharsini.C[1] and Dr. Antony Selvadoss Thanamani[2,"]An overview of knowledge discovery and data mining techniques" International Journal of Innovative Research in Computer and Communication Engineering.

[32] Mike Chapple, , "Classification in Data Mining Classification Helps Make Big Data Understandable" ,November 25, 2016

[33] Cheng Lin Coll. of Math & Comput. Sci., Jiangxi Sci. &Technol. Normal Univ., Nanchang, China, Fan Yan Coll. of Math & Comput. Sci., Jiangxi Sci. &Technol. Normal Univ., Nanchang, China. **"**The Study on Classification and Prediction for Data Mining" IEEE Xplore-14 September 2015.

[34] Arpit Baheti , Durga Toshniwal "Trend Analysis of Time Series Data Using Data Mining Techniques", Proceeding BIGDATACONGRES '14 proceedings of the 2014 IEEE international congress on big Data , Pp-430-437.

[35] Varun Chandola and Vipin Kumar ,Department of Computer Science, University of Minnesota, MN, USA.Under consideration for publication in Knowledge and Information Systems , "Summarization - Compressing Data into an Informative Representation".

[36] Chapman&Hall/CRC Data Mining and knowledge Discovery Series, edited by Charu C. Aggarwal "Data Classification: Algorithms and Applications". Pp-2, 4.

[37] Arun k pujari,"Data Mining Techniques" pp 71, 116-117

[38] V.Gayathri, Chanda Mona G.T.N Arts & Science College, Tamilnadu, India M,"A Survey Of Data Mining Techniques On Medical Diagnosis And Research "

[39] Yongjian Fu "Data mining: Tasks, Techniques and Applications", Department of computer science, University of Missouri-Rolla.

[40] Jason Brownlee "Supervised and Unsupervised Machine earning Algorithms" on March16, 2016

[41] Jerzy Stefanowski, Instytut Informatyki, "Data Visualization or Graphical Data Presentation "Data mining for SE – 2013, Pp-4