Comparative Analysis of Classifiers for Polycystic Ovary Syndrome Detection using Various Statistical Measures

Priyanka R. Lele
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Pune
Maharashtra, India

Anuradha D. Thakare
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Pune
Maharashtra, India

Abstract— **Polycystic Ovary Syndrome (PCOS) is a condition that affects girl or women during their child-bearing years and disturbs the levels of hormones. This disturbance results in problems affecting many body systems. Women having PCOS have skip or irregularity in menstrual periods as well as cysts formation in the either or both ovaries. Symptoms of PCOS are irregular periods, excess androgen, polycystic ovaries, abnormal BMI, disturbed levels of hormones (LH, FSH, DHEAS), poor insulin resistance. But as per research studies, these symptoms are not sufficient for accurate detection for diverse data.**

This article presents an approach where classification of PCOS will use physical symptoms and sonograms. The results of only physical symptoms are presented here. The sonogram analysis along with the physical symptoms of PCOS are needed for accurate detection and reducing number of outliers during analysis. Such detection of PCOS also helps in proper treatment and reducing the health loss. The performance analysis of various Machine learning algorithms like Multilayer Perceptron, K-star, IB1 instance-based, Locally weighted learning, Decision Table, M5 rules, Zero R, Random Forest and Random Tree to classify PCOS is presented. Amongst all the algorithms K-star algorithm is out performing in all the performance measure.

Keywords— Classification, Machine Learning, Polycystic Ovary Syndrome, performance measures, sonography

I. INTRODUCTION

There are many disorders related to women reproductive system which may lead to some serious health issues in future. These disorders are related to the ovaries, uterus, cervix, the vagina, etc. the cause behind occurring of these diseases is hormonal changes inside the body, hormonal imbalance, irregular living patterns, stress, etc. Polycystic ovary syndrome (PCOS) comes in the category of hyperandrogenism i.e. excess androgen production by ovaries. It is a disorder commonly found in reproductive age group(15-40 yrs). The age group is fixed as before 15 age is when the menstruation begins; so there is a huge possibility that the menstrual periods are irregular and after 40 age the menopause periods of women begins. In this condition the women's hormone levels are affected[1]. This hormonal imbalance leads to cysts formation on the outer edge of the ovaries. The cysts are like follicle or small ball of tissues which is found in both or either ovary of PCOS women. These cysts are small and harmless. The size as well as

number of these cysts is not fixed. They can vary from 2mm-9mm in size[2]. PCOS further leads to infertility in women. Infertility causes due to infrequent ovulation that is not able to release a mature egg from the ovary. This infertility affects conceive rate for a women to get pregnant. According to recent study, it is found that 18% of female in East India suffer from PCOS[3]. Women with PCOS produce higher-than-normal amount of androgen. This imbalance in androgen causes skip or irregularity in menstrual periods. It also causes hirsutism and excessive acne formation[4]. PCOS can lead to long-term health problems like diabetes and heart disease. Fig. 1[5] shows the cysts formation in the ovary of a woman having PCOS.
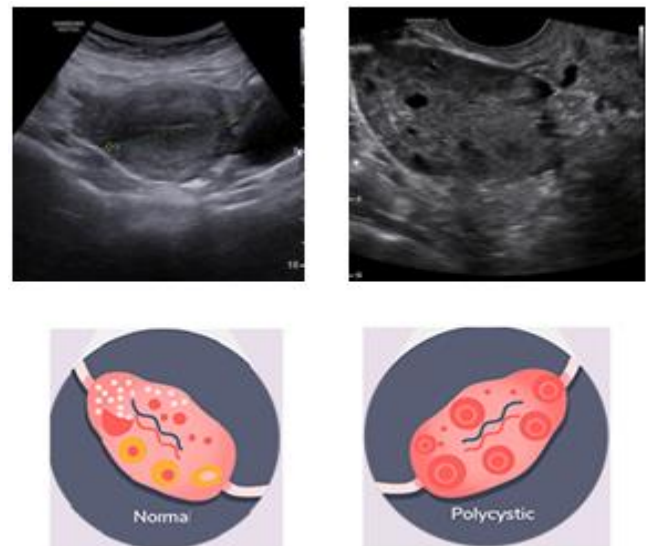

Fig. 1 Polycystic Ovary [5]

Commonly found symptoms of PCOS are: Irregularity or missed menstrual periods, Excessive Hair growth on face and unwanted body parts known as Hirsutism, Acne formation and oilier skin due to high androgen levels, abnormal Body Mass Index leading to obesity. Along with the mentioned physical symptoms; there is a need to conduct a blood test for checking the hormone levels in body. Hormone tests include increase in levels of Luteinizing hormone(LH), Follicle-stimulating hormone(FSH), Dehydroepiandrosterone(DHEAS), Fasting

blood sugar and Fasting insulin. Ratio of fasting blood sugar and fasting insulin if decreased leads to poor insulin resistance[6].

The exact cause of PCOS is not known. Some of the reasons may be: genetics, insulin resistance that leads to high testosterone, hormone imbalance that imposes negative effects on whole body [7].

There are various tests that needs to be conducted to diagnose PCOS. To diagnose PCOS and find other causes of your symptoms, a doctor may ask you about your previous medical history and do a physical exam, Pelvic ultrasound test and some blood tests [8].

II. LITERATURE SURVEY

Jayanta Pal, Barindra Nath Mallick, J. Evid," Community Screening For PCOS Amongst Adolescent Girls In A Semiurban Area In West Bengal"[9], 2015 adolescent girls of West Bengal were asked about their initiation of periods and about oligomenorrhoea. They were also examined for clinical features like excess in androgen. According to Rotterdam criteria, the cases were concluded by having PCOS or not having PCOS.

Kar Sujata, Samparna Swoyam, "2D and 3D Trans-vaginal Sonography to Determine Cut-offs for Ovarian Volume and Follicle Number per Ovary for Diagnosis of Polycystic Ovary Syndrome in Indian Women"[10], 86 women having PCOS and 45 controls/volunteers were choosen. A 2-D and 3-D trans-vaginal ultrasonography was carried out in early follicular phase (D2 – D5). Ovarian volume(OV), follicle number per ovary(FNPO), stromal volume, vascularization index (VI), vascularization flow index (VFI) and flow index (FI) were measured in PCOS and controls. Mann-Whitney test. Logistic regression model were used to compare the data between PCOS and control.

Cesare Battaglia, Bruno Battaglia, Elena Morotti, Roberto Paradisi, Isabella Zanetti, Maria Cristina Meriggiola, Stefano Venturoli, "Two and three dimensional sonographic and color Doppler techniques for diagnosis of polycystic ovary syndrome[11] ",112 lean Italian women having PCOS and 52 controls /volunteers were choosen. All participants underwent transvaginal sonographic examinations (RIC5-9H, Voluson 730 Expert sonography system) in their follicular phase for measurement of the ovarian volume, follicle count, and follicular maximum diameter. Continuous variables was analyzed using Shapiro-Wilk normality tests.

Miriam e. Silfen, michelle r. Denburg, alexandra m. Manibo, rogerio a. Lobo, Richard jaffe, michel ferin, lenore s. Levine, and sharon e.Oberfield, "Early Endocrine, Metabolic, and Sonographic Characteristics of Polycystic Ovary Syndrome (PCOS): Comparison between Non obese and Obese Adolescents[12]", 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.

Jacob P. Christ, Heidi Vanden Brink, Eric D. Brooks, Roger Pierson, Donna R. Chizen and Marla E. Lujan, "Ultrasound features of polycystic ovaries relate to degree of reproductive and metabolic disturbance in polycystic ovary syndrome[13]", 49 women (aged between 19 to 36) diagnosed with PCOS were chosen. Evaluation of menstrual cycle and also physical exam assess various parameteres(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.

III. PROPOSED SYSTEM

A. Work Flow of Proposed Approach

The patients with metabolic (physical) symptoms like acne, facial hair growth and irregular period's needs can be examined in daily routine. But, these metabolic symptoms alone are not sufficient to diagnose the PCOS. Therefore the hormone tests like LH(Luteinizing hormone), FSH(Follicle-stimulating hormone), androgen level, DHEAS, fasting insulin, fasting blood sugar should be examined. The physical as well as hormonal symptoms will be considered as a feature set for the proposed system. These features will be statistically analyzed with machine learning algorithms. Here, aim is to select efficient algorithm for feature dataset. The workflow for PCOS detection is represented in Fig. 2.
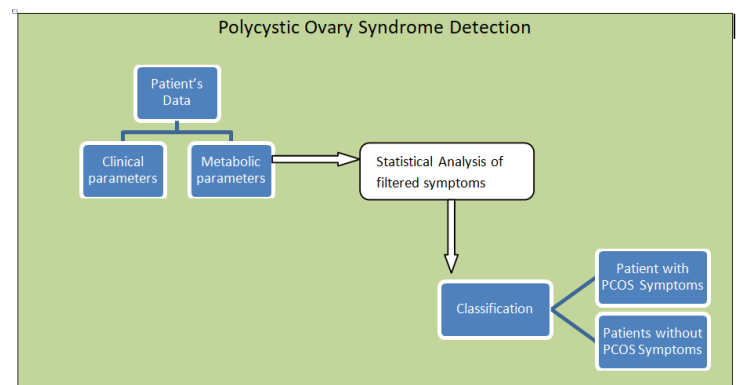

Fig 2. Workflow of proposed system

The mathematical formulation of proposed system also needs the parameters like age, height, weight, fasting insulin, fasting blood sugar, sonography along with above mentioned parameters.

B. Dataset used

The dataset for proposed system is not readily found on available repositories. Therefore, dataset is created in discussion with medical practitioner with their expertise in PCOS detection. . The dataset generated has 13 attributes and 2 classes. Total 40 instances are created. These attributes are the symptoms related to PCOS such as physical symptoms(age, height, weight, irregular periods, hirsutism,

acne), and blood test results; (LH(Luteinizing hormone), FSH(Follicle-stimulating hormone), androgen level, DHEAS, fasting insulin, fasting blood sugar) and clinical test(sonography). The class type specifies the presence of PCOS or not. The most important symptoms (that are influencing factors) as highlighted are weight, irregular periods, acne, LH and sonography. The type attribute tells about the prediction whether the women has PCOS or not. The dataset is validated based on various cases of PCOS patients and opinion of expert from medical domain.

C. Results and 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 Multilayer Perceptron, K star, IB1 instance-based, Locally weighted learning, Decision Table, M5 rules, Zero R, Random Forest and Random Tree are used for classification and performance is analyzed statistically. Statistical results like Correlation Coefficient, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error are calculated and compared using WEKA tool. Table 1 depicts the statistical results.

Table 1 Classification results with statistical measures

Algorithms	Time taken (secs)	Correla-tion coeffici-ent	Mean absolut-e error	Root mean squared error	Relative absolute error (%)	Root relative squared error (%)
Multilay-er Perceptr-on	0.56	0.9949	0.0169	0.0517	3.35	10.254
K star	0	1	0.0001	0.0003	0.01	0.0674
IB1 instance-based	0	0.9765	0.0119	0.1091	2.3631	21.649
Locally weighted learning	0	0.9339	0.0449	0.18	8.9065	35.722
Decision Table	0.13	0.959	0.03	0.1452	5.955	28.804
M5 rules	0.67	0.9294	0.0668	0.1855	13.256	36.798
Zero R	0	-0.2605	0.5038	0.504	100	100
Random Forest	0.25	0.985	0.0413	0.0904	8.1999	17.935
Random Tree	0	0.9524	0.0238	0.1543	4.7262	30.617

From the results, it is observed that in all statistical parameters, K star algorithm super sits the other algorithms, giving good classification accuracy. Other algorithms may outperform with the increase in dataset size. This analysis will be carried out further for the said research.

IV. CONCLUSION

In this paper, the Classification algorithms; Multilayer Perceptron, K star, IB1 instance-based, Locally weighted learning, Decision Table, M5 rules, Zero R, Random Forest and Random Tree algorithm are used to detect whether the patient have PCOS or not. Classification techniques are considered in this study as it enables us to predict if the patient has Polycystic Ovarian Syndrome or not based on the syndromes provided by the doctor or medical Centre. The machine learning model is developed using real time data. It has been noticed that the Root Mean Squared Error of the K – star algorithm is lowest as compared to other algorithms. These models can provide help to the doctors to recognize the disease much faster, therefore early treatment can be given to the patient.

REFERENCES

[1] https://www.healthline.com/health/polycystic-ovary-disease#what-is-pcos

[2] Neetha Thomas, Dr. A. Kavitha, "A Literature Inspection on Polycystic Ovarian Morphology in Women using Data Mining Methodologies", International Journal of Advanced Research in Computer Science, Volume 9, No. 1, January-February 2018, ISSN No. 0976-5697

[3] 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

[4] "Polycystic Ovary Syndrome, A Review of Treatment Options With a Focus on Pharmacological Approaches", Uche Anadu Ndefo, Angie Eaton, and Monica Robinson Green, P T. 2013 Jun; 38(6): 336-338, 348, 355

[5] https://www.boostthyroid.com/blog/2018/4/12/hashimotos-and-polycystic-ovary-syndrome-pcos

[6] https://www.mayoclinic.org/diseases-conditions/pcos/symptomscauses/syc-20353439

[7] https://www.endocrineweb.com/conditions/polycystic-ovary-syndrome-pcos/what-causes-pcos-how-will-it-affect-body

[8] https://www.healthline.com/health/polycystic-ovary-disease

[9] "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

[10] "2D and 3D Trans-vaginal Sonography to Determine Cut-offs for Ovarian Volume and Follicle Number per Ovary for Diagnosis of Polycystic Ovary Syndrome in Indian Women", Kar Sujata , Samparna Swoyam. J Reprod Infertil. 2018;19(3):146-151

[11] "Two- and Three-Dimensional Sonographic and Color Doppler Techniques for Diagnosis of Polycystic Ovary Syndrome", Cesare Battaglia, Bruno Battaglia, Elena Morotti, Roberto Paradisi, Isabella Zanetti, Maria Cristina Meriggiola, Stefano Venturoli. J Ultrasound Med 2012; 31:1015–1024

[12] "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

[13] "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.