i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques

Amsy Denny
Dept. of Biomedical Engineering, Sahrdaya College of Engineering and Technology (Affiliated to APJ Abdul Kalam Technological University) Thrissur, India
amsydennykallingal@gmail.com

Anita Raj
Dept. of Biomedical Engineering, Sahrdaya College of Engineering and Technology (Affiliated to APJ Abdul Kalam Technological University) Thrissur, India
anitakallery@gmail.com

Ashi Ashok
Dept. of Biomedical Engineering, Sahrdaya College of Engineering and Technology (Affiliated to APJ Abdul Kalam Technological University) Thrissur, India
ashisgashok@gmail.com

Maneesh Ram C
Dept. of Biomedical Engineering, Sahrdaya College of Engineering and Technology (Affiliated to APJ Abdul Kalam Technological University) Thrissur, India
maneeshram99@gmail.com

Remya George
Dept. of Biomedical Engineering, Sahrdaya College of Engineering and Technology (Affiliated to APJ Abdul Kalam Technological University) Thrissur, India
remyageorge@sahrdaya.ac.in

Abstract—The present world women population is widely affected by preterm abortions, infertility, anovulation etc. It is observed that polycystic ovary syndrome (PCOS), a condition seen among the women of reproductive age is having a major influence in the cause of infertility. Over five million women worldwide in their reproductive age PCOS. It is an endocrine disorder characterized by changes in the female hormone levels and the abnormal production of male hormones. This condition leads to ovarian dysfunction with increased risk of miscarriage and infertility. The symptoms of PCOS include obesity, irregular menstrual cycle, and excessive production of male hormone, acne, and hirsutism. It is extremely difficult to diagnose PCOS due to the heterogeneity of symptoms associated and the presence of a varying number of associated gynecological disorders. The time and cost involved in innumerous clinical tests and ovary scanning has become a burden to the patients with PCOS. To address this problem this paper proposes system for the early detection and prediction of PCOS from an optimal and minimal but promising clinical and metabolic parameters, which act as an early marker for this disease. The data sets required for this system development are obtained through patient survey of 541 women during doctor consultations and clinical examinations. Out of the 23 features from clinical and metabolic test results, 8 potential features are identified using SPSS V 22.0 based on their significance. Classification of PCOS with the feature set transformed with Principal Component Analysis (PCA) is done using various machine learning techniques such as Naïve Bayes classifier method, logistic regression, K-Nearest neighbor (KNN), Classification and Regression Trees (CART), Random Forest Classifier, Support Vector Machine (SVM) in Spyder Python IDE. Results revealed that the most suitable and accurate method for the PCOS prediction is RFC with an accuracy of 89.02%.

Keywords—Machine learning, polycystic ovary syndrome, Classifier, Diagnostic aid.

I. INTRODUCTION

Technology and mankind together hand in hand can make way towards better health care and services. Machine learning is a subset of artificial intelligence, in which it provides the system with the ability to automatically learn and improve without being programmed explicitly. It mainly focuses on developing algorithms that can access the datasets provided and use data for the learning purposes of the network. Applications of Machine Learning bring about huge transformation in the health industry, which includes detection, data prediction, image recognition etc.

Polycystic ovary syndrome (PCOS), is one of the relevant, most prevalent hormonal disorder seen among the women of childbearing age. This is a heterogeneous endocrine disorder which is highly prone to infertility, anovulation, cardiovascular disease, type 2 diabetes, obesity etc. PCOS is a common condition detected in nearly 12-21% of women of reproductive age and among them 70% is remain undiagnosed. PCOS condition can be treated to some extend by controlled medication and bringing alterations in life style. This includes the treatment methods with pills for birth control, diabetes, fertility, anti-androgen medicines and scanning procedures like ultrasound scan. When such interventions fail, invasive treatment procedures like surgical drilling of ovaries is also used for improving the ovulation ability of the ovary by reducing the male hormone level.

The aetiology of PCOS is underpinned by both insulin resistance and hyperandrogenism. Clinically it is characterized by reproductive, metabolic and psychological features and represents a major health burden to women. Diagnosis is recommended based on clinical or biochemical and radiological test results. PCOS is diagnosed by exclusion of irrelevant symptoms or test results, mainly because of lack of knowledge of its complex patho-mechanism. The diverse symptoms of this condition force medical practitioners to call for large number of clinical test results and unnecessary radiological imaging procedures. The early detection and diagnosis of PCOS with minimal tests and imaging procedures is of utmost importance and of great significance as the condition directly leads to ovarian dysfunction with an increased risk of miscarriage, infertility or even gynaecological cancer and mental agony for the patients due to wastage of time and money.

II. LITERATURE SURVEY

Among the in-numerous problems that exist around us, the problems that are related to the reproductive health of women was selected as an area of our interest, due to its importance in this contemporary society. A detailed survey of studies on PCOS and systems to support its diagnosis was carried out. Literature says that about 5-10% of Indian women in reproductive age are affected by the multifaceted endocrine disorder called Polycystic Ovary Syndrome (PCOS) [16]. It is a major cause of anovulatory infertility and increases the risk for insulin resistance, obesity, cardiovascular disease and psychosocial disorders [17]. The symptoms for PCOS might be varying from patient to patient. Some of them are irregularity in menstrual periods, acne, overweight, increased tendency for infertility, intense hair fall, balding of front head, increased facial hair growth [1]. Traditionally the PCOS can be suspected when number of follicles in an ovary is more than 12 per unit area and visible in radiological scan [15]. Some authors have proposed changing the cutoff from 12 follicles to 20 or abandoning ultrasound altogether in favor of other biomarkers, such as serum anti-Mullerian hormone (AMH) [1,10]. The diagnosis of PCOS is uncomplicated, requiring only the careful application of a few well-standardized diagnostic methods. PCOS diagnosis is often delayed and this affects patients' well-being negatively [21]. Escobar *et al.* [11] suggests that treatment should be symptom-oriented, long term and dynamic and adapted to the changing circumstances, personal needs and expectations of the individual patient. Joham et al. in [6] considered the relation of PCOS and infertility rate of women in this community and use of fertility hormone treatment was significantly higher in women reporting PCOS. Considering the prevalence of PCOS and the health and economic burden of infertility, strategies to optimize diagnosis of PCOS and the factors leading to fertility are important. This is because iinfertility is reported to be 15-fold higher in women reporting PCOS, independent of BMI [18]. There is a bi‑directional relationship between obesity and PCOS. Both exacerbate each other in a never‑ending cyclical manner. Essah, P.A. and Nestler, J.E suggests that the prevalence of obesity in PCOS women is 30–75% [19]. Clinical validation of PCOS is usually done by Rotterdam criteria [8] or standards set by societies involved in PCOS research. Pictorial depiction of three popular criteria can be seen in Table I.

TABLE I. CRITERIA FOR DIAGNOSIS OF PCOS

Clinical Finding	National Institutes of Health criteria,1990 (Must have both of the findings marked below)	Rotterdam Criteria, 2003(must have any two of the findings marked below)	Androgen Excess and PCOS Society, 2009(must have A plus either B or C)
*Hyperandrogenism**	X	X	A
Oligomenorrhea	X	X	B
Polycystic ovaries		X	C

* Clinical or biochemical evidence of excess androgen.

A cross sectional study by Brower *et al*. [14] suggested that the presence of clinically evident menstrual dysfunction can be used to predict the presence and possibly the degree of insulin resistance in women with PCOS. Many of the technical studies carried out in PCOS diagnosis are using features of the ultrasound scan and image processing techniques for the diagnosis of the PCOS [1-2]. Some studies used clinical and metabolic features of the disease [3]. Few recent studies are diverted in fundamental research direction, investigating the associated factors such as obesity [4] and genetic factors [5]. A summary of such studies are given in Table II. Studies are also carried out in directions of analyzing urinary steroid hormone metabolites and enzyme activities in women with and without PCOS in order to test their value for diagnosing PCOS [20].

TABLE II. SUMMARY OF FEW STUDIES ADDESSING THE ISSUE OF PCOS

Authors	Technique used	Objective of the study	Year
Cheng et al. [1]	Rule based classifier and Gradient boosted Tree classifiers	PCOS determination from Ultrasound Images	2016
Dewi, R.M. and Wisesty [2]	Gabor wavelet based feature extraction and CNN	PCOS determination from Ultrasound Images.	2018
Mehrotra et al. [3]	2 sample-test, Bayesian and Logistic Regression (LR) classifier	PCOS determination from clinical and metabolic parameters.	2011
Sachdev et al. [4]	Prospective observational study : Obese vs non-obese PCOS	Obese PCOS patients have a higher risk of adverse outcomes.	2019
Zhang et al. [5]	Machine-learning algorithms	PCOS prediction from identification of new PCOS genes.	2019
Joham et al. [6]	Logistic regression on data from cross-sectional analysis of a longitudinal cohort	Examination of factors associated with infertility and use of fertility treatment.	2015
K. Meena, M. Manimekalai, and S. Rethinavalli [7]	Information Gain Subset Evaluation and Neuro Fuzzy methods of feature selection and Decision Tree classifier	Framework for Filtering the PCOS Attributes	2015

III. METHODOLOGY

For the development of an appropriate machine learning model based diagnostic aid for PCOS, a comparison of performance of various existing algorithms in our data set need to be presented. Preparation of the model is the most crucial step that provides the outline of the research. Steps that are included in the development of an appropriate model and tuning it for obtaining possibly the best result, is detailed below with the help of a work flow diagram, Figure 1. Along with that the effective tools and available platforms utilized for the development of the system must be mentioned. The following section describe both aspects.

A. Defining problem

The most important step is to define the problem appropriately including the inputs provided into the model and the output expected out of it. It is based on the assumptions like the outputs can be predicted from the inputs provided.

B. Data collection

The critical step that will decide about how good the model will be and also as the number of data collected increases the accuracy of the model also increases and hence better will be its performance. There are many ways for data gathering like real time data gathering or from repository platforms like kaggle and UCI machine learning repository which is one of the most frequently visited one.

C. Selection of implementation platform

Tools for the efficient running of machine learning methods and also platforms for statistical analysis should be properly opted. For this research, Sypder python for model formation, HTML with SQL for designing a proper user interface, whereby the patient data can be input to the system and PCOS status can be obtained as output, and SPSS V22.0 for establishing the relevance of features, are used.

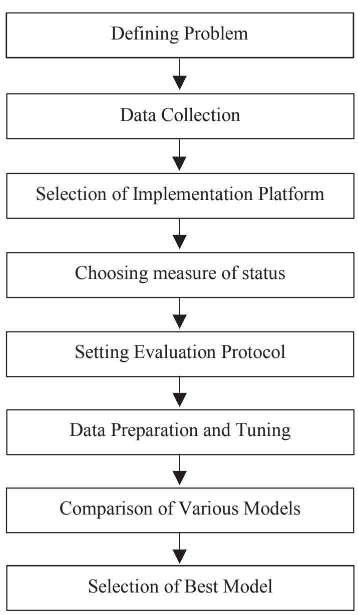

Fig. 1. Workflow of machine learning model

D. Choosing measure of status

In case classification problems, success is measured using the calculation of accuracy and precision of the model. In this study we have considered the following evaluation metrics:

- Accuracy : $\frac{(\mathrm{TP+TN})}{(\mathrm{TP+TN+FP+FN})}$
- Precision (P) : $\frac{(\mathrm{TP})}{(\mathrm{TP+FP})}$
- Sensitivity / Recall (R) : $\frac{(\mathrm{TP})}{(\mathrm{TP+FN})}$
- Specificity : $\frac{(\mathrm{TN})}{(TN+FP)}$
- F1 score : $2 * \left(\frac{P*R}{P+R}\right)$

where TP, TN, FP, FN implies True Positive, True Negative, False Positive, False Negative respectively.

E. Setting validation protocol

Maintaining a hold out validation set, i.e., in this method some portion of the data is set apart for the purpose of testing as test data and remaining as train data.

Usually the data is split in the ratio 8:2 as train data to test data. It can be depicted as in fig.2:

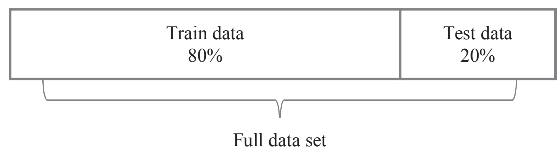

Fig. 2. Depiction of Hold-out validation set

F. Data preparation

This tiring process includes dealing with missing data, handling categorical data, feature scaling and selection of meaningful features. The missing values in the dataset is replaced by 'NaN'. Due to a model's inability to read a missing value, before confronting the model, the samples with missing values will be extirpated or else will be replaced with some pre-built estimators. Likewise before feeding the data into the model, ordinal and nominal data need to be considered accordingly. If the nature of the data demands, the dataset should undergo normalization and standardization. Finally, the overfitting can be avoided by reducing the dimensionality of data. This is done by reducing the number of feature sets present in the dataset. It is performed using Principal Component Analysis (PCA) in Spyder Python IDE, which works by identifying the patterns in the datasets and the correlations present between the features. The correlated data are then eliminated by directly removing such features. The optimal features identified by the PCA algorithm are verified for their potential in discriminating PCOS status, with SPSS V22.0. This is done by a procedure called independent sample t-test and the level of statistical significance across the two classes of PCOS and Non PCOS patients. Those features with a significance less than 0.01 are potential ones.

G. Comparison of various models

This step is to serve as a baseline. Study is carried out with selected set of features in a number of classifier algorithms. Among the existing innumerable machine learning algorithms some of them, which are proven to give best result in the detection of PCOS and Non-PCOS condition from the literature survey, is used and listed below:

- Logistic Regression (LR)
- Linear Discriminant Analysis (LDA)
- K-nearest neighbors (KNN)
- Classification and Regression Trees (CART)
- Random Forest Classifier
- Naïve Bayes Classifier
- Support Vector Machine

IV. DESIGN AND DEVELOPMENT

Data acquired from various hospitals and clinics which includes both physiological and metabolic parameters that are the contributors towards PCOS and thereby infertility, are fed into the mathematical framework of i-HOPE. The proposed system of diagnostic aid is illustrated Figure 3.

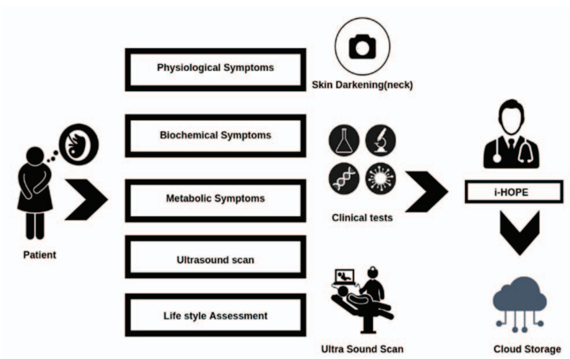

Fig. 3. Proposed model of i-HOPE

A. *Data Collection :*

More the number of features and samples, more distinguishing and accurate the model be. In this research we have considered a total of 541 samples which were collected from various clinics and hospitals in and around the district of Thrissur. Informed consent was collected from each patient, promising the anonymity of the data collected.

1) *Patient History Collection :*

One proforma is assigned for each patient so that all the data about a single patient will be in contained in a Proforma. These Proforma were carried along for data collection into clinics and for patient survey. Later on the collected data from the Proforma is consolidated together in to a single data sheet. The features which were included within the proforma was inferred from various literature studies and clinical surveys conducted, whose identity was kept anonymous during the entire process.

2) *Real-time data collection :*

For the purpose of real time detection of PCOS, a patient interface is required to input the patient data. Data to be entered may include the personal data of the patient for further references and the parameters that determine the existence of PCOS. This entered data will be inputting in to the created algorithm for data processing. To create the interface, a front end is generated using HTML and the data entered in the front end is being written in to an excel sheet using SQL.

TABLE III. FINALIZED PARAMETERS

SI No	Parameters	Value
1	Age	15-35
2	BMI	<24(normal), >24(abnormal)
3	Cycle Length and regularity	Long, normal or short Regular/Irregular
4	LH : FSH Ratio	Normal/abnormal
5	Waist : Hip Ratio	Normal/abnormal
6	Weight gain	Yes(y)/No(n)
7	Excess facial or body hair	Yes(y)/No(n)
8	Dark areas on skin	Yes(y)/No(n)
9	Pimples	Yes(y)/No(n)
10	Blood Pressure	Normal/abnormal
11	Diabetes (before and after food)	Normal/abnormal
12	Fast food intake	Yes(y)/No(n)
13	Regular exercise	Yes(y)/No(n)
14	Loss of hair	Yes(y)/No(n)
15	No. of follicles (L[a.] and R[b.])	High ,medium, low
16	Size of follicles (L[a.] and R[b.]) (mm)	>10
17	TSH (mIU/L)	0.4-4(Normal)
18	AMH	1-4(Normal) >1(Abnormal)
19	PRL	2-29 (Non-Pregnant Females) 10-209 (Pregnant Females)
20	Vit D3	20-50 (Normal) >12 (Abnormal)
21	PRG	1.5-12.4 (Normal)

[a.] Left ovary, [b.] Right ovary

B. *Parameter Selection*

The proposed model i-Hope was fed with the data obtained from the survey conducted. Parameter finalization was done with the support of expert opinions and considering the contemporary researches that in a way or other affected PCOS. The features are transformed for removing correlating among themselves which might adversely affect the classification results The finalized list of parameters before transformation are listed in Table III. These parameters include physiological, metabolic and biochemical attributes. Also considered the values of the result obtained from ultra sound scan which consists of the information about the cyst formed such as the number of cyst present in the right and left ovary and their size. The significance of parameters were studied individually based on the independent sample test, Pearson and Spearman's rho correlation of parameters with the help of SPSS software by IBM. The correlation was proved to be significant at 0.01 level (2-tailed). Table IV shows the important features for the final design of the system and their statistical significance.

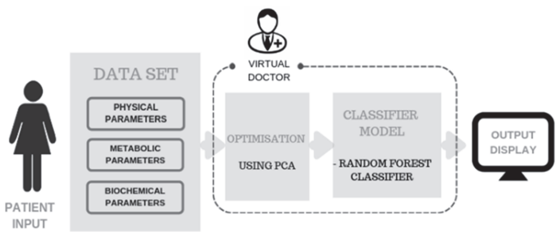

Fig. 4 The automated PCOS diagnosis aid i-Hope

C. Moulding Model

1) Setting model : Through Sypder python the evalution metrics of the model is claculated.After that the evaluation protocol is implemented from the sklearn library of the python platform. The most discriminating and contributing feature set is selected by extirpating the redundant data set through implementation of Principal Component Analysis (PCA) which is a type of unsupervised learning.

2) Model Selection : Comparing various selected machine learning algorithms to find the best performing model using spot-check algorithm, which is a precursor for the selection of best and appropriate algorithm that suites the model. Selecting the best model with better performance by analyzing the confusion matrix of each algorithm. Among the models RFC provided the best result. Consolidating all the above noted details, the final design is improvised and modeled as shown in the Figure 4.

V. RESULT AND DISCUSSIONS

A total of 541 cases were available for study, which was collected from various infertility treatment centers at Thrissur. The data comprised the women reproductive age group i.e., in between 18-40 years. Among the data collected 364 cases were normal and non-PCOS, the remaining 177 cases reported PCOS. Altogether there were 23 features, including the reports on transvaginal Ultrasound scan, hormone profile and lifestyle of the patient with impressions on physical fitness and some listed in Table III. The optimal features after PCA is statistically analyzed to see their significance and are listed in Table IV.

The algorithms opted comprises a mixture of simple linear and non-linear algorithms. The simple linear algorithms are LR and LDA. Non-linear methods are KNN, CART, RFC, NB, SVM. We reset the random number seed in each run to the data split. Accuracy estimations of each of the models were carried out with hold out validation and the estimated accuracy scores of each model was obtained. Table shows the results of evaluation metrics of the research, through this the performance of the models can be analyzed. From which we can arrive at a conclusion that the best performance was given by Random Forest Classifier model, where an accuracy of 89 % was achieved after data optimization.

TABLE IV. OPTIMAL FEATURES IDENTIFIED IN SPSS

Parameters contributing towards PCOS		**Parameters contributing towards infertility**	
Feature	***Significance***	***Feature***	***Significance***
Cycle irregularity	0.000	No. of abortion	0.002
Cycle Length	0.000		
FSH,LH Ratio	0.006	Thickness of Endometrium	0.660
AMH	0.000		
Follicle no.	0.000	Vitamin D3	0.050
Follicle size	0.002		
BMI	0.000	AMH	0.002
Weight gain	0.000		

Therefore, it can be concluded that either biochemical profile alone or USG result alone can't serve as a diagnostic tool for the treating PCOS. Because both the factors that relate PCOS and infertility falls in both categories. AMH turns out to be a very promising feature to detect PCOS and infertility as per our results. Comparing the accuracies obtained for other studies in PCOS detection, 97% is the highest accuracy obtained in [1], 82% in [2], 93.9% in [3] and 90% in [5]. Our results are lesser than these, even though a direct comparison is meaningless. With optimization of weight parameters of classifiers, performance of the system might improve.

TABLE V. ACCURACY SCORE, SENSITIVITY, SPECIFICITY AND PRECISION OF VARIOUS MODELS

Algorithm Used	**Accuracy score**	**Sensitivity**	**Specificity**	**Precision**	**F1 score**
Logistic Regression (LR)	0.8536	0.6451	0.98039	0.952380	0.3845
K- Nearest Neighbors (KNN)	0.8658	0.8064	0.90196	0.83333	0.4098
Classification and Regression Trees (CART)	0.8292	0.8387	0.82352	0.74285	0.3939
Random Forest Classifier (RFC)	0.8902	0.7419	0.98039	0.95833	0.4182
Gaussian Naïve Bayes (NB)	0.8414	0.7419	0.90196	0.82142	0.3898
Support Vector Machines (SVM)	0.8292	0.5483	1.0	1.0	0.3541

VI. CONCLUSION

Polycystic Ovary Syndrome (PCOS) is one of the most common type of endocrine disorder in reproductive age women. This may result in infertility and anovulation. The diagnostic criterion includes the clinical and metabolic parameters which are biomarker for the disease. We developed a system that automates the PCOS detection based on minimal set of potential markers. Our methodology involves the formulation of a feature vector based on real time data from patients during clinical and radiological investigation while they visit a healthcare facility. The eight metabolic and Ultrasound image features identified with the PCA feature transformatin and statistical significance is found promising for discriminating between normal and PCOS patients. Among the various algorithms used, RF algorithm is found superior in performance. This automated system can act as an assistive tool for the doctor for saving considerable time in examining the patients and hence reducing the delay in diagnosing the risk of PCOS. Implications from the clinical expert survey for this work suggests that, innovations that uphold the medical ethics are always welcome in the field of medicine and healthcare. Researches that could bring out useful innovative methodologies like the effect of Vitamin D on PCOS, studies that put forth the impact of PCOS on preterm labor/abortions, attempt to unveil the number of lean PCOS patients etc. need to be held in future.

ACKNOWLEDGMENT

We thank the authorities of Sahrdaya CET, Thrissur, for providing financial support towards the successful completion of this work. We like to extend our gratitude towards the doctors (Dr. Krishnankutty, Kare Clinic, Thrissur, Dr. Neetha George, Jubilee Mission Hospital, Thrissur, Dr. Shyni Daniel, St. James Hospital, Chalakkudy, and Dr. Rahul, St. James Hospital, Chalakkudy) for their valuable sugestions and the permissions to conduct the data collection. We thank the former guide Mr Jis Paul and the project coordinator Ms. Annsonia M. Michael whose comments have greatly influenced the implementation of the work. Any errors found are our own, and should not tarnish the reputations of the esteemed persons.

REFERENCES

[1] Cheng, J.J. and Mahalingaiah, S., 2018. Data mining and classification of polycystic ovaries in pelvic ultrasound reports. *bioRxiv*, p.254870.

[2] Dewi, R.M. and Wisesty, U.N., 2018, March. Classification of polycystic ovary based on ultrasound images using competitive neural network. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012005). IOP Publishing.

[3] P. Mehrotra, J. Chatterjee, C. Chakraborty, B. Ghoshdastidar and S. Ghoshdastidar, "Automated screening of Polycystic Ovary Syndrome using machine learning techniques," *2011 Annual IEEE India Conference*, Hyderabad, 2011, pp. 1-5.

[4] Sachdeva, G., Gainder, S., Suri, V., Sachdeva, N. and Chopra, S., 2019. Obese and non-obese polycystic ovarian syndrome: Comparison of clinical, metabolic, hormonal parameters, and their differential response to clomiphene. *Indian journal of endocrinology and metabolism*, *23*(2), p.257.

[5] Zhang, X.Z., Pang, Y.L., Wang, X. and Li, Y.H., 2018. Computational characterization and identification of human polycystic ovary syndrome genes. *Scientific reports*, *8*(1), p.12949.

[6] Joham, A.E., Teede, H.J., Ranasinha, S., Zoungas, S. and Boyle, J., 2015. Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: data from a large community-based cohort study. *Journal of women's health*, *24*(4), pp.299-307.

[7] Dr. K. Meena, Dr. M. Manimekalai, S. Rethinavalli*, "A Literature Review on Polycystic Ovarian Syndrome and Data Mining Techniques",*Volume 4,Issue 12,December 2014,International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X.

[8] Dhayat, N.A., Marti, N., Kollmann, Z., Troendle, A., Bally, L., Escher, G., Grössl, M., Ackermann, D., Ponte, B., Pruijm, M. and Müller, M., 2018. Urinary steroid profiling in women hints at a diagnostic signature of the polycystic ovary syndrome: A pilot study considering neglected steroid metabolites. *PloS one*, *13*(10), p.e0203903.

[9] Rotterdam EA-SPCWG. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril. 2004;81(1):19-25.

[10] Dewailly, D., Lujan, M.E., Carmina, E., Cedars, M.I., Laven, J., Norman, R.J. and Escobar-Morreale, H.F., 2013. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Human reproduction update*, *20*(3), pp.334-352.

[11] Escobar-Morreale, H.F., 2018. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nature Reviews Endocrinology*, *14*(5), p.270.

[12] Lawrence, M.J., Eramian, M.G., Pierson, R.A. and Neufeld, E., 2007, May. Computer assisted detection of polycystic ovary morphology in ultrasound images. In *Fourth Canadian Conference on Computer and Robot Vision (CRV'07)* (pp. 105-112). IEEE.

[13] Dumesic, D.A., Oberfield, S.E., Stener-Victorin, E., Marshall, J.C., Laven, J.S. and Legro, R.S., 2015. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocrine reviews*, *36*(5), pp.487-525.

[14] Brower, M., Brennan, K., Pall, M. and Azziz, R., 2013. The severity of menstrual dysfunction as a predictor of insulin resistance in PCOS. The Journal of Clinical Endocrinology & Metabolism, 98(12), pp.E1967-E1971.

[15] Lawrence, M.J., Eramian, M.G., Pierson, R.A. and Neufeld, E., 2007, May. Computer assisted detection of polycystic ovary morphology in ultrasound images. In Fourth Canadian Conference on Computer and Robot Vision (CRV'07) (pp. 105-112). IEEE.

[16] McCartney, C.R. and Marshall, J.C., 2016. Polycystic ovary syndrome. *New England Journal of Medicine*, *375*(1), pp.54-64.

[17] Dumesic, D.A., Oberfield, S.E., Stener-Victorin, E., Marshall, J.C., Laven, J.S. and Legro, R.S., 2015. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocrine reviews*, *36*(5), pp.487-525.

[18] Norman, R.J., Dewailly, D., Legro, R.S. and Hickey, T.E., 2007. Polycystic ovary syndrome. *The Lancet*, *370*(9588), pp.685-697.

[19] Essah, P.A. and Nestler, J.E., 2006. The metabolic syndrome in polycystic ovary syndrome. *Journal of endocrinological investigation*, *29*(3), pp.270-280.

[20] Dhayat, N.A., Marti, N., Kollmann, Z., Troendle, A., Bally, L., Escher, G., Grössl, M., Ackermann, D., Ponte, B., Pruijm, M. and Müller, M., 2018. Urinary steroid profiling in women hints at a diagnostic signature of the polycystic ovary syndrome: A pilot study considering neglected steroid metabolites. PloS one, 13(10), p.e0203903.

[21] Gibson-Helm, M., Teede, H., Dunaif, A. and Dokras, A., 2016. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. *The Journal of Clinical Endocrinology & Metabolism*, *102*(2), pp.604-612.