Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 3, pp. 553–563
DOI : 10.47974/JIOS-1414

A prediction model for poly-cystic ovary syndrome problem using computational intelligence

Dhiraj Pandey *
Department of Information Technology
JSS Academy of Technical Education, Noida
Noida 201301
Uttar Pradesh
India

Kavita Pandey †
Department of Computer Science Engineering & IT
Jaypee Institute of Information Technology Noida
Noida 201301
Uttar Pradesh
India

Budesh Kanwer §
Department of Computer Science Engineering
Poornima Institute of Engineering and Technology
Jaipur 302022
Rajasthan
India

Abstract

Women of childbearing age may face infertility owing to polycystic ovarian syndrome (PCOS).This illness causes ovarian dysfunction, which increases the chance of miscarriage and stillbirth, thus early management is necessary for a healthy lifestyle and to avoid future infections. Weight gain, irregular menstruation cycles, thinning hair, acne, dark and thick spots on the back of the neck, and anxiety disorders are the main symptoms of PCOS. A single out of five women has PCOS. Often women ignore the common symptoms of PCOS and wait until pregnancy problems emerge to get care. Considering PCOS is associated with an increased risk of developing a number of illnesses, such as glucose intolerance, , elevated

*E-mail: dhip2@yahoo.co.in (Corresponding Author)
† E-mail: Kavita.pandey@jiit.ac.in
§ E-mail: budesh82@gmail.com

cholesterol levels, and cardiovascular diseases, it should be identified as soon as feasible. The current tools and therapies are insufficient to identify and forecast PCOS at an earlier stage. To address this issue, we developed a model that will aid in the early detection of PCOS utilizing machine learning techniques and an absolute and minimal set of parameters. The Extra Tree Classifier, a forward selection approach followed by Wrapper, the Chi-square test, and Pearson Correlation was employed as selection criteria to evaluate essential characteristics. KAGGLE has a database that is used for training and testing.

Subject Classification: *68T07.*

Keywords: *Polycystic ovary syndrome, Machine learning, Extra tree classifier, Chi-square test, Wrapper method, Pearson correlation.*

1. Introduction

Clinical speculation models are used to scrutinize the link between future and basic health conditions. It enable patients and their physicians or other health care providers with information about their odds of developing a specific ailment, allowing them to make more educated decisions.

Machine-based models are not a natural extension of the old mathematical methods but provide the most efficient methods for analyzing very large data sets. A polycystic ovary insufficiency is an endocrine illness with a collection of symptoms brought on by a variety of hormonal imbalances. It is one of the most frequent endocrine illnesses in adolescent girls and the leading cause of infertility. Female reproductive malfunction and congenital adrenal hyperplasia are the two main characteristics of PCOS, which is most frequently diagnosed during adolescence.

The rest of the paper is structured as follows. In Section 2, related work on polycystic ovarian syndrome condition that affects women of childbearing surveyed. The proposed approach is elaborated on in Section 3. Section 4 covers the experiments with a discussion of the results. Finally, Section 5 concludes the paper with future directions.

2. Related Work

The topic of PCOS detection has seen a lot of research and development. [1] Suborto Bharati proposed a method[2] in which they employed a feature selection algorithm to determine the best traits that can predict PCOS. Amsy Denny [3] proposed a method that used different machine learning strategies classifiers which provided the best accuracy. Fauzia

Tabassum [4] proposed a model that indicated that PCOS is stressful and it puts people at risk of depression and anxiety disorders, and it can lead to suicidal ideation. A model was put up by Erika Marie Rodriguez [1] that examined a collection of nine people for visual testing who were identified based on their monthly bleeding cycle. For the correlation of 8 out of 9 visual test subjects, the t-test and the Pearson coefficient were used to examine differences between feature and physician ratings. Pijush Dutta [5] proposed a different way to detect PCOS through biochemical parameters. For further correctness in a variety of results, the SMOTE approach is utilized to quantify the basic attributes [6]–[9].According to the above literature review, In the past, various computer vision and deep learning methodologies have been explored to PCOS detection models[9], [10], [11]–[16]. For improved outcomes, we adopted a strategy that is a generalization of some of these techniques in this paper. Several aspects researchers have suggested to tackle the issues using computing techniques [17]-[21].

Proposed Approach

Figure 1 represents the proposed methodology. We have used the data of 541 women and performed various feature selection methods to see the important features out of 42. Then using the different classifiers we found the accuracy of each classifier and found that the ensemble classifier i.e. CatBoost Classifier gave the best accuracy.

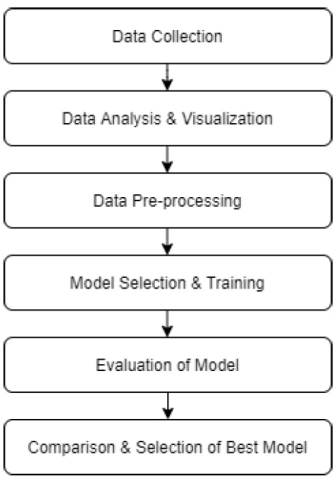

Figure 1

Proposed Methodology

Figure 2 represent the flow of the system.

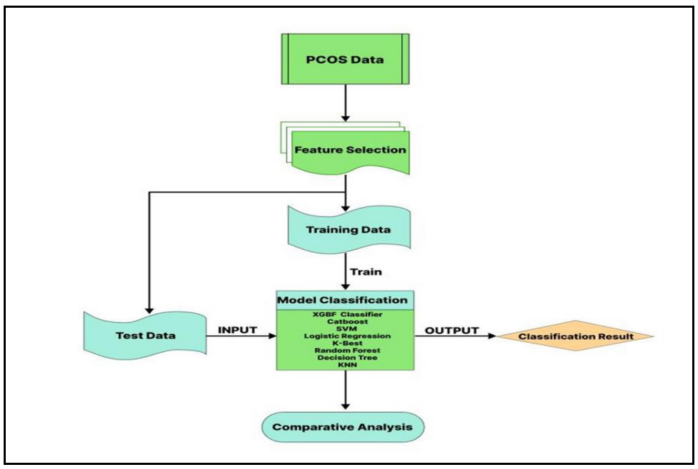

Figure 2

The proposed flow of the System

A. Data Collection

We used the Kaggle dataset, which has two files: PCOS infertility, which contains 541 rows and 5 columns of data obtained from various hospitals, and PCOS without infertility, which contains 541 rows and 42 columns of data collected from various hospitals.

B. Data Analysis and Visualization

Plotly, a robust data presentation software tool, helps us understand data through data visualization. Figure 3 shows the data visualization of the complete dataset.

C. Data Preprocessing

We've done data pre-processing and tweaking, which includes dealing with shortages. External configurations, if available, and data management, among other things, are replaced by 'NaN,' which subtracts.

D. Model Training

This information gives the class to which particular patient belongs to on the classification done based on the information of patient. In this work we have used several classifiers to check their suitability in the dataset.

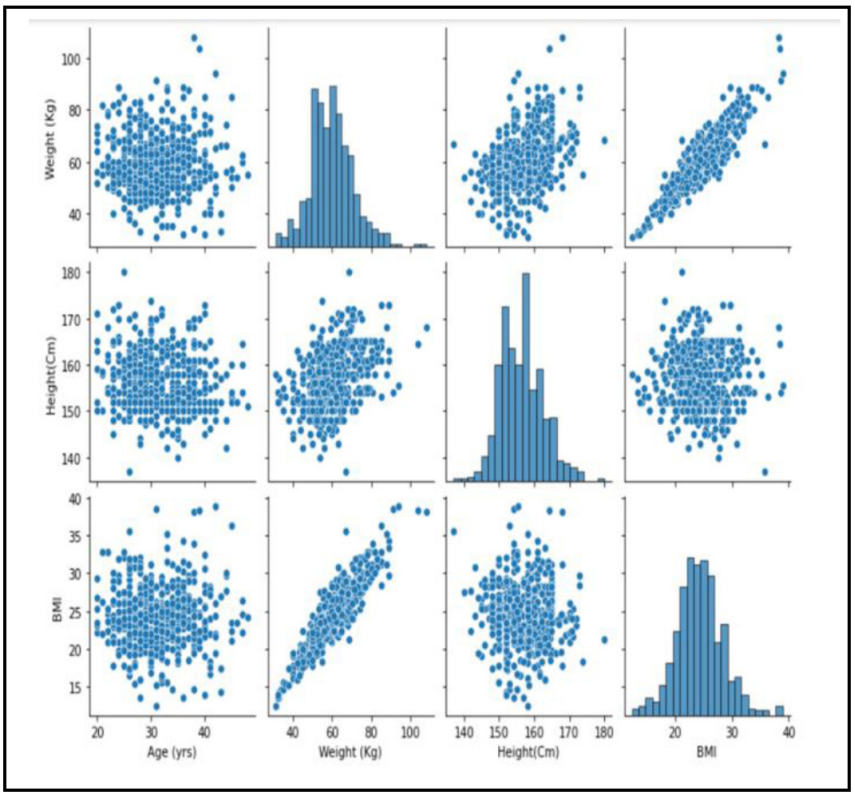

Figure 3

Data visualization

CatBoost Classifier is also giving better results in few problems. Several feature-selection techniques are also used

3. Experiment Analysis

3.1 *Dataset*

42 features include the different symptoms of a women such as period, cycle of period, and other several personal traits. Detailed analysis for this PCOS data along with several features has been analyzed.

3.2 *Results& Discussion*

We have generated reports in the form of graphs, charts, and tables for PCOS data.

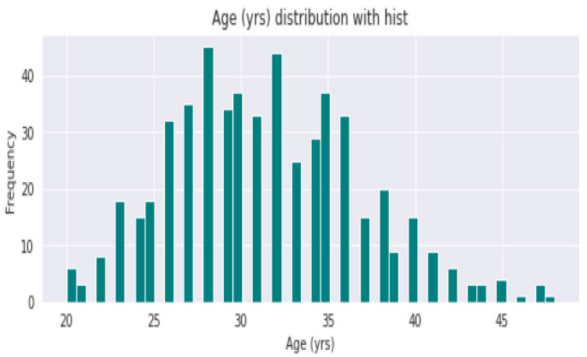

Figure 4

Age groups suffering from PCOS

Figure 4 shows the age groups which are mainly suffering from PCOS. Figure 5 shows the patterns of length of the menstrual cycle in PCOS v/s Normal.

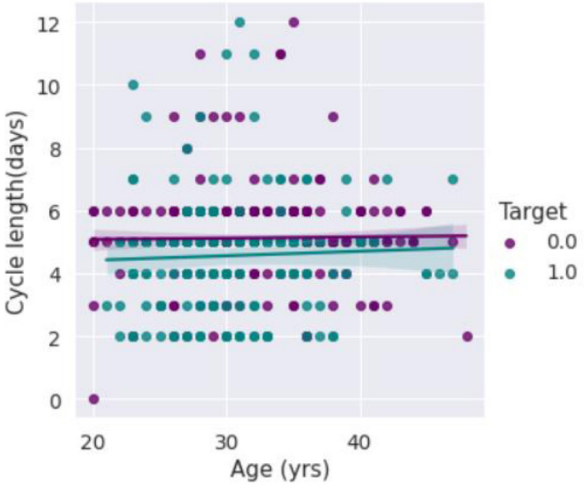

Figure 5

Patterns of the length of the menstrual cycle

Figure 6 shows the distribution of follicles in the left and right ovaries in women having PCOS v/s Normal women. We concluded that the number of follicles in women with PCOS is higher, as expected, and is unequal as well.

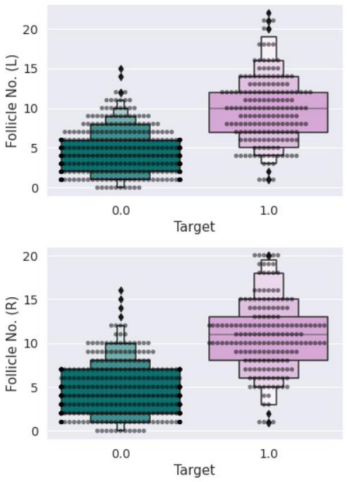

Figure 6

Distribution of follicles in ovaries

Various important features have been extracted from different feature selection methods. Figure.7 shows the Priority features extracted and methods.

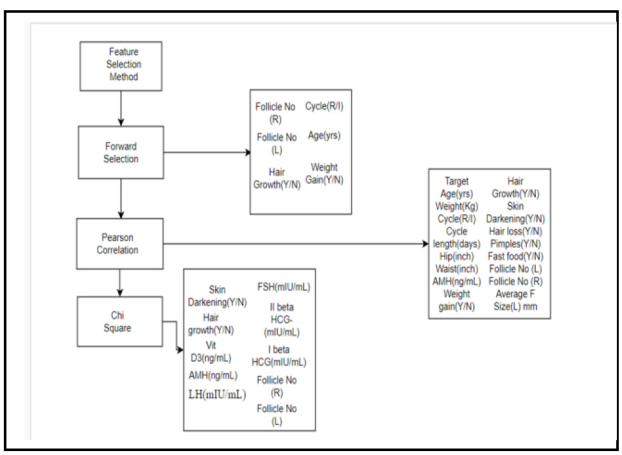

Figure 7

Priority Feature Extracted

The algorithms included 7 models: CatBoost Classifier, a Random Forest Classifier, K-Nearest Relatives, a Support Vector Machine Tree Based Classifier, Xgboost, and Logistic Regression with Aid from Support

Vector Machine. The performance measurements are displayed in table 1 below. Overall effectiveness of the predictions can be assessed based on these indicators, and the prediction fit is decided. CatBoost Classifier delivered the greatest result with a 97% accuracy rate. Fig. 8 shows the plot of different classifiers with accuracy. Table 1 shows the actual value obtained in terms of accuracy.

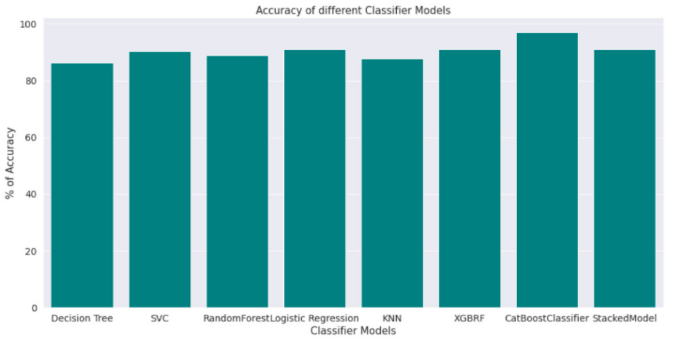

Figure 8

Accuracy of different classifiers

Table 1

Accuracy of different classifier Models

Classifier	Accuracy
Random Forest	0.89
SVC	0.91
Logistic Regression	0.91
K Nearest	0.89
Decision Tree	0.85
Stacked Model	0.91
XGBRF	0.91
CatBoost	0.97

4. Conclusion and Future Directions

The simultaneous endocrine and metabolic changes that affect one another make PCOS a complex disease at this time. Based on the proposed work, it can be concluded that PCOS depends on various factors. Even in the PCOS population, the reason why an individual is suffering from

PCOS is highly uncertain. Following the study from this method, we were able to shortlist some major features which were reflected for all patients suffering from PCOS. Various algorithms have been used to study the proximity and accuracy of the results obtained. We investigated six algorithms, including KNN, Naive Bayes, SVM, Decision Tree, LR, and CatBoost, and discovered that CatBoost had the highest accuracy of 97 percent. We created a prediction model based on this Machine Learning model. This model assists the user in predicting PCOS at an early stage and proposes that the user visit a doctor and receive adequate treatment, lowering the likelihood of PCOS diagnosis being delayed.

In the future, training will be done on a bigger dataset to train and test the model with other alternatives to check the effectiveness of the system and deploy it as a medical assistive tool to help people.

References

[1] E. M. Rodriguez, D. Thomas, A. Druet, M. Vlajic-Wheeler, K. J. Lane, and S. Mahalingaiah, "Identifying Women at Risk for Polycystic Ovary Syndrome Using a Mobile Health App: Virtual Tool Functionality Assessment," *JMIR Form. Res.*, vol. 4, no. 5, p. e15094 (2020).

[2] K. Meena, M. Manimekalai, and S. Rethinavalli, "Correlation of Artificial Neural Network Classification and NFRS attribute Filtering algorithm for PCOS data," *Int. J. Res. Eng. Technol*, vol. 4, no. 3, pp. 519–524 (2015).

[3] F. Tabassum, C. Jyoti, H. H. Sinha, K. Dhar, and M. S. Akhtar, "Impact of polycystic ovary syndrome on quality of life of women in correlation to age, basal metabolic index, education and marriage," *PLoS One*, vol. 16, no. 3, p. e0247486 (2021).

[4] A. K. Srivastava, D. Pandey, and A. Agarwal, "Extractive multi-document text summarization using dolphin swarm optimization approach," Multimed. Tools Appl., vol. 80, no. 7 (2021), doi: 10.1007/s11042-020-10176-1.

[5] N. K. Rathore, D. Pandey, R. I. Doewes, and A. Bhatt, "A Novel Security Technique Based on Controlled Pixel Based Encryption of Image Blocks for Sharing a Secret Image," Wirel. Pers. Commun., vol. 121, no. 1 (2021) doi: 10.1007/s11277-021-08630-w.

[6] A. E. Joham, H. J. Teede, S. Ranasinha, S. Zoungas, and J. Boyle, "Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: data from a large community-based cohort study," *J. women's Heal.*, vol. 24, no. 4, pp. 299-307 (2015).

[7] N. A. Dhayat et al., "Urinary steroid profiling in women hints at a diagnostic signature of the polycystic ovary syndrome: A pilot study considering neglected steroid metabolites," PLoS One, vol. 13, no. 10, p. e0203903 (2018).

[8] A. Bachelot, "Polycystic ovarian syndrome: clinical and biological diagnosis," Médecine la Reprod., vol. 19, no. 3, pp. 214-220 (2017).

[9] H. F. Escobar-Morreale, "Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment," Nat. Rev. Endocrinol., vol. 14, no. 5, pp. 270-284 (2018).

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

[11] D. A. Dumesic, S. E. Oberfield, E. Stener-Victorin, J. C. Marshall, J. S. Laven, and R. S. Legro, "Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome," *Endocr. Rev.*, vol. 36, no. 5, pp. 487-525 (2015).

[12] M. Brower, K. Brennan, M. Pall, and R. Azziz, "The severity of menstrual dysfunction as a predictor of insulin resistance in PCOS," *J. Clin. Endocrinol. \& Metab.*, vol. 98, no. 12, pp. E1967-E1971 (2013).

[13] S. Franks, "Polycystic ovary syndrome," *N. Engl. J. Med.*, vol. 333, no. 13, pp. 853-861 (1995).

[14] R. J. Norman, D. Dewailly, R. S. Legro, and T. E. Hickey, "Polycystic ovary syndrome," *Lancet*, vol. 370, no. 9588, pp. 685-697 (2007).

[15] A. Denny, A. Raj, A. Ashok, C. M. Ram, and R. George, "i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pp. 673-678 (2019).

[16] S. Bharati, P. Podder, and M. R. H. Mondal, "Diagnosis of polycystic ovary syndrome using machine learning algorithms," in *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 1486-1489 (2020).

[17] S. Farzana, Gayathri Harikumar, S. Shankaranarayanan & N. Vikram, "Study on the impact of service quality on the customer satisfaction due to e-banking services of public sectors banks in Chennai," *Journal of Statistics and Management Systems*, vol.25,no.5, pp.1205-1213 (2022).

[18] Ben Ebo Attom & Afzalur Rahman, "Empirical study of the relationship between working capital policies and firm financial performance (Profitability and market value)-Evidence from the manufacturing firms listed on the Ghana stock exchange (GSE)," *Journal of Statistics and Management Systems*, vol. 25, no.4, pp.983-1000 (2022).

[19] Kuan-Min Wang, "The contagion effect of COVID-19 on stock markets," *Journal of Statistics and Management Systems,* vol. 25, no.6, pp.1379-1398 (2022).

[20] Talha Akbar Kamal, Shagufta Naaz, Rajeev Singh Bhandari, Hitendra Shukl & Rehan Khan, "Impact of trade openness and inflation on FDI in India: ARDL Approach," *Journal of Statistics and Management System*s, vol.25, no. 5, pp.1103-1113 (2022).

[21] H. Bhardwaj, M.M. Jamal, "Effect of using Internet of Things for personalized advertisements on consumer buying behaviour," *Journal of Statistics and Management Systems,* vol. 25, no. 2, pp.1135-1146 (2022).