IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Diagnosis of Polycystic Ovary Syndrome using Machine Learning Algorithms

Mr. Abdul Khadar A[1], Shreasta M S[2], Vani N[3], Vathsala K R[4], Vidhya Shree V N[5]
Assistant Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4,5]
S J C Institute of Technology, Chickballapur, Karnataka, India

Abstract: *Artificial intelligence can be used to manage enormous amount of clinical data with great accuracy and precision in healthcare systems for diagnostic reasons. we used machine learning algorithms such as AdaBoost, Gradient Boosting, KNN, Random Forest, and Logistic Regression to diagnose PCOS based on patient clinical data. The data was analysed, and the algorithms' accuracy and precision were validated. The Random Forest algorithm in the diagnosis of PCOS on given data has the maximum accuracy, i.e., 96 percent, according to the validation metrics.*

Keywords: Polycystic Ovarian Syndrome, K-Nearest Neighbour, Random Forest, Logistic Regression.

I. INTRODUCTION

It is a common endocrine condition in females with high androgen levels. Ovulation, hyperandrogenism symptoms, and menstrual disruption are some of the cardinal aspects of PCOS that must be addressed. Pelvic pain, abundant hair, acne, hirsutism, velvety skin, male hormones, irregular periods, danker, and other symptoms are some of the most common signs and symptoms of this illness. About 5–10% of reproductive-age (15–49-year) women are affected by this. Patients with PCOS have the most important qualities, according to the feature selection approach for the provided dataset. Machine learning methods were used to derive the PCOS dataset's key properties. In terms of testing accuracy, we compared the machine learning algorithm methods.

II. PROBLEM IDENTIFICATION

PCOS is a prevalent reproductive endocrinological condition with a wide range of clinical symptoms that affects roughly 5-10% of women during their reproductive years. However, it is critical to obtain an early diagnosis in order to avoid the syndrome's early and late complications. PCOS, an exclusionary diagnosis, has been the subject of much dispute, with several different classifications emerging over time.

III. METHODOLOGY

Step 1: Data collection from data set
Step 2: Data preparation and data cleaning using min max scaler and data labelling using feature selection.
Step 3: Data transformation and splitting of the data and synthetic minority oversampling technique for increasing the number of cases in dataset
Step 4: Use machine learning algorithms like Gradient boosting, logistic regression, Random forest.
Step 5: Comparison and interpretation of results
Step 6: Providing ICT solution for replicating the work.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

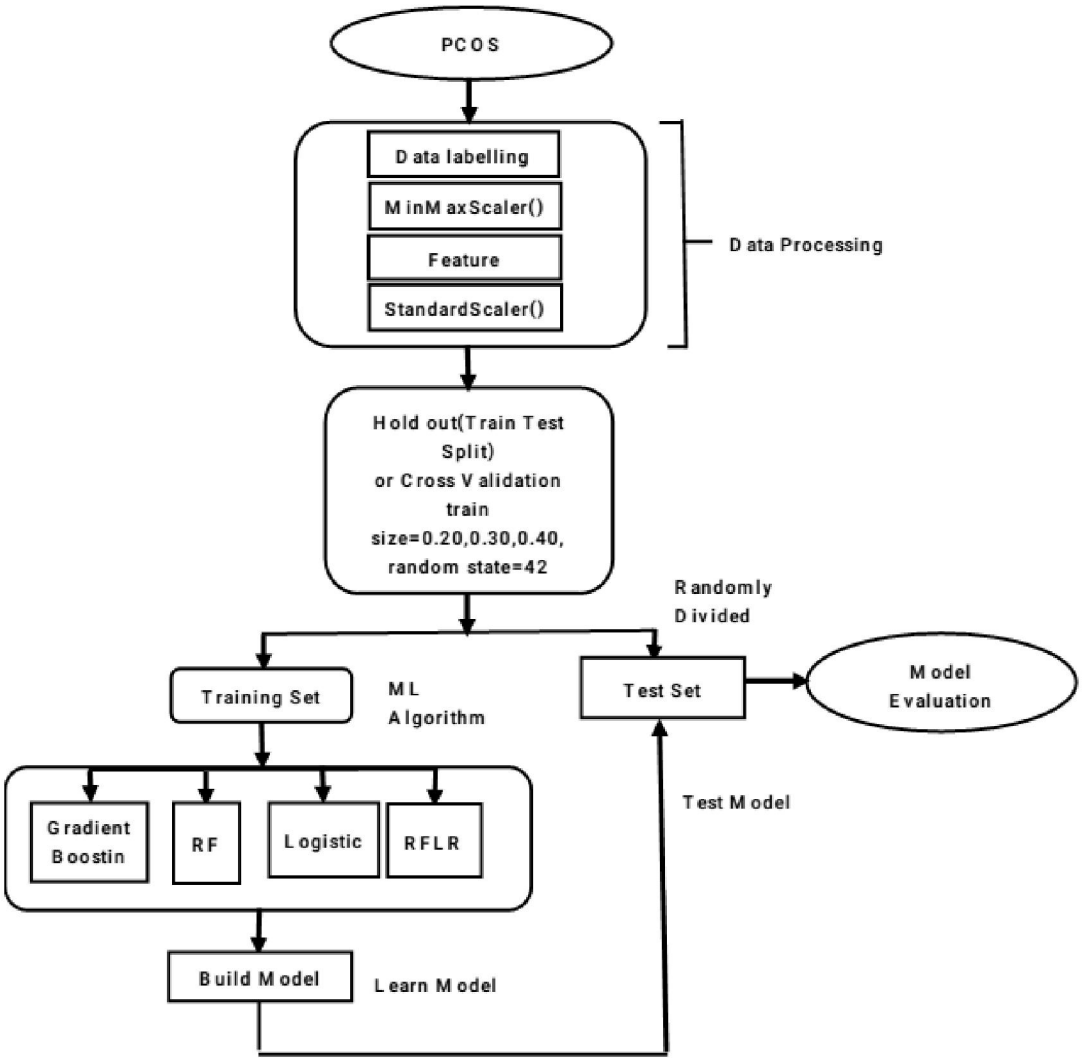


Figure 1: Methodology

IV. IMPLEMENTATION

The project is implemented by using the following modules:

- Data Preprocessing
- Prediction of POCOS
- Performance Evaluation

In this module, the raw dataset is given as input, the system will read the input data in row by row and fill any missing values by calculating the mean value of the column. The unnecessary columns can be dropped by using the drop() function.

Pseudocode:

Step 1: Read raw data using data frame
Step 2: For each row
Step 3: Find missing values
Step 4: If any missing value fill with the columns mean value
Step 5: Save the preprocessed data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Prediction of POCOS

In this module, the preprocessed data is given as input to the system. The system will apply the ADABOOST algorithm to predict the POCOS disease and generate the result according to patient data.

Pseudocode:

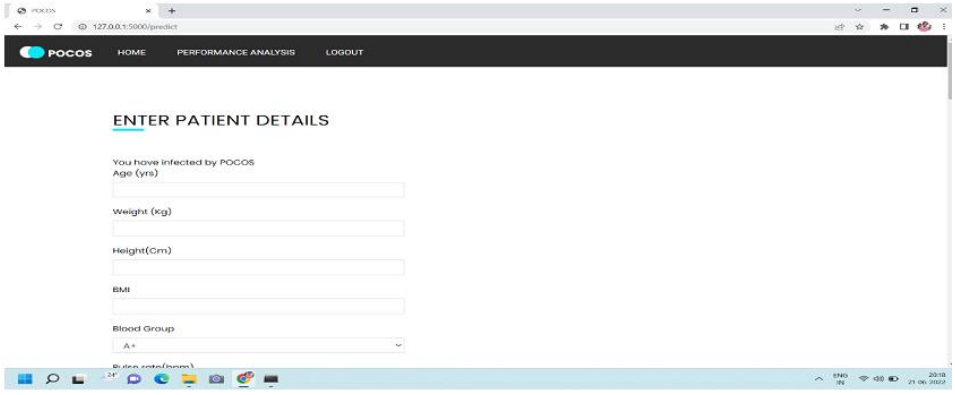
Step 1: Read the preprocessed dataset
Step 2: Read the patient values
Step 3: Split the dataset into train set and test set
Step 4: Train the model using Random Forest algorithm
Step 5: Pass the patient values
Step 6: Predict using predict() function
Step 7: Return results

V.TESTING

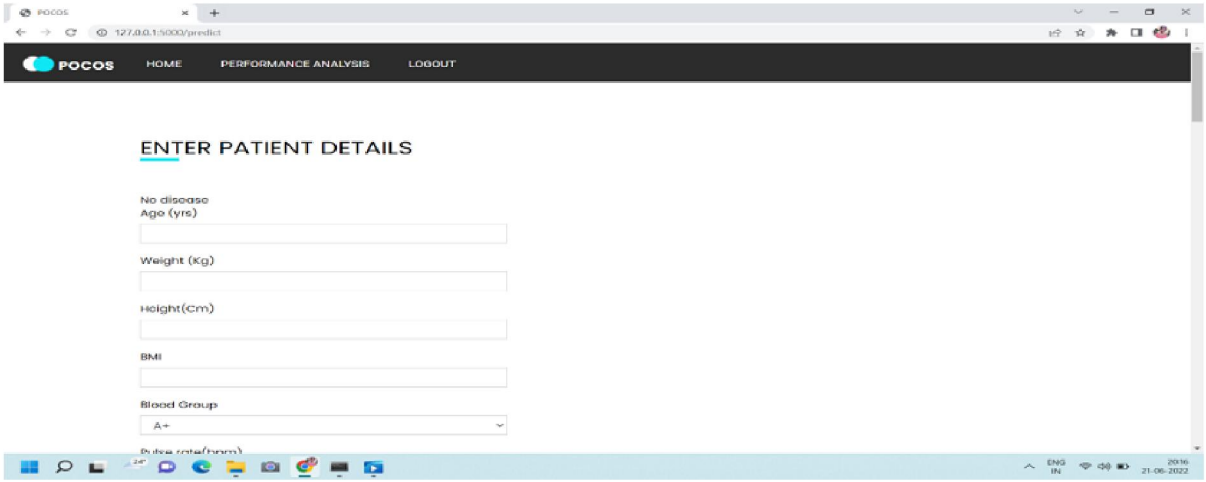
TC NO.	TEST CASE	INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
01	To test user input values	Dataset as input	The file should read and display path	The file read and displayed path	Pass
02	To test user input values	Dataset as null	Show alert message as select dataset	Shown alert message as select dataset	Pass
03	Preprocess	Dataset.csv file	Remove the null fields	Removed null fields	Pass
04	Prediction	User Data	Predict the patients is affected by POCOS or not based on the historical data using machine learning model	Predicted the disease/Normal based on historical data using machine learning model	Pass

VI. RESULTS

1. From the statistically analysed optimal and minimal set of parameters this helps in early detection of POCOS
2. The solution is to take data from the user as input and should return the output with the effective algorithm as the person/patient affected by POCOS or not.
3. While comparing the various algorithms used AdaBoost Classifier algorithm is found more accurate.
4. This is helpful for the doctors for early screening and diagnosing patients.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252



REFERENCES

[1]. V. Deepika, "Applications of Artificial Intelligence Techniques in Polycystic ovarian syndrome Diagnosis," J. Adv. Res. Technol. Manag. Sci., vol. 1, no. 3, pp. 59–63, 2019, doi: 10.1007/978-981-13-2640-0.

[2]. A. Saravanan and S. Sathiamoorthy "Detection of Polycystic Ovarian Syndrome" pp 369-374, 2015.

[3]. M. P. Lauritsen, P. F. Svendsen, and A. N. Andersen, Ugeskr. Laeger, vol. 181, no. pp. 671–679, 2019, doi: 10.1016/S1701-2163(16)32915-2.Diagnostic.

[4]. www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos. last accessed On 5 May 2020

[5]. T. Williams, "Diagnosis and Treatment of Polycystic Ovary Syndrome," Jul. 2016. Accessed: Jul. 25, 2020

[6]. R. S. Legro, "Polycystic Ovary Syndrome and Endocr. Rev., vol. 24, no. 3, pp. 302–312, Jun. 2003, doi: 10.1210/er.2003-0004.

[7]. Cardiovascular Disease: A Premature Association?," Ovary Syndrome," Jul. 2016. Accessed: Jul. 25, 2020. [Online]. Available: