International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I7P219
Volume 71 Issue 7, 201-208, July 2023

Original Article

Detection of PCOS using Machine Learning Algorithms with Grid Search CV Optimization

K. Kavitha[1], Naresh Tangudu[2], Smita Rani Sahu[3], G V L Narayana[4], V. Anusha[5]

[1,2,3,4,5]*Aditya Institute of Technology and Management, Tekkali*

[1]*Corresponding Author : kavitha.akki@adityatekkali.edu.in*

Received: 06 May 2023 Revised: 20 June 2023 Accepted: 26 June 2023 Published: 21 July 2023

Abstract *- Polycystic ovarian syndrome affects a lot of women who are of reproductive age (PCOS), a prevalent endocrine condition, develops. It has an impact on the female reproductive system, leading to polycystic ovaries, hyperandrogenism, and/or Ano/Oligo ovulation. Menstrual irregularities or high levels of androgen (male hormone) can occur in women with PCOS. The ovaries may create a great deal of small follicle clusters (cysts) and stop regularly producing eggs. Some signs of PCOS are period irregularities, an excess of androgen, polycystic ovaries, an abnormal BMI, imbalanced hormone levels, and decreased insulin sensitivity. In order to address this problem, a PCOS early detection app was developed using machine learning techniques. This study investigated the feasibility of creating an automated model to diagnose PCOS using machine learning techniques such as* Light*GBM Boost, Gradient Boost, and XGBoost, then using it with optimization methodology. For the best accuracy, grid search CV for hyperparameter tweaking. This conclusion was reached based on their statistical analysis of the data value on the earlier data set observations. The results are evaluated in terms of accuracy, recall, f1_scorings, and precision and are automated for real-life usage as web-based research.*

Keywords *- Hyperandrogenism, Ano/Oligo ovulation, Polycystic, Follicles, Statistical analysis, Grid search CV.*

1. Introduction

The development of mechanical technology has led to the healthcare sector has seen several changes. One of the key issues is PCOS, also known as functional ovarian hyperandrogenism or Stein-Leventhal syndrome. PCOS is short for polycystic ovarian syndrome, one of the primary issues. A condition known as polycystic ovarian syndrome affects women who are in their reproductive years. About 12–15% of those who are of reproductive age have PCOS. It has an effect on women's ovaries, which are reproductive organs that create the progesterone and estrogen hormones that regulate menstrual cycles. This syndrome commonly results in menstrual irregularities, hyperandrogenism, reproductive problems, and metabolic effects.

The main characteristics of PCOS are cysts in the ovaries, hormonal issues, irregular menstrual cycles, and excessive hair growth. This hormonal imbalance, which also causes irregular menstruation cycles, causes infertility. Depending on the patient's age, race, and other comorbidities, PCOS can manifest itself in various heterogeneous ways. They are more likely to associate PCOS with the most common comorbidities, such as obesity, diabetes, insulin resistance, and depression, than they are to link it to conditions including sleep apnea, fatty liver, endometrial cancer, gestational diabetes, and anxiety. Because of the aforementioned comorbidities, inconsistent diagnostic techniques have been utilized to identify PCOS in female patients. Women reportedly had to wait an average of two years for the correct diagnosis, which left them dissatisfied, irate, and perplexed. When they have trouble getting pregnant and contact their doctor, the majority of women discover that they have PCOS in their 20s and 30s.

The figure-1 shows the image of PCOS with the comparison of a normal ovary and a polycystic ovary. Menstrual cycle irregularities, slow hair development, acne, weight gain, and deepening of skin tags are all common signs of PCOS. A greater risk of miscarriage during the first trimester exists when the ovaries do not produce enough follicles. PCOS can be prevented by early detection.

As a result, the goal of this study is to develop a high-performing diagnostic tool that uses machine learning to lessen and ultimately eliminate human error in PCOS diagnosis. In order to improve accuracy, the diagnosis of PCOS used machine learning methods like XGBoost, LightGBM, and CatBoost, as well as an optimization method called Grid Search CV.

This research work includes:

- Identifying the most crucial attributes utilizing a technique-based method of dataset selection.
- Using the grid search CV optimization approach and machine learning techniques
- Comparing algorithms to evaluate accuracy.
- Lastly, this looks at the web results for a certain person has PCOS or NOT

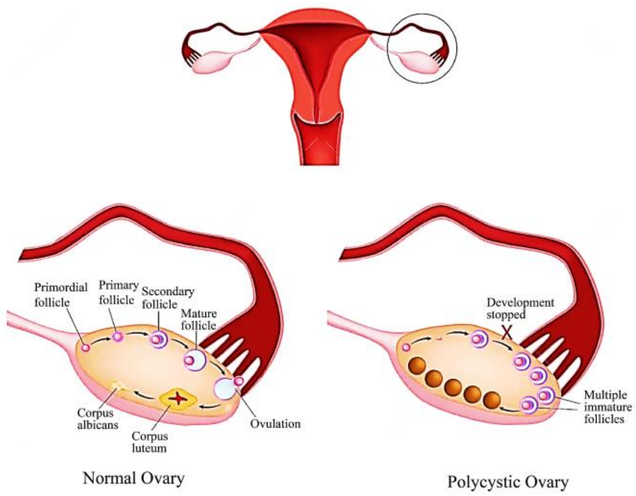


Fig. 1 Polycystic ovarian syndrome [source:]

2. Literature Survey

As technology is increasing immensely day by day, the usage of machine learning has become a great deal as a part of artificial intelligence. In the last few decades, research within the field of artificial intelligence is emphatically affecting educational and medical applications. Apart from that, detailed research on medical issues such as cardiac arrest, different types of cancers and other types of found superior medical conditions by using machine learning technology has become beneficial.

Kinjal Raut et al. (2022) used some algorithms such as Random Forest, Decision tree, support vector classifier, logistic regression, k-Nearest Neighbor, XGBRF and CatBoost classifier. Finally CatBoost Classifier (92.64%) in performance.[1] Abhishek guptha et al. (2022) used boosting ensemble learning methods such as AdaBoost, Gradient Boosting Machine, CatBoost and XGBoost. The most preferable algorithm is the CatBoost classifier, as its accuracy (99.98%) is high among all the boosting algorithms.[2] Subrato Bharathi et al. (2022) An ensemble learning approach has been developed for the Data-Driven Diagnosis of Polycystic Ovarian Syndrome. Also, the cross-validation method is used as the features are chosen, eliminated, and eliminated. The CatBoost classifier's best accuracy is 90.84%.[3] Shamik Tiwari et al. (2022) primarily concentrate on diagnosing PCOS using the clinical data set provided by the Kaggle repository. A variety of machine learning methods are assessed using non-invasive screening parameters. Random Forest performs better than the other machine learning algorithms, with an accuracy rate of 93.25 percent. [4] Wenqi Lv Ying Song et al. (2022) investigate a new technique for screening for polycystic ovarian syndrome utilizing scleral pictures. The suggested method, which is based on deep learning, consists of image processing, feature extraction, and classification processes. AUC for the non-invasive screening technique was 98% on average, with a mean accuracy of 721 participants. [5] Dana Hdaib et.al,(2022) The seven classifiers were used to implement this method for the various machine algorithms. The results showed that the KNN classifier performed best in terms of precision, while the linear discriminant classifier performed best in terms of accuracy. [6] AKM Salman Hosain et al. (2022) created two models, including the tweaked Inception V3 and PCONet- A Convolutional Neural Network (C-NN). PCONet has a superior outcome when these two models are compared, with an accuracy of 98.12%. In the test photos, Inception V3 displayed a 96.56% accuracy in terms of fine-tuning.[7] Jagruthi Vaswani et al. (2022) PCOS detection and disease diagnosis utilizing multiple ML architectures. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are the two primary deep learning methods used (RNN). These models do away with many drawbacks of modern object detection techniques.[8] Manjunath and Alagarswamy et al. (2022) offer a support vector machine, k-Nearest Neighbor, Naive Bayes, and a hybrid method for training and classifying data gathered from the Kaggle repository. The ensemble SVM plus KNN classifier had the greatest classification accuracy of 97 percent when the data was processed.[9] Preeti Chauhan et al. (2021) have experience with a variety of techniques, including Naive Bayes, Support Vector Machine, Decision Tree, and Logistic Regression. The decision tree classifier has an accuracy of 91.12% after applying the feature selection and feature deletion methods to get the most accurate features for prediction.[10] M. Yasmine A. Abu Adla et al. (2021) examined how classification algorithms could be used to detect PCOS. With the aid of cross-validation procedures, statistical analysis, and expert medical judgement, they investigated a number of classifier algorithms on a sizable data set.

The primary goal is now a linear support vector machine with 24 features, and its accuracy and precision are above 90%.[11] M.Sumathi et al. (2021) used the Convolutional Neural Network (CNN) technique to find PCOS. CNN delivers 85% accuracy when employed as an image classifier through feature extraction and segmentation techniques.[12] Namratha Tanwani et al. (2021) worked on a few machine learning models, including Logistic Regression and K-Nearest Neighbor (k-NN), to identify the presence of PCOS. More accuracy was provided using logistic regression (92%). [13] Shakoor Ahmed et al. (2021), using 541 samples with 43 attributes and the algorithms XGBRF, Random Forest, Gradient boosting, and CatBoost, it was examined how to identify PCOS. The CatBoost classifier has the greatest accuracy of all of them at 89%. [14] Homay Danaei Mehr et al. (2021) are working on diagnosing PCOS using various feature selection and machine learning techniques. Using a Kaggle dataset with both entire and reduced features, they employed ensemble AdaBoost, Ensemble Extra Tree, Ensemble Random Forest, and Multilayer Perceptron. Ensemble Random Forest has the highest accuracy at 98.89%. [15] Vikas B et al. (2021) had done research on the Convolutional Neural Network idea. The main goal of this is to examine the effectiveness of well-known deep learning techniques, including CNN, Data Augmentation, and Transfer Learning. Transfer Learning is the best algorithm overall, with an F1 score of 98% and accuracy.[25] Vaidehi Thakre et al. (2020) have employed five different machine learning classifiers, including K-Nearest Neighbors, Naive Bayes, Logistic Regression, Random Forest, and SVM. The top 30 features out of 41 features are extracted using the CHI-Square technique. A Random Forest classifier has the highest accuracy (90.0%). [17] Malik Mubasher Hassan et al. (2020) utilized methods such as a support vector classifier, CART naïve bayes, Random Forest, and Logistic Regression to classify the symptoms as 42 independent variables. the higher accuracy (96%) that logistic regression provided. [18] Anuradha Thakre et al. (2020) presented genetic clustering as a method for improving categorization. For comparing the best outcomes in this fundamental genetic algorithm (GA), additional hybrid gas is also used. With 89% accuracy, the categorization results are optimized.[19] C. Gopal Krishnan et al. (2019) examined the existing studies that are offered for detecting polycystic ovarian syndrome (PCOS) from ultrasound images of the ovaries.

It was shown that it was challenging to identify the PCOS diagnosis using ultrasound imaging accurately.[20] Jay Jojo Cheng et al. (2019) employed Rule-Based Classifiers (RBC) and Gradient Boosted Tree Model (GBT) as two NLP classifiers to identify the PCOS morphology in ultrasound reports from electronic medical records. In addition to learning from the frequency of important phrases and ovarian volume measurements with each report, RBC will measure the size of each ovary.[21] Palvi Soni et al. (2018) outlined the many PCOS symptoms, causes, and therapies. Moreover, data mining activities and approaches that are effective in predicting the condition were employed. [22] E. Setiawati et al. (2015) used the follicle segmentation on non-parametric fitness function as an image clustering technique using Particle swam Optimization. [23][24].

3. Methodology

3.1. Data Collection

This crucial step involves gathering data to evaluate the model's performance and correctness. The data set, which has 541 rows and 43 attributes, was obtained via Kaggle. The dataset's attributes include the patient's file, PCOS, age, weight, height, BMI, blood group, pulse rate, HB, RR rate, pregnancy, the number of abortions, TSH, AMH, PRL, PRG, weight gain, follicle, RBC, skin darkener, hair loss, BP, cycle, cycle length, marriage, FSH, LH, Vitamin d3, RBS, hair growth, pimples increased, fast food, regular exercise, etc.

3.1.1. Selection of Implementation Platform

The platform is used in Jupyter Notebook for this machine learning implementation, and the language used is Python for web page URL-Flask.

3.1.2. Data Preparation

This study uses feature selection to choose the features that are absolutely necessary for the prediction in order to determine whether the person has PCOS or not.

The following characteristics are listed after feature selection: age, height, BMI, pulse rate, irregular periods, respiration rate, number of abortions, ever been pregnant, waist-hip ratio, TSH level, AMH level, PRL level, vitamin D3 level, RBS level, weight gain, abnormal hair growth, darkened skin, noticed any hair loss, increased number of pimples, do you regularly eat fast food, and so forth.

Even after feature selection, the observed results remained the same, with very few variations in accuracy. After that, flaking the qualities with web-based content, such as creating html web pages for the prediction by supplying some significant data, produced the same results.

Once the findings from the feature selection dataset come in, the web page's user interface is utilized to determine whether the subject has PCOS or not.

3.2. Modelling

3.2.1. Boosting

By combining a number of weak classifiers, the "boosting" ensemble modelling technique aims to create a powerful classifier. Weak models are used frequently to build a model. The training data set is first used to create a model. Then a second model is created to remedy the original model's flaws. This process is repeated until the maximum number of models have been added, or the full training set of data has been accurately forecast.

3.2.2. Gradient Boosting

Gradient boosting is one of the most popular boosting methods. In gradient boosting, each prediction corrects the error of the one before it. Each predictor is trained using the predecessor's residual mistakes as labels rather than changing the training instance weights, as in AdaBoost.

The CART learner serves as the basic learner in the Gradient Boosted Trees method (Classification and Regression Trees).

3.2.3. XGBoosting

Using XGBoost technology, gradient-boosted decision trees are developed. XGBoost models frequently win Kaggle Competitions. This strategy calls for the gradual construction of decision trees. Weights are important in XGBoost. Each independent variable is given a weight prior to being fed into the decision tree that forecasts outcomes. Variables that the first decision tree incorrectly predicted are given more weight before entering the second one. Then, these numerous classifiers and predictors are combined to produce an accurate and reliable model. Regression, classification, ranking, and specialist prediction are among the issues it can address.

Formulae behind XGBoosting Classifier

Mathematically the model is in the form of:

$$\hat{y}_i = \sum_{k=0}^{k} f_k(x_i)\,, f_k \in F \tag{1}$$

The objective function for the model is:

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \tag{2}$$

$$= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(l-1)} + f_t(y_i)\right) + \Omega(f_t) + constant \tag{3}$$

$$\text{Obj(t)} = \sum_{j=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) \quad (4)$$

$$= \sum_{i=1}^{n} \left[2\left(\hat{y}1^{(t-1)-y_i}\right) f_t(x_i) + f_t(x_i)^2 \right] + \Omega(f_t) + constant \quad (5)$$

Applying Taylor series expansion to second order:

$$Obj(t) = \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i(t-1)\right) + g_i f_j(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_t) + constant \quad (6)$$

Where g_i and h_i can be defined as :

$$gi = \partial y_i^{(t-1)} l\left(y_i, \hat{y}_i(t-1)\right) \quad (7)$$

$$h_i = \partial_{y_i}^2 (t-1)\left((y_i)_y^{-2} i^{(t-1)}\right) \quad (8)$$

Simplifying and removing constant:

$$\sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} n_i f_t^2(x_i) \right] + \Omega(f_t) \quad (9)$$

$$\text{Here } \ \Omega(f_t) = r^T + \frac{1}{2}\lambda \sum_{i=1}^{T} \omega_i^2 \quad (10)$$

Now our objective function becomes:

$$obj(t) \approx \sum_{i=1}^{n} \left[g_i \omega q(x_i) + \frac{1}{2} h_i \omega_{q(x_i)}^2 \right] + r^T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \quad (11)$$

Simplifying the above expression:

$$0bj^{(t)} = \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2}\left(H_j + \lambda\right) \omega_j^2 \right] + \gamma^T \quad (12)$$

Where

$$G_j = \Sigma_i \in Ij^{g_i} \quad (13)$$

$$H_j = \sum_{i \in I_j} h_i \quad (14)$$

In this equation, w_j are independent of each other where

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (15)$$

$$Obj^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma^T \quad (16)$$

3.2.4. LightGBM Boosting

Model performance is improved using less memory because of the decision tree-based gradient boosting framework known as LightGBM. By utilizing two cutting-edge techniques, gradient-based one-side sampling and exclusive feature bundling, it addresses the problems with the histogram-based strategy, which is the primary method used in all GBDT frameworks (EFB).

The two GOSS and EFB approaches described here are combined to provide the characteristics of the LightGBM algorithm. Combined, they give the model an edge over rival GBDT frameworks and allow it to operate successfully. Due to how quickly the amount of data is increasing every day, traditional algorithms are finding it more and more difficult to deliver results promptly.

Formulae Behind LightGBM Classifier

$$\tilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i\right)^2}{n_r^j(d)} \right) \quad (17)$$

3.2.5. CatBoost

CatBoost is a target-based categorical encoder. A supervised encoder that considers the intended value is used to encrypt the categorical columns. Binomial and continuous goals are both supported.

Target encoding is a popular technique for category encoding. The average value of the target assigned to that category in the training dataset, combined with the target probability determined across the entire dataset, is substituted for a categorical feature. This causes a target leakage, though, because the target is required to foresee the target. When used in fresh contexts, these models usually show overfitting and perform badly.

Formulae Behind CatBoost Classifier

$$F^t = F^{t-1} + \infty \cdot h \quad (18)$$

$$h^t = \underset{h \in H}{\arg min}\, \mathbb{E}\mathcal{L}(y, F^{t-1} + h) \quad (19)$$

$$h^t = \underset{h \in H}{\arg min}\, \mathbb{E}\left(\frac{\delta \mathcal{L} y}{\delta F^{t-1}} - h\right)^2 \approx \underset{h \in H}{\arg min} \frac{1}{n}\left(\frac{\delta \mathcal{L} y}{\delta F^{t-1}} - h\right)^2 \quad (20)$$

3.2.6. Grid Search CV Optimization

A technique for finding the ideal parameter values from a set of parameters in a grid is called grid search CV. It functions basically as a cross-validation approach. The model and its parameters must be entered. The optimal parameter values are extracted before creating the forecasts. Grid Search CV is a machine-learning library for Python. A thorough search across the designated parameter values for an estimator is performed. In essence, an estimator object must have a score function; otherwise, any kind of scoring must pass. Fit and forecast are the two primary strategies that may be used with grid search CV.

3.2.7. Web-Technologies

HTML stands for hypertext markup language. Markup language is used to create web pages. Markup language and hypertext are both used to define the text document inside the tag that specifies the structure of online pages.

CSS can be used to create attractive web pages. CSS stands for cascading style sheets. Using this will simplify the process of creating attractive web pages. It can be used to lay out web pages. More importantly, it enables you to accomplish this without relying on the HTML that makes up each online page.

3.2.8. Flask

The web application of the Flask framework was developed in Python. Armin Ronacher was the developer of this and the leader of the international Python organization Pocco.

The parameter passed to the Flask function Object () [native code] is the name of the currently active module (__name__).

The route () function of the Flask class is a decorator that tells the application which URL to use to invoke the associated method.

Formulation: app.route (rule, options) The rule parameter for the function specifies the URL binding.

- The choices describe a set of inputs that should be sent to the underlying Rule object.
- The URL for() method is especially helpful for creating dynamic URLs for certain functions. The first parameters of the function are the name of the function and one or more keyword arguments corresponding to the variable component of the URL.

4. Results

Table 1 represents all the accuracy values produced by the models used in this research.

These performance results will let to know how much accurate the particular model is.

Table 1. Performance results of all algorithms

Algorithms	Accuracy	Precision	Recall	F1-score
XGBoost	88.85%	88%	94%	91%
LightGBM	88.89%	88%	97%	92%
CatBoost	88.89%	88%	95%	92%
Grid Search CV for CatBoost classifier	94.23%	89%	97%	92%
Grid Search CV for LightGBM	93%	87%	94%	91%
Grid Search CV for XGBoost	93%	88%	92%	90%

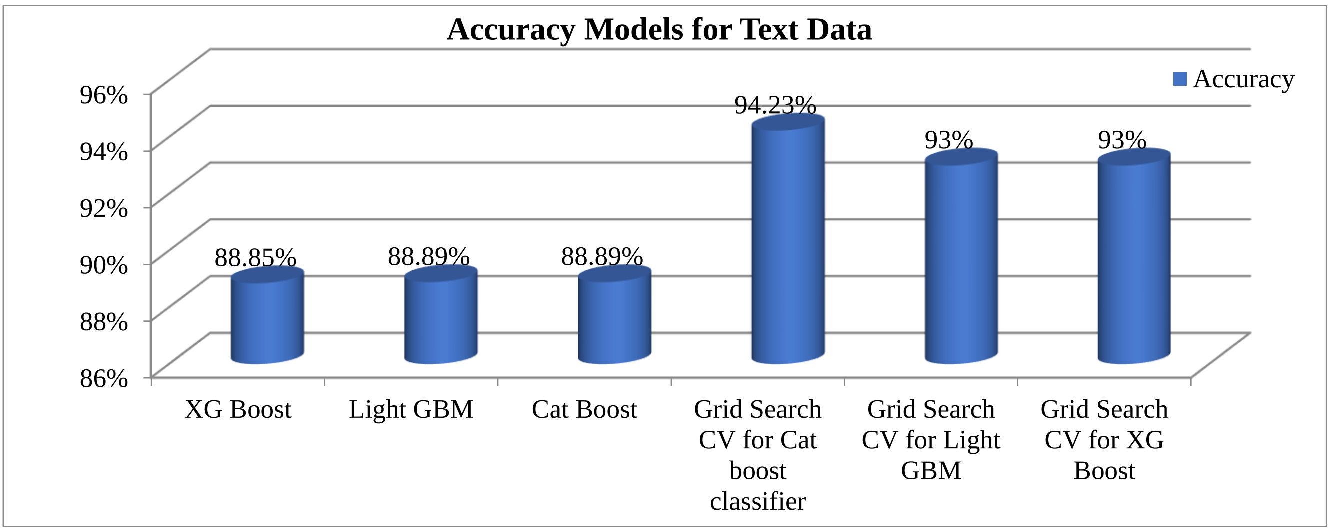
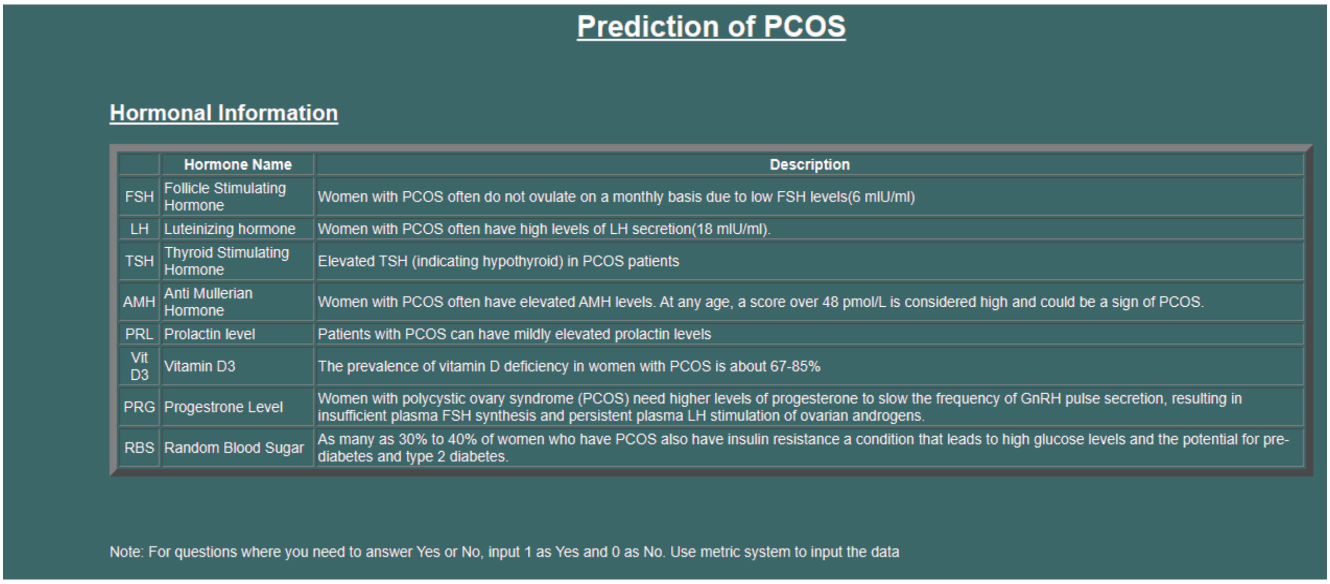


Fig. 2 Accuracy comparison



Prediction of PCOS

Hormonal Information

	Hormone Name	Description
FSH	Follicle Stimulating Hormone	Women with PCOS often do not ovulate on a monthly basis due to low FSH levels(6 mIU/ml)
LH	Luteinizing hormone	Women with PCOS often have high levels of LH secretion(18 mIU/ml).
TSH	Thyroid Stimulating Hormone	Elevated TSH (indicating hypothyroid) in PCOS patients
AMH	Anti Mullerian Hormone	Women with PCOS often have elevated AMH levels. At any age, a score over 48 pmol/L is considered high and could be a sign of PCOS.
PRL	Prolactin level	Patients with PCOS can have mildly elevated prolactin levels
Vit D3	Vitamin D3	The prevalence of vitamin D deficiency in women with PCOS is about 67-85%
PRG	Progestrone Level	Women with polycystic ovary syndrome (PCOS) need higher levels of progesterone to slow the frequency of GnRH pulse secretion, resulting in insufficient plasma FSH synthesis and persistent plasma LH stimulation of ovarian androgens.
RBS	Random Blood Sugar	As many as 30% to 40% of women who have PCOS also have insulin resistance a condition that leads to high glucose levels and the potential for pre-diabetes and type 2 diabetes.

Note: For questions where you need to answer Yes or No, input 1 as Yes and 0 as No. Use metric system to input the data

Fig. 3 Hormonal information

Fig. 2 shows the graphical representation of accuracy according to each model, giving the basic idea of the algorithm's performance.

4.1. Web-based Results

Fig. 3 gives the hormonal information of all the symptoms. It provides the functionality for FSH (Follicle stimulating hormone), LH (Luteinizing hormone), TSH (Thyroid-stimulating hormone), AMH (Anti Mullerian hormone), PRL (Protein level), VI D3 (Vitamin D3), PRG (Progesteine level), RBS (Random blood sugar).

Fig. 4 represents the questions related to symptoms based on the answers filled in the given blank; it will show whether the person is suffering from PCOS or not.

Fig. 5 and Fig. 6 represents the levels of a person who is suffering from PCOS and also the levels of a person who is not suffering from PCOS.

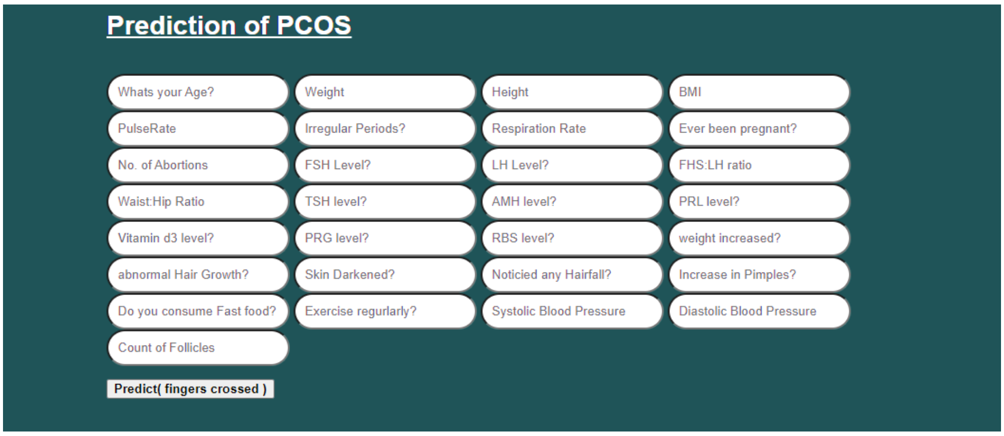


Fig. 4 Questions regarding prediction

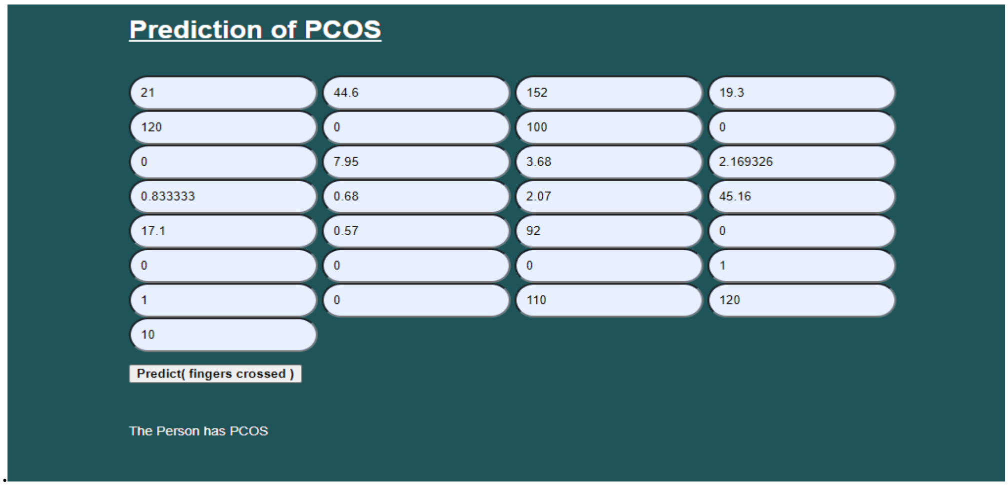


Fig. 5 Person suffering from PCOS

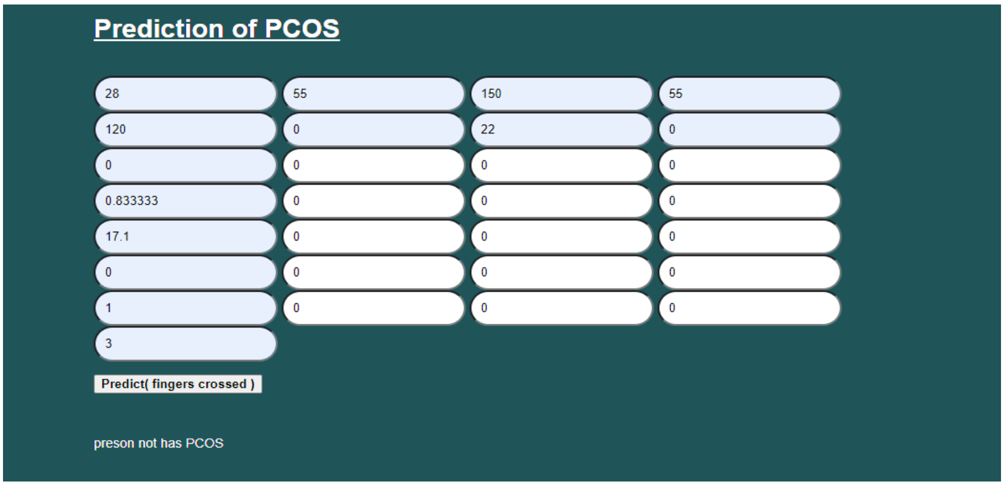


Fig. 6 Person not suffering from PCOS

5. Conclusion

A lot of research has already been drawn to machine learning. Algorithms for machine learning are able to gain knowledge from the information. Automating multiple decision-making processes is a very useful feature of machine learning algorithms. A hyperandrogenic condition known as PCOS (polycystic ovarian syndrome) affects women of reproductive age. Due to the multiplicity of probable explanations for some of its symptoms, PCOS can be challenging to diagnose. There are a lot of older studies on PCOS detection in this endeavor. The XGBoost, LightGBM, and CatBoost algorithms were used in this study. These algorithms were then each applied to the grid search CV optimization technique, and the outcomes were observed. The CatBoost classifier algorithm, which was used in this study's conclusion, has the highest accuracy among the aforementioned algorithms. The best ideal solution is produced, and the accuracy observed is 94%. For the front-end application or the static web pages uses HTML and CSS, where HTML is used for the structure of a web page, and CSS is used for the designing of a web page. It produces URLs using Flask, which was the concept of Web Technologies. The application's web page interface allows users to engage with it, provides information on the questions they are asked, and predicts whether or not they have PCOS.

References

[1] Kinjal Raut, Chaitrali Katkar, and Suhasini A. Itkar, “PCOS Detect using Machine Learning Algorithms,” *International Research Journal of Engineering and Technology*, vol. 9, no. 1, 2022. [Google Scholar] [Publisher Link]

[2] Abhishek M. Guptha et al., “Succinct Differentiation of Desperate Boosting Ensemble Learning Methods for Prognostication of Polycystic Ovary Syndrome Diagnosis,” *International Conference on Advances in Computing, Communication, and Control*, pp. 1-5, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Subrato Bharati et al., “Ensemble Learning for Data-Driven Diagnosis of Polycystic Ovary Syndrome,” *Intelligent Systems Design and Applications*, vol. 418, pp. 1250–1259, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Shamik Tiwari et al., “SPOSDS: A Smart Polycystic Ovary Syndrome Diagnostic System using Machine Learning,” *Expert Systems with Applications*, vol. 203, p. 117592, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Wenqi Lv et al., “Deep Learning Algorithm for Automated Detection of Polycystic Ovary Syndrome using Scleral Images,” *Frontiers in Endocrinology*, vol. 12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Dana Hdaib et al., “Detection of Polycystic Ovary Syndrome (PCOS) using Machine Learning Algorithms,” *5th International Conference on Engineering Technology and its Applications, IEEE*, pp. 532-536, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] A.K.M Salman Hosain, Md Humaion Kabir Mehedi, and IrtezaEnan Kabir, “PCONet: A Convolutional Neural Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images,” *Proceedings of the 8th International Conference on Engineering and Emerging Technologies,* Kuala Lumpur, Malaysia, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jagruthi Vaswani et al., “A Systematic Literature Review on the Diagnosis of PCOS using Machine Learning Algorithms,” *GIT_Journal of Engineering and Technology*, 2022. [Google Scholar] [Publisher Link]

[9] Manjunathan Alagarsamy et al., “Detection of Polycystic Syndrome in Ovary Using Machine Learning Algorithm,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1, pp. 246-253, 2023. [Google Scholar] [Publisher Link]

[10] Preeti Chauhan et al., “Comparative Analysis of Machine Learning Algorithms for Prediction of PCOS,” *IEEE International Conference on Communication information and Computing Technology*, Mumbai, India, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yasmine A. Abu Adla et al., “Automated Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques,” *Sixth International Conference on Advances in Biomedical Engineering,* pp. 208-212, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] M. Sumathi et al., “Study and Detection of PCOS Related Diseases using CNN,” *IOP Conference Series: Materials Science and Engineering,* 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Namrata Tanwani, “Detecting PCOS using Machine Learning,” *International Journal of Modern Trends in Engineering and Science,* vol. 7, no. 1, 2020. [Google Scholar] [Publisher Link]

[14] Shakoor Ahmad Bhat, “*Detection of Polycystic Ovary Syndrome using Machine Learning Algorithms*,” MSc Research Project, MSc. Data Analytics, 2021. [Google Scholar] [Publisher Link]

[15] Homay Danaei Mehr, and Huseyin Polat, “Diagnosis of Polycystic Ovary Syndrome through Different Machine Learning and Feature Selection Techniques,” *Health and Technology,* vol. 12, pp. 137-150, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] S. Aswath, and S. Valarmathi, "Obstructive Sleep Apnea Severity Prediction Model GUI using Anthropometrics," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 12, pp. 134-144, 2022. [CrossRef] [Publisher Link]

[17] Shreyas Vedhpatak, Vaidehi Sunil Thakre, “PCOS Detection and Prediction using Machine Learning Algorithms,” *Bioscience Biotechnology Research Communications,* vol. 13, no. 14, pp. 240-244, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Malik Mubasher Hassan, and Tabasum Mirza, “Comparative Analysis of Machine Learning Algorithms in Diagnosis of Polycystic Ovarian Syndrome,” *International Journal of Computer Applications*, vol. 175, no. 17, 2020. [Google Scholar] [Publisher Link]

[19] Anuradha D. Thakare, and Priyanka R. Lele, “Genetic Clustering for Polycystic Ovary Syndrome Detection in Women of Reproductive Age,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 3, pp. 1356-1361, 2020. [Google Scholar] [Publisher Link]

[20] C. Gopalakrishnan, and M. Iyapparaja, "A Detailed Research on Detection of Polycystic Ovary Syndrome from Ultrasound Images of Ovaries," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S11, 2019. [Google Scholar] [Publisher Link]

[21] Jay Jojo Cheng, and Shruthi Mahalingaiah, "Data Mining Polycystic Ovary Morphology in Electronic Medical Record Ultrasound Reports," *Fertility Research and Practice*, vol. 5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Palvi Soni, and Sheveta Vashisht, "Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques," *Proceedings of the International Conference on Communication and Electronics Systems, IEEE*, pp. 816-820, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] K.Kavitha, and Suneetha Chitttineni, "Efficient Sentimental Analysis using Hybrid Deep Transfer Learning Neural Network," *International Journal of Engineering Trends and Technology,* vol. 70, no. 10, pp. 155-165, 2022. [CrossRef] [Publisher Link]

[24] E. Setiawathi, Adiwijaya, and ABW Tjokorda, "Particle Swarm Optimization on Follicles Segmentation to Support PCOS Detection," *International Conference on Information and Communication Technology,* pp. 369-374, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[25] B Vikas, Y Radhika, and K Vineesha, "Detection of Polycystic Ovarian Syndrome using Convolutional Neural Networks," *International Journal of Current Research and Review,* vol. 13, no. 6, 2021. [Google Scholar] [Publisher Link]