



**HAL**  
open science

# Developing a System for Automatic Prediction of Polycystic Ovary Syndrome Using Machine Learning

Kemi Akanbi, Odunayo G Adepoju, Isaac Kofi Nti

► **To cite this version:**

Kemi Akanbi, Odunayo G Adepoju, Isaac Kofi Nti. Developing a System for Automatic Prediction of Polycystic Ovary Syndrome Using Machine Learning. 2024. hal-04618138

**HAL Id: hal-04618138**

**<https://hal.science/hal-04618138v1>**

Preprint submitted on 20 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# DEVELOPING A SYSTEM FOR AUTOMATIC PREDICTION OF POLYCYSTIC OVARY SYNDROME USING MACHINE LEARNING

Kemi Akanbi\*

School of Information Technology, University of Cincinnati, akanbikr@mail.uc.edu

Odunayo G. Adepoju

School of Information Technology, University of Cincinnati, adepojog@mail.uc.edu

Isaac Kofi Nti

School of Information Technology, University of Cincinnati, ntious1@gmail.com

Polycystic Ovary Syndrome (PCOS) is a condition that leads to lifelong health problems outside of infertility. The lack of a single, known cause and universal symptoms makes diagnosis challenging. The early and accurate prediction will prevent many subsequent serious and morbid illnesses that can arise from PCOS. Therefore, this study proposes a predictive Machine Learning (ML) model to identify patients at risk of PCOS and alert healthcare professionals, allowing for early intervention. The predictive performance of the Random Forest Classifier, Logistic Regression, Gradient Boost, Adaptive Boost, and XGBoost machine learning algorithms was compared based on accuracy, precision, recall, and F1-score with a publicly available dataset from Kaggle. The experiment was performed in Google Colab. Our experimental results showed a good predictive performance of 90% across all evaluation metrics. However, Random Forest outperformed all other models achieving 96% accuracy, precision, recall, and F1-score, respectively. The high accuracy (96%) obtained by this study suggests that the proposed model could effectively identify patients at risk of PCOS, potentially aiding early diagnosis and intervention.

CCS CONCEPTS • Applied computing • Life and Medical sciences • Health Informatics

**Additional Keywords and Phrases:** Polycystic Ovary Syndrome, Diagnosis, Infertility, Machine Learning

## ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

## 1 INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a condition caused by an imbalance of hormones in women within childbearing years thereby affecting the ovary and thereby causing infertility [1,2]. It can lead to other complicated health conditions such as endometrial cancer, depression, HBP, and heart disease [2]. The United States Center for Disease Control and Prevention reports that infertility in the US is often linked to PCOS, with as many as five million women affected [3]. The prevalence of the condition is not limited to the United States, PCOS is a common health problem affecting women worldwide [2]. A sizable portion of women with PCOS remain undiagnosed, with studies suggesting less than half receive a proper diagnosis [1]. Despite its widespread occurrence, the exact cause that brings about PCOS is unclear, it may be inherited or be triggered by the environment or body [4] leading to elusiveness in timely detection and effective treatment [5].

---

\* Place the footnote text for the author (if applicable) here.

In affected women, there is a development of small fluid-filled sacs known as cysts in the ovaries resulting in the inability of the ovaries to produce eggs [6]. Polycystic Ovary Syndrome (PCOS) is a complex condition marked by various symptoms like irregular periods, cysts in the ovaries, hormonal imbalances, and weight gain; A common effect is excessive hair growth (hirsutism) on the body and face, along with acne and pimples.[7]

Machine Learning, an important aspect of Artificial intelligence has emerged as a revolutionary force in driving solutions due to its ability to learn the patterns and relationships within data [1]. Due to the challenging nature of diagnosing PCOS, a diagnostic approach can be established through the utilization of Machine Learning which has demonstrated notable efficacy in accurately diagnosing diseases within the healthcare domain [8].

By seamlessly integrating Machine Learning into PCOS diagnosis, we can significantly improve the lives of women affected through early detection which unlocks the keys to managing symptoms, addressing infertility challenges, and adopting healthy habits [9]. Early detection can mitigate the effect on women suffering from PCOS and prevent the progression to more severe health situations [2]

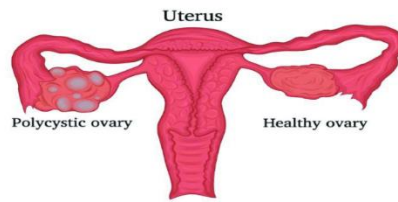


Figure 1: Healthy and Polycystic Ovary [4]

Therefore, this study aims to develop a machine-learning model for early PCOS detection. Specifically: (i) compare different algorithms, select the best performing one, and optimize its processing speed, (ii) identify significant factors influencing PCOS prediction. The current study contributes to knowledge as follows: (i) advancing PCOS diagnostic techniques for earlier diagnosis and intervention. (ii) Understanding effective computational methods for PCOS detection through comparing and selecting the best-performing algorithm. (iii) expanding knowledge of PCOS by identifying key factors influencing prediction, potentially informing future research and clinical management.

The remaining parts of this paper are organized as follows. Section 2: Related works Section 3: Methodology with subsections on Dataset Acquisition, Data Preparation, Model Training, Selection, and Model Performance Evaluation. Section 4: Result and Discussion 5. The conclusion and Future Direction.

## 2 RELATED WORKS

In study [1], the researchers compared the performance of 5 models: Naïve Bayes, KNN, SVM, DT, and Logistic Regression. The metrics utilized in the study are accuracy, precision, and specificity. DT proved to be the highest performer recording an accuracy of 81%, precision of 70%, and specificity of 94%. The researchers posit that the study can be expanded into developing an application for tracking PCOS [1]. [4] designed a mobile application featuring a machine-learning model to estimate the likelihood of a PCOS diagnosis. The study evaluated the performance of six models and discovered Logistic Regression to be the topmost performer due to its high precision rate with precision for classes 0 and 1 is 0.91 and 0.95, Recall for 0 and 1 is 0.98 and 0.91, and F1-score for 0 and 1 is 0.93 and 0.95 respectively [4].

The study [5] employed ANN in the detection of PCOS using 10 input variables. The Model proved to be an excellent performer with all metrics above 90%; Measures recorded are accuracy of 96.1%, Specificity of 90%, and Sensitivity of 96.8%. This proves that ANN can recognize patterns in PCOS. However, the authors identified the complexity in terms of computational power as a drawback because other simple models may achieve the same result [5]. The findings of [8] indicate that the linear discriminant classifier proved to be the most effective in terms of accuracy, precision, and specificity of 92.6,97.6,93.55. However, the KNN classifier showed the highest level of sensitivity of 97.22 among the tested classifiers. The author compared the performance of KNN, NN??, NB, SVM, Tree, Logistic Regression, and Linear Discriminant classifier. They indicated improvement to the models can be achieved by carefully selecting the parameters of the algorithms used in creating the models [8].

In the study [9] Linear SVM records the highest performance across the performance metrics employed given a precision of 0.93, 0.80 recall and 0.91 accuracies out of the 11 algorithms used to develop the models. The models compared for top performance are Linear SVM, Polynomial SVM, and KNN. The authors reported a need to optimize the selected model performance as a future research direction [9]. The work of [18] demonstrated that feature selection improves the accuracy of a model. The study experiment fitted LR, DT, SVM, and RF on the Kaggle PCOS dataset. RF achieved an accuracy of 88.90% with 40 input features after the removal of low-correlated features, accuracy increased to 92.66. The author indicates that future work would be to broaden the study by including more records in that dataset. Study [17] investigated the use of 40-fold cross-validation on a subset of the dataset containing the ten highest-ranking features and evaluated with RF, Logistic Regression, and RFLR. The best-performing model is RFLR (“Hybrid Random Forest and Logistic Regression”) obtaining an accuracy of 91.01%, a precision of 89.90 %, and a recall rate of 90 %.

### 3 METHODOLOGY

Figure 2 shows the study proposed framework. Figure 2 shows the study proposed framework. This section details each phase of our framework.

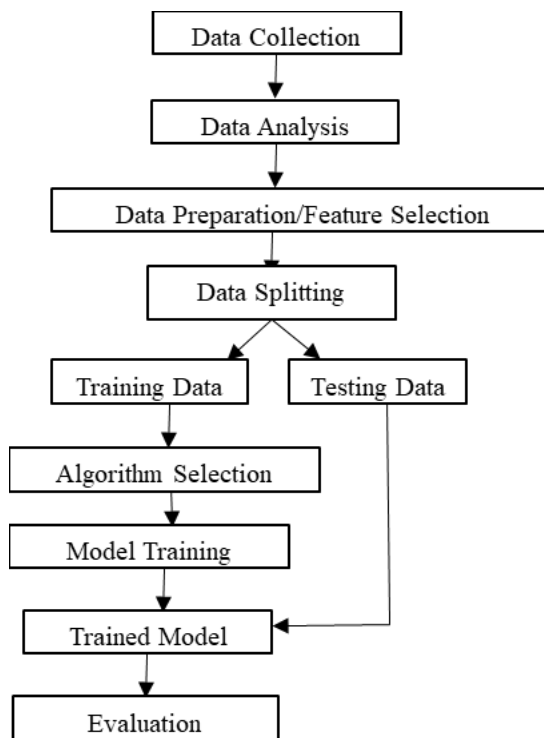


Figure 2: Proposed Methodology Framework.

#### 3.1 Dataset Acquisition

The dataset utilized in the study is a publicly available dataset from Kaggle. The PCOS dataset contains 541 instances and 44 features which are both physical and clinical. Table 1 shows the features present in the dataset.

Table 1: PCOS Dataset Features

No	Feature	No	Feature	No	Feature	No	Feature
1	Serial Number	12	Cycle(R/I)	23	Waist(inch)	34	Hair loss
2	Patient File Number	13	Cycle length(days)	24	Waist: Hip Ratio	35	Pimples
3	Polycystic Ovary Syndrome (Yes/nO)	14	Marraige Status (Yrs)	25	Thyroid-Stimulating Hormone (TSH (mIU/L) )	36	Fast food
4	Age (yrs)	15	Pregnant(Y/N)	26	Anti-Mullerian Hormone (AMH(ng/mL) )	37	Regular Exercise
5	Weight (Kg)	16	No. of abortions	27	Prolactin levels(PRL(ng/mL)	38	Blood Pressure _Systolic (mmHg)
6	Height(Cm)	17	I beta-HCG(mIU/mL)	28	Vitamin D (ng/mL)	39	Blood Pressure _Diastolic (mmHg)
7	Body Mass Index(BMI)	18	II beta-HCG(mIU/mL)	29	Progesterone Levels(ng/mL)	40	Follicle Number. (Left)
8	Blood Group	19	Follicle-Stimulating Hormone (FSH(mIU/mL) )	30	Random Blood Sugar(mg/dl)	41	Follicle Number. (Right)
9	Pulse rate(bpm)	20	Luteinizing Hormone (LH(mIU/mL))	31	Weight gain	42	Avg. F size (L) (mm)
10	Respiration Rates( breaths/min)	21	Follicle-Stimulating Hormone/ Luteinizing Hormone ((FSH/LH) )	32	hair growth	43	Avg. F size (R) (mm)
11	Hemoglobin (Hb(g/dl) )	22	Hip(inch)	33	Skin darkening	44	Endometrium

### 3.2 Data Preparation

To understand the data, we utilized Seaborn and Matplotlib Libraries for visualization, we checked for missing values and columns containing mixed data. The missing values were replaced with the mean value of the data points in the column.

Feature selection (FS) is a crucial step to improve the classifier's performance and reduce computation time. FS consists of selecting the most effective or prominent features from the complete set of features. When irrelevant or redundant features are removed, the classifier's performance is boosted, and good classification results will be achieved [9].

The correlation of the input features to the target variable was investigated. Based on the correlation, we selected features that have a 0.1 or above correlation score with the PCOS (Y/N) and they were used as the model's input variable. Using the Random Oversampling Technique, balancing the class distribution was achieved. The target variable consists of 177 instances of PCOS diagnosis coded as 1 and 364 instances with no PCOS diagnosis coded as 0. The huge difference between the number of positive and negative diagnoses may skew the performance of the models. Hence, we adopted the random oversampling method for class balance. Finally, we partition our clean data into 80% for training and 20% for testing.

### 3.3 Algorithm Selection and Model Training

The following algorithms were used, Random Forest Classifier, Logistic Regression, Gradient Boost, Adaptive Boost, and XGBoost. Ensemble Machine Learning algorithms have been identified to perform better when used for diagnosis in medicine [2]. Given the exceptional ability of ensemble-based gradient-boosting algorithms in discerning patterns, their impact on medical research and practice is becoming increasingly significant [10]. Different boosting algorithms have various strengths. Adaboost helps in transforming a weak classifier into a strong one and helps the weakest classifier boost its accuracy [11]. Gradient Boosting classifier belongs to the Gradient Boosting Decision Trees (GBDT) class which integrates multiple trees to produce a group of dependent predictors [6,12] while XGBoost employs parallel and distributed computing making learning faster and enabling quicker model exploration [13]. Likewise, Random Forest is one of the ensembles learning models [14] that assembles multiple decision trees to make decisions [6].

### 3.4 Performance Evaluation

The metrics adopted for the evaluation of the performance of the models in this study are Accuracy, Precision, recall, and f1-score. To effectively evaluate the accuracy of predictions, [6] identified some essential terms of the performance metrics components that need to be understood. The metric elements which are TP, TN, FP, and False Negative are defined by Shazia Nasim [6] as

- TP- Patients who have PCOS and are correctly classified as having PCOS by the model.
- FP: Patients who do not have PCOS and are incorrectly classified as having PCOS by the model.
- TN: Patients who do not have PCOS and are correctly classified as non-PCOS by the model
- FN: Patients with PCOS and are incorrectly classified as non-PCOS by the model.

The metric terms are then used to build the performance evaluation metrics [8]

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \text{-----(1)}$$

$$Precision = \frac{TP}{TP+FP} \text{----- (2)}$$

$$Recall = \frac{TP}{TP+FN} \text{-----(3)}$$

$$F1 - Score = \frac{2 \cdot precision \cdot recall}{precision+recall} \text{-----(4)}$$

## 4 RESULT AND DISCUSSION

This section discusses the results of the experiment conducted. The Machine Learning models were built in Google Colab utilizing the Scikit learn module. To optimize the models deployed, we leveraged the feature selection technique. A line of code was written to filter out features that are correlated with the target, then we removed those that have less than 0.1 correlation with the target and fitted our models on the remaining 19 features. The input features used in this work are Follicle No. (R) Follicle No. (L), Skin darkening (Y/N), hair growth(Y/N), Weight gain(Y/N), Cycle(R/I), Fast food (Y/N), Pimples(Y/N), AMH(ng/mL), Weight (Kg), BMI, Cycle length(days), Hair loss(Y/N), Age (yrs.), Waist(inch), Hip(inch), Avg. F size (L) (mm), Marriage Status (Yrs), and Endometrium (mm). The ranking of the selected variables based on the correlation score is shown in Figure 4. We observed that Follicle Number (Right) and Follicle Number (Left) were the most influential factors in predicting PCOS.

The obtained results show that healthcare professionals can prioritize assessing Follicle Number(Right) and Follicle Number (Left) during PCOS diagnosis, potentially leading to earlier and more accurate detection. This could refine diagnostic criteria, enhance sensitivity and specificity, reduce misclassification, and ensure appropriate management. Also, future research could focus on understanding the biological pathways linking Follicle Number (Right) and Follicle Number (Left) to PCOS, which could inform the development of targeted therapeutic or preventive strategies. Finally, PCOS patients can benefit from personalized education and counseling on Follicle Number (Right) and Follicle Number(Left) to enhance their understanding of the condition and adherence to treatment plans.

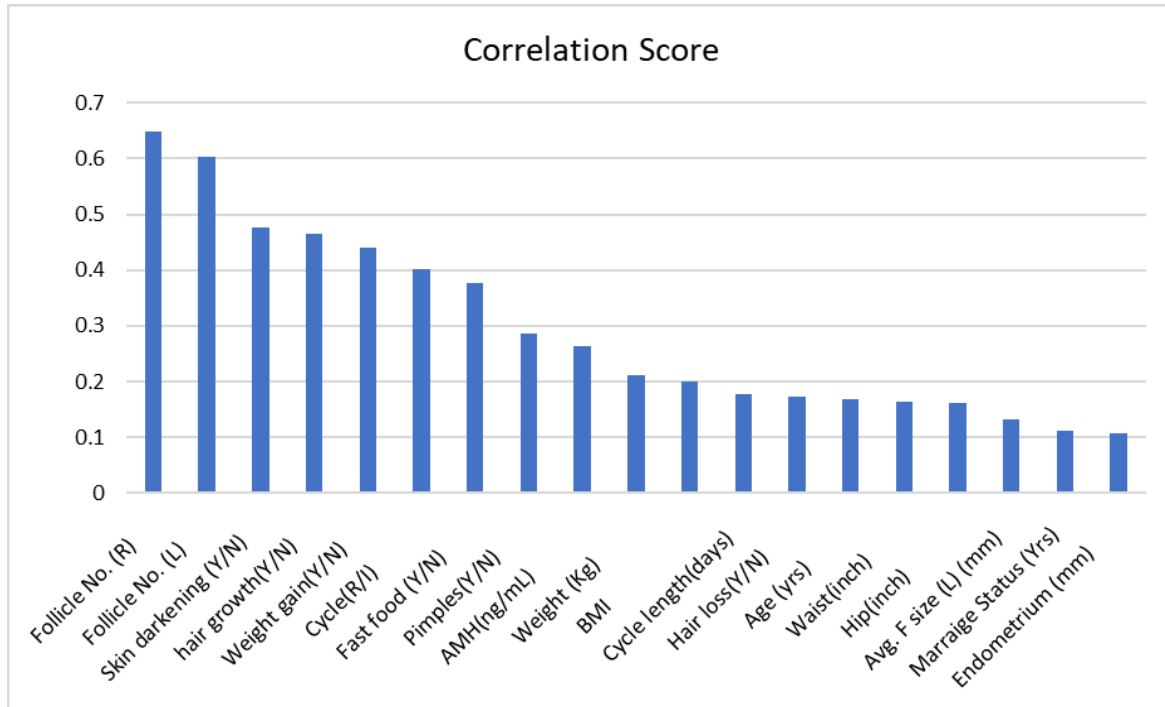


Figure 4: Input Feature ranking

Overall, the models performed well on all four metrics employed for performance evaluation in the study. Random Forest achieved the overall performance measure of 96% accuracy, precision, recall, and F1-score, respectively. Table 2 shows the performance of individual models across all the evaluation measures. This is closely followed by the performance of the Gradient gradient-boosting model which achieves 95% on all metrics except precision. The precision is at the same level as Random Forest.

We evaluated the computational time of the models fitted in this study. Random Forest recorded the least execution time of 0.035 seconds. Figure Y is a graph indicating the execution time of all models.

Table 2: Model Performance

Classifier	Accuracy	Precision	Recall	F1-Score
<b>Random Forest Classifier</b>	96%	96%	96%	96%
<b>Logistic Regression</b>	90%	91%	90%	90%
<b>XGBoost</b>	95%	95%	95%	95%
<b>Gradient Boost</b>	95%	96%	95%	95%
<b>AdaBoost</b>	91%	92%	91%	91%

Figure 5 is the confusion matrix showing how each model performed based on the number of true positives, false positives, true negatives, and false negatives. These are the elements that form the metrics upon which the overall performance of individual models is accessed.

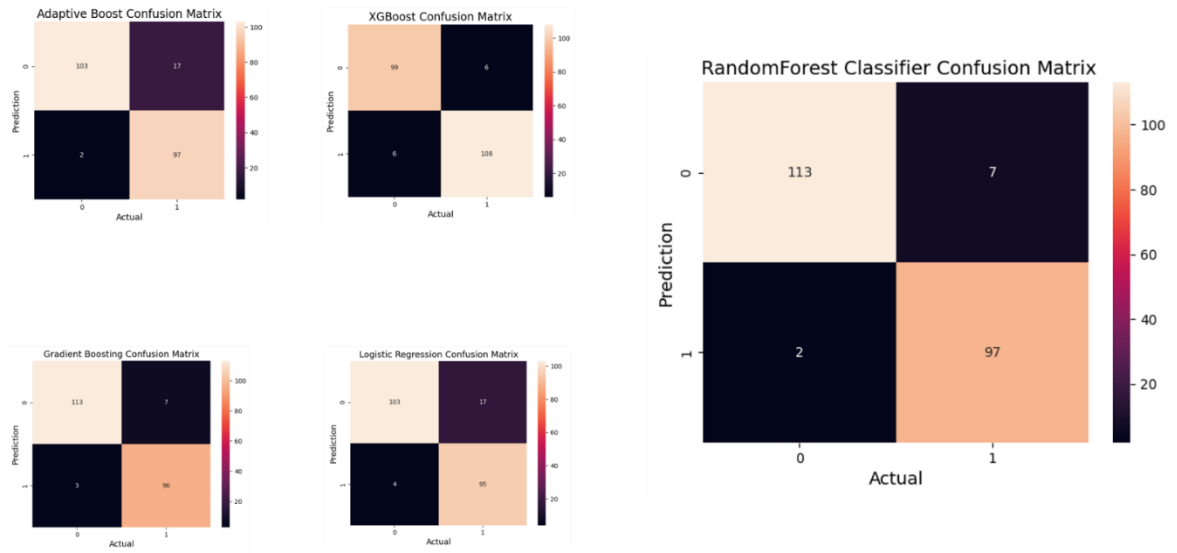
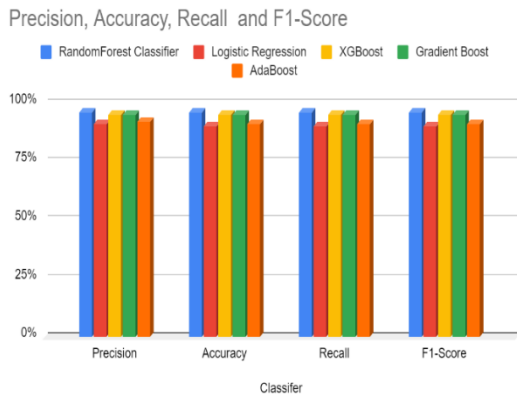
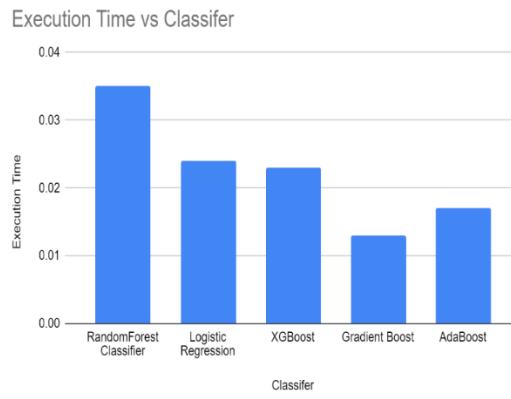


Figure 5: Confusion Matrix

Visualization is an essential technique in understanding the dataset and the results of models. Fig. 6a and b are bar graphs displaying the performance of the models based on the evaluation metric and execution time. Visualization aids in understanding the result and selecting the best-performing model to be utilized in creating the intended system.



(6a)



(6b)

Figure 6: Model Performance Graph

#### 4.1 Comparison with Existing Works

We compared the results obtained in this work with the performance of models created in previous work. The comparison is conducted on the performance of the models in predicting outcomes 1(Yes) and 0(No)



Table 3: Model Comparison

Literature	Target Category	Accuracy	Precision	Recall
<b>[10]</b>				
RandomForest Classifier	0	92.71	91.23	96.30
	1		94.87	88.10
AdaBoost	0	92.71	89.83	98.15
	1		97.3	85.71
Xgboost	0	95.83	94.64	98.15
	1		97.50	92.86
<b>[4]</b>				
RandomForsest Classifier	0	X	86.00	98.00
	1		95.00	76.00
Logistic Regression	0	X	91.00	98.00
	1		95.00	91.00
<b>[16]</b>				
RandomForest Classifier	0	86	84.00	95.00
	1		89.00	70.00
Logistic Regression	0	88	87.00	97.00
	1		93.00	74.00
<b>Proposed Model</b>				
RandomForest Classifier	0	96.00	98.00	94.00
	1		93.00	98.00
Logistic Regression	0	90.00	96.00	86.00
	1		85.00	96.00
XGBoost	0	95.00	94.00	94.00
	1		95.00	95.00
Gradient Boost	0	95.00	97.00	94.00
	1		93.00	97.00
AdaBoost	0	91.00	98.00	86.00
	1		85.00	98.00

Additionally, [15] created a hybrid model, by combining three models built on convolutional neural network (CNN) which are VGG16, AlexNet, and Inception V3, this hybrid model had its performance metrics as follows; Accuracy: 87%, Precision: 93% and recall: 81%. This result obtained in [15] indicates that the complexity of a model does not determine the performance achieved. Opting for simpler models could enhance understanding and compatibility, compared to Deep Learning models that might require more explanation to gain trust [2].

## 5 CONCLUSION AND FUTURE DIRECTION

This study is part of a larger study to develop an evidence-based application for the diagnosis of PCOS. We utilized a dataset publicly available in Kaggle to build a model that will achieve high performance with the least execution time. Early detection is crucial in the diagnosis of PCOS to avoid the condition resulting to more critical health situations. Our results indicate feature selection aids in increasing the accuracy of the models. 5 classifiers were used on the 19-feature subset of the dataset we obtained for this study.

Further improvement to this experiment can be the inclusion of more algorithms to be fitted on the dataset and modifying the test-train-split ratio.

## REFERENCES

- [1] P. Chauhan, P. Patil, N. Rane, P. Raundale and H. Kanakia, "Comparative Analysis of Machine Learning Algorithms for Prediction of PCOS," *2021 International Conference on Communication Information and Computing Technology (ICCICT)*, Mumbai, India, 2021, pp. 1-7, doi: 10.1109/ICCICT50803.2021.9510128.
- [2] Elmannai, H., El-Rashidy, N., Mashal, I., Alohali, M. A., Farag, S., El-Sappagh, S., & Saleh, H. (2023). Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. *Diagnostics (Basel, Switzerland)*, 13(8), 1506. <https://doi.org/10.3390/diagnostics13081506>
- [3] <https://www.cdc.gov/diabetes/basics/pcos.html>
- [4] Vasavi, Reddy Richa et al. "Polycystic Ovary Syndrome Monitoring using Machine Learning." *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* (2023): 1013-1019
- [5] A. Ahmetašević, L. Aličelebić, B. Bajrić, E. Bečić, A. Smajović and A. Deumić, "Using Artificial Neural Network in Diagnosis of Polycystic Ovary Syndrome," *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro, 2022, pp. 1-4, doi: 10.1109/MECO55406.2022.9797204.
- [6] S. Nasim, M. S. Almutairi, K. Munir, A. Raza and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," in *IEEE Access*, vol. 10, pp. 97610-97624, 2022, doi: 10.1109/ACCESS.2022.3205587
- [7] S. Islam, N. Nabi, S. A. Khushbu, N. J. Ria and A. K. M. Masum, "A Process of Finding Common Symptoms and Diagnosis Age among PCOS Patients Through a Survey," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9580114.
- [8] S.D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi and A. Alkhayyat, "Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms," *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, Al-Najaf, Iraq, 2022, pp. 532-536, doi: 10.1109/IICETA54559.2022.9888677
- [9] Y. A. Abu Adla, D. G. Raydan, M. -Z. J. Charaf, R. A. Saad, J. Nasreddine and M. O. Diab, "Automated Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques," *2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME)*, Werdanyeh, Lebanon, 2021, pp. 208-212, doi: 10.1109/ICABME53305.2021.9604905
- [10] M. S. Khan Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee and R. Hasan, "Improved Sampling and Feature Selection to Support Extreme Gradient Boosting For PCOS Diagnosis," *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA, 2021, pp. 1046-1050, doi: 10.1109/CCWC51732.2021.9375994David Harel. 1979. First-Order Dynamic Logic. Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, New York, NY. <https://doi.org/10.1007/3-540-09237-4>
- [11] U. N. A and K. Dharmarajan, "Diabetes Prediction using Random Forest Classifier with Different Wrapper Methods," *2022 International Conference on Edge Computing and Applications (ICECAA)*, Tamilnadu, India, 2022, pp. 1705-1710, doi: 10.1109/ICECAA55415.2022.9936172
- [12] M. Massaoudi, A. Darwish, S. S. Refaat, H. Abu-Rub and H. A. Toliyat, "UHF Partial Discharge Localization in Gas-Insulated Switchgears: Gradient Boosting Based Approach," *2020 IEEE Kansas Power and Energy Conference (KPEC)*, Manhattan, KS, USA, 2020, pp. 1-5, doi: 10.1109/KPEC47870.2020.9167534
- [13] Z. Yan and H. Wen, "Electricity Theft Detection Base on Extreme Gradient Boosting in AMI," *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Dubrovnik, Croatia, 2020, pp. 1-6, doi: 10.1109/I2MTC43012.2020.9128712
- [14] Palimkar, P., Shaw, R.N., Ghosh, A. (2022). Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach. In: Bianchini, M., Piuri, V., Das, S., Shaw, R.N. (eds) *Advanced Computing and Intelligent Technologies*. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. [https://doi.org/10.1007/978-981-16-2164-2\\_19](https://doi.org/10.1007/978-981-16-2164-2_19)
- [15] P. Chitra, K. Srilatha, M. Sumathi, C. Ishwarya and M. Jagadesh, "Automated Detection of Polycystic Ovaries using Pretrained Deep Learning Models," *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, Kanjirapally, India, 2023, pp. 1-6, doi: 10.1109/AICERA/ICIS59538.2023.10420280
- [16] V. Srinithi and R. Rekha, "Machine learning for diagnosis of polycystic ovarian syndrome (PCOS/PCOD)," *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, Coimbatore, India, 2023, pp. 19-24, doi: 10.1109/ICISCoIS56541.2023.10100490.
- [17] S. Bharati, P. Podder and M. R. Hossain Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms," *2020 IEEE Region 10 Symposium (TENSYP)*, Dhaka, Bangladesh, 2020, pp. 1486-1489, doi: 10.1109/TENSYP50017.2020.9230932. keywords: {Feature extraction;Random forests;Logistics;Boosting;Machine learning algorithms;Testing;Biochemistry;PCOS;gradient boosting;random forest;logistic regression;recall;precision},
- [18] Raghav, S., Rathore, M., Suri, A., Jain, R., Nagrath, P., Kumar, A. (2023). Prediction and Analysis of Polycystic Ovary Syndrome Using Machine Learning. In: Chatterjee, P., Pamucar, D., Yazdani, M., Panchal, D. (eds) *Computational Intelligence for Engineering and Management Applications*. Lecture Notes in Electrical Engineering, vol 984. Springer, Singapore. [https://doi.org/10.1007/978-981-19-8493-8\\_13](https://doi.org/10.1007/978-981-19-8493-8_13)

**Please fill in the all authors' background:**

**Position can be chosen from:**

**Prof. / Assoc. Prof. / Asst. Prof. / Lecture / Dr. / Ph. D Candidate / Postgraduate, etc.**

Full Name	Email Address	Position	Research Interests	Personal Website (if any)
Kemi Akanbi	<a href="mailto:akanbikr@mail.uc.edu">akanbikr@mail.uc.edu</a>	Ph.D. Student	Machine Learning and Health Informatics	
Odunayo Gabriel Adepoju	<a href="mailto:adepojog@mail.uc.edu">adepojog@mail.uc.edu</a>	M.S Student	Machine Learning and Human-Computer Interaction	
Isaac Kofi Nti	<a href="mailto:ntious1@gmail.com">ntious1@gmail.com</a>	Asst. Prof	Machine Learning, Data Science, and Cybersecurity	