Exploring Machine Learning Models for Efficient Polycystic Ovary Syndrome Diagnosis

Arshi Hussain, Anil Kumar Mahto, Kavita Sinha, Naved Alam*

Department of CSE, IEC College of Engineering & Technology, Greater Noida, India

Department of CSE & IT, Jaypee Institute of Information Technology, Noida Uttar Pradesh, India

Department of CSE (AI), IIMT College of Engineering, Greater Noida, India
Department of Computer science and Engineering, Delhi, Jamia Hamdard

Abstract:-Polycystic Ovary Syndrome (PCOS) is a common endocrine illness affecting women of reproductive age. It is distinguished by irregular menstruation, hyperandrogenism, and polycystic ovarian morphology. To lessen the long-term health effects of PCOS, early diagnosis and treatment are essential. Because of their ability to analyses complicated datasets and detect trends that human observers may miss, machine learning (ML) models have emerged as promising techniques for assisting in PCOS diagnosis. In this study, we investigate multiple machine learning models for effective PCOS diagnosis, highlighting their merits, limits, and possible uses in clinical practice. We examine the current state of ML-based PCOS diagnosis, identify obstacles, and suggest future research areas. different existing Machine Learning (ML) techniques, namely, Support vector classifier (SVC), logistic regression, Random Forest (RF) classifier, Decision Tree (DT), K-Nearest Neighbor (KNN), XGBoost classifier, and Catboost classification algorithms. We also try to implement the ensemble of XGBoost with random forest algorithm. The performance of the classifier, F1 Score is evaluated on six selected features (Follicle numbers[L/R], weight gain, skin darkening, hair growth and PCOS[Y/N]) for higher correlation value, according to evaluation metrices and results, we have found that PCOS correctly predicted with XGBoost and Random-forest ensemble method.

Keywords: *PCOS, KNN, SVM, DT, Random-forest, Logistic Regression.*

1. Introduction

Polycystic Ovarian Syndrome (PCOS) is a hormonal disorder affecting females of reproductive age, causing hormonal imbalances, irregular menstrual cycles, and the growth of cysts. It can lead to health complications like infertility, acne, weight gain, insulin resistance, and metabolic disturbances. The cause is unknown, but it is believed to involve genetic and environmental factors. Diagnosis involves medical history, physical examination, blood tests, and ultrasound. Treatment aims to manage symptoms, regulate menstrual cycles, improve fertility, and reduce long-term health risks. Early diagnosis is crucial for optimal outcomes [1,2]. clinician also need to be focus on taking a detailed menstrual history of the patient on any anomalies, any changes in the patient’s dimensions and state of health in terms of skin tags, acne, alopecia, and terminal hair. Early diagnosis may help women with family planning, to reduces health risk associated with the PCOS.

Polycystic Ovary Syndrome (PCOS) is a complex endocrine disorder affecting women of reproductive age. Early detection and accurate diagnosis are crucial for effective management and prevention. Artificial Intelligence (AI) and Machine Learning (ML) techniques have shown potential in improving disease detection and diagnosis. A study by the National Institutes of Health (NIH) found that AI/ML-based programs effectively detect and diagnose PCOS. These programs use advanced computational methods to analyze data, identify patterns, and provide personalized interventions. The findings suggest further research in this field for improved patient outcomes. However, rigorous clinical testing is needed to ensure their reliability and accuracy [3].

Typically, the diagnosis is performed via accepted, standardized criteria that have developed over time. The diagnosis involves evaluating clinical, laboratory, and radiological criteria. Clinical features include acne, hair growth, irregular menstrual cycles, elevated androgen levels, and insulin resistance. Laboratory findings include elevated androgen levels and insulin resistance. Radiological findings reveal characteristic features like multiple small follicles and increased ovarian volume. Other clinical parameters include BMI, metabolic syndrome, and family history of PCOS or related conditions. Healthcare providers must carefully evaluate each criterion and consider the patient's individual characteristics and medical history when making a diagnosis. however, PCOS is often overlooked because some of its symptoms can co-occur with other conditions like obesity, diabetes, and cardiometabolic illnesses.

Artificial Intelligence has a potential to enhance the precision and efficacy of PCOS diagnosis, hence facilitating prompt intervention and improving patient outcomes. For that Machine learning models are used to analyses many data sources, including clinical and hormonal factors, ultrasound images, blood test parameters, and patient demography. These data points are used to train and validate the algorithms that determine if a female has PCOS or not. The goal is to develop a Machine learning model [4] as a tool that can help medical professionals diagnose patients more quickly and accurately as compare to other methods. With their strong performance ratings in PCOS detection [5].

These approaches represent a diverse set of techniques utilized to analyze and interpret data for the purpose of detecting PCOS. Each method offers unique advantages and may be suitable for different types of data and research objectives. Through a review of the literature, this paper aims to provide insights into the efficacy and applicability of these AI and ML algorithms in the context of PCOS detection.

1.1 Related work

In this section existing literature is conducted and discussed along with the approaches using Artificial Intelligence and Machine learning Algorithms for PCOS. Table give brief analysis of review on PCOS and summarizes the study with their remarks. this review explores AI and ML algorithms for detecting Polycystic Ovary Syndrome, examining various methodologies like CNN, SVM, K-NN, Random Forest, Logistic Regression, Decision Tree, and Naïve Bayes.

Samia Ahmed et al. [6] introduced a PCOS detection framework employing various machine learning algorithms, including Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Random Forest, Logistic Regression, Decision Tree, and Naïve Bayes. The performance of these algorithms in PCOS detection was evaluated using both quantitative and qualitative approaches.

Hela Almannai et al. [7] conducted a study aimed at enhancing the efficiency and effectiveness of PCOS diagnosis using various machine learning models, including Logistic Regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), XGBoost, and Adaboost. Their approach involved providing model explanations to ensure optimal feature selection and the selection of the best model. To improve model performance, Bayesian optimization was employed to optimize machine learning models. Additionally, the authors addressed the class imbalance issue by combining Synthetic Minority Oversampling Technique (SMOTE) and Edited Nearest Neighbors (ENN). experimental results were obtained using a benchmark PCOS dataset with two different ratios of data splitting: 70:30 and 80:20. The findings of the study revealed that the Stacking Machine Learning model with Recursive Feature Elimination (REF) for feature selection achieved the highest accuracy of 100%, outperforming other models eva

Varada Vivek Khanna et al. [8] introduced an innovative approach leveraging heterogeneous Machine Learning (ML) and Deep Learning (DL) classifiers to predict Polycystic Ovary Syndrome (PCOS) among fertile patients. The data for the study was sourced from the open-source Kaggle platform, and various features were collected using eXplainable Artificial Intelligence (XAI) techniques, including SHapley Additive Values, LIME (Local Interpretable Model Explainer), ELI5, and Qlattice, in combination with a Random Forest classifier and tree-based classifiers. the proposed AI-based framework demonstrates high accuracy in detecting PCOS among patients. Additionally, the study presents an automated screening architecture equipped with explainable machine learning tools. These tools aim to provide medical professionals with interpretable insights into the model's decision-making process, thereby aiding them in making informed clinical decisions.

Ashwini Kodipalli and Susheela Devi [9] conducted a significant study aimed at developing an automated early detection and prediction model for Polycystic Ovary Syndrome (PCOS) that could accurately predict and estimate the likelihood of associated mental health issues. The study proposed a Fuzzy Model approach to address the linguistic nature of the mapping between symptoms and diagnosis in PCOS. In their approach, the Fuzzy Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method was utilized and evaluated for its performance in PCOS detection and prediction. The researchers collected a local yet specific dataset comprising a spectrum of women, which was used for comparison between the Fuzzy TOPSIS method and the widely used Support Vector Machines (SVM) algorithm. the results of the study demonstrated the effectiveness of the Fuzzy TOPSIS method, achieving an impressive accuracy of 98.20% in PCOS detection and prediction. In comparison, the SVM algorithm yielded an accuracy of 94.01% on the same dataset. This highlights the potential of the Fuzzy TOPSIS method as a promising approach for early detection and prediction of PCOS, particularly in association with mental health issues.

Namitha T S and Meera Rose Mathew [10] developed a model for detecting and predicting Polycystic Ovary Syndrome (PCOS) using machine learning models like K-Nearest Neighbor (K-NN) and logistic regression. They used multiple models to identify the best-performing model for real-world datasets, considering factors like data nature and distribution. K-NN and logistic regression are popular for classification tasks due to their simplicity and effectiveness. Their goal was to create an effective framework for accurately identifying and classifying PCOS cases based on symptomatic profiles.

Palak Mehrotra et al. [11] developed a method for automating the detection of Polycystic Ovary Syndrome (PCOS) using early markers. The algorithm involves formulating a feature vector based on clinical and metabolic features, identifying relevant features through statistical analysis, and using two classification algorithms: Logistic Regression and Bayesian Classifier. The Bayesian Classifier outperformed Logistic Regression in accuracy, with an overall accuracy of 93.93%. This method effectively automates PCOS detection, enabling accurate discrimination between normal and PCOS individuals.

Kinjal Raut et al. [12] implemented various machine learning algorithms, including Random Forest, Logistic Regression, Support Vector Classifier (SVM), K-Nearest Neighbors (K-NN), Decision Tree, XGBoost, and CatBoost Classifier, for the early detection and prediction of Polycystic Ovary Syndrome (PCOS). These algorithms were applied to identify an optimal and minimal set of parameters, which were statistically analyzed and evaluated to determine their effectiveness in PCOS detection. the study found that the CatBoost Classifier achieved higher accuracy compared to the other algorithms evaluated. By leveraging machine learning techniques, the researchers aimed to develop a robust and accurate model for PCOS detection.

B. Yamini et al. [13] proposed a machine learning model utilizing several algorithms, including Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbors (K-NN), and XGBoost. These algorithms were employed to predict PCOS, and the outcomes of the study demonstrated the promising predictive ability of machine learning models. specifically, the random forest model achieved a 90% accuracy rate, which was reported to be higher than any other model tested in the study. This indicates the effectiveness of the random forest algorithm in accurately predicting PCOS.

The studies highlight the potential of machine learning in diagnosing and managing PCOS, emphasizing the need for appropriate algorithms and optimization of parameters. They also emphasize the importance of robust evaluation methodologies, such as cross-validation and performance metrics, to ensure the model's generalizability and scalability. As these models continue to be refined and validated, there is hope for improved diagnostic accuracy, personalized treatment strategies, and improved outcomes for PCOS patients.

This paper presents a review on technological advances to detect PCOS. A critical analysis of state of art approaches has been conducted. Remaining part of the paper is organized as follows. Section 2 presents the state of the paper gives the methodology and section 3 concluded with results followed by the references.

Methodology

The process of predicting PCOS in women using machine learning models involves several steps, as illustrated in Figure 1.

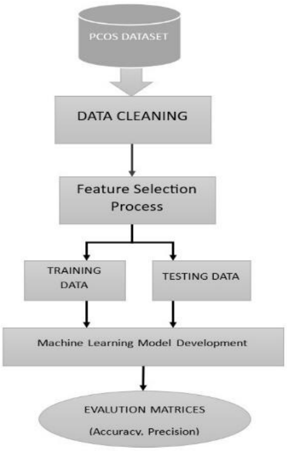


Fig. 1.Machine Learning Model development diagram

1.2 Data Set

First, we obtained the dataset from Kottarathil's open-source repository on Kaggle, which contains information on 541 fertile women and includes 43 attributes [14]. This multicentric dataset was compiled from ten hospitals in Kerala, India. The target variable in this dataset is "PCOS (Yes/No)," with 177 out of 541 patients receiving a PCOS diagnosis. For categorical features, the values were encoded as 1 for "Yes" and 0 for "No." Among the 43 attributes, 24 were non-invasive measurements, while the remaining attributes consisted of gynecological or hormonal data collected through fluid samples and invasive vaginal ultrasonography.

1.3 Data Pre-processing

Next, we performed data pre-processing on the Kaggle dataset to ensure its suitability for model training. Despite the dataset having undergone pre-processing previously, we conducted additional steps to address any potential issues., and data encoding.

1.4 Features Selection Methods

Improvement of the model can be possible after selecting appropriate features from the data set, filtering process is helpful to get knowledge about the best feature after giving statistical test score [15]. Table 1. Shows that features correlation values greater than 0.4 with the target variables of PCOS as Yes/No and boxplot is shown in figure for each selected features of follicle number for left and right, skin darkening, hair growth, and weight gain.

Table 1. Feature correlation with PCOS

PCOS Features	Features weights
PCOS (Y/N)	1.000000
Follicle Number. (R)	0.648327
Follicle Number. (L)	0.603346
Skin darkening (Y/N)	0.475733
hair growth(Y/N)	0.464667
Weight gain(Y/N)	0.441047

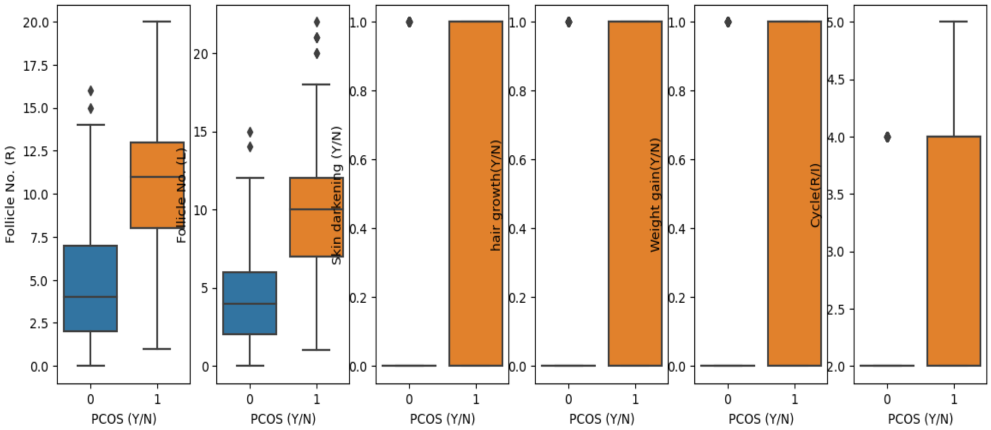


Fig. 2. selected feature for pcos detection on different conditions

1.5 Splitting Dataset

This PCOS datasets is split into training and testing data as 80% training data and 20% testing data. When the data preprocessing and exploratory data analysis is completed, then data is clean data, ready to be processed by the machine learning models. Dataset is splitted with the ratio of 8:2 as the training and testing dataset.

1.6 Machine Learning Model

The following models were utilized in the development of the model: Random Forest, Logistic Regression, Linear SVC, Decision Tree, Cat-boost, Gradient Boosting, and XG-Boost with the Random Forest algorithm employed to predict better results. The evaluation was conducted by measuring accuracy, precision, recall, and F1-score values separately. As presented in Table 2, the accuracy values for the top features derived using Random Forest, Cat-Boost, XG-Boost, Logistic Regression, Linear SVM, Decision Tree, and XG-Boost+Random-Forest are 0.82, 0.88, 0.89, 0.86, 0.84, 0.88, 0.84, and 0.89, respectively. The accuracy figure as shown in Fig. 3. illustrates that both XG-Boost and Random-Forest are renowned for their stability and excellent performance. Compared to certain other algorithms, they are less prone to overfitting and are capable of handling intricate relationships in the data.

These models outperform other classifiers in terms of accuracy, indicating that our model has an accuracy of 0.89. We thoroughly examined our models.

Logistic regression [16] Assign the features values X and target value as y, split the dataset into training and testing dataset. Logistic regression is a predictive analytics tool used for regression as well as classification.

➢ Decision tree is a supervised learning approach, used to address regression and classification problems. decision tree graph is defined as a root node and the leaf node. Leaf nodes are used to define the decision's output, which is the result of the decision, while decision nodes are used to make decisions [17]. The characteristics of the given dataset are used to inform tests and judgements. It is a graphical representation used to find the answer to a problem given certain criteria. Because of its tree-like appearance and the fact that it has a root node in addition to extending, it is called a decision tree.

➢ Random forest algorithm [18] can be used for the continuous variables in terms of regression and for the categorical variables in terms of classification. Breiman et. al created random forest model [19], This model starts the bagging concept process as well as the random feature selection process. The bagging model uses sampling to create new data by substituting the existing dataset with n sample size, which is taken from the training data. However, during tree splitting, the random feature selection process permits random feature subsets in each node so that diversity of base technique can be seen. During prediction, accuracy is increased by both bagging and random feature selection.

➢ KNN [20] is an instance-based classification algorithm, KNN uses the training set to make the prediction directly. Predictions are made for the new instance (x) by looking for the closest instance in the training data set that is like the new instance; knn is a k nearest neighbors refer to them as neighbors. The neighbours are summarised to provide the outcome. The mean or median of the comparable classes is the result of regression. The most common class value, or the mode, is also used to derive the categorization. We compute the Euclidean distance for the input variables in real time. In cases where the input variables are of a similar type, Euclidean distance measurement is preferable. The computational complexity of KNN grows with larger training data sets. The KNN performs well when the data is of small size. KNN is a classification algorithm used for the non-parametric estimation. Prediction is performed based on nearest neighbour's k fold classifier. In this study we are getting 91.95 training accuracy and 84.23 testing phase accuracy using KNN Algorithm. Figure shows Confusion matrix for the actual and predicted values.

➢ Support Vector Classification (SVC) [21] is a potent supervised learning algorithm, used for both classification and regression problems. It is often referred to as Support Vector Machine (SVM) in the context of classification. SVM is a Supervised learning used for Regression analysis and classification. In svm, support vectors are calculated based on minimum separation by generating a hyperplane. The best hyperplane is the algorithm's output. It determines the greatest minimum distance, and the separation between is determined by Euclidean distance.

➢ CatBoost [22] is a categorical boosting classification algorithm which works especially well for categorization tasks. It is made to efficiently handle categorical characteristics and can manage missing data on its own without preprocessing. In PCOS detection this algorithm can be utilized as a classifier to examine various pcos categorical features.

➢ XGBoost [23] is used for large datasets for performing optimization and parallelization techniques, it can handle massive datasets with millions of samples. XGBoost offers a wide range of loss functions and objective functions, training and optimising models may be done with flexibility dependent on the needs and features of the dataset. because of its ability to handle a high number of features and observations, it can be used to analyse medical datasets that may include a wide variety of factors.

1.7 Evaluation Metrics

Standard classification performance measures, including confusion matrix, accuracy, precision, recall, F-1 scores, AUCROC score (Area Under the Receiver Operating Characteristics curve), and precision-recall curve, have been used to assess and compare our suggested methodology. The amount or measurement of separability is denoted by AUC, while a probability curve is represented by ROC. True Positive Rate (TPR) is usually used to generate the curve. False Positive Rate (FPR) patients who do not have the ailment can be distinguished from one another by the model more successfully if the AUC is higher [24].

Additional measures, especially the AUC (Area Under the Curve)-ROC (Receiver Operating Characteristics) curve as shown in Table 3, were used to assess the model performances in this work.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2\,\frac{Precision\ \times recall}{precision + recall}$$

The model performances in this study were evaluated using additional metrics, namely the AUC (Area Under the Curve)-ROC (Receiver Operating Characteristics) curve. It is a well-known metric of how well classification systems function at various threshold levels, where ROC represents a probability curve and AUC represents the degree or measurement of separability. True Positive Rate (TPR) is usually used to generate the curve. The more accurately the model can distinguish between people who have the ailment and those who do not, the greater the

AUC., FPR, or False Positive Rate. We have calculated Accuracy score and F1 metrices value as shown in table 2. for the PCOS detection.

Table 2. Machine learning Algorithms and their accuracy/F1 Score valuers

Machine Learning Algorithms	Training Accuracy	F1 Score
Decision Tree	84	74
CatBoost	88	83
XGBoost	89	84
LogisticRegression	86	77
Random Forest	84.7	83
Linear SVC	88	89
KNN	84	73
Xgboost+random forest	89.6	85

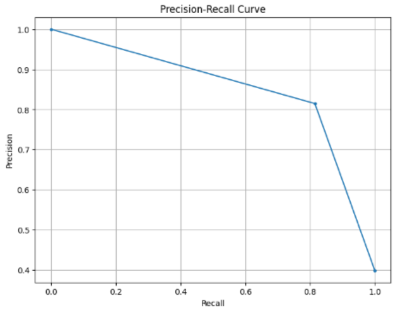
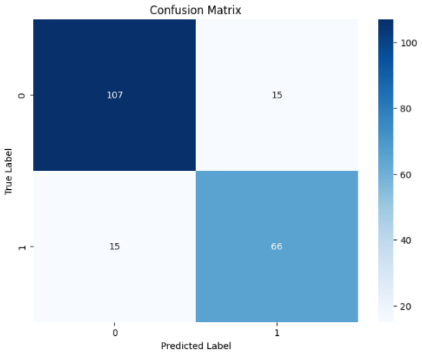
2. Results

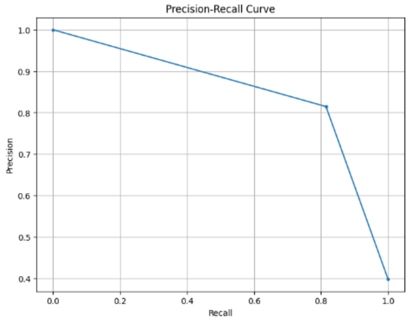
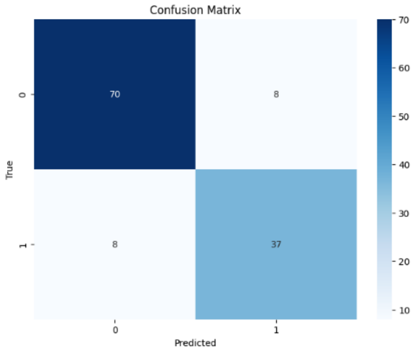
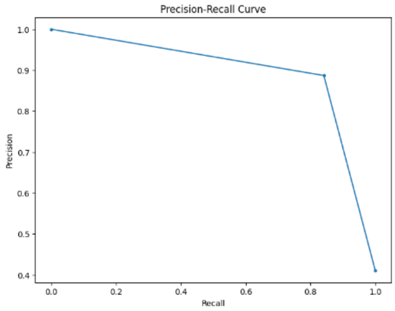
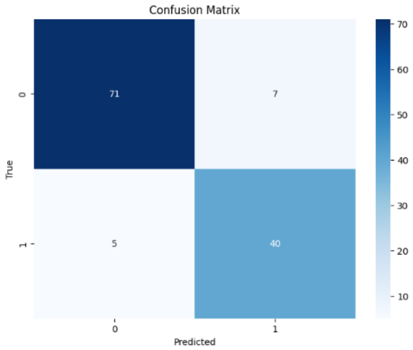
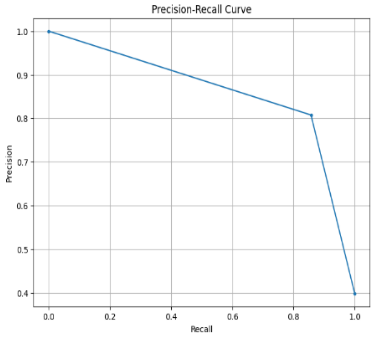
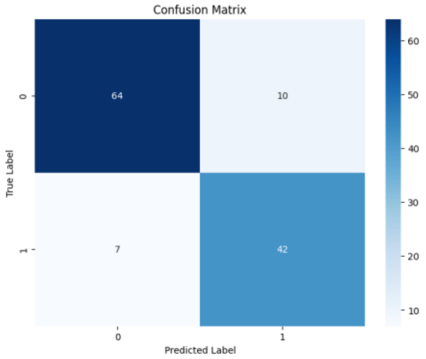
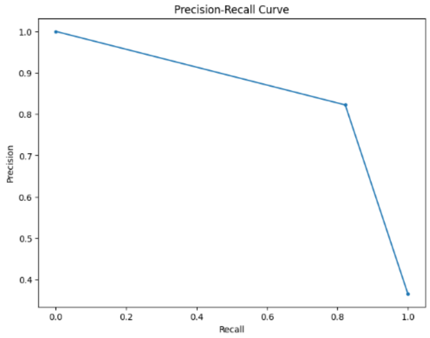
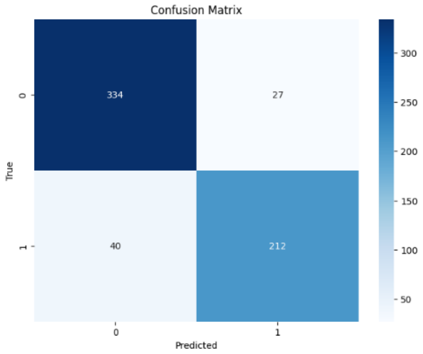
In this section we have implemented supervised and unsupervised machine learning algorithm using Logistic regression, svc, Random-Forest, xgbost, catboost and knn, after implementing all the algorithm using jupyter notebook we have calculated the values of confusion matrix and evaluated true negative, true positive, false negative and false positive rate.

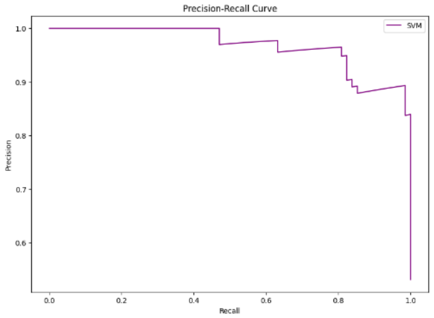
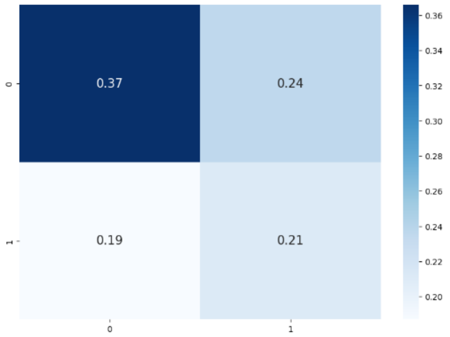
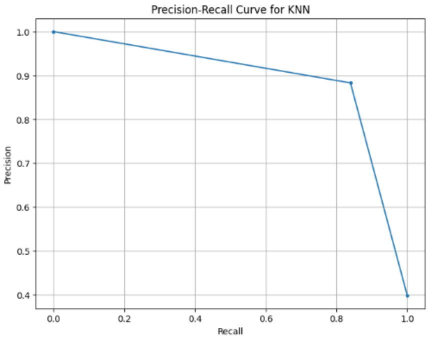
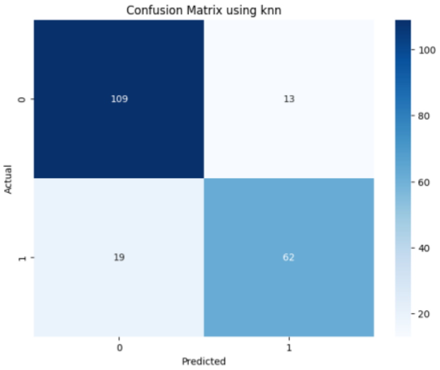
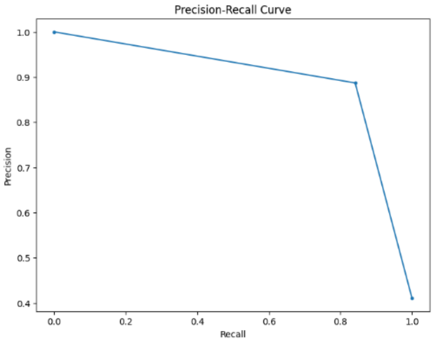
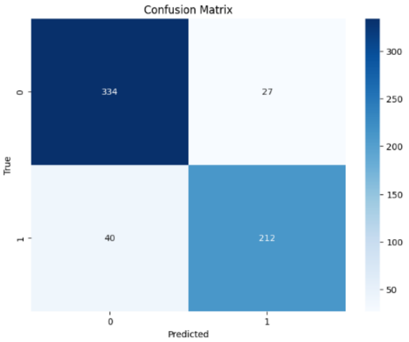
In this experimentation the accuracy of Decision tree, Logistic regression, Random-forest, Linear support vector classifier, KNN, XGBoost, CatBoost and ensemble of XGBoost with random forest were calculated for both training and testing before applying Recursive Feature Elimination (RFE) and represented in table 2. The confusion matrices for are represented in table 3. The comparison of all algorithms represented in Fig 3.

AUC/ROC AUC, or area under the receiver operating characteristic curve as shown in table 3. This shows us results of various machine learning algorithms for the regression and classification, is a helpful statistic for assessing and visualizing classification ability [25].

Table 3. Machine learning algorithms with precision recall curve and confusion matrix

NAME OF ALGORITHM	PRECISION RECALL CURVE	CONFUSION MATRIX HEAT MAP
Decision Tree		

CAT-BOOST		
XG-Boost		
Logistic Regression		
Random-forest		

SVC		
KNN		
XG-Boost and Random-forest		

2.1 Best Accuracy Score

This section presents an overview of the experimental outcomes. We also talk about the optimal model for each feature selection technique as shown in fig. 3.

Previous research is also contrasted with the suggested paradigm. Additionally, the explainability of the model is examined.

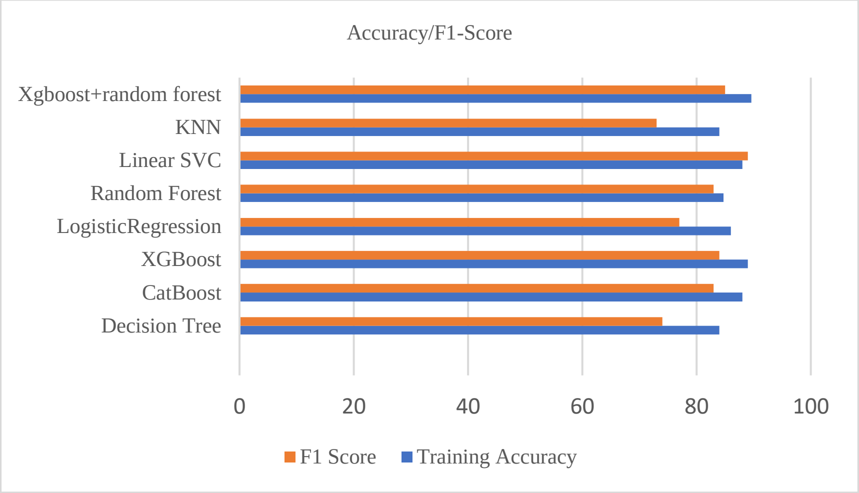


Fig. 3. Accuracy and F1-Score bar chart

Our study's findings show that, by utilizing a limited set of putative markers, we have successfully created a method for automatically identifying Polycystic Ovary Syndrome (PCOS). PCOS is a complex endocrine disorder primarily affecting women who are ready to have children and is frequently linked to anovulation and infertility. we included clinical and metabolic variables that act as disease biomarkers in our diagnostic criteria. These markers are essential for correctly recognizing and diagnosing PCOS. our objective was to develop a model that could accurately and efficiently identify PCOS by utilizing these characteristics. Follicle Number (R), Follicle Number (L), skin darkening, hair growth, weight gain, and Cycle (R/I) are among the inputs that our suggested model uses.

These markers are frequently evaluated during clinical examinations since they are known to be linked to PCOS. we experimented with different machine learning models and discovered that the Random Forest and optimizable XG-Boost algorithms worked best for PCOS detection. Based on the input indicators, these algorithms performed better than others in correctly categorizing women having PCOS or not. the possible influence of these findings on clinical practice makes them significant. our technology finding can help healthcare providers save time during patient examinations by automating the PCOS identification procedure. Because of this automation, people who are at risk of PCOS may be identified more quickly, enabling prompt management and intervention measures.

3. Conclusion

By incorporating a variety of clinical and biological data collecting from kaggle, machine learning models have the potential to significantly increase the effectiveness and precision of PCOS diagnosis. To solve current issues and confirm the therapeutic usefulness of ML-based diagnostic tools in practical contexts, more study is necessary. Working together, researchers, physicians, and data scientists may fully utilize machine learning (ML) concept to improve PCOS diagnosis and provide more individualized therapy for patients. This paper reviewed the state-of-the art approaches in PCOS detection using AI and ML Algorithms. Various studies employing different Machine learning techniques for PCOS diagnosis were analyzed, highlighting their methodologies and results. In the end, our automated system has the potential to enhance patient outcomes and the standard of care in the field of reproductive health by reducing the time required to diagnose PCOS risk.

References

[1] Shetty D, Chandrasekaran B, Singh AW, Oliverraj J. Exercise in polycystic ovarian syndrome: An evidence-based review. Saudi Journal of Sports Medicine. 2017 Sep 1;17(3):123.

[2] Leon LI, Anastasopoulou C, Mayrin JV. Polycystic Ovarian Disease. InStatPearls [Internet] 2021 Jul 21. StatPearls Publishing. Available:https://www.ncbi.nlm.nih.gov/books/NBK459251/ (accessed 21.8.2022)

[3] Barrera FJ, Brown EDL, Rojo A, Obeso J, Plata H, Lincango EP, Terry N, Rodríguez-Gutiérrez R, Hall JE, Shekhar S, 2023. Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. Frontiers in Endocrinology.

[4] Subrato Bharati, Prajoy Podder, and M. Rubaiyat Hossain Mondal. Diagnosis of polycystic ovary syndrome using machine learning algorithms. In 2020 IEEE Region 10 Symposium (TENSYMP), pages 1486–1489, 2020. doi:10.1109/TENSYMP50017.2020.9230932.

[5] Bichler, Martin and Kiss, Christine, "A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management" (2004). AMCIS 2004 Proceedings. 230.

[6] S. Ahmed et al., "A Review on the Detection Techniques of Polycystic Ovary Syndrome Using Machine Learning," in IEEE Access, vol. 11, pp. 86522-86543, 2023, doi: 10.1109/ACCESS.2023.3304536

[7] Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El-Sappagh S, Saleh H. Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. Diagnostics (Basel). 2023 Apr 21;13(8):1506. doi: 10.3390/diagnostics13081506. PMID: 37189606; PMCID: PMC10137609.

[8] Khanna, V.V.; Chadaga, K.; Sampathila, N.; Prabhu, S.; Bhandage, V.; Hegde, G.K. A Distinctive Explainable Machine Learning Framework for Detection of Polycystic Ovary Syndrome. Appl. Syst. Innov. 2023, 6, 32. https://doi.org/10.3390/asi6020032

[9] Kodipalli A, Devi S. Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM. Front Public Health. 2021 Nov 30;9:789569. doi: 10.3389/fpubh.2021.789569. PMID: 34917583; PMCID: PMC8669372.

[10] Namitha T. S, Meera Rose Mathew, “Polycystic Ovary Syndrome Analysis Using Machine Learning Algorithms” Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022 Vol.4, Issue.1 DOI: 10.5281/zenodo.6362178

[11] Mehrotra, P., Chatterjee, J., Chakraborty, C., Ghoshdastidar, B. and Ghoshdastidar, S. Automated screening of polycystic ovary syndrome using machine learning techniques. Annual IEEE India Conference (INDICON), 2011, 1-5.

[12] Kinjal Raut, Chaitrali Katkar, Prof. Dr. Mrs. Suhasini A. itkar. Journal, I. R. J. E. T. (2022). PCOS Detect using Machine Learning Algorithms. IRJET., volume:09, issue: 01.

[13] B.Yamini, Venkata Ramana Kaneti, Prema.P, Ambhika C, M.Nalini, Siva Subramanian.R, "Machine Learning-Driven PCOS Prediction for Early Detection and Tailored Interventions," SSRG International Journal of Electrical and Electronics Engineering, vol. 10, no. 9, pp. 61-75, 2023. Crossref, https://doi.org/10.14445/23488379/IJEEE-V10I9P106

[14] Polycystic Ovary Syndrome (PCOS) 2023. [(accessed on 17 March 2023)]. Available online: https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos

[15] 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.

[16] Tiwari S, Kane L, Koundal D, Jain A, Alhudhaif A, Polat K, Zaguia A, Alenezi F, Althubiti SA (2022) SPOSDS: a smart polycystic ovary syndrome diagnostic system using machine learning. Expert Syst Appl 203:117592. https://doi.org/10.1016/j.eswa.2022.117592

[17] Zhu R, Wang Y, Liu JX, Dai LY (2021) IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier. BMC Bioinform 1–17

[18] J. Madhumitha, M. Kalaiyarasi, and S. Sakthiya Ram. Automated polycystic ovarian syndrome identification with follicle recognition. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC), pages 98–102, 2021. doi:10.1109/ICSPC51351.2021.9451720.

[19] Subrato Bharati, Prajoy Podder, and M. Rubaiyat Hossain Mondal. Diagnosis of polycystic ovary syndrome using machine learning algorithms. In 2020 IEEE Region 10 Symposium (TENSYMP), pages 1486–1489, 2020. doi:10.1109/TENSYMP50017.2020.9230932.

[20] 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. Kodipalli A, Devi S. Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM. Front Public Health. 2021 Nov 30;9:789569. doi: 10.3389/fpubh.2021.789569. PMID: 34917583; PMCID: PMC8669372.

[21] Y. Rathod et al., "Predictive Analysis of Polycystic Ovarian Syndrome using CatBoost Algorithm," 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/TENSYMP54529.2022.9864439.

[22] Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El-Sappagh S, Saleh H. Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. Diagnostics (Basel). 2023 Apr 21;13(8):1506. doi: 10.3390/diagnostics13081506. PMID: 37189606; PMCID: PMC10137609.

[23] A. Bansal and A. Singhrova, "Performance Analysis of Supervised Machine Learning Algorithms for Diabetes and Breast Cancer Dataset," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 137-143, doi: 10.1109/ICAIS50930.2021.9396043.

[24] S. Shajahan and T. Poovizhi, "A Novel Approach to Estimation Precision and Recall for Star Rating Online Customers Based on Negative Hotel Reviews using Multinomial Naive Bayes over Multischeme Classifier," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ICBATS54253.2022.9759081.