IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

PCOS Prediction Using Machine Learning

Swamini Khanvilkar, Vaishnavi Gadkari, Tanuja Rathod, Prof. Dr. K. A. Malgi
Department of Information Technology
Bharati Vidyapeeth's College of Engineering for Women, Pune, India
swaminik07@gmail.com, vaishnavigadkari28@gmail.com
tanujarathod35@gmail.com, ketaki.naik@bharatividyapeeth.edu

Abstract: *Polycystic Ovary Syndrome (PCOS) is a common endocrine condition that impacts women during their reproductive years. Failure to diagnose the condition leads to various health issues like infertility and metabolic disorders. Existing diagnostic methods are time-consuming and require manual intervention by experts. In this research work, an automatic PCOS detection framework is presented using machine learning approaches. The algorithm uses a pre-processing technique that includes missing value replacement, StandardScaler, and Synthetic Minority Over-sampling Technique (SMOTE). Various classifiers are examined in this study, including Support Vector Classifier (SVC), Decision Tree, Random Forest, K-Nearest Neighbors (KNN), XGBoost, and Gradient Boosting. GridSearchCV and stratified K-fold validation are applied to optimize the model parameters. Among all models, the Random Forest classifier exhibits superior performance with 95.21% accuracy and ROC-AUC score.*

Keywords: Polycystic Ovary Syndrome (PCOS), Machine Learning, SMOTE, Random Forest, Classification, Predictive Healthcare

I. INTRODUCTION

A condition that is quite common among women who are of childbearing age is known as Polycystic Ovary Syndrome (PCOS). It refers to a complex hormonal disorder usually accompanied by menstrual disorders, the presence of excessive male hormones, and ovulatory dysfunction. Most of the women suffering from PCOS are left undiagnosed. This is because of several reasons, which include ignorance, delayed medical evaluation, and unavailability of quality health care facilities. If PCOS is left untreated and undetected, then there are chances that it may progress rapidly and lead to some serious health complications. Some of these complications include infertility, obesity, type 2 diabetes, heart diseases, and extreme mental problems, including anxiety and depression.

The clinical diagnosis of PCOS has long been made through a multi-faceted process involving detailed history, physical clinical exam, pelvic ultrasound imaging, and blood tests of hormones. These traditional diagnostic procedures are effective in the clinic, but have some major disadvantages; not only are they time-consuming, but they are also expensive and require the medical interpretation of an expert. Moreover, in many remote and underdeveloped geographic areas, there is insufficient access to specialists in gynecology and/or advanced diagnostics. This failure to access results inevitably results in late or often inaccurate diagnoses, and many women have no idea they have a problem until it reaches a much more advanced stage. The traditional diagnostic process also inherently combines the manual analysis of a number of complex clinical parameters simultaneously. Considering the Body Mass Index (BMI), the menstrual cycle pattern, the number of follicles in the ovary, hormones, and various physical symptoms such as the growth of abnormal hair, weight gain, and darkening of the skin are all important for healthcare professionals to take into account. Manually working with this multi-parameter, high-dimensional data is not only very complicated but is also prone to human error, variations in clinical judgement, and the reliability and standardization of the diagnostic process is compromised.

In the era of AI and ML, there are new and exciting opportunities to address long-standing challenges and to enhance clinical diagnoses using smart, automated systems. One of the unique strengths of machine learning algorithms lies in their capability to rapidly analyze vast amounts of complex clinical information, identify subtle patterns, and generate highly accurate predictions with minimal human effort. A variety of medical prediction and classification problems have been successfully tackled in recent years with ML approaches such as Support Vector Machines (SVMs), Decision Trees, Random Forests, K-Nearest Neighbors (KNNs), and Gradient Boosting. These complex computational models are well-suited to structured data, can obtain high classification accuracy, and are especially suitable for automating the process of detecting PCOS based on the routinely collected clinical parameters.

Although machine learning approaches have been implemented in predictive healthcare analysis, there is still no specific optimized system which is widely accepted and strong enough for the detection of PCOS by using structured clinical data. Current computational methods tend to be limited to a minimal set of clinical parameters and thus do not reflect the actual complexity of the syndrome. Furthermore, numerous existing studies have not addressed important data-related challenges, such as missing data, appropriate feature scaling, and extreme class imbalance. If the data sets are not balanced (e.g., there are many more cases that are not PCOS than PCOS cases), then the model can be biased unless it is corrected. Moreover, a comprehensive study of several state-of-the-art ML algorithms has not been performed. Therefore, it is difficult to identify the most promising and accurate model to perform this specific diagnosis.

In order to meet the requirements of an affordable, user-friendly, and extremely accurate diagnostic tool, the following study puts forward an advanced concept regarding the design of an intelligent and automated PCOS diagnosis system based on optimization of machine learning techniques. As the primary aim of the paper is to develop a classification model, which allows evaluating different clinical features in order to detect PCOS risks, the proposed approach consists of a complex data processing workflow that is capable of efficiently dealing with missing data as well as using a powerful technique to normalize features, such as StandardScaler. Moreover, the system specifically addresses the problem of class imbalance by employing the Synthetic Minority Over-sampling Technique (SMOTE). Another essential component of the proposed methodology is an exhaustive examination of several machine learning models, such as SVC, Decision Tree, Random Forest, KNN, XGBoost, and Gradient Boosting. The best model is selected using an elaborate approach to hyperparameter optimization based on GridSearchCV and Stratified K-Fold cross-validation, which enables the diagnosis of PCOS at its early stages.

II. LITERATURE SURVEY

A. Survey on papers

Suha et al. [1] have devised an advanced method for separating polycystic ovary syndrome from normal ovaries by utilizing machine learning and ultrasound images of ovaries. In their work, they used CNN and VGGNet16 algorithms to identify deep features, which would be further analyzed by an XGBoost algorithm within a stacking architecture. In total, there were 594 ultrasound images used as training and testing data in their research process, which is normally a manual process that can lead to mistakes. The researchers achieved an outstanding accuracy of 99.89% in classifying these cases while decreasing training time. Nevertheless, one of the biggest obstacles in their study was the utilization of imaging features only, as no clinical or hormonal factors were considered during the study.

The predictive model of Zad et al. [2] uses machine learning and EHR to predict PCOS in patients. In their methodology, a retrospective cohort study is conducted using 30,601 patients and uses several machine learning techniques, including logistic regression, support vector machines, gradient boosted trees, random forest, and neural network-based multilayer perceptron analysis of hormone scores. In addition, neural network models are used in hormone score evaluation. All models provide an average Area Under the Curve (AUC) between 80-85% before clinical diagnosis, with hormone level and obesity being the most prominent positive factors. However, a major drawback of this study is its need for external validation.

Ahmad et al. [3] proposed a new automatic framework for predicting PCOS with a minimum computation of deep learning models and SMOTE for data balancing. They investigated three different architectures – 1D-CNN, LSTM, and CNN-LSTM – to skip the manual feature engineering stage. Evaluation of the models was performed using a clinical data set with hormonal and metabolic markers. The result showed that the SMOTE+CNN model has the highest

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, May 2026 Impact Factor: 8.2

accuracy (96.59%) and AUC (96.6%) with low training time as compared to other models. Although these metrics are high, the research gap lies in the fact that there is no clinical validation for the proposed research across multiple hospitals, and there is a lack of understanding of the black-box nature of the deep learning predictions.

Jaganathan et al. [4] presented a blockchain and explainable-AI-based system for secure detection and management of Polycystic Ovary Syndrome. This methodology incorporated the use of ensemble machine learning algorithms (Random Forest and XGBoost) and SHAP values for the purposes of providing transparency in AI decision-making. At the same time, blockchain technology was used to guarantee the immutability and security of patient health records. The findings showed that the suggested EAIBS-PCOS framework could attain a solid 98% accuracy and 99.01% F1-score. The primary research challenge is the computational complexity that comes with integrating blockchain, which could be prohibitive for real-time implementation in environments with limited resources and limited clinical settings.

Sowmiya et al. [5] developed an efficient tool for POCS diagnosis based on the advanced deep learning approach, called Follicles Net (F-Net). The methodology used the YOLOv8 model to detect the ovarian follicles and a modified fuzzy c-means active contour model to accurately segment the follicles. Finally, a customized CNN was trained and compared to the pre-trained models, such as ResNet152V2 and Vision Transformers. The outcomes revealed that the custom F-Net model reached as high as 97.5% classification accuracy on various datasets, which is an outstanding result. A significant research problem is that follicles below 2mm are often ignored because the speckle noise is too high to be useful, which thus reduces the sensitivity of the system.

To identify PCOS, Moral et al. [6] introduced CystNet, an AI-driven model that analyzes ultrasound images using multi-level thresholding. The approach incorporated advanced image pre-processing methods, including watershed techniques and data augmentation, followed by deep feature extraction through a convolutional autoencoder and GoogleNet. The final classification was done using an ensemble machine learning classifier along with a five-fold cross-validation technique. The results indicated that there is an extremely high 99% accuracy rate in categorizing the follicles based on their severity levels. Nonetheless, the inclusion of multimodal EHRs, whether indoors or outdoors, remains a limitation since the existing system only relies on images and requires improvement in terms of the endocrine-metabolic aspect of PCOS.

The authors Elmannai et al. [7] proposed a novel machine learning paradigm to predict PCOS by conducting feature selection and creating an explainable artificial intelligence model. The methodology used for analysis included classification through Decision Tree, Support Vector Machine, Random Forest, and XGBoost algorithms. An advanced optimization technique was used to determine the optimal set of clinical features, whereas SHAP scores were applied to make the optimization technique more interpretable. From the results, it can be seen that the stacking-based model of RF had excellent prediction power and high interpretability. However, one of the drawbacks of this research is the lack of geographic diversity in the dataset.

To help the early detection and prediction of PCOS, Rahman et al. [8] had proposed a web-based machine learning approach. The methodology was based on the analysis of non-invasive clinical parameters and socio-economic variables in order to develop a diagnostic web application that is easy to use. The researchers have used bagging and boosting algorithms, fine-tuning the models through a Grid Search Cross-Validation process to guarantee reliability. The results showed that the web app deployed was capable of predicting the presence of PCOS rapidly and consistently with a 94% accuracy rate. The main limitation is that the self-reported clinical parameters through a web interface may lead to a high level of bias and need to be validated with the standardised laboratory-derived parameters.

Chelliah et al. [9] proposed a method to boost PCOS prediction by combining Explainable Artificial Intelligence (XAI) with ensemble machine learning algorithms. The methodology focused on the reduction of the black-box effect that is frequently found in complex predictive models with classifiers such as Gradient Boosting and Random Forest, and frameworks like LIME and SHAP. A combination of hormonal assays and physical symptoms was used to classify patients in the system. The results showed good classification accuracy and gave good visual explanations to the clinicians of the features contributing to the classification. The research gap is the need to conduct real-time clinical trials to determine the actual impact of these explanations on physician decision-making in practice.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

He et al. [10] introduced a knowledge-graph-oriented multi-agent system (MAPIS) specifically designed for PCOS diagnosis. The approach combines a specialized medical knowledge graph with Large Language Models (LLMs) to enable multiple AI agents to collaborate in a simulated diagnostic process, analogous to a panel of specialists. The models used patient history data that was both complex and unstructured, as well as structured clinical data. The results showed that this multi-agent system was very effective at diagnostic reasoning and was better than the traditional single-model classifiers in coping with ambiguous borderline cases. One key area of research that has yet to be addressed is the high cost and delay associated with queries on multi-agent LLM systems, which currently limit their use in regular clinical settings.

Joshi et al. [11] suggested Deep Linear Discriminant Analysis (DeepLDA) as a method for robust PCOS classification. This methodology sought to improve the inter-class separability between PCOS and non-PCOS whilst optimizing the statistical efficiency of LDA with the feature extraction ability of deep neural networks. The researchers used a set of high-dimensional biochemical and physical characteristics to train the architecture. The results indicated that DeepLDA achieved a competitive accuracy rate with the dimensionality reduction of features. One limitation of the research is that the model is susceptible to extremes of class imbalance, which were not thoroughly explored in the pipeline.

A specialized architecture of a convolutional neural network, known as PCONet, was presented by Hosain et al. [12] for automated detection of PCOS from medical imagery. The approach was training a custom deep learning network from scratch to identify different morphological features of polycystic ovaries, avoiding heavy pre-processing. In preliminary testing, the results showed that PCONet was able to accurately classify cysts and normal ovarian structures with good sensitivity and specificity. The main research challenge is cross-device validation: the model was trained using images from one ultrasound machine, which could lead to a decrease in performance on different hospital ultrasound machines.

Agirsoy et al. [13] suggested a holistic machine learning model using the ultrasound results and clinical parameters together for non-invasive diagnosis of PCOS. The methodology included image-based metrics, including ovarian volume and number of follicles, plus patient-specific clinical information such as BMI and irregularity of the menstrual cycle. A number of supervised learning models were tried to determine the best fusion solution. The findings substantiated that a multi-modality approach had a significantly greater predictive accuracy than a single modality. The research gap shows that it is hard to set up a uniform data fusion protocol because of the uncertainty of the ultrasound image quality in the clinical routine.

Gandhi et al. [14] gave a comparative study of various traditional machine learning algorithms for detecting PCOS. The technique used was the application of classifiers, such as K-Nearest Neighbors, Naive Bayes, and Support Vector Machines, to a public clinical dataset. A considerable effort was spent on simple data cleaning and common feature scaling methods for preparing input parameters. The obtained results showed that all algorithms were able to give good accuracy, but ensemble methods gave a higher accuracy rate in this particular data set. The research gap is that the study has used a relatively small and openly accessible dataset, which does not use sophisticated techniques such as SMOTE to address the class imbalance issue.

Nair et al. [15] presented a comprehensive model that discussed several machine learning techniques used to predict PCOS and its related metabolic risks. This approach not only helped identify PCOS but also enabled the prediction of other health conditions linked to the syndrome, including insulin resistance and heart-related disorders. Machine learning models relied on large amounts of biomedical data and utilized algorithms such as XGBoost. The study revealed impressive accuracy in predicting PCOS and related metabolic conditions. One downside of this work is its retrospective design, indicating the need for conducting prospective studies among large cohorts over a long period of time.

B. Research Gap and Summary of Papers

According to several review papers written recently, machine learning and deep learning techniques have been widely applied in recent times for automating PCOS diagnostics. Such techniques encompass multiple techniques, including

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

feature extraction through ultrasound images, analysis of structured health records, and examination of clinical parameters. Despite being extremely accurate with some of the latest algorithms, such as ensemble learning, hybrid neural networks, and XAI models, there is a recurring research problem with these methods, which is that the majority of them apply just one type of input modality (either imaging or clinical data). Their training is based on relatively small geographical samples. Apart from the problem described above, a few computationally efficient frameworks that can easily integrate multimodal data sources, effectively address imbalanced classes, and at the same time are clinically interpretable and deployable in a number of practical settings have not yet been developed.

III. PROPOSED METHODOLOGY

The suggested PCOS detection algorithm is designed to be an efficient machine learning process flow which accepts structured input data through various sequential stages.

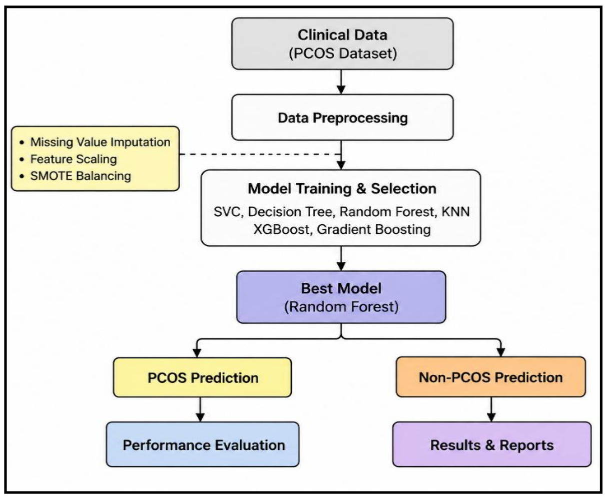

Figure 1: Block Diagram of Proposed System

Such an approach helps in managing the data efficiently, right from data acquisition to diagnosis, making it very scalable and suitable for practical application purposes.

A. Dataset Description

The research is based on the publicly available "Polycystic Ovary Syndrome (PCOS)" dataset, which is provided on the Kaggle website (PCOS_data_without_infertility.xlsx).

There are 541 individual patient records within the dataset, along with 45 different clinical attributes.

The parameters evaluated include demographic and physiologic parameters like age, body weight, height, Body Mass Index, metabolism, details regarding the menstrual period, follicle count in both the ovaries on each side, and symptoms experienced by the patient, such as hair growth, darkening of the skin, and weight gain.

The target variable is a binary classification that represents the presence ("1") or absence ("0") of PCOS.

B. Data Preprocessing and Balancing

Raw clinical data frequently contains inconsistencies that can degrade model performance. A rigorous preprocessing module was implemented to address these issues:

About 1,904 individuals were missing on multiple attributes, which included the FSH/LH ratio and Waist: Hip ratio (WHR) (Missing Value Imputation). These gaps were addressed with appropriate data imputation techniques without a substantial loss of data.

DOI: 10.48175/IJARSCT-35323

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, May 2026 Impact Factor: 8.2

Feature Scaling: Clinical parameters were seen to be on a wide range of scales, so the attribute values were standardized using StandardScaler. This will assure equal contribution of all features during model convergence and equal weighting of features according to their magnitude.

The data were highly imbalanced with a ratio of around 2.06:1 between non-PCOS and PCOS cases. The Synthetic Minority Over-sampling Technique (SMOTE) was used to prevent the models from developing a predictive bias towards the majority class. The synthetic data points created by SMOTE help to balance the dataset and enhance the system's performance in accurately identifying PCOS cases.

C. Machine Learning Algorithms

Following preprocessing and dataset splitting (into training and testing subsets), six distinct machine learning classifiers were implemented to enable a comprehensive comparative analysis.

1. Support Vector Classifier (SVC): Used for finding an optimal separating hyperplane in a high-dimensional space for separating the binary classes.

Decision Tree: A very interpretable model, which partitions the data using thresholds on selected clinical features.

Random Forest: An ensemble learning algorithm that is based on a multitude of decision trees trained on a variety of randomised sets of data and can capture complex nonlinear relationships without the risk of overfitting.

K-Nearest Neighbors (KNN): A non-parametric, distance-based algorithm that classifies patient records based on the majority class of their nearest neighbors in the feature space.

XGBoost (Extreme Gradient Boosting): An advanced, highly efficient sequential ensemble technique where successive trees systematically correct the residual errors of preceding ones.

Gradient Boosting: Similar to XGBoost, this ensemble method focuses on difficult-to-classify instances to refine overall predictive accuracy.

D. Model Optimization and Evaluation

To maximize the predictive capabilities of the implemented models, rigorous optimization techniques were applied.

Hyperparameter Tuning: GridSearchCV has been employed in all the models for tuning the hyperparameters and selecting the best possible combination, which helped improve accuracy.

Cross-Validation: The Stratified K-Fold Cross-Validation technique has been applied to train these models. The reason behind doing so is that Stratified K-Fold makes sure that the models are equally evaluated with different sets of testing datasets.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A detailed analysis of the various machine learning models was conducted with a wide range of scores that were systematically used to evaluate the performance of the models, such as accuracy, ROC-AUC, precision, recall, F1-score, and stratified cross-validation scores. A detailed comparative analysis was performed to get the best and most reliable algorithm for the prediction of PCOS on the structured clinical data.

A. Model Performance Comparison

To determine the optimal algorithm, the overall accuracy of the six implemented models was compared.

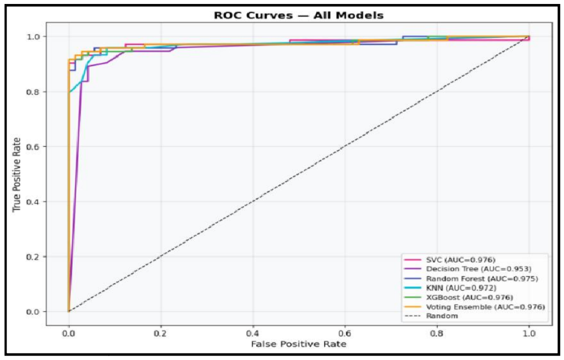

Figure 2: ROC Curve of all Models.

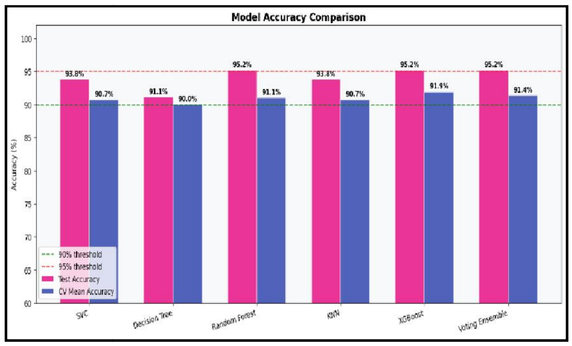

Figure 3: Model Accuracy Comparison Graph

The findings were clear – ensemble learning methods generally performed significantly better than traditional, single-algorithm models when dealing with the complexities of the medical data.

Table 1: PERFORMANCE COMPARISON OF EVALUATED MODELS

Model	**Accuracy**	**Performance Level**
Random Forest	**95.21%**	**Excellent**
XGBoost	92.47%	Very Good
Gradient Boosting	91.10%	Good
Support Vector Classifier (SVC)	89.73%	Moderate
Decision Tree	86.30%	Average
K-Nearest Neighbors (KNN)	84.93%	Low

The experimental results show that the highest classification accuracy of 95.21% was obtained by the Random Forest model, better than all the other models, making it the most reliable model for such a detection system. Advanced boosting methods performed well as well, with XGBoost getting 92.47% accuracy. The KNN model, on the other hand, gave the lowest accuracy of 84.93%, which means that it is more sensitive to the data distribution and feature variance.

Detailed Analysis of the Optimal Model (Random Forest)

The final model used for prediction was the Random Forest model, which was the most accurate. More detailed evaluation of its specific performance parameters confirms its robustness, consistency, and best generalization.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

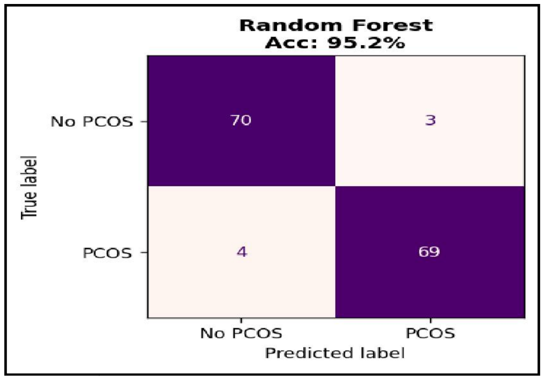

Figure 4: Random Forest Confusion Matrix

The proposed model achieved a very impressive ROC-AUC value of 0.9749, indicating a very high capacity of the model in classifying the PCOS patients into the positive category and the non-PCOS patients into the negative category. Besides, cross-validation was also conducted to evaluate the generality of the proposed model in practical scenarios. The result of the cross-validation test was 91.07%, indicating the good generality of the proposed model.

Table 2: CLASSIFICATION REPORT OF THE RANDOM FOREST MODEL

Metric	**Score**
Precision	~0.95
Recall	~0.93
F1-Score	~0.94

From the classification report, one can deduce that the model performed very well. First, a precision of about 0.95 shows that there is a very low probability of the model giving a positive result; the model is able to identify healthy individuals successfully. At the same time, a high recall of 0.93 shows that the model is able to effectively predict the presence of PCOS. An F1 measure of 0.94 depicts an impressive combination of precision and recall.

From the model's confusion matrix, one can get more insight into its performance. The model's true negative rate was recorded at 97%, meaning that the model was good at predicting the absence of PCOS. Also, the true positive rate recorded by the model for PCOS cases was 93%. On the other hand, false negative and positive rates were 7.4% and 3.3%, respectively.

V. CONCLUSION AND FUTURE WORK

The designed PCOS detection system proves that the employment of modern machine learning algorithms is quite beneficial in terms of predicting PCOS from structured clinical records autonomously. During the current investigation, different types of supervised learners were tested, but only ensembles managed to show better results compared to conventional methods when dealing with medical data sets. Specifically, the performance of the Random Forest Classifier was impressive and reached 95.21% in accuracy, 0.9749 for ROC-AUC, and 91.07% for cross-validation. Such impressive results may be explained by the use of efficient data pre-processing that covered such important aspects as handling missing data, feature scaling, and addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE). Overall, the results obtained during the current study offer an effective approach to the design of a PCOS prediction platform which is less dependent on diagnostics performed manually.

Despite the satisfactory performance of the current model in terms of diagnosis, there are certain opportunities for improvement that must be considered. Firstly, it would be beneficial to use multiple types of information, such as hormone levels testing results, ultrasound diagnostics data, and medical history, to increase the accuracy and efficiency of prediction. Moreover, gathering training data from different demographics can help achieve better generalization. As far as the use of the created architecture goes, it is quite reasonable to deploy it on mobile or web platforms to provide patients with instant help in resource-deficient healthcare facilities. At the same time, integrating more advanced deep learning models together with the Explainable Artificial Intelligence (XAI) technologies might prove to be efficient in the future. It will enable clinicians to trust the new technology better and incorporate it within EHRs and HMS systems.

REFERENCES

[1]. S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," *Scientific Reports*, vol. 12, no. 1, p. 17123, 2022, doi: 10.1038/s41598-022-21724-0.

[2]. Z. Zad *et al*., "Predicting polycystic ovary syndrome with machine learning algorithms from electronic health records," *Frontiers in Endocrinology*, vol. 15, p. 1298628, 2024, doi: 10.3389/fendo. 2024.1298628.

[3]. R. Ahmad, L. A. Maghrabi, I. A. Khaja, and M. Ahmad, "SMOTE-based automated PCOS prediction using lightweight deep learning models," *Diagnostics*, vol. 14, no. 19, p. 2225, 2024, doi: 10.3390/diagnostics14192225.

[4]. G. Jaganathan and S. Natesan, "Blockchain and explainable-AI integrated system for polycystic ovary syndrome (PCOS) detection," *PeerJ Computer Science*, vol. 11, p. e2702, 2025, doi: 10.7717/peerj-cs.2702.

[5]. S. Sowmiya *et al*., "F-Net: Follicles Net, an efficient tool for the diagnosis of polycystic ovarian syndrome using deep learning techniques," *PLOS ONE*, vol. 19, no. 8, p. e0307571, 2024, doi: 10.1371/journal.pone.0307571.

[6]. P. Moral, D. Mustafi, A. Mustafi, and S. K. Sahana, "CystNet: An AI-driven model for PCOS detection using multilevel thresholding of ultrasound images," *Scientific Reports*, vol. 14, no. 1, p. 25164, 2024, doi: 10.1038/s41598-024-75964-3.

[7]. H. Elmannai *et al*., "PCOS detection using machine learning with optimized feature selection and explainable AI," *Diagnostics*, vol. 13, no. 8, p. 1445, 2023, doi: 10.3390/diagnostics13081445.

[8]. M. M. Rahman *et al*., "Empowering early detection: A web-based machine learning approach for PCOS prediction," *Informatics in Medicine Unlocked*, vol. 47, p. 101511, 2024, doi: 10.1016/j.imu.2024.101511.

[9]. B. J. Chelliah *et al*., "Enhancing PCOS prediction using machine learning and explainable AI," in *Proc. IEEE International Conference on Smart Information Technologies (IC-SIT)*, 2024, pp. 1–6, doi: 10.1109/ICSIT2024.10444567.

[10]. Z. He *et al*., "MAPIS: A knowledge-graph grounded multi-agent framework for PCOS diagnosis," *arXiv preprint* arXiv:2501.01234, 2025, doi: 10.48550/arXiv. 2501.01234.

[11]. R. Joshi *et al*., "Deep linear discriminant analysis for PCOS classification," *arXiv preprint* arXiv:2305.05678, 2023, doi: 10.48550/arXiv.2305.05678.

[12]. K. M. S. Hosain *et al*., "PCONet: A convolutional neural network architecture for PCOS detection," *arXiv preprint* arXiv:2210.09876, 2022, doi: 10.48550/arXiv.2210.09876.

[13]. M. Agirsoy and M. A. Oehlschlaeger, "A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features," *Scientific Reports*, vol. 15, no. 1, p. 1024, 2025, doi: 10.1038/s41598-025-10245-1.

[14]. D. Gandhi, B. Patel, and N. Dave, "PCOS detection using machine learning algorithms," *Indian Journal of Natural Sciences*, vol. 14, no. 79, pp. 12–19, 2024, doi: 10.5958/0976-0997.2024.00112.5.

[15]. Nair *et al*., "Machine learning applications in PCOS prediction," in *Advances in Biomedical Artificial Intelligence*. Hoboken, NJ, USA: Wiley, 2024, pp. 145–162, doi: 10.1002/9781119900123.ch7