Grenze International Journal of Engineering and Technology, Jan Issue

Assistive PCOS Detection Platform with Clinical Data and Ultrasound Scan using Machine Learning

Sridevi B. S[1], Varna V Bhat[2], Shreyas K. V[3], Tejas R[4], Maya B. S[5] and Asha T[6]
[1-6]Computer Science & Engineering, Bangalore Institute of Technology, Bengaluru, India.
Email: mayabs@bit-bangalore.edu.in, sridevibs2014@gmail.com, varna.bhat.2001@gmail.com, shreyaskv94@gmail.com, tejasvirat72@gmail.com, ashat@bit-bangalore.edu.in

***Abstract*— Polycystic ovary syndrome (PCOS) is an endocrine disorder characterized by excess androgen hormone and symptoms like obesity, irregular cycles, acne, hirsutism, and abnormal ovarian follicle development. Treatment targets anovulation, hyper androgenism, and menstrual dysfunction. This study developed machine learning models using clinical data to predict PCOS presence. The dataset included FSH/LH values and observed symptoms. Random Forest, Decision, SVM, Naïve Bayesian, Gradient Boosting, and KNN classifiers were evaluated. Random Forest achieved 90% accuracy. Additionally, a doctor model based on CNN deep learning will provide a secondary diagnostic opinion to professionals. This model will enhance the accuracy of PCOS diagnosis and support healthcare providers in decision-making. An assistive platform is proposed to detect PCOS, provide prevention measures, and offer education on menstrual health and well-being. It aims to benefit rural communities and address specific PCOS needs in a concise manner.**

***Index Terms*— PCOS, follicles, Machine Learning models, Random Forest, clinical value, FSH/LH value, Assistive platform.**

I. INTRODUCTION

Polycystic ovary syndrome (PCOS) is hormonal problem that occurs during the reproductive years. Women affected by PCOS may not experience regular menstruation. One may experience periods lasting many days. One may end up having too much of the male hormone called androgen in their body. With PCOS, many small sacs of fluid develop along the outer edge of the ovary. These are called cysts. The small fluid-filled cysts contain immature eggs. These are called follicles. The follicles fail to release eggs during ovulation. The collection of many such follicles causes PCOS. The exact cause of PCOS is quite difficult to pinpoint. Common symptoms of PCOS is shown in figure 1 below. Early detection and treatment along with lifestyle changes and a healthy diet may lower the risk of long-term complications such as heart diseases and prediabetes. Since the exact cause of PCOS is not concrete, doctors have studied a range of patients and listed out few common factors that might be causing PCOS. They include: Insulin resistance, Low-grade inflammation, Heredity - Excess androgen.
There are numerous complications of PCOS women, most of which are highly problematic. This is the reason why awareness of PCOS is vital. Some of these issues can include like Difficulty in conceiving, Pregnancy-induced high blood pressure, Miscarriage or premature delivery, Liver inflammation caused by fat build-up in the liver, Metabolic syndrome- a cluster of conditions including high blood pressure, high blood sugar, and unhealthy cholesterol, Type 2 diabetes , Sleep apnoea , Depression, anxiety and eating disorders, Cancer of the endometrium,

Grenze ID: 01.GIJET.10.1.804

Obesity is a frequent consequence with PCOS and can worsen complications of the disorder.

II. Related Works

The detection and diagnosis of Polycystic Ovary Syndrome (PCOS) have been the focus of several studies that leverage machine learning techniques. Rachana et al. [1] proposed a follicle recognition technique on ultrasound images to detect PCOS. By identifying characteristic ovarian follicles associated with PCOS, their approach demonstrated promising results in the detection of the syndrome. Kaur et al. [2] developed a diet recommendation framework for PCOS management using deep convolutional neural networks. Bharati et al. [3] explored the diagnosis of PCOS using machine learning algorithms. They applied feature selection techniques and evaluated various machine learning models to identify key features associated with the disease.

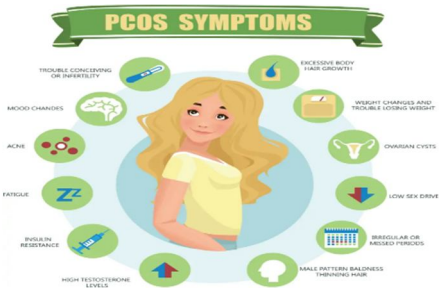


Figure. 1. PCOS Symptoms

Their study highlighted the importance of feature selection in improving diagnostic accuracy. Jaralba et al. [4] proposed a machine learning-based initial screening approach for PCOS. They employed five classifiers including Naïve Bayes, KNN, ANN, SVM, and RF to identify potential cases of PCOS. Mehr and Polat [5] compared different machine learning and feature selection techniques for PCOS diagnosis. Vedpathak and Thakre [6] developed a machine learning-based system for accurate PCOS detection and prediction of optimal treatment options. Emphasizing the significance of early intervention, their study highlighted the importance of precise detection for effective PCOS management. Salman Hosain et al. [7] proposed a fine-tuned InceptionV3 convolutional neural network for PCOS detection from ovarian ultrasound images. Sumathi et al. [8] utilized the watershed algorithm and Open CV for feature extraction from ultrasound images in PCOS detection. Denny et al. [9] developed i-HOPE, a PCOS detection and prediction system employing machine learning techniques. Their model incorporated clinical and metabolic parameters as biomarkers for PCOS diagnosis, aiming to provide accurate detection and personalized treatment recommendations. Kodipalli and Devi [10] focused on predicting PCOS and associated mental health conditions using fuzzy inference and SVM. Gupta et al. [11] proposed a PCOS detection system based on a combination of clinical features and machine learning algorithms. Their approach involved feature selection and employed classifiers such as Random Forest and Support Vector Machine to achieve accurate PCOS detection. Singh et al. [12] developed a PCOS detection model using a hybrid approach that combined deep learning and machine learning techniques. Das et al. [13] proposed a PCOS prediction model based on data-driven machine learning algorithms. They employed various feature selection techniques and utilized classifiers such as Decision Trees and Random Forest to predict the likelihood of PCOS development in individuals. Sharma et al. [14] developed an intelligent PCOS prediction system using machine learning algorithms. They employed feature selection and classification algorithms to achieve accurate prediction results. Agarwal et al. [15] presented a PCOS detection framework using a combination of clinical data and machine learning techniques. Their system utilized feature engineering and dimensionality reduction techniques to extract relevant features, and applied classifiers such as Gradient Boosting and Support Vector Machine for accurate PCOS diagnosis.

III. Methodology

Inspired by the existing research works in predicting PCOS in patients, it is proposed to implement a cross platform application that aids both the patients and doctors in diagnosing PCOS. The model would also suggest preventive methods to the patients. The overview of the architecture of the PCOS Assistive platform is shown in figure 2.
The clinical datasets and ultrasound images are collected by the system to preprocess, analyze and extract the key features and image of any significant cysts to further train the model using Machine Learning and Deep Learning algorithms. The trained model is later deployed on the cross-platform application that will suggest to the user to

consult a doctor if she's likely to be suffering with PCOS and also provide aid to the professional as a secondary opinion. The platform would also provide methods like Food-Image based Diet recommendation, Video tutorials on Yoga and lifestyle through which the user can prevent PCOD.
The Cross-Platform consists of two major models - Patient and Doctor.

A. Patient Model

In this model, the Chabot takes the clinical data from the user to predict if the Patient has PCOS or not. The model is proposed to be implemented using the Random Forest classifier. The model would then primarily suggest to the Patient to consult a professional for further diagnosis. The overview of the patient model is shown in figure 3.

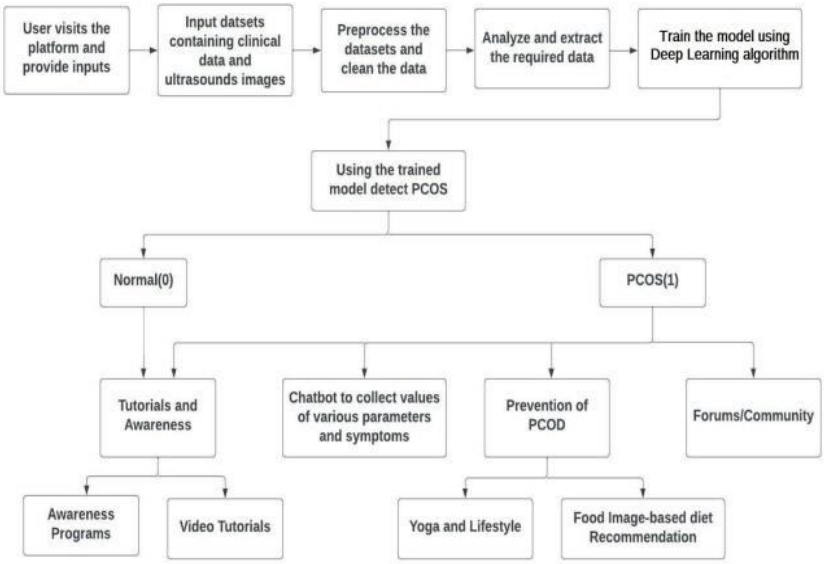


Figure. 2. Architecture of Assistive PCOS Detection platform

The prediction in patient model is done in the following steps:

B. Training phase

The model is trained on a large dataset of clinical and biochemical test data to accurately predict the presence of PCOS. During this process, the machine learning algorithms are trained to identify patterns and features within the input data that are indicative of PCOS. This training phase is essential to ensure that the PCOS-Assist system provides accurate results and assists medical professionals in diagnosing and treating PCOS effectively.

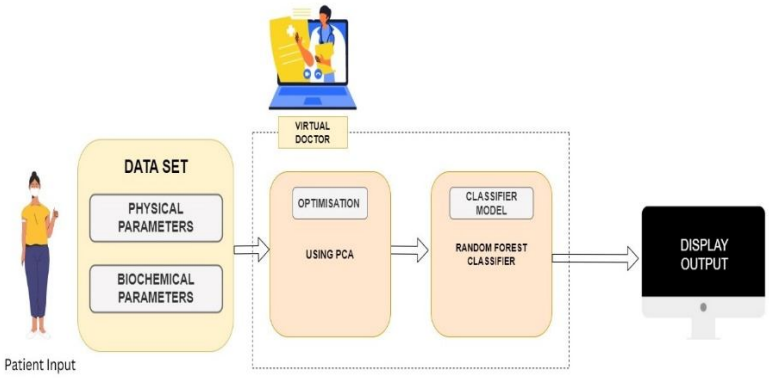


Figure. 3. Overview of patient model in PCOS cross platform

C. Loading of trained model

Once the model is trained, it is loaded into the system and is ready for use. Loading the trained model is essential because it allows the system to utilize the learned knowledge and expertise of the model when processing new data. By loading the trained model into the system, the PCOS-Assist system can provide accurate predictions and detections of PCOS based on the knowledge and experience gained during the training phase.

D. Patient input Clinical Test and Biochemical Test data

The patient inputs their clinical and biochemical test data into the system for analysis. The clinical test data may include information such as age, body mass index (BMI), menstrual history, and symptoms associated with PCOS. The biochemical test data may include information such as levels of hormones like testosterone, luteinizing hormone (LH), follicle-stimulating hormone (FSH), and insulin.

E. Feature Extraction

The feature extraction process involves identifying the most relevant features from the patient's clinical and biochemical test data, which can help to identify whether the patient is affected by PCOS. By extracting the most relevant features from the patient's input data, the system can more accurately predict the presence or absence of PCOS, which can help medical professionals make better-informed decisions when diagnosing and treating patients.

F. Apply the prediction model

The extracted features are analyzed using machine learning algorithms to predict the presence of PCOS in the patient. The prediction model is Random Forest algorithm that has been trained on a large dataset of clinical and biochemical test data to accurately predict the presence of PCOS. The model then generates a probability score that indicates the likelihood of the patient having PCOS.

G. Display Output

After the machine learning model analyzes the extracted features, the system displays the results of the analysis as output. The system displays the results of the analysis as output, providing the patient and medical professional with an accurate diagnosis of PCOS. The output displayed includes the probability score generated by the prediction model, indicating the likelihood.

H. Doctor Model

This model detects ultrasound scans to identify cysts as simple cysts, PCOS, or cancer masses. To accomplish this, an Ensemble Deep Learning model has been implemented using TensorFlow. This model can effectively detect and classify cysts present in ultrasound scans with a good degree of accuracy. By identifying the type of cyst present in the scan, the system can help medical professionals determine the most appropriate course of treatment for the patient. This module is a crucial component of the PCOS-Assist system as it provides medical professionals with an effective tool for diagnosing and treating patients with cysts, improving their overall health outcomes. The doctor model overview is shown in the Figure 4

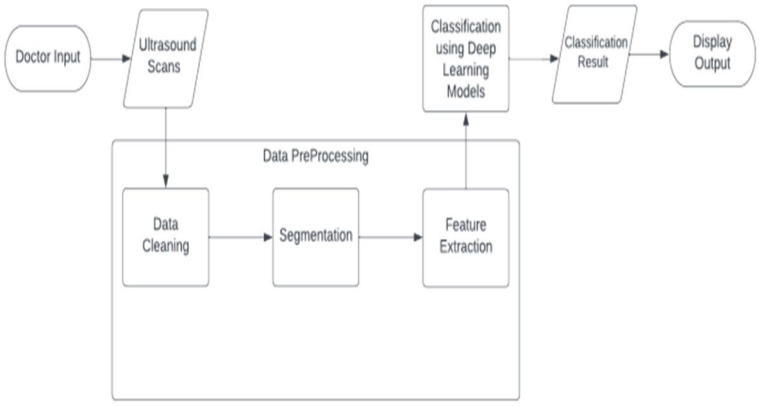


Figure. 4. Overview of Doctor Model

Doctor model detects PCOS in the following steps:

I. Loading Datasets

Large dataset of ultrasound scan images are loaded to the system with specified labels of PCOS and normal ovaries. The model is trained based on this loaded dataset.

J. Preprocessing data

Data preprocessing is transforming raw data into a format that is suitable for further analysis or training machine learning models. The preprocessing step often includes several sub-steps. This model has the following sub-steps as Converting image to Grayscale, Removing noise using erode and dilate, Thresholding, Determine extreme points, Cropping original image.

K. Data loading to a list

The preprocessed data is loaded to a list to facilitate further processing and analysis. Lists allow for easy manipulation of collections of data, such as image features or attributes. They can be easily sorted, filtered, or modified as needed to prepare the data for use in machine learning algorithms or other applications.

L. Split data for training and testing

The training set is used to train the model, while the testing set is used to evaluate how well the model generalizes to new, unseen data.

M. Building the model using CNN layers

CNNs [8] are highly effective in image classification tasks due to their ability to extract meaningful features from images. They can automatically learn important features such as edges, corners, and shapes from the input images, making them ideal for image-based tasks. This model uses 7 layers as Zero padding 2D layer, Conv 2D layer , Batch Normalization layer, Activation layer, Max Pooling 2D layer, Flatten layer, Dense layer .

N. Training the model

This involves feeding the training data into the model and adjusting the weights of the CNN layers over multiple epochs with a predefined learning rate to optimize the model's performance.

O. Output

After the model is trained, it is evaluated on the test dataset to calculate the accuracy, precision, recall, and F1 score. The classification report is then displayed to provide a detailed summary of the model's performance.

IV. Results And Discussions

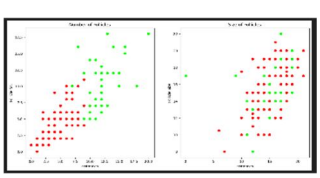
A. Patient Model

A model is proposed here which predicts the possibility of a patient having PCOS based on collected data from their FSH/LH test results such as FSH/LH hormone levels and also the symptoms collected during their doctor visit such as details regarding the patient’s period cycle, hair growth and other factors. Considering all of the above factors, a dataset is developed which consists of 44 attributes of 541 patients obtained from Kaggle. A prediction system is implemented which predicts the probability of a person having PCOS given the test data. The model is first cleaned through manual methods like removing null and irrelevant values and removing NaN values. The dataset was trained using various classifiers such as: Random Forest Classifier, Decision Classifier, Support Vector Machine Classifier, Naïve Bayesian Classifier, Gradient Boosting Classifier, KNN Classifier. When comparing Random Forest classifier for the patient model and Convolutional Neural Networks (CNN) for the Doctor model with the state-of-the-art approaches for PCOS diagnosis, there are several factors to consider.After training all the models, it was found that Random Forest classifier yielded the best results compared to all others. It gave an accuracy of 90 %. The results and the comparison of all the applied classifiers are shown 5.

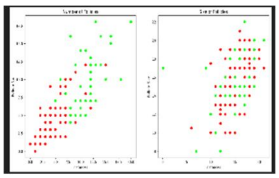
B. Doctor Model

PCOS doctor model was trained and evaluated using a dataset consisting of 1923 examples. The above table shows the different cases with accuracy and confusion matrix. The results obtained during the evaluation with Test Loss is 0.063. This value indicates the average loss (error) of the model on the test data, lower value represents better performance with Test Accuracy: 0.97 shown in figure 6a.
Further analysis and validation of the model's performance in real-world clinical settings would be necessary to

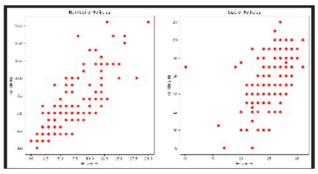
assess its practical utility and generalizability. The confusion matrix for testing images is shown in fig 6b,c. Home screen and awareness of PCOS assist shown in fig 7.



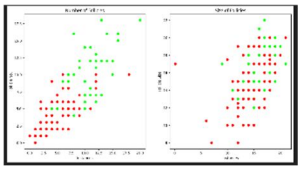
a. RF



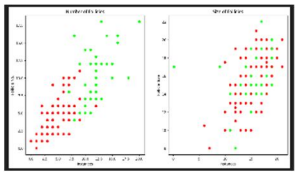
b..DT



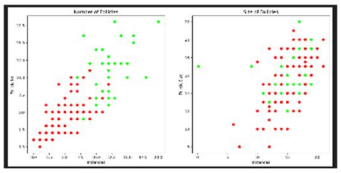
c. SVM



d. NB



e. GB



f. KN

Figure. 5. Different Classifier follicle size graph

TABLE I. DIFFERENT TEST CASES WITH RANDOM FOREST CLASSIFIER

Test No.	***Accuracy***	***Confusion Matrix***	**Test No.**	***Accuracy***	***Confusion Matrix***	**Test No.**	***Accuracy***	***Confusion Matrix***
1	0.911111111	[[98 5] [8 24]]	5	0.903703704	[[98 5] [8 24]]	9	0.888888889	[[97 6] [9 23]]
2	0.896296296	[[97 6] [8 24]]	6	0.896296296	[[97 6] [8 24]]	10	0.888888889	[[96 7] [8 24]]
3	0.903703704	[[97 6] [7 25]]	7	0.896296296	[[96 7] [7 25]]	11	**0.901481481**	**RFC**
4	0.896296296	[[96 7] [7 25]]	8	0.933333333	[[100 3] [6 26]]			

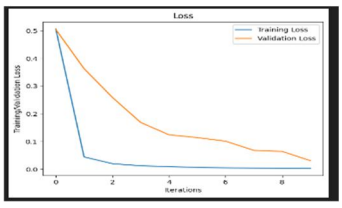
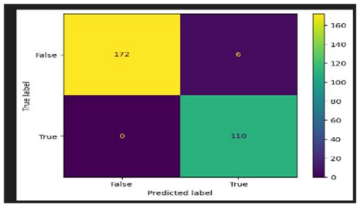
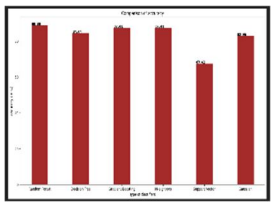


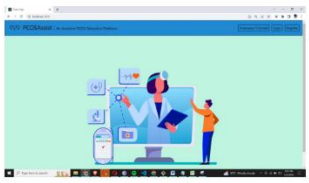
Figure 6a. Loss function



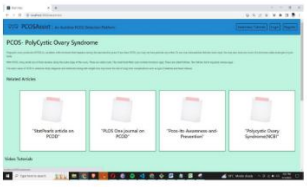
6b. Confusion Matrix



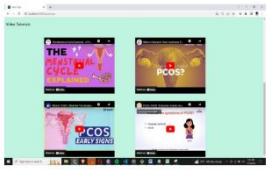
6c. Different Accuracy



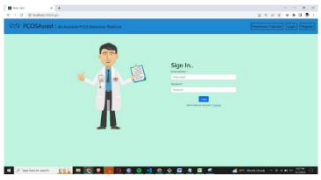
7a. Home Screen



7b.Awareness screen of PCOS



7c.Video tutorial Awareness



7d. Login page

7e. Preventive measures

7f. Doctor Model in PCOS

V. Conclusion

In this paper, we aimed to develop an assistive PCOS detection platform using machine learning techniques, clinical data, and ultrasound scans. Our platform focused on providing guidance and assistance to rural women in the context of women's health to predict PCOS early stages. Additionally, we developed an open forum within the platform, allowing users to communicate, express their feelings, resolve doubts, and share experiences related to PCOS. There are certain limitations, the accuracy of PCOS detection heavily relies on the quality and availability of clinical data and ultrasound scans. In future scope, platform's recommendations for preventive measures, such as yoga, lifestyle modifications, and diet recommendations, are proposed to be customized to each user.

References

[1] B. Rachana, T Priyanka, et al, “Detection of polycystic ovarian syndrome using follicle recognition tech-nique”, Glo. Tran. Proc., vol.2, pp.304-308, 2021.

[2] Rajdeep Kaur, Rakesh Kumar, Meenu Gupta, “Food Image-based diet recommendation framework to overcome PCOS problem in women using deep convolutional neural network”, Com. and Ele, Engg.,Elsevier, vol.103, pp.0045-0064, 2022.

[3] S. Bharati, P. Podder and M. R. Hossain Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Ma-chine Learning Algorithms“, IEEE Region 10 Symp., pp.1486-1489, 2020

[4] Joshua Rei Jaralba , Renann Baldovino, and Homer Co, “A Machine Learning Approach for Initial Screening of Polycystic Ovarian Syndrome”, Smart City Applications, Lecture Notes in Net.s and Sys., vol.183, pp. 517–529, 2021.

[5] Homay Danaei Mehr, Huseyin Polat, “Diagnosis of polycystic ovary syndrome through different ma-chine learning and feature selection techniques”, J.Hel. and Tech., vol.12, pp. 137–150, 2022.

[6] Vedpathak, Shreyas & Thakre, Vaidehi, “PCOcare: PCOS Detection and Prediction using Machine Learning Algorithms”, J.Bio. Biotec. Res. Comm., vol.13(14), pp. 240-244, 2020.

[7] A.K.M. Salman Hosain, Md Humaion Kabir Mehedi, Irteza Enan Kabir."PCONet: A Convolutional Neu-ral Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images", 8th Inter.Conf. on Engg. and Emer. Tech., vol.4, pp.24-30, 2022.

[8] Sumathi, M., P. Chitra, R Sakthi Prabha and K. Srilatha. “Study and detection of PCOS related diseases using CNN.”, IOP Conf. Series: Materials Science and Engg., 1070 , vol.1, 2021.

[9] A. Denny, A. Raj, A. Ashok, C. M. Ram and R. George, "i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques“, IEEE Region 10 Conf., pp.673-678,2019.

[10] Kodipalli, Ashwini & Devi, Susheela, “ Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM ”, Frontiers in Public Health, vol. 9, 2021.

[11] Gupta A, Sharma S, Joshi A. "PCOS detection using machine learning algorithms”, 7th Inter.Conference on Comp., Com. and Sec. , 2022. p. 1-5.

[12] Singh A, Giri D, Chakraborty C. "PCOS detection using deep learning and machine learning techniques”, Inter. Conf. on Inte.Comp. and Sust. Sys. . 2022. p. 178-183.

[13] Das A, Choudhury SR, Bhattacharyya S. "PCOS prediction using machine learning techniques." In: 2022 Inter. Conf. on Smart Sys. and Inv. Tech., 2022. p. 892-897.

[14] Sharma M, Arora P, Choudhary A, Kumar S. "An intelligent system for PCOS prediction using machine learning”, Inter. Conf. on Com., Power and Com. Tech., 2022. p. 1-5.

[15] Agarwal S, Yadav R, Yadav S, Kumar A, Verma S. "PCOS detection framework using machine learning algorithms”,7th Inter. Conf. on Comp., Comm. and Sec., 2022. p. 1-6.