Identification And Classification of PCOD &PCOS Using Machine Learning Techniques

Assistant Prof. Sharanya P S, Megha, Ramitha Shetty ,Sharanya Rai B, Sushmitha
(Email: sharanya@ajiet.edu.in,meghamahesh264@gmail.com,shettyramitha693@gmail.com, sharanyarai2848@gmail.com,sushmibirwas@gmail.com)

***Abstract*—** Polycystic Ovarian Disease (PCOD) is a medical condition in which a woman's ovaries generate immature or partially mature eggs in large numbers during reproductive age. These eggs develop into cysts in the ovaries over time. Due to the cysts, ovaries become large & secrete large amounts of male hormones (androgen) which causes infertility problems, irregular periods, unwanted weight gain, and other health issues. Polycystic Ovary Syndrome (PCOS) is a condition that affects girl or women during their child-bearing years and disturbs the levels of hormones. This disturbance results in problems affecting many body systems. Women having PCOS have skip or irregularity in menstrual periods as well as cysts formation in the either or both ovaries. Symptoms of PCOS and PCOD are irregular periods, excess androgen, polycystic ovaries, abnormal BMI, disturbed levels of hormones (LH, FSH, DHEAS), poor insulin resistance. But as per research studies, these symptoms are not sufficient for accurate detection for diverse data.

This proposed system will verify the possibilities of PCOS and PCOD based on their symptoms and confirms it by classifying Ultrasound ovary images using deep learning methodology. The ultra-sound ovary image analysis along with the physical symptoms of PCOS and PCOD are needed for accurate detection and reducing number of outliers during analysis. Such detection of PCOS and PCOD also helps in proper treatment and reducing the health loss.

Keywords—Symptoms based form for PCOD/PCOS, Ultrasound image processing ,Precautions.

INTRODUCTION

There are many disorders related to women reproductive system which may lead to some serious health issues in future. These disorders are related to the ovaries, uterus, cervix, the vagina, etc. the cause behind occurring of these diseases is hormonal changes inside the body, hormonal imbalance, irregular living patterns, stress, etc. Polycystic ovary syndrome (PCOS) and PCOD comes in the category of hyperandrogenism i.e. excess androgen production by ovaries. It is a disorder commonly found in reproductive age group(15-40 yrs). The age group is fixed as before 15 age is when the menstruation begins; so there is a huge possibility that the menstrual periods are irregular and after 40 age the menopause periods of women begins. In this condition the women's hormone levels are affected. This hormonal imbalance leads to cysts formation on the outer edge of the ovaries. The cysts are like follicle or small ball of tissues which is found in both or either ovary of PCOS women. Commonly found symptoms PCOS are Irregularity or due to missed menstrual periods, Excessive Hair growth on face and unwanted body parts known as Hirsutism, Acne formation and oilier skin due to high androgen levels, abnormal Body Mass Index leading to obesity. Along with the mentioned physical symptoms; there is a need to conduct a blood test for checking the hormone levels in body. The symptoms associated with PCOD often leave a lasting impact on women's physical and mental well-being. It is estimated that about 34% of the women suffering from PCOD problem also suffer from depression, and nearly 45% suffer from anxiety. This makes it essential to diagnose and manage the symptoms early to prevent PCOD problems in future. Many of the women diagnosed with PCOD problem also experience poor quality of life-related to mood swings, negative social relationships, low self-confidence, negative self-image, disruption of eating and sleep patterns, low motivation

RELATED WORK

Detection of polycystic ovary syndrome has become a new topic for researchers since the last decade. Researchers have implemented various techniques to diagnose PCOS / PCOD at an early stage. Detection of polycystic ovary syndrome using machine learning algorithm the major contributions of this paper is pre-processing, which will include a selection of salient features using the feature selection method , plotting an architecture for hybrid machine learning mode, implementing the Hybrid Extreme Gradient Boosting with Random Forest (XGBRF) ensemble method and Cat Boost Model and comparing their results with baseline approaches by providing the reasonable comparison. Then used Gradient Boosting, Random Forest, Logistic Regression, HRFLR, SVM ,Decision Tree, MLP individually to produce the precision, recall, f1-score. Shakoor Ahmad Bhat (2021)[8]. Semantic segmentation and classification of polycystic ovarian disease using attention UNet , Pyspark and ensemble learning model the dataset used in this paper is obtained from SDM College of Medical Sciences and Hospital, Dharwad. ML algorithms used for dataset to check the speed and accuracy of computation, then optimize it using Randomized Search, Grid Search, and Bayesian methods. Results obtained from Random Forest provides better accuracy after optimization. This is the base model that is being used to process the data in Hadoop .The metrics used to measure the performance of the model are Accuracy, Precision, Recall, and F1 score[2]. Detecting PCOS using Machine Learning this paper focuses on two models used for supervised machine learning are K-NN and Logistic Regression, where Logistic Regression is more accurate with accuracy 92% while K-NN's is 90.74% and Filter method is used to find the weights of the features in order to determine which of them have high correlation with the target .The classification included are Precision, Recall and Support[1]. Empowering early detection: A web-based machine learning approach for PCOS prediction in this paper it uses 13 classifiers, Bernoulli NB, Support Vector Machines, Random Forest, K- Nearest Neighbor, Naïve Bayes, Gradient Boosting, and

Multilayer perceptron where AdaBoost Classifier achieved the highest accuracy of 94 %.The user interface of the PCOS categorization system is constructed utilizing the Django framework[3]. Early detection of PCOD using machine learning techniques in this paper two types of object detection algorithms, Decision Trees, and Deep Neural Networks. In Decision Tree, we used XGBoost, LGBM, and CatBoost which are based on gradient boosting framework .In Deep Neural Network we used the USG image to YOLOv2 and tinyYOLOv3 to detect and draw the bounding box around the follicles on the USG image. The results achieved from the Decision Tree-based ML models when trained and tested on the Tabular data consisting of actual 540 records. The confusion matrix, precision, recall, and F1-score values obtained[5].

OBJECTIVE

To develop a machine learning model that process patient's symptoms to determine the possibilities of PCOS and PCOD. To develop a deep learning model using CNN to confirm the possibilities of PCOS and PCOD by extracting the features of ultra-sound ovary images. To apply the concept of machine learning and deep learning in the medical field to accurately detect the diseases. To enable the common user to detect PCOS and PCOD diseases without any difficulties and confusions.

PURPOSE

Main purpose of the project is to develop a reliable, efficient, and automated system that leverages machine learning algorithms to detect, differentiate, and classify Polycystic Ovarian Disease (PCOD) and Polycystic Ovary Syndrome (PCOS). These are common hormonal disorders affecting women, and their early diagnosis is critical for effective treatment and prevention of complications.

PROPOSED SYSTEM

The Proposed System is developed to make the life of affected women easy. This is a machine learning based system in which the women can check their health issues with respect to pcos/pcod by uploading the symptoms and further by uploading the ultrasound image of the ovaries, Through the help of the system it identifies and classifies whether its pcod / pcos/ normal ,as its very difficult for the affected women's to distinguish between pcod/pcos. In order to use this system , all the users can fill their particular symptoms in the form of yes or no, by calculating these symptoms through classifiers like random forest for pcos and k-nearest neighbor for pcod. After attaining positive result, ultrasound images can be uploaded using Convolution neural network .This system will also provide precautions based on the result. This paper can be put into use in any hospital, clinics.

Proposed System Consists of 3 Main Module:

1. Symptom based input form for Pcos using Random forest classifier.
2. Symptom based input form for Pcod using K-Nearest Neighbor classifier.
3. Ultrasound image processing using Convolution neural network.

A. Symptom based input form for PCOS using RandomForest classifier

Identify symptom-based input form for PCOS diagnosis using a Random Forest classifier, need symptoms associated with PCOS such as irregular periods, excessive hair growth (hirsutism), acne or oily skin, weight gain (especially around the abdomen), thinning hair, infertility, darkened skin areas (e.g., around the neck or armpits), sleep apnea, depression or anxiety, and high blood pressure or insulin resistance. The input form can be designed with a series of questions where the user answers in a binary format (Yes/No) these collected the data, can be trained in Random Forest classifier using a library like scikit-learn. The model will output a prediction, such as the patient having PCOS, based on their provided symptoms .

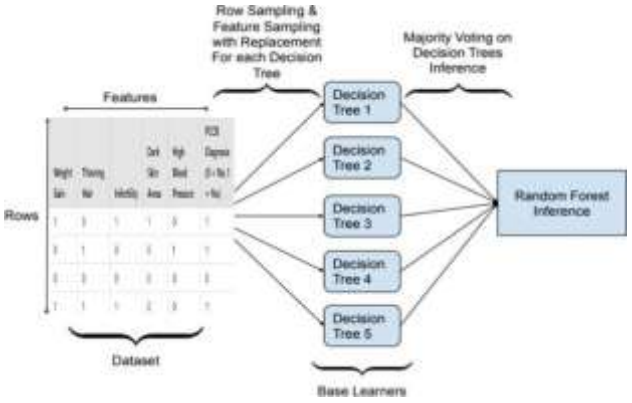

Figure 1:Idendification of PCOS using Random Forest classifier

B. Symptom based input form for Pcod using K-Nearest Neighbor classifier.

Sytom- based input form for PCOD (Polycystic Ovary Disease) using a K-Nearest Neighbors (KNN) classifier, dataset that includes patients symptom like acne, scalp hair thinning, weight gain, and difficulty in pregnancy. This dataset is processed by encoding categorical data (e.g., "Yes" or "No") into numerical values, and then splitting it into training and test sets for model evaluation, use the KNeighbors Classifier from sklearn to build the model. The user's details are processed and transformed into a numerical format, which is then fed into the trained KNN classifier to make a prediction. The prediction result, indicating whether the user might have PCOD or Normal

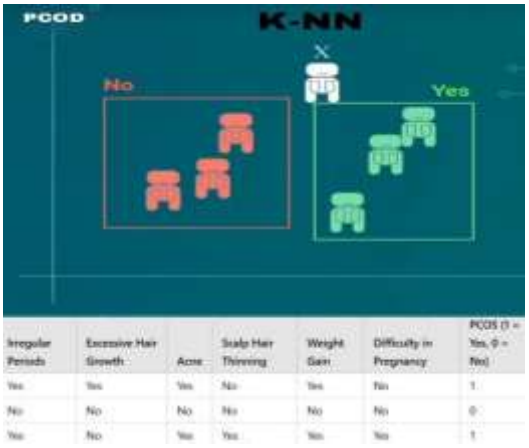

Figure 2:Identification of PCOD using K-Nearest Neighbor classifier

C. Ultrasound image processing using Convolution neural network.

Image processing for diagnosing PCOD (Polycystic Ovarian Disease) and PCOS (Polycystic Ovary Syndrome) using Convolutional Neural Networks (CNNs) it analyse medical images such as ultrasound scans or MRI images of ovaries. CNNs is a deep learning algorithm which is commonly used for image recognition

because they can automatically detect patterns, edges, and objects in images.

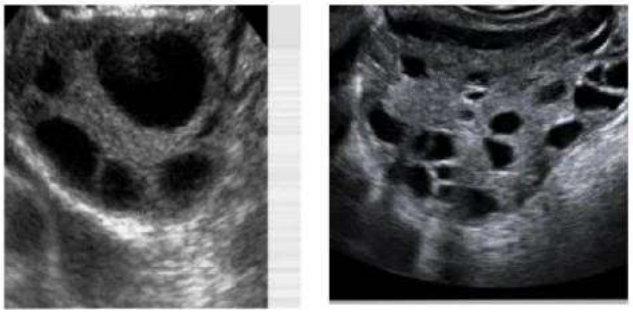

Figure 3:Infected Images of Ultrasound scans

A CNN model is constructed by consisting of multiple layers such as convolutional layers to detect patterns, pooling layers for dimension reduction, and connected layers for classification. After training the model using dataset, it can be used to make predictions on medical images, outputting the patients PCOS/PCOD based on the image features. CNNs consist of multiple layers that work together to detect features in images, it needs to be resized to a consistent shape (e.g., 224x224 pixels). This method automates the detection process, potentially assisting healthcare professionals, but it requires high-quality, annotated data and careful evaluation to ensure accuracy.

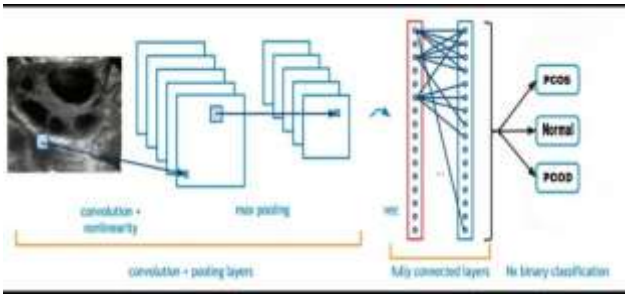

Figure 4 : Proposed CNN architecture with details layered view

The figure 4 illustrates the working of a Convolutional Neural Network (CNN), which is commonly used for image classification tasks. The process begins with a convolutional layer followed by max pooling, which helps extract important features while reducing the dimensionality of the input image. After this, a non-linearity function, such as ReLU, is applied to introduce non-linearity into the model, enabling it to learn complex patterns. The extracted features then pass through multiple convolutional and pooling layers, refining the learned representations. Finally, the processed data is fed into fully connected layers, which analyze and interpret the extracted features to make predictions. The final step involves binary classification, where the network categorizes the input into one of two possible classes. This structured approach allows CNNs to effectively recognize patterns and classify images with high accuracy.

METHODOLOGY

The system contains four main modules called Pre-processing, Model creation, System Training and Classification. Pre-processing module is used to pre-process the dataset images by resizing them to required dimension, removes inconsistencies, missing values, and outliers to ensure high-quality data for model training, normalize numerical values such as age, BMI, etc., to ensure that the model doesn't favor certain features with higher ranges, Dividing the dataset into training and testing sets. Model Creation module is used to construct a machine language model using desired number of layers. System training phase is used to train the system with dataset images and store the model weight. Classification module is used to classify the input test image to determine the image class.

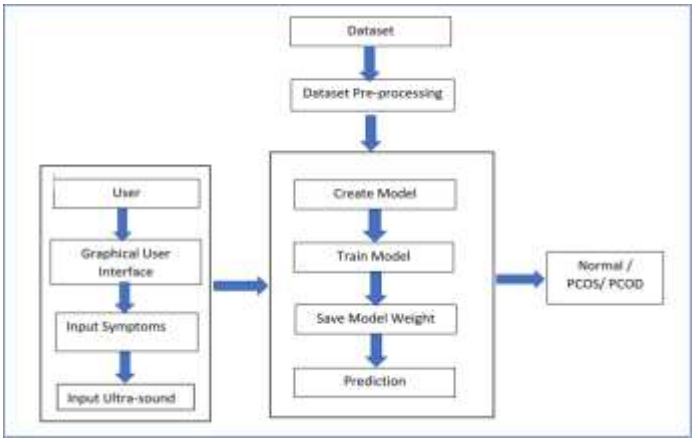

Figure 5:Architecture of proposed system

The system is designed for the diagnosis of Polycystic Ovary Syndrome (PCOS) or Polycystic Ovarian Disease (PCOD). It utilizes a machine learning approach. Dataset preparation is a critical step in building a PCOS/PCOD detection system and involves collecting and preprocessing both clinical and ultrasound image data. Clinical data includes patient information such as age, BMI, menstrual cycle history, hormonal levels (e.g., LH, FSH, testosterone, and insulin resistance), and symptoms like hirsutism, acne, and irregular periods. Ultrasound images of the ovaries are gathered to assess ovarian structure, cyst formation, and follicle size. The training and testing interface is a critical component of the PCOS/PCOD detection system, designed to facilitate efficient model development and evaluation. During the training phase, the interface processes preprocessed datasets, which include clinical and ultrasound image data, and splits them into training and validation sets. . The Convolutional Neural Network (CNN) is trained on the image data to extract key features, while clinical data is processed in parallel and integrated with the CNN features. This combined feature set is used to train the classification model, such as K-Nearest Neighbors (KNN) algorithm. The training interface provides real-time visualizations of key metrics like accuracy, and precision through graphs, enabling developers to monitor progress and identify issues like overfitting or underfitting. Once the training process is complete, the optimized model and its weights are saved for future use. After the model is trained and optimized, it is saved in a format suitable for deployment, such as a serialized file containing the model weights and architecture. The system is then integrated into a user-friendly application, typically accessed through a graphical user interface.

IMPLEMENTATION

The system implementation integrates both the Random Forest and K-

Nearest Neighbors (KNN) classifiers for robust performance. Random Forest is employed for its ability to reduce overfitting and improve accuracy by aggregating decisions from multiple decision trees. The KNN classifier complements this by using a distance-based approach to classify data points, making it effective for PCOD diagnosis by comparing new cases with the closest data points in the training set.

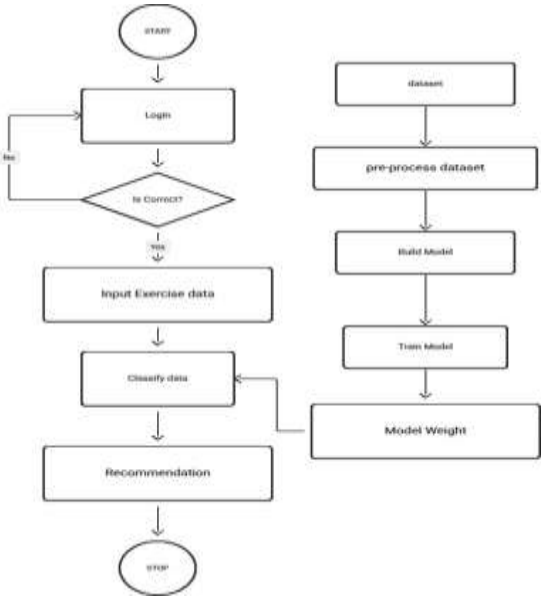

Figure 6: Flow Chart of PCOD&PCOS

The pseudo code for the proposed model is shown below:

```
Polycystic_ovarian_syndrome-old > Polycystic_ovarian_syndrome_App > train_symtoms.py > ...
1  import pandas as pd
2  import numpy as np
3  from sklearn.model_selection import train_test_split
4
5  from sklearn.ensemble import RandomForestClassifier
6
7  from sklearn.metrics import accuracy_score
8  from sklearn.metrics import confusion_matrix, accuracy_score
9  from sklearn import svm
```

The above Python code imports essential libraries for handling data, training machine learning models, evaluating their performance. Pandas and NumPy to handle structured data, such as patient records .The train_test_split function from sklearn.model_selection is used to divide the dataset into training and testing sets. RandomForestClassifier, which is an method that builds multiple decision trees and combines their outputs for better accuracy and stability. It is particularly useful in medical diagnosis applications because it can handle large datasets and provide interpretable results. It is an learning method based on multiple decision trees, is imported from sklearn.ensemble for classification tasks. Evaluation metrics such as accuracy_score and confusion_matrix from sklearn.metrics help assess the model's performance, The final import, svm from sklearn, refers to the Support Vector Machine algorithm, which is supervised learning model used for classify.

```
Polycystic_ovarian_syndrome-old > Polycystic_ovarian_syndrome_App > predict.py > ...
1  import numpy as np
2  from os import getcwd
3  from  tensorflow.keras.preprocessing import image
4  from tensorflow.keras.models import load_model
```

The getcwd() function from os gets the current folder path, which helps find files without needing to type the full path. The image module from TensorFlow Keras is used to load and prepare images so they can be used in a deep learning model. By combining these tools, the project can efficiently handle image input, process it into a suitable format, and use a pre-trained model to make accurate predictions.

RESULT

The results obtained from the implementation of various machine learning models and deep learning approaches for the identification and classification of PCOD/PCOS. The effectiveness of each model is evaluated based on key performance metrics, accuracy and a comparison is made to determine the most suitable approach.

Figure 7:Home Page

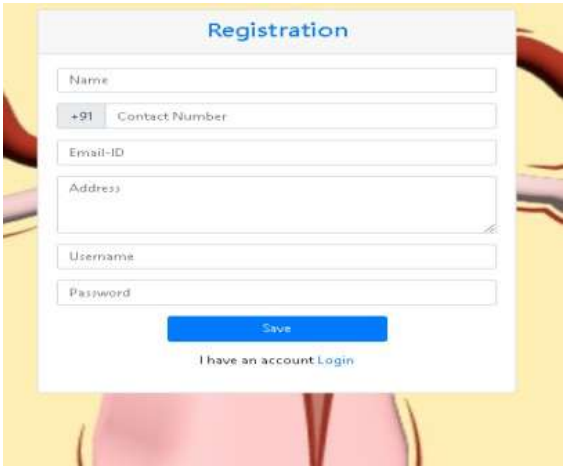

Figure 8: Registration Page

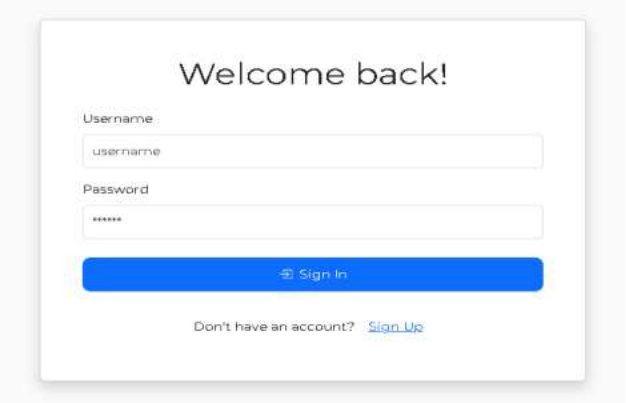

Figure 8:Login Page

Figure 9: Upload Details(PCOD)

Figure 10: Upload Image(If PCOD is Found)

Figure 11: Result Page(PCOD)

Figure 12: History Page(PCOD)

Figure 13:Upload Details(PCOS)

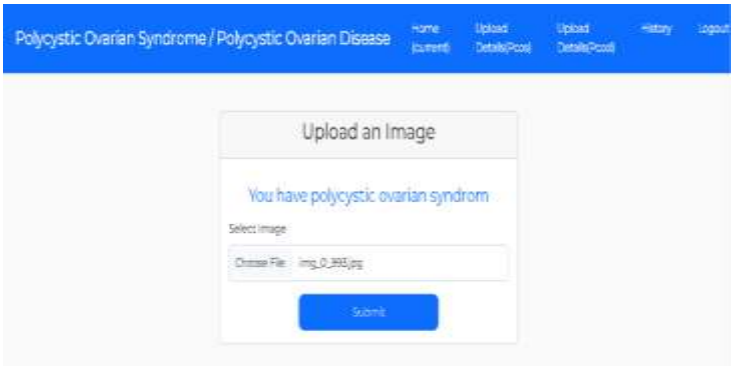

Figure 14: Upload Image(If PCOS is Found)

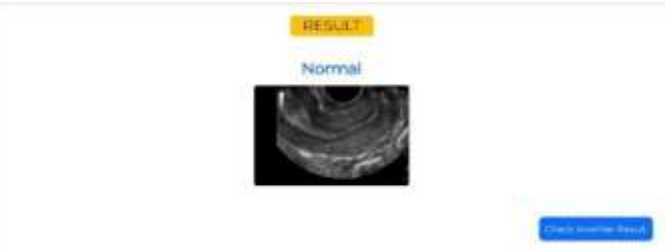

Figure 15: Result Page(PCOS)

Figure 16: History Page(PCOS)

The user interface for the PCOD and PCOS diagnosis system is designed to be user-friendly and interactive. It includes features like a home page with system introductions, a secure login/registration system, and an intuitive symptom input form for diagnosis.

CONCLUSION

This project has demonstrated the power of machine learning in healthcare by successfully identifying and classifying PCOD and PCOS with high accuracy. The use of advanced preprocessing techniques, effective feature selection, and robust classification algorithms ensured the system's reliability in diagnosing these disorders. By analyzing patient data such as hormonal levels, symptoms, and lifestyle factors, the project has provided an innovative approach to assisting medical professionals in making accurate and timely diagnoses. The findings from this project highlight the potential of artificial intelligence in addressing complex medical challenges. The system not only aids in diagnosis but also reduces the dependency on invasive tests, thereby providing a non-invasive, cost-effective solution for patients. This contributes to faster clinical decision-making and improved patient outcomes. Moreover, the project underscores the importance of explainable machine learning models in healthcare, ensuring that predictions are interpretable and actionable for medical practitioners

ACKNOWLEDGMENT

I thank our mentor Prof. Sharanya P S for analysing the data and advised on all aspects related to machine learning models, other concept and validated experimental results and reviewed the paper

REFERENCES

[1] 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.

[2] N. Nabi, S. Islam, S. A. Khushbu and A. K. M. Masum, "Machine Learning Approach: Detecting Polycystic Ovary Syndrome & It's Impact on Bangladeshi Women," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-7.

[3] S. Bharati, P. Podder and M. R. Hossain Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 1486-1489.

[4] Kermanshahchi J, Reddy AJ, Xu J, Mehrok GK, Nausheen F. Development of a Machine Learning-Based Model for Accurate Detection and Classification of Polycystic Ovary Syndrome on Pelvic Ultrasound. Cureus. 2024 Jul 22;16(7):e65134.

[5] A. Denny, A. Raj, A. Ashok, C. M. Ram and R. George, "Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 673-678.

[6] Suha SA, Islam MN. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Sci Rep. 2022 Oct 12;12(1):17123.

[7] Md Mahbubur Rahman, Ashikul Islam, Forhadul Islam ," Empowering early detection: A web-based machine learning approach for PCOS prediction 27 April 2024.

[8] Gülhan P. G., Özmen G ve Alptekin H., "CNN based determination of Polycystic Ovarian Syndrome using automatic follicle detection methods " Journal of Polytechnic, 27(4): 1589-1601, (2024).