iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49079

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

PCOS Detection: A Study of the Literature

Sayali Deodikar[1], Asma Shaikh[2], Shraddha Jadhav[3], Aishwarya Joshi[4], Sejal Mutakekar[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, MMCOE, Savitribai Phule Pune University*

Abstract: ***PCOS or Polycystic ovary syndrome is a syndrome that happens during the reproductive stage. This occurs when the women's ovaries produce a higher androgen hormone level. Many women face the issue of infertility. This problem usually occurs due to PCOS. Higher levels of androgen hormone stop the process of ovulation leading to infertility and*** *gynaecological* ***cancer. Though the rate of arising this disease is lower as compared to other diseases, it can cause serious health issues when remaining untreated. Doctors manually review the ultrasound reports and check the number of follicles and their geometrical properties. The detection process of PCOS may consume a lot of time. Technology has changed the health- care sector. Deep learning is a field of machine learning which works on a principle similar to how the human brain works. Neural networks and machine learning algorithms are one of the technologies that are used in critical disease detection. This paper focuses on reviewing the literature and gives insights into the previous methodologies for PCOS detection systems.***
Keywords: PCOS, Convolutional Neural Networks, Deep Learning, Machine Learning, Ultrasound

I. INTRODUCTION

Polycystic ovary syndrome is a disorder involving a prolonged menstrual cycle. PCOS is a condition in the female body that causes multiple sacs in the ovaries. One in five that is almost 20% of the women population suffer from this syndrome. In most women suffering from PCOS missed or irregular menstrual periods, excess hair growth, acne, infertility, and weight gain are the common symptoms. Women suffering from PCOS may be at higher risk for diseases like type 2 diabetes, depression, high blood pressure, anxiety, heart problems, and endometrial cancer.

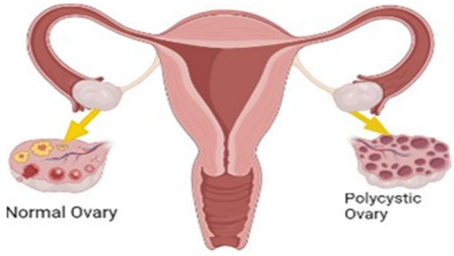


Fig 1. Polycystic ovary syndrome

The Diagnosis of PCOS is mostly done with three types- 1) Ultrasound test- This test is performed with sound waves over the ovaries to find out the size of the ovaries and if there are cysts in the ovaries. It also checks the thickness of the lining of the uterus.2) Self-diagnosis: Most women start to notice the major symptoms of PCOS like irregular periods, excess hair growth, and weight gain in the early stages, and with the help of a gynaecologist it is confirmed. 3) Blood test- In the blood tests for PCOS they look for high levels of androgens and other hormones. It is checked if the reason behind symptoms is not other diseases like thyroid. In most cases, doctors manually examine ultrasound images and conclude the affected ovary but are unable to find whether it is a simple cyst or PCOS. As cysts are very fine, doctors take time to diagnose PCOS with high accuracy. Along with that, many women suffer from conditions such as heavy bleeding, weight gain, excessive hair growth, etc. But due to the lack of knowledge, they don't understand that these are the symptoms of PCOS.

II. MOTIVATION

The exact prevalence of PCOS is not clear but 1 out of 10 women is affected, which is ranging between 2.5% to 25%. Not everyone comes up with this issue and prefers to keep this as personal suffering. They prefer keeping this as confidential information. In recent years, research has proved this to be harmful and suggests its early diagnosis. Hence, this makes it the need of an hour to make use of technology to detect PCOS and have a proper treatment to mitigate its effects and future consequences. To avoid this disorder being a base for many other complementary diseases, there is a need to diagnose this disease and gain a proper treatment.

©IJRASET: All Rights are Reserved

This may turn out to be the first successful step for completely healing this disorder. A user-friendly environment can help women as well as doctors, easily detect the results within a few seconds. This system will enable them to upload images or enter the information to get the report. The Images will be classified by Convolution Neural Networks, which are considered to be the best option in recent times. Moreover, taking user inputs (Eg: a questionnaire related to symptoms) can help a system come to a conclusion. Not only will the women be able to detect if they are suffering from the syndrome, but also they will be introduced to the nearest possible hospitals, where they can have further treatment.

III. LITERATURE REVIEW

Islam (2022) et. al [1], suggested an approach for the detection of PCOS using ovary ultra-sonography (USG) scans. An extended machine-learning classification technique for PCOS prediction has been proposed on over 594 images for training and testing purposes. In it, the CNN algorithm with feature extraction technique is used for extracting features from an image then stacking ensemble machine learning technique using conventional models as base learners and bagging and boosting ensemble model as a meta-learners was used for classifying PCOS and NON-PCOS with better accuracy and less time complexity. from that base learner model VGG-16 and from meta learner model XGBoost model as image classifier gives the highest accuracy of classification with is 99. 89%. Four different techniques are used for extracting the features in which the first technique is the traditional approach of ML training which applies relevant digital image processing, the second is chi-square and PCA, and the third technique is using the DNN algorithm and stacking ensemble training. Out of which the DNN algorithm gives better results. Performance analysis was done on the basis of accuracy, precision, recall, and F1 score.

A.K.M. Salman Hosain (2022) et al [2], calculated the accuracy of PCONet- a CNN model to classify polycystic ovarian ultrasound images and then compared it to pre-trained Incep- tionV3. The same training set and image preprocessing procedure was followed for both models. Both models were trained for thirty epochs. Steps per epoch were determined by dividing the number of images by the batch size. The batch size for training both models was 16. The PCONet showcased an accuracy of 98.12%, which was higher than the fine-tuned InceptionV3, which showed 96.56% accuracy.

Shazia Nasim and Younas (2022) et al [3], have suggested a technique for PCOS detection using a novel feature selection approach based on the chi-squared(CH-PCOS) mechanism. Using this approach, the gaussian naive Bayes (GNB) outperformed the ML model and state-of-art-studies. The GNB achieved 100% accuracy, precision, recall, and f1-score with a minimum time complexity of 0.002 seconds. The K-Fold cross-validation of GNB achieved a 100% accuracy score. The study says that the GNB model gives accurate results for the classification of PCOS on the basis of dataset features prolactin(PRL), blood pressure systolic, blood pressure diastolic, thyroid stimulating hormones, relative risk, and pregnancy are the prominent factor having high involvement in PCOS prediction to validate the overfitting of employed ML models they had applied the k-folds cross-validation techniques, the 10 folds of the dataset are used for validation. The technique shows that k-folds validation achieved 100% accuracy and the MLP models achieved 99% accuracy and by k-fold 98% accuracy had achieved. The SGD and KNC models achieved the lowest accuracy.

Subha R (2022) et al [4], have used swarm intelligence (SI) for feature selection and machine learning to develop a robust and efficient diagnostic model to detect PCOS conditions. The authors have used various methods like correlation and the Chi-Square test for optimal feature selection. They claim, having done a comparative analysis of the results and validation have done based on the parameters accuracy of training and testing, precision, recall, F1-score, and AUC-ROC. They conclude that the feature ML models with different feature selection algorithms are the best for different feature dimensions and the model with PSO-based feature selection gives the highest performance with minimum feature size.

Angela Zigarelli (2022), et al [5], have used Machine learning and deep learning techniques to analyze health data and improve diagnostic accuracy and precision, disease treatment, and prevention. The goal of their proposed study is to develop a machine-aided self-diagnostic tool that predicts the diagnosis of PCOS with and without any invasive measures, using Principal Component Analysis (PCA), k-means clustering algorithm, and CatBoost classifier. The work is well aligned with emerging artificial intelligence and digital health care. They claim to have achieved 81% to 82.5% prediction accuracy of PCOS status without any invasive measures in the patient models and achieved 87.5% to 90.1% prediction accuracy using both non-invasive and invasive predictor variables in the provider models. Their proposed prediction models are ultimately expected to serve as a convenient digital platform with which users can acquire pre- or self-diagnosis and counsel for the risk of PCOS, with or without obtaining medical test results. Their model may enable women to conveniently access the platform at home without delay before they seek further medical care. Clinical providers can also use this proposed prediction tool to help diagnose PCOS in women.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Kinjal Raut (2022) et al [6], detected PCOS using various Machine Learning Algorithms like Random Forest, Decision Tree, Support Vector Classifier(SVC), Logistic Regression, K- Nearest Neighbour(KNN), XGBoost with Random Forest (XGBRF), CatBoost Classifier, and Cross-Validation. The accuracies obtained by different algorithms are Decision Tree – 82.79%, SVC – 69.05%, Random Forest – 89.42%, Logistic Regression – 83.32%, K-Nearest Neighbors –74.34%, XBRF – 85.89%, CatBoostClassifier – 92.64%. Therefore, from the above results, the conclusion is that CatBoostClassifier has outperformed and obtained the highest accuracy. A DCNN algorithm with python programming can be a good option for easy identification of PCOS at an earlier stage.

Shubham Bhosale (2022) et al [7], used the DCNN algorithm for detecting PCOS on the basis of ultrasound images. Before applying the CNN algorithm, the data was preprocessed with image segmentation, which is used for reducing the image's noise. The univariate feature selection method was used for selecting the most suitable features. The time complexity of this algorithm is O(nˆn). The above mathematical model is NP-Complete. The space complexity depends on the presentation and visualization of discovered patterns. More the storage of data, the more the space complexity.

The proposed method by Wenqi Lv and Huang. (2022) et al [8], is composed of image preprocessing, feature extraction, and classification steps based on deep learning. Used an improved U-Net embedded with an attention module to segment the sclera from full-eye images, a Resnet18 to extract deep features, and a multi-instance learning model to classify PCOS and normal samples are made. Results show that the non-invasive screening method achieved a mean AUC of 98%, a mean accuracy of a dataset that contains 721 subjects.

The paper presented by Bhat (2022) et al [9], describes the various methods used in medical image preprocessing for the detection of PCOS. No of follicles, their size, shape, and properties are important factors that can be grasped from ultrasound images. The ultra-sound images can be noisy and can degrade the image quality. The authors analyzed the image enhancement methods such as Histogram equalization, Adaptive histogram equalization, Contrast stretching. Histogram equalization is based on the principle that for a better image, its histogram will be normally distributed. This method is well-suited for grayscale images. Adaptive histogram equalization is an enhanced version of normal HE. HE works with a single histogram of an image whereas it divides the image into several parts and constructs the histogram. In contrast stretching, the contrast values are fitted into the desired range. The authors also discussed Particle swarm optimization and hybrid principle component analysis for better classification.

Kodipalli and Devi (2021) et al [10], studied women's health in women under the age of 25. The majority of research was carried out using traditional methods like the T-square test, and the Chi-Square test. The implementation used is an approach to apply machine learning and fuzzy systems logic to the data and perform a comparative study of the two. The data collection for this study included both physical assessments (menstrual cycle, regularity of the cycle, length of the cycle, duration of the cycle, recent weight gain, hair loss, family history of having diabetes and hypertension, and eating and sleeping habits) and psychological assessments (anxiety, depression, body image dissatisfaction). For this research work the comparison matrix Mij is constructed, then the fuzzified geometric mean value is calculated, Calculate the fuzzy weights, Any defuzzification method can be used to calculate the defuzzified weights, From the weights (wi), calculate normalized weights. Then these weights (wi), calculate normalized weights. Naive bias, Decision tree, and Random forest have an accuracy of 97.65%, 96.27%, and 89.02% respectively. The presented study proved that 66.07% of women with PCOS have associated mental health issues.

In a comparative study of different denoising techniques by Shruti Bhargava Choubey (2021) et al [11], noises that are used for evaluation are standard as they give their mere presence in almost every case of imaging. The study of dual-stage filtering images for the medical field has been executed efficiently. The MSE, PNSR, and WPSNR were used to evaluate the system. The noise was analyzed with a focus on its ill effects in PCOS Images that can lead to the evaluation of diseases. The PCOS test images had some development in most of the parameters in deliberation.

Vikas B (2021) et al [12], followed the iterative process. In every iteration, accuracy is compared with the previous iteration. The models used are the basic CNN (benchmark) model, a model with more hidden layers and dropout layer, transfer learning model trained on the augmented dataset with hyperparameter tuning. The accuracy significantly improved by 10% from the initial model. The highest accuracy obtained was 94%.

VGG-19, DenseNet-121, ResNet-50, and inception V3 and model stacking, the GAN(Generative Adversarial Network) architecture is used by Kumari (2020) et al [13], to produce artificial images for better performance. The model was with VGG-19, DenseNet-121, ResNet-50, and inception V3 and model stacking, out of this highest accuracy with better sensitivity and specificity is achieved by VGG-19 i.e. approximately 70%. Due to less amount of dataset, a technique of synthetic image generator along with the data augmentation is used.

Tanwani [14] (2020) compared machine learning algorithms K-Nearest Neighbor (K-NN) and Logistic Regression. The method to find accuracy was an F1 score for both of the algorithms. The F1 score helped determine the best model between the two. The F1 score for KNN was found to be 0.90 and for that of Logistic Regression is 0.92, Therefore, the model of Logistic Regression was selected for the diagnosis of polycystic ovary syndrome detection in ovaries.

Holger H. (2018) et al [15], presented an up-to-date overview of semi-supervised learning concepts considering earlier and recent advancements in machine learning. Semi-supervised learning attempts to improve the performance of models. This technique has been proven to be best for computer-aided disease detection, part of speech tagging, and drug discovery. The authors have discussed assumptions about the data necessary for being able to apply the SSL. Further, the taxonomy and different methods of SSL are described with analogies. Pseudo-labeling is one of the algorithms that first train the model using a labeled dataset. The model with better accuracy is used to label the unlabeled dataset. The model is again trained on pseudo-labeled data to attain higher performance.

R M Dewi (2019) et al [16], focused on PCOS detection using the Gabor Wavelet Method of feature selection and competitive neural networks. Gabor filters are used to extract features directly from gray-scale images. A competitive neural network is a combination of hamming net and max net. The highest accuracy obtained was 80.84% with 32 feature vectors with a processing time of 60.64 seconds

IV. CONVOLUTIONAL NEURAL NETWORK MODELS

This section contains a brief overview of three variants of convolutional neural networks.

A. VGG16

VGG16 is a 16-layered (convolution and fully connected) setup on the Imagenet database for image recognition and classification. It has widespread use in feature learning. The input for a VGG16 model may have a fixed size of 224x224 RGB and can represent each input as a tensor with the dimensions (224, 224, 3) and label it x. X shows a pixel representation as x(I, j, k). Here, i is the first dimension, which is the location of a pixel value within the image, j is the second dimension, which is the location of a pixel value within the image and k is the third dimension, which is the channel that pixel value belongs to. The output for it is a vector probability for each of the 1000 classes for the input image. The accuracy of a model is calculated using the following formula:

$$e = \frac{1}{2}(min_i d(c_i, C_1) + min_i d(c_i, C_2))$$

B. INCEPTION V3

The Inception V3 is a popular deep learning model based on Convolutional Neural Networks, which is used for image classification. The basic model of the Inception model is made up of four parallel layers.1. 1×1 convolution, 2. 3×3 convolution, 3. 5×5 convolution 4.3×3 max pooling.

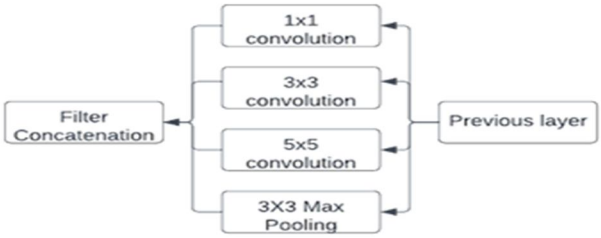


Fig 2. InceptionV3 Model

To fetch the cyst present in affected ovaries right kernel size must be used. Restnet prefers the larger kernel for information. The large convolutions are then followed by small convolutions finding the cyst in the affected ovaries. In the case of PCOS-affected ovaries, cysts are extracted from images as a feature. The accuracy of this model is up to 80% on the image dataset.

C. RESNET 50

ResNet, short form for Residual Networks, is a classic neural network. ResNet50 is a variant of the Residual Networks model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. Deep residual networks like the popular ResNet-50 is a convolutional neural network (CNN) that has 50 layers. Restnet network architecture is inspired by VGG-19. It simultaneously finds an optimized number of layers with features. The result shows that the model can achieve up to 92% accuracy and it outperforms most of the existing neural network models

V. DATASET DESCRIPTION

The dataset is collected from the kaggle. It is the freely available dataset having ultrasound images. The dataset comprises two classes of images - 'Infected and Non-Infected'. 1,562 images for the infected class and 2,284 images for Non Infected class. For each class, the images are divided into training, testing, and validation sets. The figures show the difference between the two classes. An image with PCOS has a net-like structure which is blood cells grouping or the follicles and the normal ovary

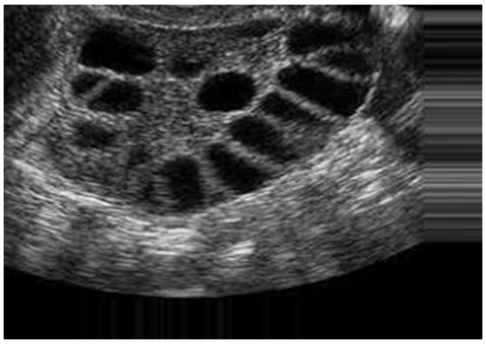


Fig 3. PCOS Ultrasound

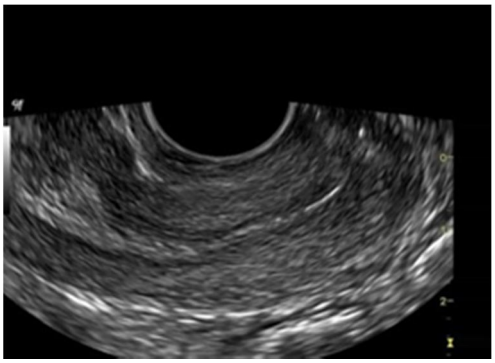


Fig 4. Non-PCOS Ultrasound

images do not have a net-like structure or have less number of follicles. For the second module, records from the dataset are collected by floating the form. Till now it contains 167 responses. The form contained 23 questions and answers of which will result in various features. The features are age, marital status, body description, no. of kids, menstrual cycle duration, pain during the menstrual cycle, etc. Based on the features of the previous response, the new response gets classified into PCOS and Non - PCOS.

VI. CONCLUSION

Research and technology have been playing an important role when it comes to medical imaging and disease diagnosis.PCOS is one of the syndromes that puts women's life at high risk. Many techniques are used for detecting PCOS. In this paper, we focused on some relevant research papers and journals for our project. This study includes various approaches to medical image classification, different algorithms used for disease detection, and an overview of convolutional neural network models. The performance of the model depends on the preprocessing of data, feature selection technique, and selection of hyper-parameters. As a result, we can claim that machine learning and deep learning can be used for the accurate detection of PCOS. The presented study will help us for the advancement of our project titled design and development of a system for PCOS detection.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

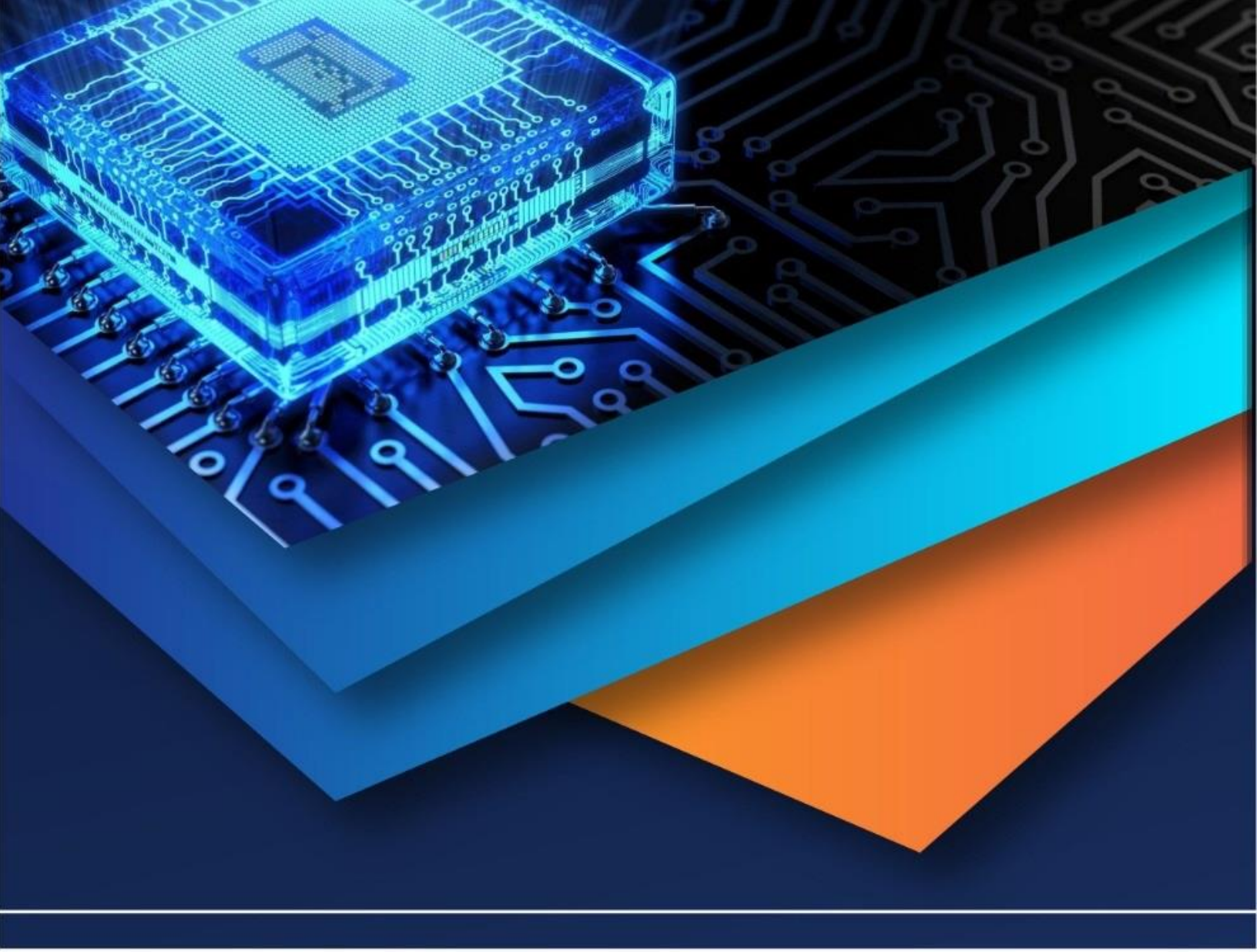
VII. FUTURE WORK

The detection of PCOS has now become important. Hence, the ease to diagnose this dis- order also plays a very vital role. Making this system available only as a website could restrict its scope. Nowadays, not only websites but also android applications have gained huge importance, due to the people's (younger as well as the older generation's) preference towards cell phones. People may prefer using an android application over a website. There- fore, developing an android application would complement spreading awareness and getting quick diagnoses and speedy results.

Getting a disorder diagnosed freely through an application or a website is possible through this project. One can get the recommendation of the nearby hospitals as well, which can help them diagnose the syndrome. As an additional contribution to this, if the doctors and experts support the project by suggesting treatments online for the minor diseased individual, this would help the individual heal the syndrome using home remedies and at a lower budget.

REFERENCES

[1] Sayma Alam Suha Muhammad Nazrul Islam. "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image", Scientific Reports, 2022
[2] Irteza Enan Kabir A.K.M. Salman Hosain, Md Humaion Kabir Mehedi. "PCOnet: A convolutional neural network architecture to detect polycystic ovary syndrome (PCOS) from ovarian ultrasound images" 2022.
[3] Kashif Munir Ali Raza Shazia Nasim, Mubarak Almutairi, and Faizan Younas. "A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics", 10, 2022.
[4] Rekha Radhakrishnan Sumalatha P. Subha R, Nayana B R. "Computerized diagnosis of polycystic ovary syndrome using machine learning and swarm intelligence technique", 2022.
[5] Hyunsun Lee. Angela Zigarelli, Ziyang Jia. "Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: Observational study" 2022
[6] Prof. Dr. Mrs. Suhasini A. Itkar Kinjal Raut, Chaitrali Katkar. "PCOS detect using machine learning algorithm", 09, 2022.
[7] Arun Shivsharan. Shubham Bhosale, Lalit Joshi. "Pcos (polycystic ovarian syndrome) detection using deep learning" 04, 2022.
[8] Rongxin Fu Xue Lin Ya Su Xiangyu Jin Han Yang Xiaohui Shan Wenli Du Qin Huang Hao Zhong Kai Jiang Zhi Zhang Lina Wang Wenqi Lv, Ying Song and Guoliang Huang. "A deep learning algorithm for automated detection of polycystic ovary syndrome using scleral images", 2022.
[9] Siji Jose Pulluparambil 1 Subrahmanya Bhat. "Medical image processing: Detection and prediction of PCOS – a systematic literature review". 5, 2022. ISSN 2581-6411
[10] Ashwini Kodipalli and Susheela Devi. "Prediction of PCOS and mental health using fuzzy inference and SVM" 9, 2021.
[11] Durgesh Nandan Anurag Mahajan. Shruti Bhargava Choubey, Abhishek Choubey. "Polycystic ovarian syndrome detection by using two-stage image denoising", 38, 2021.
[12] Vineesha K. Vikas B, Radhika Y. Detection of polycystic ovarian syndrome using convolutional neural networks. 13, 2021.
[13] Sweta Kumari. "Classification of PCOS/PCOD using transfer learning and gan architectures to generate pseudo ultrasound images" 2020.
[14] Namrata Tanwani. "Detecting PCOS using machine learning", 07, 2020. ISSN 2348-3121.
[15] Holger H. Hoos Jesper E. van Engelen1. "A survey on semi-supervised learning",2019
[16] U N Wisesty, Jondri R M Dewi, Adiwijaya. "Classification of polycystic ovary based on ultrasound images using competitive neural networks", 2018



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)