# A Systematic Literature Review on Diagnosis of PCOS using Machine Learning Algorithms

Jagruti Vaswani[a]*, Harsha Mulchandani[b], Rahul Vaghela[c], Dr.Rajan Patel[d]

*[a][b]M.E. Student, Gandhinagar Institute Of Technology, Moti bhoyan, Gandhinagar, Gujarat,382721, India*
*[c]Assistnat Professor, IT, Gandhinagar Institute Of Technology, Moti bhoyan, Gandhinagar,Gujarat, 382721, India*
*[d]Professor, CE, Gandhinagar Institute Of Technology, Moti bhoyan, Gandhinagar,Gujarat, 382721, India*

**Abstract**

The healthcare industry is very different from other industries. This industry is crucial and people expect the most important level of care and services. In terms of medical diagnostics by human specialists, it is very limited due to its subtlety, the complexity of the disease itself, and the variety that exists in all different interpretations. Machine Learning provides exciting and highly accurate solutions to medical diagnosis and is considered a critical alternative to future applications in the healthcare environment. The paper provides a review of the research done on diagnosis of diseases using different ML architectures and diagnosis of PCOS. Based on the review and its observation, research gaps are found.

*Keywords*: Medical Diagnosis, PCOS, Machine learning, Deep learning, CNN, Faster-RCNN .

## 1. Introduction

Diagnosis plays a significant role for not only patient care but research as well. If the diagnosis is accurate and timely, the patient has a better chance of getting a better health result because clinical decisions will be made to better understand the patient's health problem. Smart programs are increasingly being used in medicine and health care, but there is still a need for a robust and purposeful approach. Machine learning has shown great potential in targeting medical experts or patients. ML provides methods, techniques, and tools that can help solve diagnostic and predictive problems in a variety of medical fields.

It is argued that the successful implementation of ML methods could help the integration of computer-based programs into the healthcare environment that provide opportunities to assist and improve the work of medical professionals and ultimately improve the efficiency and quality of medical care. The Diagnosis can be made more accurate and the system can be more simplified with the help of Machine learning Algorithm like deep learning using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN).These methods can be widely used to model or simulate an intelligent system or process using annotated training data.

## 2. Background

- In terms of health care, A disease can be defined as a health condition that has a clearly defined reason behind it. A syndrome (from the Greek word meaning 'run together') however, may produce a number of symptoms without an identifiable cause. A syndrome refers to a group of symptoms, while a disease refers to an established condition.
- Medical conditions are often categorized as acute or chronic. Acute illnesses generally develop suddenly and last a short time, often only a few days or weeks. Chronic conditions develop slowly and may worsen over an extended period of time—months to years.
- One such type of illness is PCOS; it is a chronic disease with manifestations across the lifespan. Polycystic ovary syndrome (PCOS) is the most common endocrine disorder with a prevalence of 5~10 % among women with reproductive age and among them 70% remain undiagnosed. PCOS condition can be treated to some extent by controlled medication and bringing alterations in lifestyle.
- A polycystic ovary (PCO) can be characterized by twelve or more follicles with a diameter of 2-9 mm this causes the development of enlarged ovaries. PCOS affects both health and the quality of women's life. One place in particular, where machine learning has a widespread effect on the community is in the area of health care. In a growing industry of smart watches, fit bits, and devices that constantly gather a plethora of health data, the prevalence of using machine

---

* Jagruti Vaswani
*E-mail address: jiyavasvani2202@gmail.com*

learning to analyze this data is gaining momentum. Machine learning has many applications in health care that involve diagnosis, image recognition, identification and prediction of data, etc.



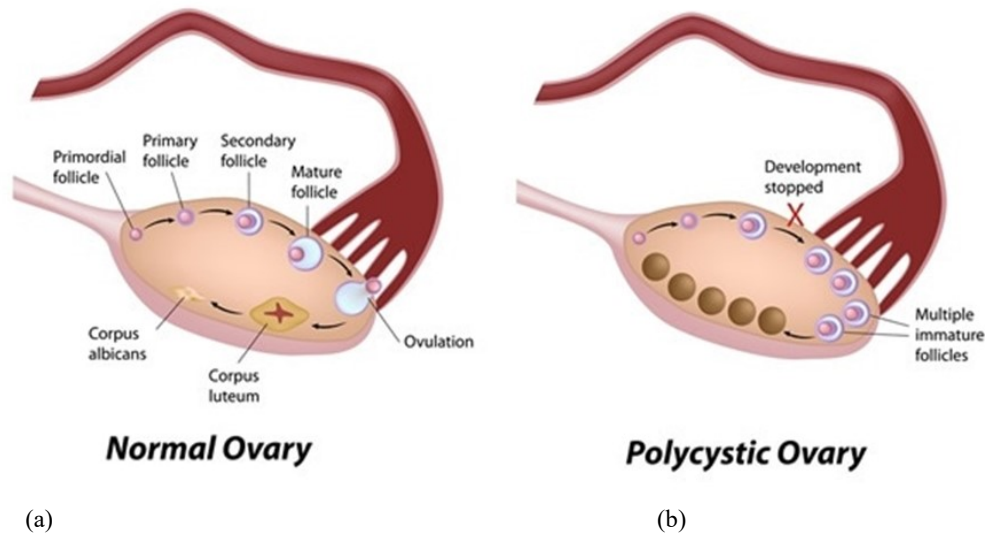(a)                                          (b)

Fig. 1. Illustration of Ovaries for (a) Normal Ovary and (b) Polycystic Ovary

- The symptoms include cardiovascular diseases, infertility, type 2 diabetes, acne, baldness, hair loss, obesity, anxiety, depression, and stress. The early diagnosis and treatment can be used to control based on the symptoms and by the prevention of long-term problems. PCOS can be detected through ultrasound and sonography by a doctor by reckoning the number and size of follicles situated in the ovaries. However, this process takes a protracted interval, and needs good image quality and high accuracy to detect the presence of PCOS.

- Machine learning (ML) classification and feature selection algorithms have been used by researchers and clinicians for the prediction of diseases as a non-invasive method. PCOS datasets which consist of heterogeneous attributes related to biochemical, clinical, medical history, symptoms of the patients and ultrasound images are used to build predictive models.

## 3. Literature Review

### 3.1    *Medical diagnosis using Deep Learning*

- *Deep Learning in Medical Image Analysis by Lee Gobert, Fujita Hiroshi- 2020*

  The paper is divided into 3 parts. Part 1 explains issues and efforts needed to develop robust deep-learning-based CAD tools and integrate these tools into the clinical workflow. Also introduced medical image synthesis methods, mainly focusing on CNNs and GANs models. Part 2 compares the performance of deep learning algorithms for classification, Detection and segmentation on Images. Part 3 explains their proposed Computer-Aided Diagnosis system for Breast Lesion in Digital Mammogram based on deep learning detection, segmentation, and classification.

- *Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review by Essam H. Houssein, Marwa M. Emam, Abdelmgeid A. Ali and Ponnuthurai Suganthan- Oct 2020*

  In this study, authors review the latest studies focused on the detection and classification of breast cancer. The review includes SVM, DT, Nearest Neighbor, Naive Bayesian Network, and ANN. It also focused on the Convolutional Neural Network and its Deep Learning architectures used to detect and classify breast cancer from different image modalities. This review provides a description of the medical imaging as well; Mammograms, Ultrasound, MRI.

### 3.2    *Diagnosis of PCOS using Machine Learning*

- An Efficient SMOTE Based Machine Learning classification for Prediction & Detection of PCOS by Pijush Dutta, Shobhandeb Paul and Madhurima Majumder – Nov 2021.

In this research a novel prediction model using Synthetic Minority Oversampling Technique (SMOTE) with five machine learning algorithms like Logistic Regression, Random Forest, Decision Tree Support vector machine and K-Nearest Neighbor (KNN) model which automates the PCOS detection in an early stage with higher degree of efficiency.In terms of execution time, SMOTE based Random Forest has taken remarkably less time with 0.10 seconds and both SMOTE based SVM & KNN has taken maximum area under ROC. The overall dataset contains 178 instances of the positive class (1) and 363 instances of the negative class (0). Unequal number of positive classes and negative classes within the dataset is one of the major causes of waning exactness of classification models.

- Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms by Subrato Bharati, Prajoy Podder and M. Rubaiyat Hossain Mondal – June 2020.

  It can be seen that the ratio of Follicle-stimulating hormone (FSH) and Luteinizing hormone (LH) Indicated as the best ranking feature. It is shown in this paper that gradient boosting, random forest, logistic regression and RFLR exhibit good accuracy and recall values. RFLR has the best testing accuracy of 91.01%.

- Detection of Polycystic Ovarian Syndrome Using Follicle Recognition Technique by Rachana B, Priyanka T, Sahana K N, Supritha T R, Parameshachari B D, Sunitha R-July 2021.

  In this paper the approach taken to detect PCOS is using ultrasound images studying various features by combining segmentation, feature extraction and classification process. The various approaches for segmentation and classification were studied to enhance the past research to obtain a model with greater accuracy. For classification KNN method is used.

- Study and detection of PCOS related diseases using CNN by M Sumathi, P Chitra, R Sakthi Prabha and Srilatha K – Dec 2020.

  In this paper, CNN is used as an image classifier, by segmentation and feature extraction methods algorithm is capable of detecting cysts in the dataset. This process uses some input ultrasound images as train data and with their reference, it will classify test data in the dataset to know whether the ovary is affected and the parameters like area, solidity, extent, perimeter where exactly affected.

- Polycystic Ovarian Syndrome (PCOS) Classification and Feature Selection by Machine Learning Techniques by Satish C. R Nandipati, Chew XinYing and Khaw Khai Wah– Dec 2020.

  Radip Miner and Python are used to perform feature selection on various features like BMI: body mass index, hair growth and comparison between various algorithms like KNN, SVM, Random Forest, Naïve Bayes and Neural Network.The performances of each classifier and average performances show that RapidMiner can be used as an alternative machine learning tool. However, this cannot be a general rule since the performances depend on the nature of the dataset, sampling, and pre-processing steps.

3.3     *Comparison of Object detection algorithms*

- Some object detection algorithms are YOLO (You only look once), R-CNN, Fast R-CNN.
- YOLO is orders of magnitude faster (45 frames per second) than other object detection algorithms. The limitation of the YOLO algorithm is that it struggles with small objects within the image.
- R-CNN still takes a huge amount of time to train the network as one would have to classify 2000 region proposals per image.
- Both of the above algorithms (R-CNN & Fast R-CNN) use selective search to find out the region proposals. Selective search is a slow and time-consuming process affecting the performance of the network.
- Selective Search is a region proposal algorithm. It is based on computing hierarchical grouping of similar regions based on color, texture, size and shape compatibility.
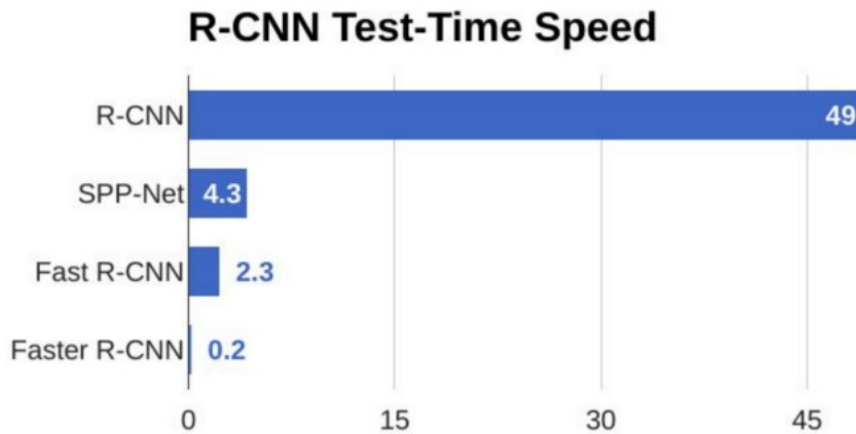
Fig. 2. Time speed graph of object detection algorithms

- Faster R-CNN is an object detection algorithm that eliminates the selective search algorithm and lets the network learn the region proposals.
- The image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using a selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals.
- The predicted region proposals are then reshaped using a RoI pooling layer which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes.
- Faster R-CNN is much faster than its predecessors. Therefore, it can even be used for real-time object detection.

## 4. Observations

- CAD by the use of R-CNN showed high performance for detection.
- In image-based CADs by the use of CNN, classification of lung abnormalities was superior to feature-based CADx
- Analysis of types of images used for medical diagnosis.
- CNN used with a layer of faster R-CNN improved the accuracy in feature extraction.
- In term of execution time, SMOTE based Random Forest has taken remarkably less time with 0.10 seconds and both SMOTE based SVM & KNN has taken maximum area under ROC.
- Image processing before classification can improve the accuracy and performance of the model.
- The principle to detect PCOS is to find if the ovaries have a volume greater than or equal to 10 cm3, or having follicles around 10 or more in numbers with each 2-9 mm diameter.
- KNN provides accuracy as good as 97% but takes up a lot of computational time.
- Performance can be increased and time complexity can be reduced by an algorithm which combines image pre-processing and 'feature selection process' and only the selected features are trained.

## 5. Conclusion

- Repositories like kegal can be used to obtain a data set of medical images to train and test the model.
- Faster R-CNN has not been used yet for the diagnosis of PCOS.
- Deep learning models are faster than other models in the field of medical diagnosis.
- Other object detection algorithms can also be used for detection of cysts in ovaries such as YOLO, CNN , SPP-Net, R-CNN and fast R-CNN.
- Faster R-CNN eliminates many limitations of its contemporary object detection algorithms.
- Faster R-CNN has the tendency to detect smaller objects which will lead to more accurate and timely diagnosis of PCOS.

**References**

1.  Subrato Bharati, Prajoy Podder and M. Rubaiyat Hossain Mondal "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms". In 2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh.
2.  B Rachana , T Priyanka , K N Sahana , T R Supritha , B ParameshachariD and R Sunitha "Detection of polycystic ovarian syndrome using follicle recognition technique". In Global Transitions Proceedings Volume 2, Issue 2, November 2021, Pages 304-308.
3.  Satish C. R Nandipati, Chew XinYing and Khaw Khai Wah2 "Polycystic Ovarian Syndrome (PCOS) Classification and Feature Selection by Machine Learning Techniques". In Applied Mathematics and Computational Intelligence Volume 9, Dec 2020 [65-74]
4.  M Sumathi, P Chitra, R Sakthi Prabha and Srilatha K "Study and detection of PCOS related diseases using CNN". In IOP Conference Series: Materials Science and Engineering, Volume 1070, International Conference on Recent Innovations in Engineering and Technology (ICRIET 2020) 4TH-5TH December 2020, Tamil Nadu, India.
5.  Pijush Dutta, Shobhandeb Paul and Madhurima Majumder "An Efficient SMOTE Based Machine Learning classification for Prediction & Detection of PCOS". In Research Square – November 8th, 2021.
6.  Amsy Denny, Anita Raj, Ashi Ashok, C Maneesh Ram and Remya George "i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques". In TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).
7.  Xinyi Zhang, Bo Liang, Jun Zhang, Xinyao Hao, Xiaoyan Xu, Hsun-Ming Chang, Peter C.K. Leung and Jichun Tan "Raman spectroscopy of follicular fluid and plasma with machine-learning algorithms for polycystic ovary syndrome screening". In Molecular and Cellular Endocrinology Volume 523, 1 March 2021, 111139.
8.  Chi-Hong Ho , Chia-Ming Chang , Hsin-Yang Li , Heng-Yi Shen , Fu-Kong Lieu and Paulus Shyi-Gang Wang "Dysregulated immunological and metabolic functions discovered by a polygenic integrative analysis for polycystic ovary syndrome". In Reproductive BioMedicine Online Volume 40, Issue 1, January 2020, Pages 160-167.
9.  Bhat Shakoor Ahmad (2021) "Detection of Polycystic Ovary Syndrome using Machine Learning Algorithms". Masters thesis, Dublin, National College of Ireland.
10. Dr.V.Krishnaveni "A ROADMAP TO A CLINICAL PREDICTION MODEL WITH COMPUTATIONAL INTELLIGENCE FOR PCOS". In Positive Journal Vol 14, Issue 5, 2021, Pages 29-37.
11. Lee Gobert, FujitaHiroshi (2020). [Advances in Experimental Medicine and Biology] Deep Learning in Medical Image Analysis Volume 1213 (Challenges and Applications).
12. Essam H. Houssein, Marwa M. Emam, Abdelmgeid A. Ali and  Ponnuthurai Suganthan  "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review".
13. Malik Haris and  Adam Glowacz (2021) "Road Object Detection: A Comparative Study of Deep Learning-Based Algorithms" Ashish Kumar (2020) "Artificial Intelligence In Object Detection - Report".