

PAPER • OPEN ACCESS

## Study and detection of PCOS related diseases using CNN

To cite this article: M Sumathi *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1070** 012062

View the [article online](#) for updates and enhancements.

### You may also like

- [An implementation of Elman neural network for polycystic ovary classification based on ultrasound images](#)  
I F Thufailah, Adiwijaya, U N Wisesty et al.
- [Ion Conduction and Electrochemical Performance of Poly\(oxetane\)-Based Electrolytes with Tri\(cyanoethoxymethyl\) Moiety As a Side Chain](#)  
Yohei Nakano and Hiromori Tsutsumi
- [Correlation between anti-Müllerian hormone serum level and Bax/Bcl-2 mRNA expression ratio from granulosa cells in patients with PCOS](#)  
B Wiweko, A Beelonie, Asmarinah et al.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

## 241st ECS Meeting

Vancouver, BC, Canada. May 29 – June 2, 2022



ECS Plenary Lecture featuring  
**Prof. Jeff Dahn,**  
**Dalhousie University**



Register now!



# Study and detection of PCOS related diseases using CNN

M Sumathi, P Chitra, R Sakthi Prabha and Srilatha K

Sathyabama Institute of Science and Technology, Chennai

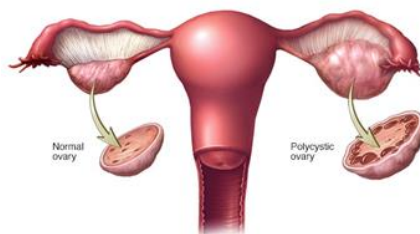
*sumathi.ece@sathyabama.ac.in, chitraperumal@gmail.com*

**Abstract.** Polycystic ovary syndrome (PCOS) is a group of symptoms caused by high levels of androgens in women. The cause of PCOS is a group of genetic and environmental factors that are common pathologies, often associated with clinical symptoms of arteries, hirsutism, acne, and hyperandrogenism, along with chronic infertility. Recent studies show that about 18% of Indian women suffer from this syndrome. Doctors were manually examining ultrasound images and conclude the affected ovary but unable to find whether it is a simple cyst, PCOS, or cancer cyst. In this paper, CNN based algorithms proposed and coding developed in Python programming for classification of cysts, and they are filled with blood or fluid using ultrasound images. The study is performed on CNN based image processing feature extraction to classify cysts in the dataset. That is the study is carried out using an independent trained dataset of the same PCOS related diseases. Finally, the test dataset is used for performing the feature extraction process and the results are met with 85% accuracy using performance factors.

**Keywords:** CNN, PCOS related diseases, Deep learning, Medical image processing.

## 1. Introduction

The term Polycystic Ovarian Disease (PCOD) was first defined by Michael Leventhal and Irving Stein as a Triple term of 'Obesity', 'Hirsutism', and 'Amenorrhea'. In 1935 they realized the relationships between reproductive disorders and obesity. Hence it is also called as the 'Hyperandrogenic Anovulation' (HA) or 'Stein-Leventhal Syndrome'[13]. It is the most popular disorder of the endocrine system in the ovaries, affecting approximately 2-8% of women giving birth worldwide[14]. Now it is also called "Syndrome O", i.e. Over nourishment, Over insulin production, Ovarian confusion, and Ovulation disorder, so PCOD is also called polycystic ovary syndrome (PCOS). Figure 1 displays PCOS in the ovary.



**Figure 1.** Poly Cystic Ovary vs Normal Ovary.

Nidhi R. et al conducted a review on 460 girls aged about 15 to 18 years at the Andhra Pradesh residential university. And the review reported that about 9.13% prevalence among adolescents in India[15]. Although it is a universal disease, its symptoms are difficult to diagnose. Whereas



syndrome can't be judged with a single symptom as it has multiple aetiologies and presentations. PCOS is diagnosed by gynecologists and good and better understanding required analyzing the disorder with a bunch of symptoms. The calculation of PCOS prevalence depends on the parameters used to determine this syndrome. Because the PCOS symptoms emerge surreptitiously and correlate with normal puberty progression, ingenious features may not be examined in the early stages. This may be due to the inability to identify the young girl's disorder.

## 2. Related Works

Considerable research work is done in the detection of Poly Cystic Ovarian Syndrome related diseases[4-12]. Deep learning algorithms were performed in this area. Convolutional neural networks, classification, image segmentation, and feature extraction are performed.

Joseph, Carolina, and Arnauet al., proposed a study on the application of a convolutional neural network (CNN) to segment medical X-rays suitable for small data sets [2]. Its application provides an overall accuracy of 92%, F1 0.92, and AUC 0.98, bypassing traditional image processing techniques such as aggregation and entropy methods while preserving the existing neuron networks used for segmentation in non-medical environments. They developed a fully automatic method for segmenting medical radiograms using a small set of data. They have also presented a two-module decoding architecture as a compromise between a deep network for extracting high-level features and subtle details and a network that avoids excessive transformation into small data sets.

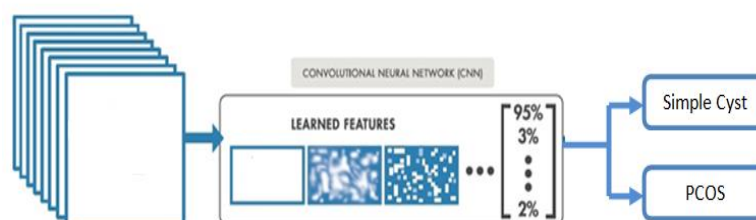
S. Khazendar, et. al., a proposed automatic model which characterize the ultrasound ovarian tumors images [10] using support vector machine and local binary pattern operator. The input images are first pre-processed and for each 2 x 2 blocks of the input image, the Local Binary Pattern Histograms are extracted. A Support Vector Machine (SVM) was used for training by using cross-validation with random sampling. This procedure was repeated 15 times and 100 images were randomly selected in each round. If the training performed on multiple masses improves test performance and this test is maintained during external validation, this machine intelligence can be incorporated into a program of ultrasound instruments that describes ovarian mass characterization.

Pim, Jelmer M., was carried out the deep learning research on multitasking segmentation of medical images with multiple methods [4]. The CNN network was trained to divide six tissues on the MRI image of the brain, the pectoral muscle in MR breast image, and the coronary artery in CTA of the heart. The results show that for training the CNNs, a single CNN structure can be used that can obtain accurate results for images from different methods and show different anatomical structures. Furthermore, it is possible to train a single unit of CNN that cannot segment multiple tissue categories into a single anatomical structure, but also multiple structures.

However, the algorithm for classification of cysts may be implemented which are not suitable for all kinds of cysts. This classification is based on segmentation, feature extraction using the CNN algorithm which is an emerging and useful technique in the medical field. The paper is organized as follows. Chapter 1 deals with the Introduction followed by related works in Chapter II. Chapter III discusses the proposed methodology with results and discussion in Chapter IV. Finally, the efficiency of the detection is expressed in Chapter V.

## 3. Proposed Methodology

The proposed work is composed of an efficient model of CNN based image processing for the classification of diseases. The below figure 2 shows the basics of CNN.



**Figure 2.** Process of CNN

The watershed algorithm is used for the extraction of features and OpenCV is used for measuring parameters like area, perimeter, extent, solidity, orientation. In CNN, code for train model, the test model is built, and a convolution model which has several layers in it. For a given dataset (consists of both train data and test data) of two classes, train data is used to train the model and test data is used to test the results of the trained model. Train data is used to train models, and test data is used to test train model results. An image is given as input to the CNN model and its resolution is decreased using kernel (convolution matrix) given in equation.1.

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (1)$$

Where  $f$  is the input image in the form of a matrix,  $g$  is the kernel matrix or convolution matrix. Each time an image is sent through a convolution layer, the image is desolated and for each layer, one feature is extracted. To get noise removed from the image, the morphological transformations such as erosion, dilation, etc. are used in the process of feature extraction. These transformations are simple image-dependent operations. Usually, a grayscale image is used on which these operations can be performed. The erosion process breaks down the boundaries of the foreground. It performs a 2D convolution with the kernel. Then consider a characteristic pixel image (original image) and if each pixel is below kernel are 1, it is eroded differently. Therefore, the thickness of the front body is thin. Dilation does exactly opposite operation of erosion. The erosion process erodes the white noises but narrows the object. As the noise is removed, the surface of the object is increased. So the lost area of the object due to erosion is undone by dilation. Therefore, usually, the erosion process is followed by dilation. OpenCV module is used for measuring the parameters like area, perimeter, extent, solidity, orientation. Using a morphological gradient, an image is simplified, which is the dissimilarity between erosion and dilation of an image. For the morphologically enhanced image, it is possible to calculate the contours and find properties using the basic formulae that represented in equations. (2) to (4).

$$\text{Aspect Ratio} = \text{Width/Height} \quad (2)$$

$$\text{Extent} = A_0/A_{ba} \quad (3)$$

where,  $A_0$  is Object Area

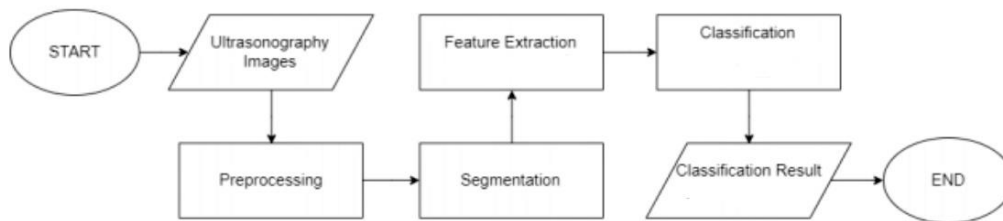
$A_{ba}$  is Bounding Rectangle Area.

$$\text{Solidity} = A_{ca}/A_{cha} \quad (4)$$

Where  $A_{ca}$  is Contour Area

$A_{cha}$  is Convex Hull Area

The workflow of the proposed methodology is presented in figure 3.



**Figure 3.** Overall Process of Proposed Methodology.

The step by step process of the proposed methodology is as follows,

Step 1. Data collection. In this work, ovarian ultrasound images are collected.

Step 2. Next is preprocessing. As ultrasound images are mainly affected by speckle noise, it is necessary to remove the speckle noise.

Step 3. After preprocessing, the region of interest i.e cyst portion is segmented using a watershed algorithm.

Step 4. Once the cyst is segmented, features are extracted which is necessary for the classification process.

Step 5. Using the extracted features, the CNN model is trained using training ultrasound images.

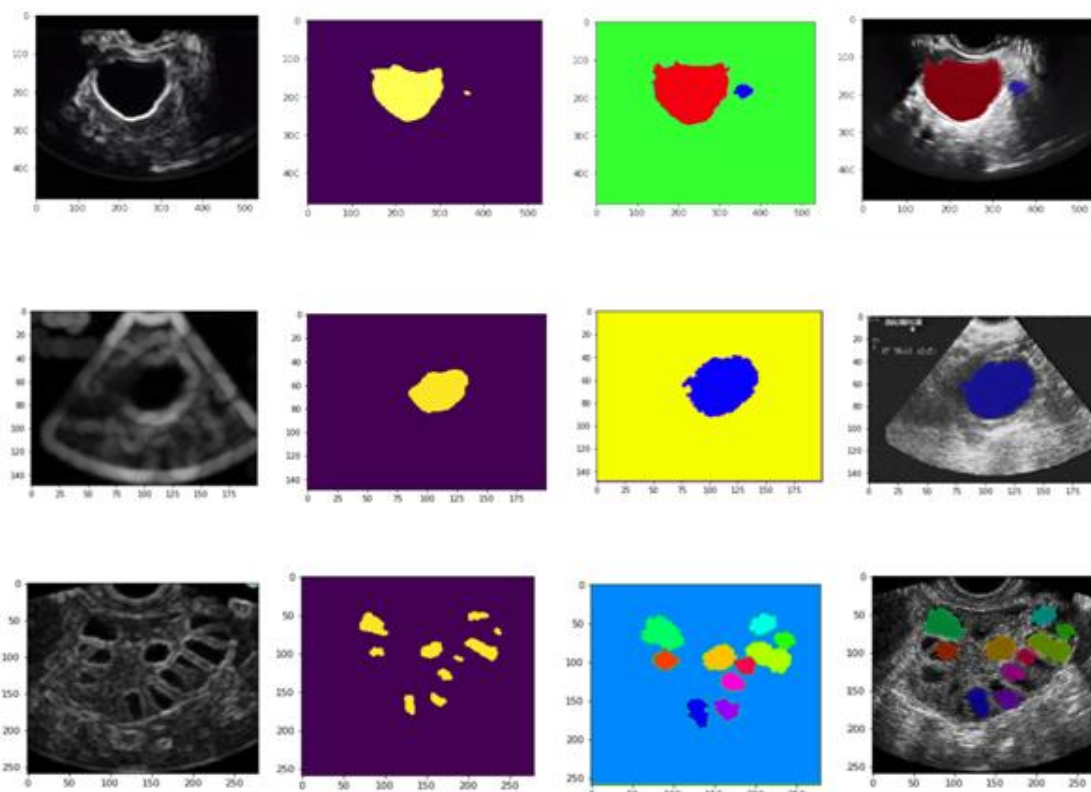
Step 6. To validate the model, the test images are applied to classify the cyst.

The performance of the proposed methodology is explained in the next section.

#### 4. Results and Discussion

The complete research work is carried out on Anaconda Navigator, Jupyter Notebook as a platform using Intel Core i5@2GHz processor. The cyst in an ovary is first segmented using a watershed segmentation algorithm that segments the cyst very accurately. The following Figure 4.1 shows the performance of the watershed algorithm.


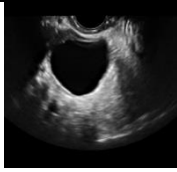



After the segmentation process, the most needed parameters area, perimeter, aspect ratio, extent, and solidity are measured to assist the clinicians. Using these parameters, the clinicians can able to identify



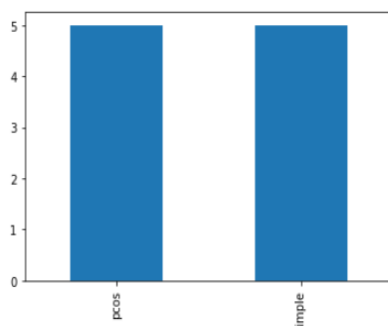
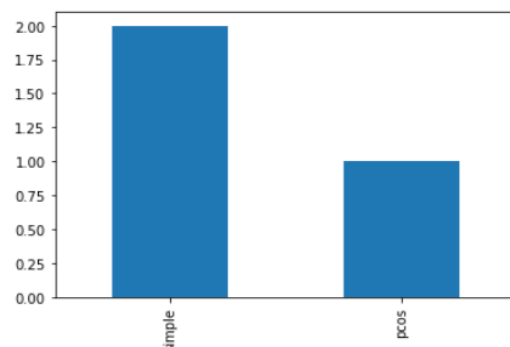
**Figure 4.** Segmentation of cyst using the watershed algorithm

the type of cyst, location, and size easily. Table 1 gives information about the cyst.

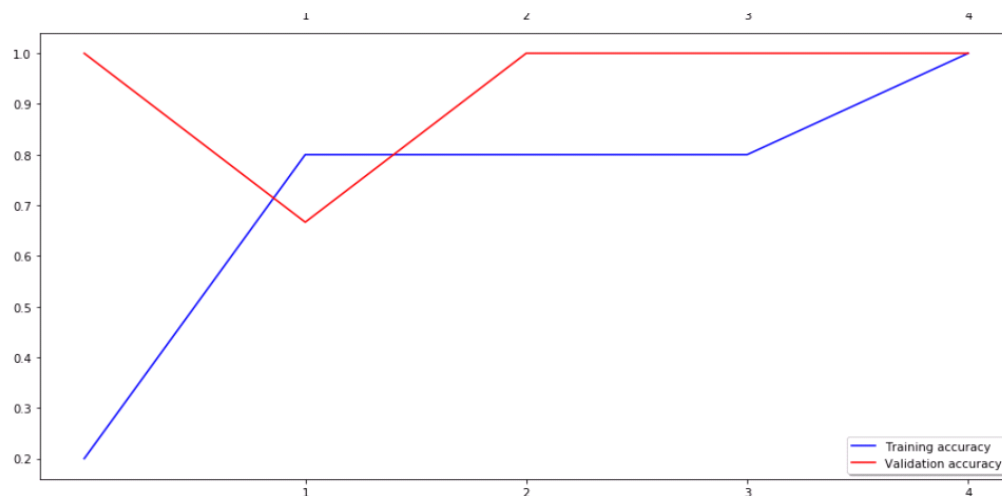
**Table 1.** Dimensions of the cysts.

S.no.	Input Image	Kernel	Area(sq.cm)	Perimeter(cm)	Aspect Ratio	Extent	Solidity
1		(3,3)	4.524	1.3920	1.0	0.8622	1.25
2		(11,11)	14.978	3.2419	1.125	0.8637	0.989
3		(13,13)	60.147	4.856	1.276	0.7696	0.981
4		(7,7)	59.592	15.732	1.643	0.674	0.891
5		(9,9)	39.679	9.725	0.982	0.782	0.973

Using these parameters, CNN is trained using the input images from ultrasoundimages.com and sonoworld.com. Then, the test image is applied to CNN to validate its performance. The CNN model developed with two classes in each dataset. Each class is identified either by 1 or 0. In training, the model is fed an image in the category of 1 or 0. While validating and testing the features of cysts in each image are compared with that of classes in train dataset and are identified under category 1 or 0. In this case, Simple Cyst images are '1' and PCOS images are '0'. The figure 5 and 6 shows the classification performance of the CNN model. Also, the validation accuracy and training accuracy is shown in the figure 7.

**Figure 5.** The plot of train data with classes.**Figure 6.** The plot of validation data with classes.





**Figure 7.** The plot of Validation and Training Accuracy.

## 5. Conclusion

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.

## 6. References

- [1] Zhe Wu, Shuang Zhao, Yonghong Peng, Xiaoyu He Xinyu Zhao et al, "Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images" IEEE (May 2019).
- [2] Laura J. Brattain, Brian A. Telfer et al, "Machine learning for medical ultrasound: status, methods, and future opportunities", Abdominal Radiology (February 2018).
- [3] R M Dewi et al, "Classification of polycystic ovary based on ultrasound images using competitive neural network", Journal of Physics (2018).
- [4] Amsy Denny et al, "i-HOPE: Detection and Prediction System for Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques", IEEE (TENCON 2019).
- [5] E.Setiawati et al, "Particle Swarm Optimization on Follicles Segmentation to Support PCOS Detection", (ICoICT) 2015.
- [6] Mahmoud M. Abu Ghosh; Ashraf Y. Maghari, "A Comparative Study on Handwriting Digit Recognition Using Neural Networks", IEEE, 2017.
- [7] Rahul Chauhan et al, "Convolutional Neural Network (CNN) for Image Detection and Recognition", (ICSCCC) 2018.
- [8] S. Khazendar et al, "Automated characterization of ultrasound images of ovarian tumors: the diagnostic accuracy of a support vector machine and image processing with a local binary pattern operator", J no.7 (1): 7-15 (2015).
- [9] Nivetha.M et al. 2016 "Survey of Poly Cystic Ovarian Disease (PCOD) Among the Girl Students of Bishop Heber College, Trichirapalli, Tamil Nadu, India".
- [10] Knochenhauer ES.et al. Prevalence of Polycystic Ovarian Syndrome. Journal of Clinical Endocrinology & Metabolism 1998; 83(9):3068-3082.
- [11] Nidhi R et al. Prevalence of Polycystic Ovarian Syndrome in Indian adolescents. Journal of Paediatric Adolescent Gynaecology 2011 Aug;24(4):223-227.
- [12] PalviSoni et al, "Image Segmentation for detecting Polycystic Ovarian Disease using Deep Neural Networks", International journal of computer science and technology (March 2019).
- [13] B. Cahyono et al, "An Implementation of Convolutional Neural Network on PCO Classification based on Ultrasound Image". (ICoICT) 2017.

- [14] V. Deepika, “Applications of Artificial Intelligence Techniques in Polycystic ovarian syndrome Diagnosis”, *Journal of Advanced Research in Technology and Management Sciences*. (November 2019).
- [15] Sheela, S., M. Sumathi, and G. Nirmala Priya. 2016. “Comparative Analysis of Various Filtering Techniques for Speckle Noise Suppression in Ultrasound Images.” *International Journal of Control Theory and Applications* 9(2): 419–33.
- [16] Sheela, S., V. Subashini, and A. Karunakaran. 2016. “A Robust Image Matching Using FDCT and Local Textures for Image Retrieval Application.” *International Journal of Control Theory and Applications* 9(5): 2377–83.
- [17] Sheela, S., and M. Sumathi. 2016. “Study and Theoretical Analysis of Various Segmentation Techniques for Ultrasound Images.” *Procedia Computer Science* 87: 67–73.
- [18] P.Chitra, Jinthu Maua.,(2017),“Sharpening of edges in radiographic images using edge preserving filter”, *Proceedings of the IEEE International Conference on Inventive Systems and Control, ICISC 2017, Vol 1, PP 702-705, Jan 2017*.
- [19] P.Chitra.,”Quantitative Characterization of Radiographic Weld Defect Based on the Ground Truth Radiographs Made on a Stainless Steel Plates”, *Advances in Intelligent Systems and Computing, Springer publication, Volume 433, pp 157-166. Feb 2016*