



# Machine Learning Assisted Medical Diagnosis for Segmentation of Follicle in Ovary Ultrasound

Eliyani<sup>1,2</sup>, Sri Hartati<sup>1</sup>(✉), and Aina Musdholifah<sup>1</sup>

<sup>1</sup> Universitas Gadjah Mada, Sekip Utara Bulaksumur,  
Yogyakarta 55281, Indonesia  
shartati@ugm.ac.id

<sup>2</sup> Universitas Muhammadiyah Gresik, Gresik 61121, Jawa Timur, Indonesia

**Abstract.** Machine learning can be applied to the diagnosis of polycystic ovarian syndrome (PCOS), one of criteria for PCOS patients is presence polycystic ovary (PCO). PCO is the presence of the least 12 follicles in the ovary or follicular diameter between 2 and 9 mm and/or increased ovarian volume 10 cm<sup>3</sup>. In this research, a computational model for the detection of follicles of various sizes and extracting relevant features of the follicle and calculate the diameter and the number of follicles is proposed. The proposed model consists of pre-processing, speckle noise reduction, follicular segmentation, feature extraction, feature selection, and calculate the diameter of number follicles. The segmentation method uses active contour to divide objects base on the similarity of follicle shape feature so that it is more accurate in calculating the number and diameter of follicles. The performance of this method is tested on a dataset of ovarian ultrasound images of patients at Sardjito Hospital, Yogyakarta using Probabilistic Rand Index (PRI) and Global Consistency Error (GCE).

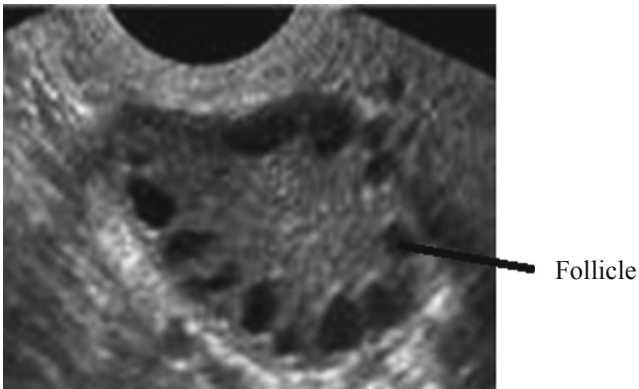
**Keywords:** Segmentation · Active contour · Feature extraction

## 1 Introduction

Machine learning has developed rapidly in various fields to help humans make decisions. This development has great potential for medical imaging technology, medical analysis, medical diagnostics, general health care, one of which is machine learning-assisted medical diagnosis. Some machine learning studies that have helped in the health sector, such as diagnosing leukemia [1], classification of malaria [2] and evaluation of the administration of hypertension drugs [3]. This paper highlights the direction of new research on machine learning to help detect follicles in ovarian ultrasound images. Detecting the number of follicles will be useful as one of means in determining the diagnosis of patients with polycystic ovary syndrome.

Polycystic ovary syndrome (PCOS) is experienced women of reproductive age with excessive androgen production, usually ovulation is irregular, sometimes even does not have ovulation, ovulation disorders cause difficulty in getting pregnant. Women experiencing PCOS have higher risk of type 2 diabetes mellitus, hypertension, cardiovascular, ovarian hyperstimulation syndrome (OHSS) and endometrial cancer [4, 5].

PCOS can be diagnosed by various biochemical criteria such as blood tests to check hormone levels, glucose and thyroid levels. Another method is to see the presence of follicles in the ovary by ultrasonography (USG). This method is often used because it is cheap, fast and low risk for patients [6, 7]. Ultrasound images of ovarian morphology with polycystic ovary (PCO) are characterized by 12 or more follicles measuring 2–9 mm and/or increasing ovarian volume of more than 10 cm<sup>3</sup> [8] Diagnosis of PCOS patients with calculating the number of the follicle is more reliable than calculating ovarian volume [6]. Follicles such as fluid-filled bags to form eggs, follicles produce estrogen which is needed in the development of eggs, so monitoring follicles in the ovary are very important for women for planning pregnancy. The characteristic appearance of the follicle by ultrasound [9] (see Fig. 1).



**Fig. 1.** PCOS [9]

PCOS is the most common female endocrine disorder, affecting 10% of women in the reproductive age [10, 11]. Diagnosis of PCOS is a pleasant experience because it has an explanation of the ovarian state and gives hope for treatment, but sometimes makes a woman frightened by the results of the diagnosis. Therefore it is essential to make an accurate diagnosis so that there is no diagnostic error [9, 12].

## 2 Data Collection

Follicle images are taken from ultrasound tests in accordance with specific rules, in women with regular menstrual cycles, examinations are carried out in the initial follicular phase (days 3–5), for women with irregular menstrual cycles (oligo-amenorrhic) can choose random days, or more precisely, the first 3–5 days after bleeding caused by progesterone. Ultrasonography should be arranged to achieve the best contrast to distinguish follicular fluid from the ovarian stroma [7, 13]. This research uses digital images of ovarian ultrasound from Permata Hati patients in Sardjito Hospital by converting analog images into digital images using the scanner.

### **3 Method**

In general, machine learning for detection and calculation of follicle's number and diameter in ovarian ultrasound images consists of several stages, including pre-processing, reduce speckle noise, segmentation, feature extraction, feature selection, calculate number of follicle, calculate diameter of follicle and model performance evaluation.

#### **3.1 Preprocessing**

The initial stage of image processing a pre-processing is an essential stage that needs to be done before image processed; it serves to improve image quality to produce better features in the next step. This research will use pre-processing method of histogram equalization and reduce speckle noise.

#### **3.2 Segmentation**

This research uses an active contour segmentation method, and active contour divides ovarian ultrasound image into several regions and separate object area and background from an ovarian ultrasound. Segmentation of active contour has the advantage of being able to adjust to object pattern according to its input parameters so that it can move broader or narrower to look for ovarian ultrasound objects.

#### **3.3 Features Extraction**

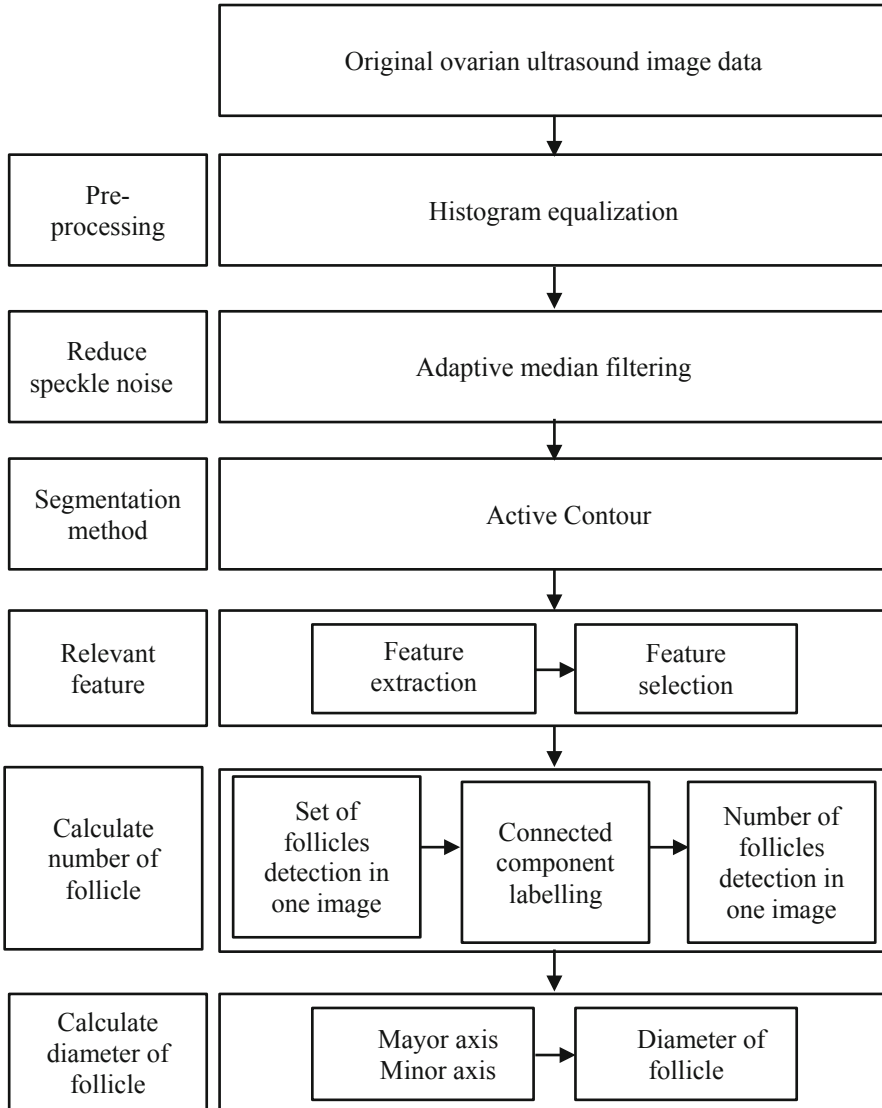
Feature extraction is a step to create a feature and reduce dimensions of dimensions from high dimensions to lower dimensions. Reliable feature extraction techniques are the main key in solving pattern recognition problems. Chain code is one shape feature extraction algorithm whose value does not change with the treatment of rotation, translation, reflection, and scaling. This method results in a value that shows the direction of the pixel, making up the object. This research will use geometric feature extraction, such as area, perimeter, major axis, minor axis, eccentricity, extent, circularity, and tortuosity produce measurements to recognize correct follicles.

#### **3.4 Feature Selection**

Once the feature extraction process is complete, then the next process is the selection of important feature that able to differentiate between very similar objects. The number of elements from the feature extraction process is too much possible to have the same values. The Feature selection is needed to improve accuracy. This research uses a chi-square feature selection technique to extract the most relevant features.

### 4 Proposed Method

In general, a computational model for detection and calculation of the number and diameter in ovarian ultrasound images are shown in Fig. 2 which consists of stages of image acquisition, pre-processing, follicle detection, feature extraction, feature selection, calculating follicle counts, calculating follicular diameter and model performance evaluation.



**Fig. 2.** Proposed method

The initial pre-processing is carried out to improve the quality of the image so that it can produce better features for the next stage. Pre-processing method used here is histogram equalization and speckle noise reduction.

One attempt to improve the distribution of ovarian ultrasound images that irregular shape of circles, is using histogram equalization method. There is a lot of speckle noise in ovarian ultrasonography, so to reduce the speckle noise used with the adaptive median filtering. This method can be used to handle filter operations in damaged images with impulse noise and smooth noise so that the image output is much better than the median filtering result.

The adaptive filter works on a rectangular region  $S_{xy}$ . The adaptive median filter changes the size of  $S_{xy}$  during the filtering operation depending on specific criteria as listed below. The output of the filter is a single value which replaces the current pixel value at  $x, y$  the point on which  $S_{xy}$  centered at the time. The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image with its neighboring pixels around it. The neighborhood size is adjustable, and the comparison threshold is adjustable. The Adaptive median filters work on two levels, namely [14].

Level A:         $A1 = Z_{median} - Z_{min}$   
                    $A2 = Z_{median} - Z_{max}$   
                   If  $A1 > 0$  and  $A2 < 0$  then do level B  
                   Else increase window size  
                   If  $window\ size \leq maximum\ allowed\ size\ of\ S_{xy}$  repeat level A  
                   Else,  $output = Z_{xy}$

Level B:         $B1 = Z_{xy} - Z_{min}$   
                    $B2 = Z_{xy} - Z_{max}$   
                   If  $B1 > 0$  and  $B2 < 0$  then  $output = Z_{xy}$   
                   Else,  $output = Z_{median}$

With

$Z_{min}$  = Minimum gray level in  $S_{xy}$

$Z_{max}$  = Maximum gray level in  $S_{xy}$

$Z_{median}$  = Median gray level in  $S_{xy}$

$Z_{xy}$  = Gray level at coordinates  $(x, y)$

## 5 Proposed Method Segmentation of Follicle

This research uses an active contour segmentation method to divide the ovarian ultrasound image into several regions and separate the object area and background from an ovarian ultrasound. Active contour for dividing ovarian ultrasound images into several regions and separating the object area and background from an ovarian ultrasound. The segmentation steps using active contour are shown in Fig. 3.

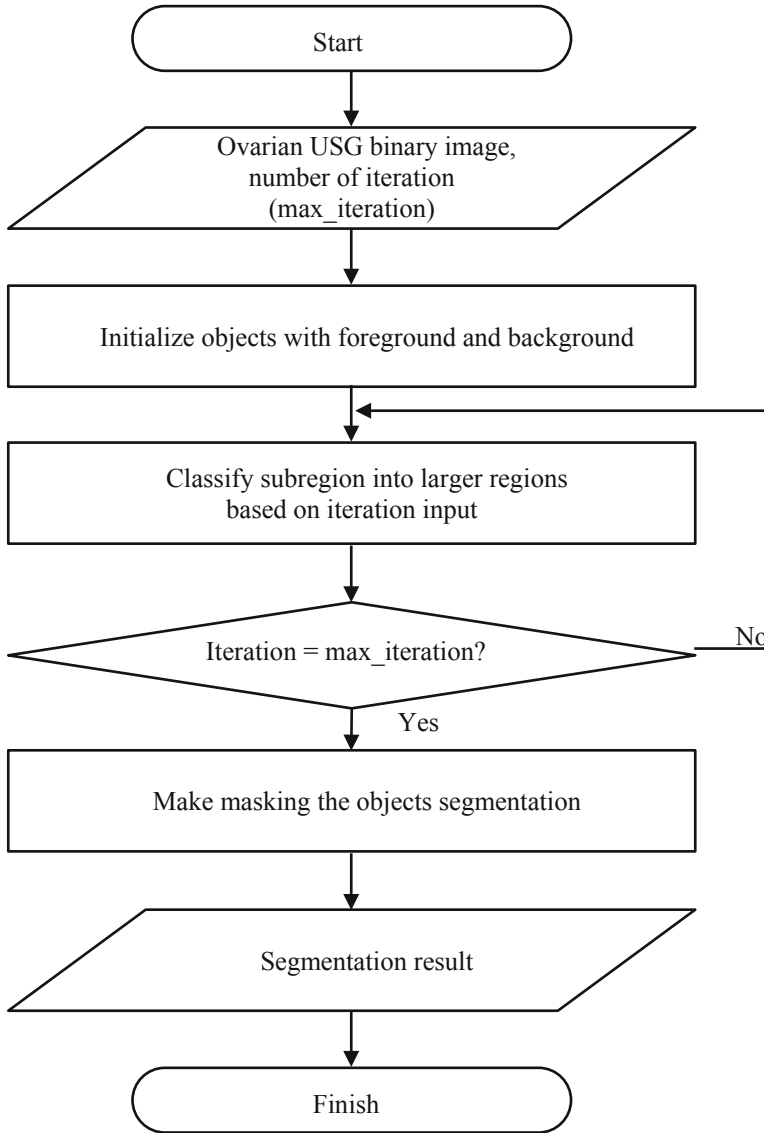


Fig. 3. Active contour

This method has the advantage of being able to adjust the object's pattern with its input parameters so that it can move broader or narrower to look for ovarian ultrasound objects. The active contour is developed to form a contour curve, the edge of the object in the image. Active contour model can be roughly classified as parametric active contour model, with snake parametrically by  $z(s) = (a(s), b(s))$ .

Active contour model is the snake model proposed by Kass et al. [15]. The snake model is an energy functional is in Eq. (1) which consists of two part called the inner and outer energy.

$$\begin{aligned} E_{snake}^* &= \int_0^1 E_{snake}(z(s))ds \\ &= \int_0^1 E_{internal}(z(s)) + E_{imageforce}(z(s)) + E_{constraint}(z(s))ds \end{aligned} \quad (1)$$

Where  $E_{internal}$  represent the internal energy of spline due to bending,  $E_{imageforce}$  gives rise to the image force, and  $E_{constraint}$  gives rise to the external constraint force [15].

The segmentation of the Chan-Vese model is usually dependent on the placements of initial contours, Chan Vese was formulated by minimizing the energy functional is in Eq. (2) [16].

$$\begin{aligned} F(c_1, c_2, C) &= \mu \text{Length}(C) + v \cdot \text{Area}(\text{inside}(C)) \\ &+ \lambda_1 \int_{\text{inside}(C)} |u_0(x, y) - c_1|^2 dx dy \\ &+ \lambda_2 \int_{\text{outside}(C)} |u_0(x, y) - c_2|^2 dx dy \end{aligned} \quad (2)$$

Parameters for determining evolution are used  $\mu$ ,  $C$  is any variable curve,  $v$  is the parameter to increase speed, and the constants  $c_1$ ,  $c_2$ , depending on  $C$ , are the averages of  $u_0$  inside  $C$  and respectively outside  $C$ ,  $\lambda_1$  and  $\lambda_2$  parameter to adjust the intensity inside  $c_1$  and  $c_2$  Active contour model with  $v = 0$  and  $\lambda_1 = \lambda_2 = \lambda$  is a particular case of the minimal partition problem, the best approximation  $u$  of  $u_0$  as a function taking only two value, namely [16].

$$u = \begin{cases} \text{average}(u_0) \text{ inside } C \\ \text{average}(u_0) \text{ outside } C \end{cases} \quad (3)$$

Calculate the related Euler-Lagrange equation for unknown functions  $\phi$ , slightly regularized from functions  $H$  and  $\delta$ , denoted here by  $H_\epsilon$  and  $\delta_\epsilon$ .  $F_\epsilon$  is represented associated regularized functional [16].

$$\begin{aligned} F_\epsilon(c_1, c_2, \phi) &= \mu \int_{\Omega} \delta_\epsilon(\phi(x, y)) |\nabla \phi(x, y)| dx dy \\ &+ v \int_{\Omega} H_\epsilon(\phi(x, y)) dx dy \\ &+ \lambda_1 \int_{\Omega} |u_0(x, y) - c_1|^2 H_\epsilon(\phi(x, y)) dx dy \\ &+ \lambda_2 \int_{\Omega} |u_0(x, y) - c_2|^2 (1 - H_\epsilon(\phi(x, y))) dx dy \end{aligned} \quad (4)$$

The segmentation method will be evaluated using Probabilistic Rand Index (PRI) and Global Consistency Error (GCE) for evaluating the performance of segmentation method. The PRI value must be higher than the GCE value.

Probabilistic Rand Index (PRI) is rand index that is a fasimilarity function that converted the problem of comparing two partitions with possibly differing number of classes into a problem of computing pairwise label relationships.

PRI calculates pixel pair fractions whose labels are consistent between calculated and actual segmentation. A measure of similarity between two groups of data. PRI gives a value between zero and one. If two segmented results do not have a similarity, the results are zero and if segmented images are identical, result is one. Formula for PRI is defined in Eq. (5) [17].

$$PR(S_{test}, \{S_{GT}\}) = \frac{1}{\binom{n}{2}} \sum_{\substack{i,j \\ i < j}} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})] \quad (5)$$

Where  $S_{GT}$  is ground truth images which are grouped manually  $\{S_1, S_2, \dots, S_{GT}\}$  according to segmented ovary images by the algorithm. Set of all perceptually correct segmentation is defined by random numbers  $p_{ij}$ ,  $c_{ij}$  shows the event that a pair of pixels  $i$  and  $j$  have the same label on  $S_{test}$  image [17].

$$c_{ij} = I(L_i^{S_{test}} = L_j^{S_{test}}) \quad (6)$$

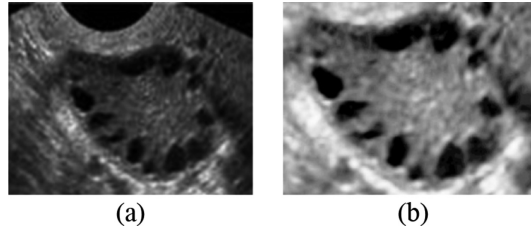
Global Consistency Error (GCE) is a measure to the extent of the results of segmentation. If one segment is the appropriate subset of the other, then the pixel is located in the area of repair and the error must be zero. If there is no subset relationship, then it indicates that the two regions overlap inconsistently. Let  $S$  and  $S'$  be two segmentations of an image  $X\{x_1, x_2, \dots, x_N\}$  consisting of  $N$  pixels. The Local Refinement Error (LCE) allows fo refinement to occur in either direction at different locations in the segmentation [17].

$$GCE(S, S') = \frac{1}{N} \min \left\{ \sum_i LRE(S, S', x_i), \sum_i LRE(S', S, x_i) \right\} \quad (7)$$

## 6 Experiment Results

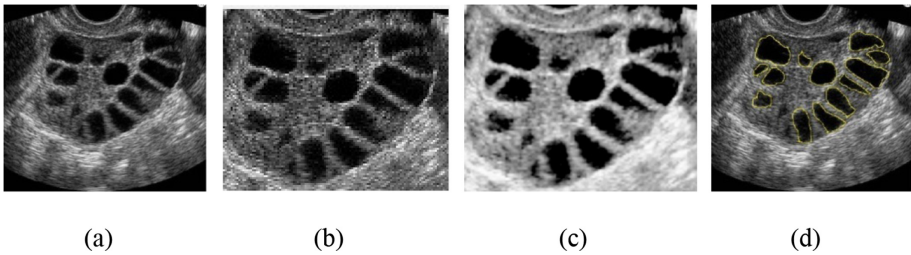
Proposed segmentation method will be tested on 100 images of ovarian ultrasound from Sardjito Hospital Yogyakarta, Indonesia. This research right now is in the stage of implementation of the segmentation method using active contour. The Fig. 4(a) depicts an original ultrasound image of the ovary and the result adaptive median filtering images are shown in Fig. 4(b).





**Fig. 4.** Original ultrasound image of the ovary and result image of proposed method. (a) Original image, (b) Adaptive median filtering image

Images may have poor contrast and can't be used directly. Therefore, it is necessary to get rid of speckle noise presented in an image. In order to remove speckle noise, adaptive Median filters are commonly used. An additional benefit of the adaptive median filter is that it seeks to preserve detail while smoothing speckle noise, the adaptive algorithm performed quite well. The Fig. 5(a) depicts an original ultrasound image of the ovary and the result images of proposed method are shown in Fig. 5(b) and (c).



**Fig. 5.** Original ultrasound image of the ovary and result image of proposed method. (a) Original image, (b) Histogram equalized image (c) Adaptive median filtering image, (d) Manual segmentation of follicles by medical expert.

## 7 Conclusion

Previous stage has detected presence of follicles in ovarian ultrasound images, then next step in calculating the number of follicles using connected component labeling. The expected result of this process is to be able to label and color each follicle in an ovarian ultrasound image making it easier to calculate number of follicles in one ovarian ultrasound image. Calculation of follicular diameter that is shaped like a circle is produced from area of follicle. Area is the number of pixels that make up a follicle (region), result of calculating the diameter in pixel values will be converted into millimeters.

**Acknowledgment.** The author would like to thank the Research Directorate of Universitas Gadjah Mada for funding this research in the RTA (Rekognisi Tugas Akhir) scheme 2019 program.

## References

1. Bagasjvara, R., Candradewi, I., Hartati, S., Harjoko, A.: Automated detection and classification techniques of Acute leukemia using image processing: a review. In: 2016 2nd International Conference on Science and Technology-Computer (ICST), pp. 35–43 (2016)
2. Hartati, S., Harjoko, A., Rosnelly, R., Chandradewi, I., Faizah, : Performance of SVM and ANFIS for classification of malaria parasite and its Life-Cycle-Stages in blood smear. In: Yap, B., Mohamed, A., Berry, M. (eds.) *Soft Computing in Data Science, SCDS 2018, Communications in Computer and Information Science*, vol. 937, pp. 110–121. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-13-3441-2\\_9](https://doi.org/10.1007/978-981-13-3441-2_9)
3. Soetanto, H., Hartati, S., Wardoyo, R., Wibowo, S.: Hypertension drug suitability evaluation based on patient condition with improved profile matching. *Indones. J. Electr. Eng. Comput. Sci.* **11**(2), 453 (2018)
4. Bilal, M., Haseeb, A., Rehman, A.: Relationship of polycystic ovarian syndrome with cardiovascular risk factors. *Diab. Metab. Syndr. Clin. Res. Rev.* **12**(3), 375–380 (2018)
5. Fernandez, E.D.T., Adams, K.V., Syed, M., Maranon, R.O., Romero, D.G.: Long-lasting androgen-induced cardiometabolic effects in polycystic ovary syndrome. *J. Endocr. Soc.* **2** (8), 949–964 (2018)
6. Ali, H.I., Elsadawy, M.E., Khater, N.H.: Ultrasound assessment of polycystic ovaries: ovarian volume and morphology; which is more accurate in making the diagnosis?! *Egypt. J. Radiol. Nucl. Med.* **47**(1), 347–350 (2016)
7. Coelho Neto, M.A., et al.: Counting ovarian antral follicles by ultrasound: a practical guide. *Ultrasound Obstet. Gynecol.* **51**(1), 10–20 (2018)
8. Azziz, R.: Definition, diagnosis, and epidemiology of the polycystic ovary syndrome. In: Azziz, R. (ed.) *The Polycystic Ovary syndrome, Current Concepts on pathogenesis and Clinical care, Endocrine Updates*, vol. 27, pp. 1–15. Springer, Boston (2007). [https://doi.org/10.1007/978-0-387-69248-7\\_1](https://doi.org/10.1007/978-0-387-69248-7_1)
9. Karakas, S.E.: New biomarkers for diagnosis and management of polycystic ovary syndrome. *Clin. Chim. Acta* **471**, 248–253 (2017)
10. Al-Shaikh, S.F.M.H., Al-Mukhtar, E.J., Al-Zubaidy, A.A., Al-Rubaie, B.J.U., Al-Khuzae, L.: Use of clomiphene or letrozole for treating women with polycystic ovary syndrome related subfertility in Hilla city. *Middle East Fertil. Soc. J.* **22**(2), 105–110 (2017)
11. Caburet, S., Fruchter, R.B., Legois, B., Fellous, M., Shalev, S., Veitia, R.A.: A homozygous mutation of GNRHR in a familial case diagnosed with polycystic ovary syndrome. *Eur. J. Endocrinol.* **176**(5), K9–K14 (2017)
12. Skiba, M.A., Islam, R.M., Bell, R.J., Davis, S.R.: Understanding variation in prevalence estimates of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum. Reprod. Update.* **24**(6), 694–709 (2018)
13. Porru, C., Fulghesu, A.M., Canu, E., Cappai, A.: Ultrasound diagnosis of polycystic ovarian syndrome : current guidelines, criticism and possible update. *Austin J. Obstet Gynecol.* **4**(2), 1074 (2017)
14. Gonzalez, R.C., Woods, R.E.: *Digital image processing*. p. 976, Nueva Jersey (2008)
15. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. *Int. J. Comput. Vis.* **1** (4), 321–331 (1988)
16. Chan, T.F., Vese, L.A.: Active contours without edges. *Br. Dent. J.* **142**(2), 73 (2001)
17. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 929–944 (2007)