Identification of Polycystic Ovarian Syndrome Using Follicle Recognition

[1]D Dinesh Kumar, [2]S Sakthiyaram

[1]Assistant Professor, Department of Information Technology,
St. Joseph's College of Engineering, OMR, Chennai – 600119
dineshkumard@stjosephs.ac.in

[2]Assistant Professor, Department of Electronics & Instrumentation Engineering,
Bannari Amman Institute of Technology, Sathyamangalam – 638401
Sakthiyaram92@gmail.com

Abstract— Sonography, also known as ultrasound, has helped in identifying and treating infertile patients. Ultrasound imaging of the ovary's follicles reveals the type of cyst, the diversity of follicles, and the size of the follicles' response to hormonal imbalance. Image segmentation enriches the image's region of interest with additional data and accurately identifies the object and its background. However, because segmentation on ultrasound images is difficult due to noise, follicle identification can be streamlined and made more effective by combining photograph preprocessing with morphological operations. In order to classify PCOS Ovaries and Normal Ovaries, the machine learning algorithms Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression, and the proposed method are used. Physical identification is used to compare classification results. The proposed algorithm yields an accuracy of 98%.

Keywords— ***Polycystic Ovary, Segmentation, Classification, SVM, KNN***

I. INTRODUCTION

Polycystic ovary syndrome, or PCOS, is a complicated condition marked by high levels of androgen, irregular periods, and/or small cysts on one or both ovaries. The disorder can have a physical cause (polycystic ovaries) or a chemical cause (most of the time) (hyperandrogenemia). Hyperandrogenism, which is a sign of PCOS, can stop follicular development, cause microcysts in the ovaries, stop ovulation, and change the way your periods work [1].

Alternatives to polycystic ovary syndrome include functional ovarian hyperandrogenism or Stein-Leventhal syndrome, a group of endocrine diseases linked to long-term ovulation and a lack of excess androgens (PCOS). Multiple follicular cysts form in the ovaries, which aids PCOS, which is believed to be caused by an egg's inability to exit the ovary. Along with PCOS, the main signs of this endocrine cycle are high androgen levels, chronic failure to ovulate, bloated ovaries with swollen follicles, and insulin resistance [2].

Most of the time, PCOS affects 5 to 10% of women between the ages of 18 and 44. When there is too much insulin in the body, there is more fat in the blood, more male hormones are made, and the ovaries don't work right. When PCOS is found in women, the risk of diabetes, heart problems, being overweight, prostate disease, and breast cancer goes up by a lot. Because of this, it is very important to figure out what PCOS is. Morphology has made a lot of pictures of polycystic ovaries, and as image technology improves, these pictures get better and better.

In the 18th century, physiology and anatomy were first described, and PCOS could be diagnosed based on certain symptoms. PCO is characterized by a hardening of the ovary's connective tissue and the presence of a large sac containing numerous cystic vesicles and tunica albuginea. For a number of years, wedge resection has been the only treatment for PCOS, and the standard procedure has been considered a histological examination. Since wedge resection is no longer utilized, it is still difficult to obtain these samples for PCO testing. Histopathological requirements include signs of follicular cells going in reverse and follicular atresia, lutropin with inner sac overgrowth, and a thicker ovarian enveloping membrane [3].

The studies examined reproductive tissues removed during wedge resection or hysterectomy. They discovered a strong correlation between PCO histopathology and the ultrasound diagnosis of polycystic morphology. According to Hughesdon's histological data, the number of follicles in the PCO has increased two to threefold. The cystic systems are known as "follicles with cysts that are pathological."

II. LITERATURE SURVEY

Polycystic Ovary Syndrome (PCOS) is an endocrine disease in women that has serious effects on their health and can only be diagnosed with medicine or even surgery. When PCOS is looked at by hand, there are often false polycysts. In recent years, many automated algorithms have been worked on to find polycysts in ultrasound images. This paper talks about how to find and recognise cysts in ultrasound images of the ovary with more than 90% accuracy [4].

In the real modern world right now. The goal of a data warehouse is to look at data from different sources and put it in the right place. It also collects useful types of data. Data mining uses different sets of rules to set up classifications, correlations, forecasts, clustering, etc. Women today get sick with many different things. PCOS, or polycystic ovary syndrome, is the most common problem women have these days. It can make it hard for women to have children. This is about women having miscarriages and different kinds of illnesses. The main signs of

this condition are irregular periods, dry skin, acne, asthma, and an anxiety disorder. In this article, the authors talk about the symptoms and risks of PCOS. It also tells what treatments work best for this disease and how different data processing methods had to be put together to accurately predict this disease [5].

Polycystic ovary syndrome is the most common hormonal problem that women have when they have their period (PCOS). This has caught the attention of couples who have been having trouble getting pregnant. One of the ways the doctor makes a diagnosis is by looking at the USG image of the ovary by hand to see how many and how big the follicles are. This study can have an effect on how low uncertainty, reproducibility, and performance are. To deal with these problems, it is suggested that there should be an automatic way to find the follicle on the USG picture so that PCOS can be diagnosed. The first system decides which type of follicle is in the original area that was all the same. The next step is to choose follicle parameters for the relevant areas and then use the segmented region's attribute to figure out what the follicle is. This diagnosis was always based on the number and size of the pixels in the image to determine if the person had PCOS or not. The methods used in region growing are based on the area and the seeds. To figure out the size of the follicle, you would use a different method, such as stereology or Euclidean distance. Using area growth and Euclidean distance on follicle quantification is the best way for a device to find PCOS [6].

Polycystic ovarian syndrome (PCOS) is a common condition that affects the hormones. Many women get it when they are of childbearing age. This disease is caused by infertility, which is linked to no periods and hair growth. PCOS is thought to be a mix of physiological, endocrinological, and biochemical problems, especially ones that have to do with how oestrogen and androgens are used in the body. The edge of the ovaries shows the normal size of immature follicles, which is less than 10 mm. It is a hard barrier to get past because doctors now use ultrasound images that show the number, size, and position of follicles and other details needed to look into these kinds of problems in women. In real testing, PCOS is a big problem because follicles come in different sizes and are connected to blood vessels and tissues, which can lead to mistakes. Some studies suggest using ultrasound pictures of the ovaries to look at PCOS in different ways. In this trial, USG samples were taken from a woman's ovary and used to look at the key variables and known procedures in the study of PCOS and compare them to the results of other studies [7].

Polycystic Ovary Syndrome is a reproductive disorder that causes periods to come and go at odd times (PCOS). Insulin and androgen hormones are a big part of why this is happening. This condition will lead to more serious health problems like heart disease, diabetes, and obesity, which should be found earlier. This condition can be seen by looking at the structure of the ovary and doing hormone tests. However, ultrasonography is the most cost-effective way to classify the shape of the ovary. A gynaecologist will do the study by hand to see if an ovary is normal or if it has any polycystic ovary (PCO) follicles. This paper could set up a PCOS detection system using Gabor Wavelet feature extraction, and Elman Neural Network would be used to tell the difference between PCOS and other conditions. Since the context layer can tell what the previous state was, Elman Neural Network is used. From this article, we can see that layer latency, training feature, and hidden layer are all related to the quality and processing schedule of each dataset. Based on the studies mentioned in the article, the highest valid value is 78.1%, which includes 32 characteristics [8].

III. PROPOSED METHODOLOGY

The first step in this process is to gather data. Trans-Vaginal Ovarian USG images of about 30 patients with PCOS signs are collected, along with their Body Mass Index, length of cycle, and post-menstrual Luteinizing Hormone (LH) and Follicle stimulating hormone (FSH) values [9][10].

People with diseases like thyroid problems or hypercortisolism are not included in this study. Methods like Gray scaling and Histogram equalisation are used in preprocessing to make the images more useful. For Feature Extraction, a multiscale morphological technique is used to pull out dark or light parts of the original image. Canny edge detection is the method used to divide things up. It gives the maximum number of follicles that have been found, from which the needed follicles are separated using different threshold values, such as larger and smaller follicle size, area, weirdness, and density [11].

Classification is done based on how many follicles were collected during the last step and on clinical and laboratory evidence from doctors. Fig. 1a) shows Ultrasound Images of PCOD and Fig.1b) shows a Feature extracted PCOD Image

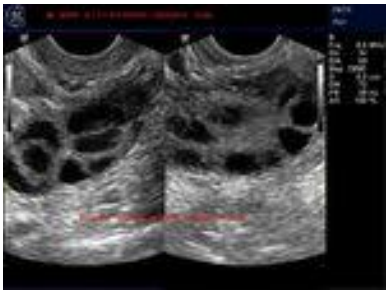

Fig. 1.a

Fig. 1. b

Fig. 1. a) Ultrasound Images of PCOD, b) Feature extracted PCOD Image

Fig.2 shows the flowchart of the process.

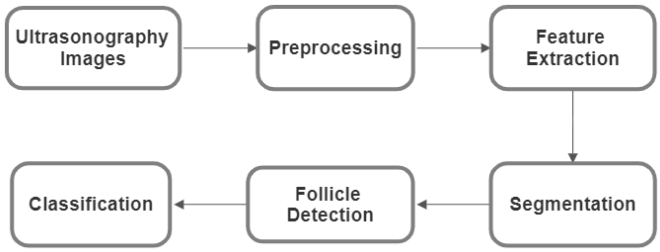

Fig. 2 Flowchart of the process

IV. PREREPROCESSING

Preprocessing is an important step that must be done before the data can be analysed. This step would lead to data that is easier to access, has better quality, and stores all the important details in the image. There are numerous steps in preprocessing, including histogram equalisation, grey scaling, image bi-leveling, image inversion, and data cleaning. Grey scaling has been used to convert colour images to grayscale images. Bi-level is the method by which the rates of the figure array are changed to 0 or 1, thereby transforming the figure from black and white to colour. The terms used were deterioration and expansion. Erosion is used to remove "noise" objects from a binary image by making them less distinct [12]. Then, the sedation method is used to harden artefacts that were lost during the process of erosion and were represented by binary images. The invert image command transforms black into white and white into black. This is done because it is simpler to identify an object when its background is black and its subject is white. The last step of data cleaning is to get rid of objects that aren't important. Fig. 3a shows a grayscale image, 3b shows an image with the brightness changed, and 3c shows an image with the column contrast changed.

Fig 3.a

Fig. 3. b

Fig. 3.c

Fig. 3 a) Gray Scale Image, b) Brightness changed Image, c) Column Contrast changed Image

V. FEATURE EXTRACTION

Fuzzy Feature extraction is the step of getting information about an object from an image that can be separated from information about other objects. On the ultrasound picture of the ovary, you can see follicles, endometrial arteries, swollen glands, axons, and acinus. In order to reduce false detection, a multi-scale morphological technique is used to extract features. To make an improved variance image, you have to pull out the dark and light parts of the picture and then subtract the result from the first number. Giving dark attributes a matrix size of 0 and white attributes a matrix size of 255 creates a two-level image. This provides you with black-and-white images that clearly demonstrate the limitations that must be examined in additional tests, such as classification and segmentation, in order to obtain detailed performance [13]. Figure 4a depicts a distorted image, while Figure 4b depicts a stretched image.

Fig. 4.a

Fig. 4.b

Fig 4. a) Damaged Image, b) Stretched Image

VI. SEGMENTATION

Segmentation is the step that separates an image from its source. Edge detection and Follicle cutting are the two steps in this phase. Edge detection tries to figure out what parts of the image are there. The goal of this research is to find follicles and edges using modern algorithms and a clever edge detection method [14].

Also, each follicle that is found will be marked, and each marked follicle will be cut to make a new figure that will be used in the next step. Threshold Image is shown in Figure 5.

Fig. 5 Threshold Image

VII. Detection of the Follicle

After binarization and edge detection, a vast number of features are found. In order for the programme to determine which follicles are correct, it must include criteria such as upper and lower follicle measurements, follicle area, freakishness, and density. According to the study, PCOS affects the size of the ovarian follicles, which can range in diameter from 2 to 9 mm. A typical ovary has a diameter of approximately 20 millimetres. Consequently, if follicles are round, the PCOS ovary would have a surface area of approximately 4–80 mm^2, whereas a normal ovarian follicle would have a surface area of approximately 314 mm^2. As these follicles are intended to resemble spheres, the strangeness is comparable to considering all of these origin levels, the zone of each detected region or function, and the number of similar follicles observed [15].

VIII. Classification

Classification is the final step, and separating images into distinct groups can become a habit. The classification was performed using the SVM (Support Vector Machine), KNN (K-Nearest Neighbors), and Logistic Regression algorithms, as well as the proposed method, which combines all three algorithms [16].

Algorithm for the Support Vector Machine (SVM):

SVMs, which stand for "support vector machines," are a set of learning methods for classification and regression that are guided by a teacher. They belong to the same family as the generalised linear classifier. In other words, a Support Vector Machine (SVM) is a method for classifying and predicting regression that uses machine learning theory to make predictions as accurate as possible without overfitting to the data. Support Vector Machines (SVMs) are systems that use the hypothesis space of linear functions in a high-dimensional feature space. They are trained with an optimization theory-based learning algorithm that includes a learning bias that comes from statistical learning theory.

Support vector machines became popular in the NIPS community and are now used in machine learning research around the world. In the process of recognising handwriting, SVM achieves the same level of accuracy as complex neural networks with well-developed features, even though it only uses pixel maps as data. It is also used for things like analysing handwriting, faces, and other things, with a focus on pattern classification and regression.

The theory uses the Structural Risk Minimization (SRM) concept, which has been shown to work better than the standard Empirical Risk Minimization (ERM) concept used by traditional neural networks. ERM tries to minimise the error in the training data, while SRM tries to minimise the expected risk's upper limit. Because of this difference, SVM is better at generalising, which is the goal of statistical learning. SVMs were first made to solve problems with classification, but they are now also used to solve problems with regression.

Algorithm K-Nearest Neighbor (KNN):

K Nearest neighbour classification works well in many machine learning applications when the relationships between the features and the target class are many, complex, and hard to understand. Another way to say this is that nearest neighbour works well when it's hard to explain something but you know it when you see it. KNN is easy to use and works well. It makes no assumptions about how the data is spread out and has a fast training phase.

The kNN algorithm starts with a training set of examples that have been put into multiple categories using a nominal variable. Let's say we have a test dataset with examples that aren't labelled but have the same features as the training dataset. kNN looks for the k records in the training data that are "closest" to each record in the test data, where k is a number that has already been set. Most of the classes of the k closest neighbours are given to the unlabeled test case.

The K-NN algorithm thinks of the features as coordinates in a multidimensional space of features. Since we only have two features in our dataset, the feature space is two-dimensional, with the x dimension representing how warm an ingredient is and the y dimension representing how crisp it is. After we build the dataset, we plan to use it to answer the question of whether an ovary is PCO or Normal. We can use the nearest neighbour method to figure out which class is a better match.

Distance calculation: To find an object's closest neighbours, you need a distance function or a formula that measures how similar two things are.

The kNN algorithm has always used the Euclidean distance.

Algorithm for Logistic Regression:

Logical regression, which is similar to neural networks and is the most popular supervised machine learning algorithm for classifying data, is the most popular way to use machine learning to sort data. A neural network is made up of a stack of logistic regression classifiers.

Statisticians made the logistic function, which is also called the sigmoid function, to explain how population growth works in ecology. For example, population grows quickly and reaches the carrying capacity of the environment quickly..

IX. PROPOSED METHOD

Support vector machine (SVM) and K-Nearest Neighbour (KNN) is a relatively new classification machine learning algorithm, while logistic regression (LR) is an old traditional statistical classification process. Despite the fact that there have been numerous detailed studies comparing SVM, KNN and LR, many new improvements have been made to them since their development, such as bagging and ensemble. Bagging and ensemble learning have recently gained popularity as methods for improving the generalisation efficiency of single learning algorithms. As a result, comparing SVM, KNN and LR classification output using bagging and ensemble is a fascinating subject. The results of the SVM, KNN and LR algorithms on different data sets are subjected to a rigorous statistical analysis in this report, which sets it apart for better accuracy in the classification.

PCOS screening allows the disease to be diagnosed and treated prior to it causing infertility problems. The process of early detection involves examining the PCOS through USG provides the early treatment for patients. If a cyst is found in the ovary, it will be diagnosed with size of the follicle and the type of cyst. A clinician then examines the cells under a USG to determine whether the ovary is likely to be PCO Ovary or Normal Ovary.

If machine learning could automate the identification of PCOS, it would greatly benefit the healthcare system, increase the efficiency of the detection process, and free up physicians to focus on treatment rather than diagnosis.

This analysis will classify the process as either class '0' implying non-Polycystic Ovary or class '1' implying Polycystic Ovary.

X. RESULT AND DISCUSSION

Images of PCOS taken with ultrasound are taken from 30 patients. The original Ultrasound PCOS image is then split into two levels so that it can be used in a morphological process. Greyscale and brightness-adjusted images make the ultrasound image clear, and the threshold imaging process is then used to separate the follicles. Feature Extraction is done with a multiscale morphological technique that takes out eroded and stretched parts of the original image to find PCOS. Figure 6 shows the Original Image (a), the Gray Scale Image (b), the Brightness Adjusted Image (c), and the Column Contrast Adjusted Image (d). It also shows the Threshold Image (e), the Eroded Image (f), and the Dilated Image (g).

Fig. 6. a

Fig. 6. b

Fig. 6. c

Fig. 6. d

Fig. 6. e

Fig. 6. f

Fig. 6. g

Figure 6 a) Original Image, b) Grayscale Image, c) Brightness-Adjusted Image, d) Column Contrast-Adjusted Image, e) Threshold Image, f) Eroded Image, and g) Dilated Image.

Figure 7 shows how the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values are used to test the classification model to see if a patient in the dataset has PCOS.

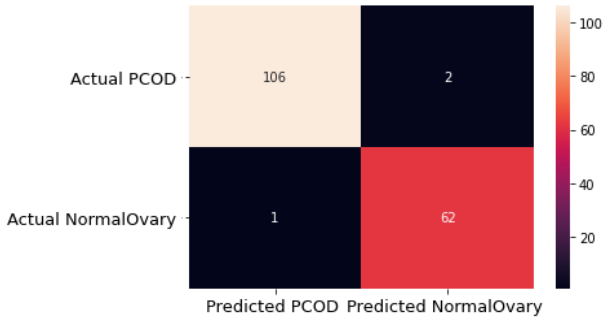

Fig. 7 Matrix predicting TP, TN, FP and FN values

Three standard machine learning algorithms are used to define Polycystic Ovarian Syndrome, and the results are compared to the suggested stacking ensemble process using metrics. In this study, the metrics used were Accuracy, Precision, and Recall. The percentage value of metrics is used in Fig. 8's bar graph to compare different algorithms.

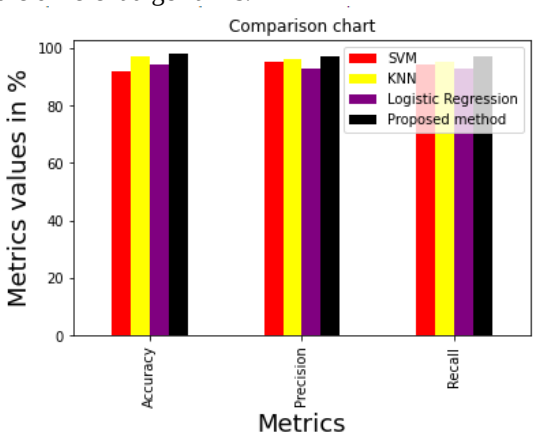

Fig 8: Performance comparison of Algorithm

TABLE 1: Performance comparison in numeric value

METHOD/ METRICS	SVM	KNN	Logistic Regression	Proposed Method
Accuracy	92	97	94	98
Precision	95	96	93	97
Recall	94	95	93	97

Table 1 shows clearly what Accuracy, Precision, and Recall are and how important they are. From this table, we can tell that the KNN algorithm does better than the other traditional algorithms.

XI. CONCLUSION

In this paper, four distinct machine-learning algorithms are used to compare a system that uses follicle recognition and classification to detect Polycystic Ovarian Syndrome. Image preprocessing is a method for enhancing an image's clarity. The images are divided into two levels, and a method for intelligent edge detection is used to separate them. Using the area and the freakishness value, we are able to differentiate between necessary follicles and areas. Using the Support Vector Machine algorithm, the KNN algorithm, the Logical Regression algorithm, and the proposed method, all of the data is grouped. The outcomes are contrasted with those of physical classification. When all algorithms are considered together, the proposed algorithm is approximately 98% accurate. Therefore, this algorithm can be used to automatically assess the health of Polycystic Ovarian Syndrome patients.

References

[1] Joham, Anju E., et al. "Polycystic ovary syndrome." The Lancet Diabetes & Endocrinology (2022).

[2] Louwers, Yvonne V., and Joop SE Laven. "Characteristics of polycystic ovary syndrome throughout life." Therapeutic Advances in Reproductive Health 14 (2020): 2633494120911038.

[3] P. Soni and S. Vashisht, "Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 816-820, doi: 10.1109/CESYS.2018.8724087.

[4] Adiwijaya *et al*, "Follicle Detection on the USG Images to Support Determination of Polycystic Ovary Syndrome", Journal of Physics Conference Series 622(1):012027

[5] C. Gopalakrishnan, M. Iyapparaja , "A Detailed Research on Detection of Polycystic Ovary Syndrome from Ultrasound Images of Ovaries", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-2 S11, September 2019.

[6] R M Dewi, Adiwijaya, U N Wisesty and Jondri, "Classification of polycystic ovary based on ultrasound images using competitive neural network", Journal of Physics: Conference Series, Volume 971, International Conference on Data and Information Science 5–6 December 2017, Telkom University, Indonesia

[7] I F Thufailah, Adiwijaya, U N Wisesty and Jondri, "An implementation of Elman neural network for polycystic ovary classification based on ultrasound images", Journal of Physics: Conference Series, Volume 971, International Conference on Data and Information Science 5–6 December 2017, Telkom University, Indonesia

[8] Yinhui Deng, Yuanyuan Wang and Ping Chen, "Automated detection of Polycystic Ovary Syndrome from ultrasound images," 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 2008, pp. 4772-4775, doi: 10.1109/IEMBS.2008.4650280.

[9] B. Potocnik D. Zazula, "Automated analysis of sequence of ovarian ultrasound images, part I: segmentation of single 2d images", Image vision and Computing, vol. 20, no 3, 2002, pp 217-225.

[10] S. Rethinavalli and M. Manimekalai, ―A Hypothesis Analysis on the Proposed Methodology for Prediction of Polycystic Ovarian Syndrome", IJCSET, vol. 6, no. 11, pp. 396–400, 2016.

[11] Sahil V. Khedaskar, Mohit A. Rokade, Bhargav R. Patil, Tatwadarshi P. N.," A Survey of Image Processing and Identification Techniques", VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 1 (2018), ISSN(Online): 2581-7280 Article No. 10, PP 1-10.

[12] S. Rihana, H. Moussallem, C. Skaf and C. Yaacoub, "Automated algorithm for ovarian cysts detection in ultrasonogram," 2013 2nd International Conference on Advances in Biomedical Engineering, Tripoli, Lebanon, 2013, pp. 219-222, doi: 10.1109/ICABME.2013.6648887.

[13] A. Denny, A. Raj, A. Ashok, C. M. Ram and R. George, "i-HOPE: 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, doi: 10.1109/TENCON.2019.8929674.

[14] R.C. Gonzalez and R .E. Woods: Digital Image Processing, Addison-Wesley Publishing Company, 1992

[15] M Kalaiyarasi, R Dhanasekar, S Sakthiya Ram, P. Vaishnavi. "Classification of Benign or Malignant Tumor Using Machine Learning", IOP Conference Series: Materials Science and Engineering, 2020

[16] J. Madhumitha, M. Kalaiyarasi and S. S. Ram, "Automated Polycystic Ovarian Syndrome Identification with Follicle Recognition," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 98-102, doi: 10.1109/ICSPC51351.2021.9451720.