International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

DIAGNOSINS OF POLYSYSTIC OVARY SYNDROME USING MACHINE LEARNING ALGORITHMS

[1] G. SINDHU,[2] N. RANADHISH,[3] Y. SHAHSI KIRAN,[4] Dr. J. PRAVEEN KUMAR
[123] Students, [4] Associate Professor
Department Of Computer Science and Design
Teegala Krishna Reddy Engineering College, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

This paper centers on the data-driven determination of polycystic ovary disorder (PCOS) in ladies utilizing progressed machine learning methods connected to a freely accessible dataset from the Kaggle store. The dataset contains restorative and clinical records of 541 ladies, out of which 177 are affirmed PCOS cases. It incorporates 43 properties that cover a wide extend of clinical, hormonal, and way of life highlights such as BMI, affront levels, FSH, LH, menstrual cycle inconsistencies, nearness of facial hair, skin break out, and other side effects commonly related with PCOS. To distinguish the foremost important indicators of the condition, univariate include determination strategies were utilized to rank the significance of each property. The examination uncovered that the proportion of Follicle-Stimulating Hormone (FSH) to Luteinizing Hormone (LH) is the foremost persuasive include in deciding PCOS, which adjusts with clinical discoveries from endocrinology. For show preparing and assessment, both holdout and 40-fold cross-validation procedures were utilized to guarantee vigor and maintain a strategic distance from overfitting. A few classification calculations counting Slope Boosting, Arbitrary Timberland, Calculated Relapse, and a cross breed demonstrate combining Irregular Woodland with Calculated Relapse (RFLR) were executed and compared based on their precision and review scores. The comes about show that utilizing as it were the beat 10 positioned highlights, high- performance classification is achievable, with the RFLR demonstrate beating others by accomplishing a testing exactness of 91.01% and a review esteem of 90%. These measurements propose that the RFLR show not as it were accurately classifies most PCOS patients but too minimizes untrue negatives, which is basic in therapeutic conclusion. The consider highlights the viability of machine learning in therapeutic diagnostics and proposes that such prescient models can serve as decision-support devices in clinical settings to help healthcare experts within the early and accurate diagnosis of PCOS, eventually progressing understanding results.

Received: 10-7-2025 Accepted: 18-8-2025 Published: 25-8-2025

I. INTRODUCTION

Polycystic ovary disorder (PCOS) could be a customary endocrine disorder in females influenced by lifted androgen. Treatment of PCOS is based on a few cardinal highlights counting anovulation, hyperandrogenism signs, and menstrual brokenness. A few of the critical indications of this clutter are pelvic torment, abundance hair, skin break out, hirsutism, smooth skin, male hormone, unpredictable periods, danker, etc. Around 5–10% of regenerative age (15–49 a long time) ladies are endured from this issue. The flare-up of PCOS was recorded to be 8% and 4.8% in African American ladies and white ladies, individually [1–2]. Ladies with this ovarian brokenness are related with hypertension, expanded chance of cardiovascular illness, corpulence, gynecological cancer, sort 2 diabetes mellitus, etc. Additionally, later inquire about has appeared that PCOS causes higher hazard of to begin with trimester premature delivery. PCOS causes in ovaries unseemly development of follicle that are anticipated at a primary stage and prematurely deliver to develop. This can be one of the causes for fruitlessness. Subsequently, it is noteworthy to screen the patients at a essential arrange to anticipate any serious moment of the PCOS illness.

(i) selecting the foremost imperative traits of PCOS patients for the given dataset using feature choice strategy.

(ii) applying machine learning calculations on the imperative highlights of the PCOS dataset.

(iii) comparing the machine learning calculations with those detailed within the writing in terms of testing precision and review.

The rest of the paper is organized as takes after. Segment II gives a brief writing audit on the application of measurable instruments and machine learning on the determination of PCOS. Area III depicts the in general technique connected to conduct this investigate work. The method of finding the foremost noteworthy highlights of PCOS is depicted in Area IV. The comes about gotten utilizing the application of machine learning on PCOS dataset are portrayed in Segment V. At last, Area VI gives the concluding comments.

1.1 MOTIVATION

There are a few inspirations for creating a PCOS (Polycystic Ovary Disorder) forecast framework utilizing machine learning and clinical information:

- Early Determination: PCOS regularly goes undiscovered due to unclear or covering side effects. A data-driven framework can help in recognizing the clutter at an early arrangement, permitting opportune restorative intercession and decreasing long- term complications.
- Women's Wellbeing Mindfulness: Numerous ladies endure from PCOS without understanding its effect on richness, hormonal adjust, and metabolic wellbeing. A prescient framework can offer assistance raise mindfulness and empower early meetings with healthcare suppliers.
- Moved forward Precision: Conventional demonstrative strategies can change between specialists. Machine learning models prepared on clinical information can offer more steady and exact forecasts based on key therapeutic markers.
- Healthcare Availability: In regions with constrained get to to gynecologists or endocrinologists, a machine learning-based expectation apparatus can bolster essential screening and offer assistance bridge the hole in healthcare administrations.
- Highlight Distinguishing proof: Utilizing include choice calculations makes a difference recognize the foremost basic pointers of PCOS (e.g., LH/FSH proportion, BMI, affront levels), contributing to superior clinical understanding and more centered symptomatic criteria.
- Cost-Effective Screening: A computerized apparatus utilizing open-source libraries and accessible datasets can offer a low-cost arrangement for people and little clinics to screen for PCOS without the require for costly tests at first.

1.2 PROBLEM STATEMENT

The most inspiration behind this PCOS prediction venture is to assist identify the clutter early using machine learning and clinical information. PCOS could be a common condition among ladies, but many go undiscovered due to its complex and vague side effects. By employing a data-driven framework, we will recognize imperative wellbeing markers like hormone levels and BMI to anticipate PCOS more precisely and rapidly. This not as it were progresses early determination and treatment but too raises mindfulness around women's wellbeing. The framework can be a low- cost, supportive apparatus for both specialists and patients, particularly in zones where get to to masters is constrained. It moreover gives understudies a chance to apply machine learning in real-world healthcare issues, making it a significant and instructive extend.

1.3 SCOPE AND OBJECTIVE

The most objective of a PCOS Expectation Framework utilizing machine learning is to bolster early determination of Polycystic Ovary Disorder by analyzing restorative information and recognizing designs related to hormonal and physical wellbeing indicators. This framework points to help healthcare suppliers and people in recognizing PCOS indications expeditiously, empowering opportune restorative mediation and progressing in general women's wellbeing results.

Goals:

- Utilize machine learning calculations to analyze clinical highlights such as hormone levels (LH, FSH), BMI, affront levels, and other significant parameters to detect patterns related with PCOS.
- Recognize the foremost vital traits influencing PCOS utilizing highlight choice strategies, progressing symptomatic center and demonstrate interpretability.
- Create a prescient demonstrate able of conveying quick and exact comes about, reasonable for integration into healthcare instruments or individual wellbeing checking stages.
- Optimize the framework for tall forecast precision, accuracy, and review through nonstop testing and approval utilizing real-world clinical datasets.

II. LITERATURE SURVEY

The literature on PCOS detection presents a range of methods for follicle detection, with two main approaches: stereology and sequential classification/feature extraction processes. In one study [3], polycystic ovary syndrome (PCOS) was identified using stereology to compute the number and size of follicles, employing Euclidean distance to measure their diameter. Other studies [4-5] applied Gabor Wavelet for feature extraction and utilized optimization techniques like Conjugate Gradient-Fletcher Reeves and Levenberg-Marquardt as variations of backpropagation for classification.

In [6], particle swarm optimization (PSO) was used to adjust the fitness function for follicle segmentation. Another study [7] utilized three classifiers: support vector machine (SVM) with RBF kernel, k nearest neighbors (kNN) with Euclidean distance, and neural network-based learning vector quantization (LVQ). Region growing schemes were employed for follicle segmentation in [8], where the technique checks whether neighboring seeds should be added to the segmentation region.

For ultrasound image segmentation, edge detection methods were used in studies [9, 10]. A median filter was applied to reduce noise in the images, and Otsu's global thresholding was used to identify pixel similarity. The Canny edge detection algorithm, as detailed in [11], was also found effective for detecting follicles in ultrasound images for PCOS diagnosis.

Despite these advancements, limited work has been done on applying machine learning techniques to clinical parameters for PCOS screening. This research aims to focus on early detection of PCOS using such techniques.

EXISTING SYSTEM

The current PCOS discovery framework basically employments conventional strategies like restorative imaging (ultrasound) and manual assessment of clinical parameters (follicle check, sore nearness, and hormones). Machine learning methods, counting SVM, CNN, neural systems, and PSO, have been connected for errands such as picture preparing, follicle division, and highlight extraction. Whereas these strategies appear guarantee, they regularly need integration of comprehensive clinical information and are not optimized for real-time conclusion, showing the require for more productive frameworks.

Drawbacks:

- Ultrasound picture handling is resourceintensive.

- A few strategies center on a contract extend of traits or picture investigation.
- Complex neural systems require expansive datasets and longer preparing.
- Precision changes, with comes about extending from 68.8% to 93.92%, regularly requiring numerous highlights.

PROPOSED SYSTEM

The proposed approach utilizes a machine learning-driven demonstrative framework that centers on clinical parameters instead of ultrasound pictures, utilizing cross breed calculations such as Irregular Woodland and Calculated Relapse (RFLR) for progressed precision. A dataset comprising 541 ladies with 43 clinical traits was utilized, which was refined to the 10 most critical highlights (e.g., FSH/LH proportion, BMI, follicle number) through a univariate include determination strategy, essentially diminishing computational complexity and handling time. Key classifiers executed within the framework incorporate Angle Boosting, Irregular Woodland, Calculated Relapse, and the half breed RFLR show. The framework accomplished a tall testing precision of 91.01% and a review esteem of 90% utilizing RFLR with 40-fold cross- validation. By centering on clinical parameters instead of picture handling, the framework offers more prominent versatility over diverse datasets and guarantees lower computational request, making it more proficient, open, and commonsense for broad utilize in real-world clinical setting.

III. MODULE DESCRIPTION

Information Collection Module: This module is mindful for gathering crude input information from different sources, counting customer-provided information and outside datasets like those from Kaggle. It extricates client statistic, monetary information, and chronicled advance execution information to form a organized dataset for preprocessing.

Information Preprocessing Module: This module handles the planning of crude information by guaranteeing it is clean and prepared for investigation. It oversees assignments like taking care of lost information, ascribing values, recognizing and evacuating exceptions, and normalizing or scaling information to guarantee it is appropriate for machine learning show preparing.

Highlight Examination Module: This module analyzes the highlights within the dataset, deciding their significance and how they relate with advance endorsement results. It distinguishes key highlights such as wage, business status, and credit history, and creates visualizations to delineate these connections.

Machine Learning Preparing Module: This module is dependable for preparing machine learning models utilizing the preprocessed dataset. It parts the information into preparing and testing sets and applies classification calculations like Calculated Relapse, Choice Tree, and Arbitrary Timberland to create prescient models for advance endorsement.

Forecast Module: Utilizing the prepared machine learning models, this module predicts advance endorsement results for modern client information. It yields the credit choice (approve/reject) beside a certainty score or likelihood, making a difference banks make educated choices.

Evaluation and Measurements Module: This module assesses the execution of the prepared models by calculating key measurements such as exactness, exactness, review, specificity, and F1-score. It too produces perplexity lattices to evaluate how well the models are foreseeing advance endorsement.

User Interface Module: This module gives an interface for bank agents to associate with the framework. It permits them to input client information, see loan approval choices, and get it the significance of different highlights within the expectation handle.

Reporting and Notice Module: This module produces point by point reports summarizing the credit endorsement choices and sends notices to important divisions or stakeholders

IJERST International Journal of Engineering Research and Science & Technology

with respect to the result of advance applications.

Admin: This part within the framework permits clients to log in and oversee the credit holder data, counting setting credit limits. Admins can moreover screen suspicious exercises and piece or enact client accounts when vital..

User: The Client part permits people to enlist and perform exercises related to credit card utilization, such as obtaining items, making installments, and managing their credit card limits inside the framework.

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

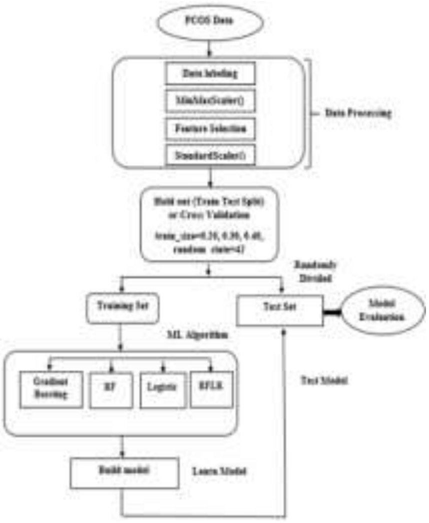

Fig. System Architecture

V. RESULTS AND DISCUSSION

This area talks about the results of applying different machine learning calculations to the PCOS dataset. Utilizing Calculation 1, the positioning of 43 highlights was decided, with the FSH/LH proportion positioning most noteworthy, taken after by FSH and Anti-Mullerian hormone (AMH). For assessing the adequacy of diverse classifiers, 10 highlights with 20-fold cross-validation were tried. Among the classifiers, irregular woodland, angle boosting, and RFLR accomplished the leading exactnesses of 87.83%, 88.75%, and 90.01%, separately.

TABLE II. RANKING OF BEST 10 FEATURES

Features name	Ranking
FSH/LH	1
FSH (mIU/mL)	2
AMH (ng/mL)	3
BMI	4
Weight gain (Y/N)	5
Follicle No. (L)	6
Follicle No. (R)	7
Avg. F size (L) (mm)	8
Cycle (R/I)	9
Cycle length (days)	10

Table 9.1 Ranking Of Best 10 Features

When analyzing diverse numbers of highlights, it was found that utilizing 10 or 14 highlights brought about in competitive execution, adjusting both accuracy and computational effectiveness. RFLR accomplished the finest classification exactness of 90.01% with 10 highlights and 89.27% with 14 highlights, which demonstrates that less highlights can surrender comparative prescient execution whereas lessening computational complexity.

TABLE III. PERFORMANCE USING 30-FOLD CROSS VALIDATION FOR DIFFERENT NUMBER OF FEATURES

Classifier	Number of features	Precision	Recall	Accuracy
Gradient Boosting	10	85%	85%	88.72%
	14	85%	85%	88.72%
Random Forest	10	88%	88%	87.99%
	14	88%	88%	87.98%
Logistic regression	10	85%	85%	88.35%
	14	83%	83%	88.35%
RFLR	**10**	**89.90%**	**90%**	**90.01%**
	14	89.20%	89.30%	89.27%

Table 9.2 Performance Using 30-Fold Cross validation For Different Number of Features

Furthermore, the execution of RFLR was compared with past considers. In spite of utilizing less highlights, the comes about accomplished in this think about, particularly the 91.01curacy and 90% review utilizing RFLR, are competitive with those from thinks about that depended on ultrasound pictures and bigger include sets. Previous works such as [4] and [7] detailed correctnesses of 93.92% and 82.55%, separately, but these utilized distinctive information sorts, making coordinate comparison troublesome. Besides, the strategies in this consider are more computationally effective, requiring less time to achieve comparable comes about, which makes them more commonsense for real-time

International Journal of Engineering Research and Science & Technology

applications.

Generally, the consider illustrates that machine learning methods, particularly RFLR, are promising for PCOS location, advertising tall precision, moo include complexity, and effective execution over distinctive approval methodologies.

VI. OUTPUT SCREENS

Fig: Open Anaconda Prompt

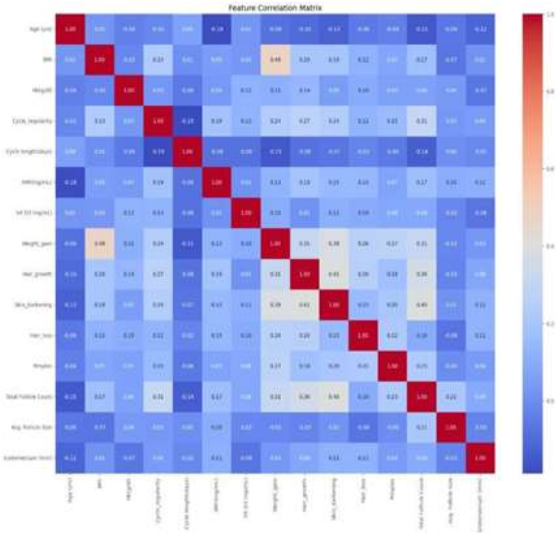

Fig: Feature Correlation Matrix

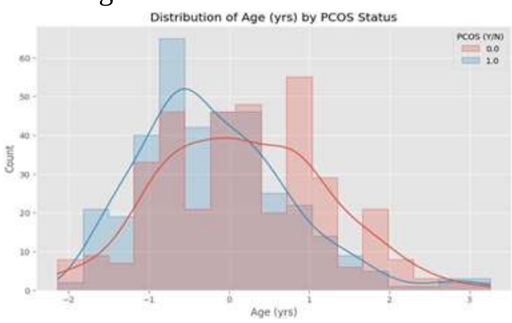

Fig: Distribution of Age(yrs) by PCOS Status

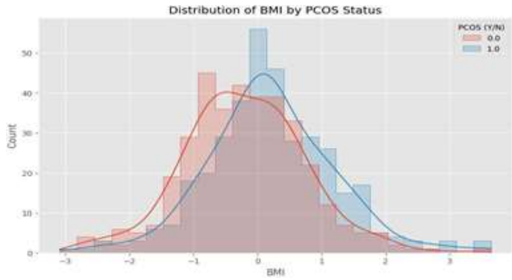

Fig: Distribution od BMI by PCOS Status

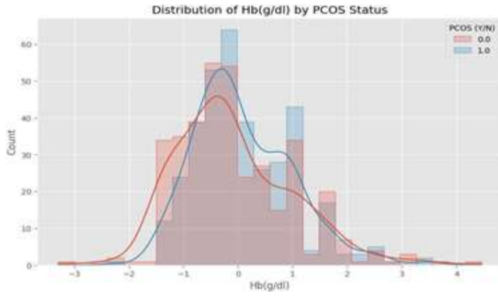

Fig: Distribution of Hb(g/dl) by PCOS Status

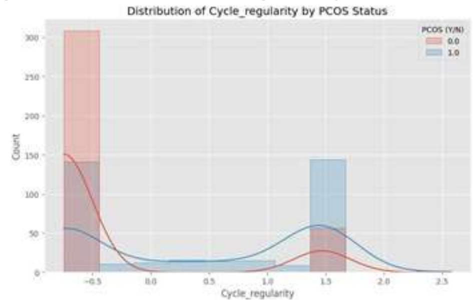

Fig: Distribution of Cycle_regularity by PCOS Status

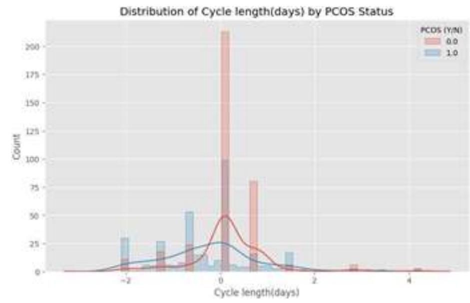

Fig: Distribution of Cycle Length(days) by PCOS Status

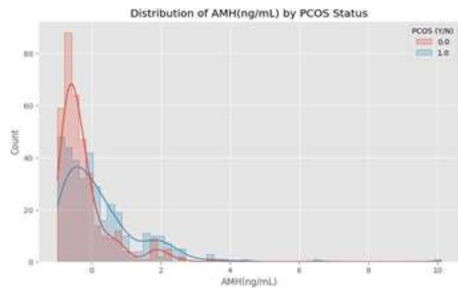

Fig: Distribution of AMH(ng/mL) by PCOS Status

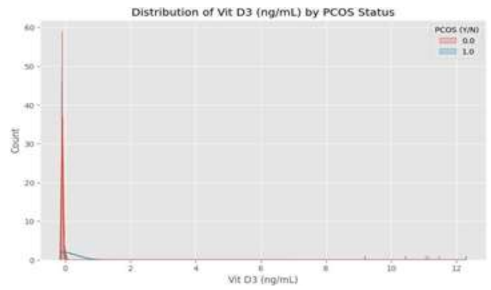

Fig: Distribution of Vit D3(ng/mL) by PCOS Status

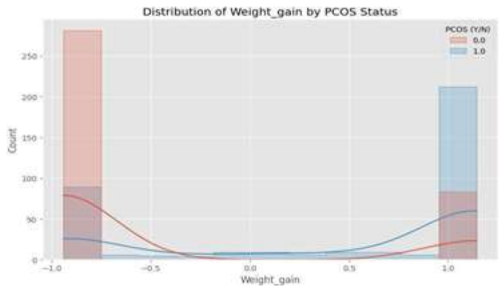

Fig: Distribution of Weight_gain by PCOS Status

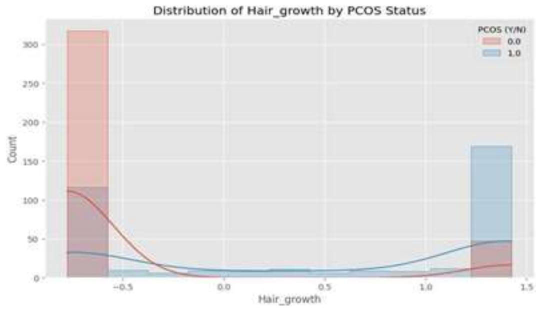

Fig: Distribution of Hair_growth by PCOS Status

VII. CONCLUSION

This ponder presents a data-driven approach to anticipate Polycystic Ovary Disorder (PCOS) in ladies utilizing machine learning calculations and include determination methods on a dataset of 541 patients collected from Kaggle. Through highlight positioning, it was watched that the proportion of Follicle Invigorating Hormone (FSH) to Luteinizing Hormone (LH) is the foremost critical indicator among the 43 highlights considered. By selecting the best 10 highlights, the models accomplished tall exactness with diminished computational complexity, demonstrating the adequacy of employing a littler, more important highlight set. Different classifiers, counting Slope Boosting, Arbitrary Woodland, Calculated Relapse, and a crossover Irregular Timberland and Calculated Relapse (RFLR), were trained and assessed. Among them, RFLR accomplished the most excellent execution, getting a testing precision of 91.01% and a review esteem of 90% with 40-fold cross-validation, beating numerous past considers whereas utilizing less highlights, in this way guaranteeing quicker computation and lower asset utilization. The consider highlights that proficient include determination combined with strong classification methods can essentially upgrade demonstrative exactness for PCOS without depending on computationally overwhelming imaging information. For future work, it is suggested to approve the discoveries on different, assorted datasets and to investigate the advancement of more modern crossover models to advance move forward classification precision, unwavering quality, and generalizability over diverse populaces.

REFERENCE

1. R. Pasquali et al., "PCOS Gathering: Investigate in Polycystic Ovary Disorder Nowadays and Tomorrow", Clin Endocrinol (Oxf), vol.74, no. 4, pp. 424–433, 2011. doi:10.1111/j.1365-2265.2010.03956.x.
2. A. S. Laganà, S.G. Vitale, M. Noventa, and A. Vitagliano, "Current Administration of Polycystic Ovary Disorder: From Seat to Bedside", Worldwide Diary of Endocrinology, 2018. doi:10.1155/2018/7234543
3. Adiwijaya, B. Purnama, A. Hasyim, M. D. Septiani, U. N. Wisesty and W. Astuti, "Follicle location on the usg pictures to back assurance of polycystic ovary syndrome", 3rd Universal Conference on Science & Building in Science, Chemistry and Material science 2015 (ScieTech 2015), vol. 622, 2015.
4. U. N. Wisesty, J. Nasri and Adiwijaya, "Modified backpropagation calculation for polycystic ovary disorder discovery based on ultrasound images", Later Propels on Delicate Computing and

Information Mining - The Moment Worldwide Conference on Delicate Computing and Information Mining (SCDM-2016), Bandung, Indonesia, pp.144-151, Admirable 18-20, 2016.
5. Adiwijaya, M. Maharani, B. Dewi, F. Yulianto and B. Purnama, "Digital picture compression utilizing chart coloring quantization based on wavelet-svd", 2013 Universal Conference on Science & Building in Arithmetic, Chemistry and Material science (ScieTech 2013), vol. 423, 2013.
6. E. Setiawati, Adiwijaya and A. Tjokorda, "Particle swarm optimization on follicles division to bolster pcos detection", 3rd Universal Conference on Data and Communication Innovation (ICoICT), pp 369-374, 2015.
7. B. Purnama, U. N. Wisesti, Adiwijaya, F. Nhita, A. Gayatri and T. Mutiah, "A classification of polycystic ovary disorder based on follicle location of ultrasound images", 3rd Universal Conference on (IEEE) Data and Communication Innovation (ICoICT), pp 396-401, 2015.
8. E. Setiawati, Adiwijaya, T. A. B Wirayuda, W. Astuti, "A Classification of Polycystic Ovary Disorder Based on Ultrasound Pictures Utilizing Directed Learning and Molecule Swarm Optimization", Progressed Science Letters, vol. 22, pp.1997- 2001, 2016.
9. P. Mehrotra, C. Chakraborty, B. Ghoshdastidar, S. Ghoshdastidar and K. Ghoshdastidar, "Automated ovarian follicle acknowledgment for polycystic ovary syndrome", Worldwide Conference on Picture Data Preparing (ICIIP), pp 1-4, 2011.
10. S. Rihana, H. Moussallem, C. Skaf and C. Yaacoub, "Automated calculation for ovarian blisters location in ultrasonogram", 2nd Worldwide Conference on Progresses in Biomedical Building (ICABME), pp 219-222, 2013.
11. J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Design Butt-centric. Mach. Intell., pp. 679-698, 1986. doi:10.1109/TPAMI.1986.4767851
12. I. F. Thufailah, Adiwijaya, U. N. Wisesty and Jondri, "An usage of Elman neural arrange for polycystic ovary classification based on ultrasound images", Worldwide Conference on Information and Data Science, vol. 971, 2017. www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos. Final gotten to on 5 May 2020.
13. B Kundan, Sangaralingam P. Combining Machine Learning and Deep Learning in the Retinopathy Diagnostic Algorithm for Enhanced Detection of DR and DME. J Neonatal Surg [Internet]. 2025Apr.2 [cited 2025Apr.9];14(5):128-40. Available from: https://www.jneonatalsurg.com/index.php/jns/article/view/2914