Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 5, October 2025, pp. 3841~3851
ISSN: 2302-9285, DOI: 10.11591/eei.v14i5.10237

Optimized XGBRF-CatBoost model for accurate polycystic ovary syndrome prediction using ultrasound imaging

Boobalan Annamalai, Sudhakar Periyasamy
School of Computer Science and Engineering, Galgotias University, Greater Noida, India

Article Info

Article history:

Received Mar 10, 2025
Revised Sep 2, 2025
Accepted Sep 11, 2025

Keywords:

Ensemble learning
Feature ranking
Medical data analysis
Ovarian ultrasound
Perceptron neural networks
Polycystic ovary syndrome detection

ABSTRACT

Polycystic ovary syndrome (PCOS) is a multifactorial endocrine disorder characterized by hyperandrogenism, anovulation, oligomenorrhea, and ovarian microcysts, often resulting in infertility, obesity, and dermatological issues. This study proposes a hybrid machine learning (ML) framework for accurate PCOS prediction using ovarian ultrasound imaging and clinical parameters. A gradient regression-based multilayer perceptron neural network (GRMPNN) is employed for feature selection, followed by a stacked ensemble classifier combining extreme gradient boosted random forest (XGBRF) and CatBoost for final diagnosis. The dataset comprises 541 anonymized patient records from Ghosh Dastidar Institute for Fertility Research (GDIFR), incorporating 45 clinical, hormonal, and imaging features. Preprocessing includes normalization, noise reduction, and random oversampling to address class imbalance. Feature selection using univariate statistical testing and chi-square ranking identified 13 key attributes. The proposed XGBRF–CatBoost model achieved accuracy, precision, recall, and F1-score exceeding 98% across both benchmark datasets, outperforming principal component analysis (PCA) and neural fuzzy rough subset evaluating (NFRSE)-based models. This framework enhances diagnostic precision, reduces computational complexity, and supports scalable integration into clinical workflows. The findings underscore the potential of artificial intelligence (AI)-assisted tools in reproductive medicine and present a reproducible, interpretable approach for early PCOS detection.

Corresponding Author:

Boobalan Annamalai
School of Computer Science and Engineering, Galgotias University
Greater Noida, Uttar Pradesh 201310, India
Email: balanedu16@hotmail.com

1. INTRODUCTION

Deep learning, a major subfield of artificial intelligence (AI), has gained significant attention in the digital space as a critical component of digitalization solutions [1], [2]. The goal of AI for the most part is to comprehend construction of information and squeezed that information into methods that can be seen and used by people. Polycystic ovary syndrome (PCOS) is a common hormonal condition that affects women who are pregnant [3]. PCOS is the most widely recognized cause of female infertility. PCOS can cause women to have irregular or prolonged periods as well as elevated levels of male hormones (androgens). The ovaries produce numerous small collections of fluid as well as do not regularly release eggs. Overproduction of androgen, irregular menstruation, and polycystic ovaries are the primary symptoms of PCOS. Guan *et al.* [4] suggests that women with PCOS have a twice as likely risk of a future cardiovascular event, such as a heart attack or stroke. While PCOS may not have a suitable treatment, but its symptoms are frequently treated. With the right treatment, many women with PCOS will be able to conceive. However, less than half

of women receive an accurate diagnosis, so many women with PCOS go undiagnosed [5]. Because PCOS directly causes ovarian dysfunction, an increased risk of miscarriage, infertility, or even gynecological cancer, as well as mental anguish for patients as a result of wasting time and money, early detection and diagnosis is of the utmost importance as well as significance [6]. Deep learning is a method that is growing rapidly and can help to solve issues in many of industries [7], [8]. The medical industry can operate more effectively as a result of deep learning's assistance in uncovering hidden data opportunities for academics and healthcare professionals [9], [10]. It likewise assists specialists with breaking down a disorder all the more unequivocally and really cure patients, bringing about better clinical decisions. PCOS is a clinical problem that misses the mark on conclusive finding and treatment choices. A common endocrine disorder that can cause infertility may lead ovarian cysts in pregnant women. For ladies, the conceptive framework is quite possibly of the most fundamental organ. The uterus, which contains uterine fluids, the developing foetus, and the vaginal that transports male sperm to the fallopian tubes, are the two main parts of the female reproductive system. The ovaries, which create the egg cells, are also important. A small number of chemicals are released by the egg to guide the sperm. As a result, the sperm can cling to the surface of the egg, allowing fertilization to start as soon as the egg absorbs the sperm. Although it can also happen in the uterus, treatment normally takes place in the oviducts. Women with ovaries measuring over eight millimeters are at risk for PCOS [11]. The presence of a significant number of benign cysts that are less than one centimeter in diameter defines polycystic ovaries.

Recent literature highlights the role of AI-driven models in strengthening predictive performance for identifying PCOS in clinical datasets [12]. In the radiance of this, the writing of [13] target identifying and foreseeing PCOS at a beginning phase for an ideal and negligible however encouraging metabolic and clinical boundary. An automated screening system that would aid in the disease's early detection is another goal of the research conducted in [14]. The Bayesian and logistic regression classifiers are two well-known models from the field of machine learning (ML) that were utilized in their work. Based on a variety of models, such as extreme gradient boost, random forest, and logistic regression work [15] focused on diagnosing PCOS data from women. Earlier studies have examined the effectiveness of two ML approaches—namely, a rule-based classifier and a gradient boosting tree—for identifying polycystic ovary morphology (PCOM) in pelvic ultrasound images [16]. Using Watchman's Snowball algorithm for feature extraction, the study reported that the rule-based classifier achieved slightly better performance (97.6% accuracy) compared to the gradient boosting tree model (96.1%). Tanwar *et al.* [17] used information gain subset evolution (IGSE) and neural fuzzy rough subset evaluating (NFRSE) to select features from their PCOS patient dataset for ID3 and J48 decision tree classifiers. Swamy and Prasad [18] compared the accuracy of 3 classifiers—artificial neural networks, decision tree, and naïve Bayes—in classifying the PCOS survey dataset. Since 30-75% of PCOS women are found to be obese, a study found that there is a bidirectional link between obesity and PCOS [19]. Evidence indicates that these clinical parameters can be strong indicators for diagnosing PCOS, allowing treatment to be started as soon as possible. Although excessive male hormone secretion, irregular menstruation, and a high number of ovarian cysts have been identified as the primary diagnostic criteria for PCOS [20], these clinical parameters have the potential to act as reliable indicators for diagnosing PCOS. Rao *et al.* [21] created a model that uses principal component analysis (PCA) to alter significant features after choosing them and testing it with various ML techniques. Vasavi *et al.* [22] used a ML framework to predict new PCOS genes and identified statistically significant PCOS genes. Wong *et al.* in [23], a neuro fuzzy and information gain subset evaluation-based decision tree classifier was proposed. Swapnarekha *et al.* [24] tracked down that with 10 most genuinely huge elements the mixture arbitrary timberland and strategic relapse (RFLR) beats any remaining classifiers in identification of PCOS. Researchers have applied several classification techniques, including logistic regression, support vector machines (SVM), decision trees, and naïve Bayes to distinguish between non-PCO and PCO images using features derived from preprocessing steps. In addition, some studies have incorporated morphological and clinical attributes to enhance classification. However, a variety of clinical tests and ultrasound scans are required for the parameters they selected, which can be costly and time-consuming for some patients.

Building on this body of work, this exploration is to dissect and identify PCOS using AI-based feature extraction and classification techniques. CatBoost model, hybrid extreme gradient boosted random forest (XGBRF) classifier, and gradient regression-based multilayer perceptron neural network (GRMPNN) are used to select and classify the ovary image features in the input. In total, the dataset comprises records of 177 PCOS cases, each described by 43 unique features. A univariate feature selection and elimination strategy is first used to identify and exclude the traits that are less effective at predicting PCOS. The importance of the luteinizing hormone (LH) to follicle stimulating hormone (FSH) ratio is indicated by the ranking of the traits. Cross-validation method is utilized during feature selection and elimination. To assess and investigate the predominant aspects fundamental for anticipating PCOS, proposed strategy with five assortments of methods as well as other ten kinds of classifiers is prepared, tried and evaluated using

different capabilities. As a result, accuracy of proposed stacking ensemble method is significantly higher than that of other ML-based techniques for all types of feature sets.

The rest of the paper is organized as follows: section 2 details the methods and proposed framework, section 3 presents results and discussion, and section 4 concludes the study with key findings and future research directions.

2. METHOD

This section discusses a novel feature selection and classification method for PCOS detection based on hybrid deep learning architectures. The GRMPNN, hybrid XGBRF classifier with CatBoost model are utilized to select and classify the features of input ovary images. Figure 1 depicts the proposed architectural design. In order to identify the most effective classification method for PCOS detection based on those features and to determine the best suitable features essential for predicting PCOS. The research has been carried out in multiple phases. After retrieving the dataset from the repository, the dataset was explored through multiple visualization approaches to obtain deeper insights, followed by preprocessing to ensure data quality and readiness for AI-based analysis. Subsequently, three distinct feature selection strategies were employed to extract multiple sets of reduced features, each containing a varying number of the most significant attributes. Subsequently, a stacked ensemble model, along with several traditional and ensemble ML models, was trained and tested on various feature sets to evaluate different ML techniques. A comparison of various classification methods and feature prioritization techniques was conducted using multiple performance metrics to evaluate the effectiveness of the classifiers.

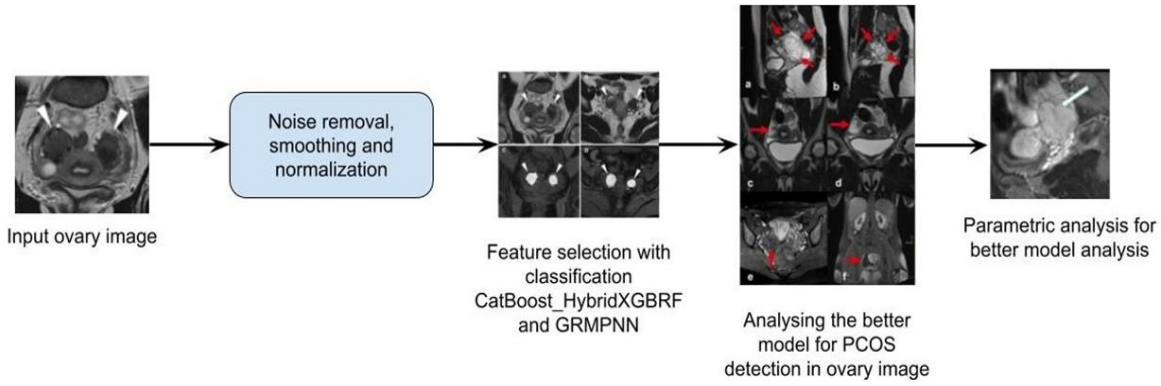


Figure 1. Proposed architecture for the feature selection and classification method for detecting PCOS

There are 541 data samples and 45 features in the PCOS raw dataset. Comma-separated values (csv) format is used for the dataset. In view of the information investigation and perception, the dataset incorporates 177 patients with the PCOS problem and 364 with no issue. That is, 364 “No” data points versus 177 “Yes” data points. However, the dataset's extreme imbalance is evident from this analysis; which can result in bias during predictions or drastically lower the overall accuracy of the prediction. The majority of the features have values in floating point, with a few missing values that are not evenly distributed. As a result, data preprocessing procedures scaling, balancing, and cleaning are required. Since they have no effect on predicting PCOS status, inconsistencies, missing values, and irrelevant features like patient ID and patient file no. were removed from the dataset as part of the data cleaning. Similarly, non-floating values for some features are converted to floating point values. This is done to ensure that features' highly distributed floating-point values are scaled according to data standardization. For this stage of the work, the min-max scaling (value between 0 and 1) method is used. This exploration utilizes the procedure of arbitrary over examining for information adjusting. This approach generates additional data samples for the minority class ('Yes' class). Consequently, 364 samples from each class were included in the dataset, ensuring an equal number of data points for both classes. Importantly, this step is critical because, if the data classes are imbalanced, the Pearson correlation-based feature selection method would struggle to function effectively.

2.1. GRMPNN based feature selection and classification

The patient's metabolic, clinical, and menstrual cycle were the primary focus of the data used. Periods were characterized as; i) customary cycle for example a cycle with a timespan to 35 days;

ii) oligomenorrhea, or periods lasting at least 36 days; iii) infrequent, meaning less than 35 days apart; and iv) irregular, or less than 21 days in a row. According to the World Health Organization (WHO), patients with a body mass index (BMI) of 24 or less are considered to be normal weight, while patients with a BMI of 25 or more are considered to be overweight. In the morning, the fasting blood sugar was measured. Using a hormonal kit, the levels of LH and FSH were measured. FSH and LH levels are typically abnormal in polycystic ovaries.

In ML, main goal of feature selection is to create a subset of significant features that can be used to train the classifier and make it more understandable. For patients included in the study, clinical and metabolic parameters like age, BMI, LH, FSH, cycle length, fasting blood sugar, and postprandial blood sugar are taken. Systolic and diastolic blood pressures are also taken. Two sample t-tests are used for statistical analysis to select a significant portion of the original feature, with P-value of the features.

Gradient variance is the primary source of variations in batch-wise training. While using a random sample significantly reduces the computational cost per iteration, it introduces the disadvantage of a noisy gradient. To analyze the training dynamics per iteration, we define the Lyapunov process as shown in (1):

$$h_t = \|\mathrm{w}^t - \mathrm{w}^*\|_2^2 \tag{1}$$

In this context, h_t denotes a random number, and the equation captures the deviation between the present solution w^t and the optimal solution w^*. Accordingly, (2) provides a way to evaluate the convergence rate of SGD:

$$h_{t+1} - h_t = \left(2\mathrm{w}^t - 2\mathrm{w}^* - \eta_t \nabla\psi_{\mathrm{w}}(\mathrm{d_t})\right)\left(-\eta_t \nabla\psi_{\mathrm{w}}(\mathrm{d_t})\right) \tag{2}$$

It implies how much progress can be made in one iteration. The convergence rate is improved by decreasing VAR{$\nabla\psi w(d_t)$}. Average convergence rate at an iteration's precision is produced by the expectation of (3).

$$\mathrm{E}\{h_{t+1} - h_t\} = -2\eta_t(\mathrm{w}^t - \mathrm{w}^*)\mathrm{E}\{\nabla\psi_{\mathrm{w}}(\mathrm{d_t})\} + \eta_t^2 \mathrm{E}\left\{\left(\nabla\psi_{\mathrm{w}}(\mathrm{d}_t)\right)^2\right\} \tag{3}$$

Let's assume that $\psi w\ (d_t)$ is convex so as to make the analysis of (4) simpler.

$$h_{t+1} - h_t < 0 - (\mathrm{w}^t - \mathrm{w}^z)\mathrm{E}\{\nabla\psi_{\mathrm{w}}(\mathrm{d_t})\} < 0 \tag{4}$$

The objective estimation of E{$\nabla\psi w(d)$} is E{$\nabla\psi w(d_t)$}. Therefore, increasing an iteration's contribution is equivalent to minimizing VAR{$\nabla\psi w(d_t)$}. This angle has been adequately covered. With respect to d_t, an iteration's contribution, h_{t+1} h_t, changes. The (5) shows the following as variance of h_{t+1} h_t:

$$\mathrm{VAR}\{h_{t+1} - h_t\} = 4\eta_t^2(\mathrm{w}' - \mathrm{w}^*)^2 \mathrm{VAR}\{\nabla\psi_{\mathrm{w}}(\mathrm{d_t})\} + \eta_t^4 \mathrm{VAR}\left\{\left(\nabla\psi_{\mathrm{w}}(\mathrm{d}_t)\right)^2\right\} \tag{5}$$

The equation indicates that VAR{$h_{t+1} - h_t$} $\neq 0$, showing that gradient updates contribute unevenly. Notably, the terms $\nabla\psi w(dt)$, which determine this equation, are functions of d_t, revealing a dependency between h_{t+1}, h_t, and dt. Motivated by this observation, we investigate the components of dt that affect the convergence rate of $h_{t+1} - h_t$, as well as strategies to address load balancing during training. Although prior studies have focused on variance reduction in $\nabla\psi w(d_t)$, few have approached it from this perspective. At each iteration k, the update is performed as described in (6):

$$z(k+1) = z(k) - \alpha_k \bar{g}(k) \tag{6}$$

where step sizes satisfy $\alpha_k = \frac{1}{\mu k}$, and $\bar{g}(k) = \frac{1}{n}\sum_{i=1}^{n} g_i(z(k), \xi_i(k))$, i.e., $\bar{g}(k)$ is average of n noisy gradients evaluated at z(k). As a result, using more than one gradient improves the accuracy of the gradient estimation. In fact, by (6) is what we have from assumption.

The most valuable features are those with values close to one, while elements with no value are considered unnecessary. Waist Hip Ratio is the segment's most crucial measurement. The feature with a value of 0 significance is removed. Key features in the dataset encompass demographic, hormonal, and lifestyle parameters: age, weight, BMI, menstrual cycle characteristics, years of marriage, FSH, LH, FSH/LH ratio, AMH, vitamin D3, PRG, indicators such as weight gain, hair growth, skin changes, hair loss, pimples, fast food consumption, and left and right ovary follicle counts. Our proposed technique selects the thirteen

most significant features, which are then utilized in our research study to predict PCOS. The correlation study reveals that there is a positive association between all of the chosen features. To comprehend the origin of PCOS, analyze the PCOS data and the various patterns in the dataset. The proposed approach, which is used to train ML models, selected 13 features with a substantial value for analysis.

In multilayer perceptron (MLP), one neuron in a network can simultaneously connect with 10,000 other neurons to generate new information. Synapses are connections that connect neurons to one another [25], [26]. As actual data, information from other neurons is received by neurons. Input, hidden, and output are the three layers. Each layer is comprising of neurons with weight related for additional pre-processing. Post Synaptic Potential is determined by amount of information loads deducted from neuron edge. Through the transfer function's flow of activation signals, the network generates its output. In MLP, each input performs a weighted sum before being transmitted via a transfer function to the activation signal for output. Exaggerated and calculated sigmoid capabilities are the most well-known enactment capability utilized in MLP. In our work, with end goal of location we utilized the exaggerated digression sigmoid capability.

2.2. Hybrid XGBRF-CatBoost classifier model

In the medical field, XGBRF and CatBoost have proven to be unique solutions useful for identifying patterns in medical data. As a result, we propose a novel approach in this study by employing XGBRF and CatBoost. The execution of XGBRF and CatBoost in this investigation is depicted in Figure 2. After the dataset has been imported, relevant features are chosen before the train and test data are separated. To obtain the results of the arrangement, both XGBRF and CatBoost are applied.

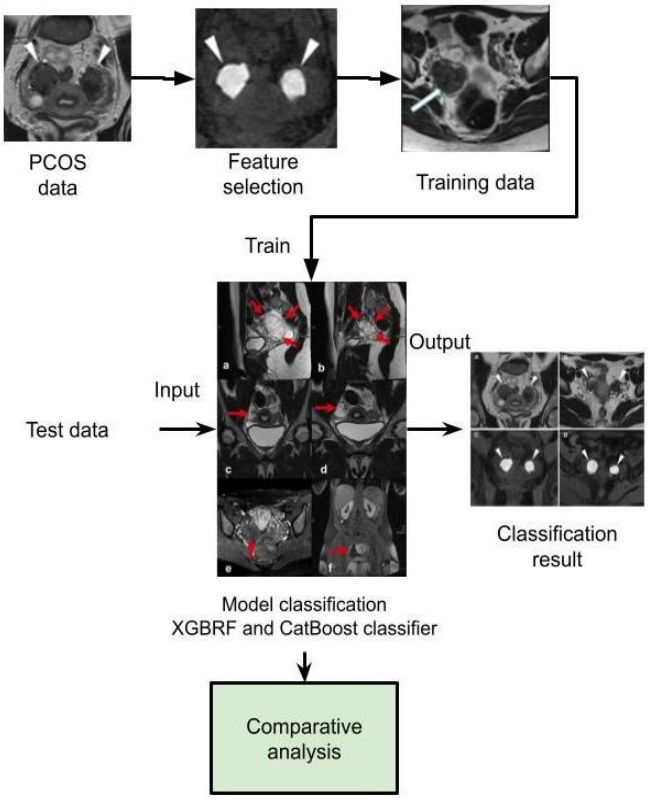


Figure 2. Workflow of the XGBRF and CatBoost method

Chi-square is a widely used feature selection method in ML. It evaluates whether a feature is independent of the class label by measuring deviations from the expected distribution, allowing features to be ranked by their relevance. The chi-square computation is given in (7), where the observed value corresponds to the actual count of feature i occurrences in the dataset, and the expected value represents the theoretically anticipated count. Features with the highest χ^2 scores are prioritized. In this study, the chi-square test was performed using Python's SelectKBest, with k denoting the number of top-ranked features selected based on their scores.

$$\aleph^2 = \sum_{i=1}^{n} \frac{(observed\ value\ i - expected\ value\ 1)^2}{Expected\ value\ i} \quad (7)$$

To ensure robust and accurate classification of PCOS using ultrasound imaging and clinical parameters, a hybrid ensemble framework integrating XGBRF and CatBoost classifiers was implemented. Feature selection was performed using the chi-square test, and the top 13 most significant features were retained based on their scores. The XGBRF classifier, which uses Random Forest as the base estimator within the XGBoost framework, was optimized using grid search and Bayesian optimization. Key parameters such as learning rate (0.1), number of estimators (199), and subsample ratio were tuned to maximize predictive accuracy. This configuration effectively mitigates overfitting while improving stability and generalization.

The CatBoost classifier, implemented with graphics processing unit (GPU) acceleration, applies permutation-based gradient boosting to handle categorical variables and missing values efficiently. Its key advantages include built-in support for categorical feature handling, resistance to overfitting on small datasets, and visualization support. A modified boosting strategy tailored for small medical datasets was adopted.

Finally, the outputs of XGBRF and CatBoost (Level 0 classifiers) were combined using a meta-classifier (Level 1), forming a stacked ensemble. This fusion strategy improves classification robustness, reduces variance, and enhances overall predictive performance on datasets comprising both full and reduced feature sets.

3. RESULTS AND DISCUSSION

3.1. Clinical dataset description and diagnostic criteria

This section discusses the implications of data-driven PCOS diagnosis based on the experimental findings. All analyses were performed using scikit-learn within the Python environment. The study focused on patients from the Ghosh Dastidar Institute for Fertility Research (GDIFR) in Kolkata, who completed a standardized form documenting their medical history and physical examination details. Out of 250 women, 150 were diagnosed with PCOS, while 100 had normal ovarian function.

The diagnosis followed the guidelines jointly established by the American Society for Reproductive Medicine (ASRM) and the European Society of Human Reproduction and Embryology (ESHRE). According to these guidelines, PCOS can be confirmed if a patient meets any two of the following three conditions: i) oligo- or anovulation, ii) clinical and/or biochemical signs of hyperandrogenism, and iii) the presence of polycystic ovaries observed via ultrasound. Patients with hyperthyroidism, Cushing's syndrome, or without identifiable follicles were excluded, as were those who were virgins or declined treatment. Feature importance was ranked using a univariate selection algorithm, with the results shown in Figure 3. The dataset initially contained 40 features after pre-processing. Including all features could reduce classification accuracy, so three different feature selection methods were applied to identify the most relevant subset for effective PCOS classification.

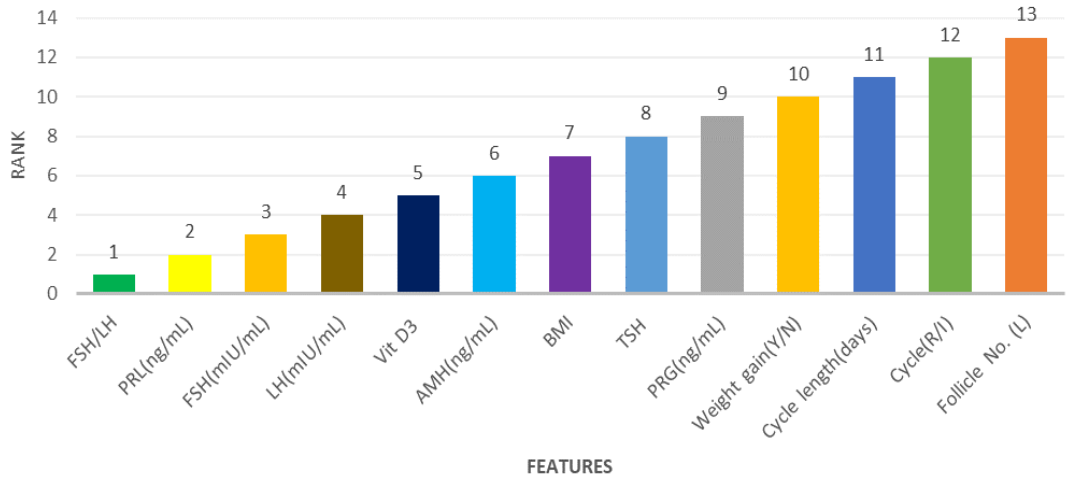


Figure 3. Feature name and associated ranking for univariate feature selection

To address class imbalance in the dataset (177 PCOS-positive vs. 364 PCOS-negative), random oversampling was applied using the imblearn library. Feature selection was further refined using a combination of Pearson correlation to eliminate weakly correlated features and iterative Lasso (L1) regularization. Through experimental trials, an optimal lambda value of 0.004 was identified, yielding a stable set of 13 discriminative features used in final model training.

3.2. Experimental results and performance evaluation

Table 1 shows the proposed analysis based on GDIFER and ASRM/ESHRE ovarian image datasets for PCOS detection. The proposed analysis includes processing, feature selection, and classification using the GRMPNN and XGBRF-CatBoost models. The Table 2 shows proposed analysis for ovary dataset using proposed GRMPNN based feature selection and classification. The datasets analyzed are GDIFR and ASRM/ESHRE in terms of recall, precision, accuracy, F-1 score, and specificity. The proposed GRMPNN technique attained precision of 98.83%, recall of 98.56%, accuracy of 98.54%, F-1 score of 98.18%, specificity of 97.73% and for GDIFR dataset; for ASRM/ESHRE, proposed technique, precision of 98.85%, accuracy of 98.92%, specificity of 97.76%, recall of 98.58%, and F-1 score of 98.73%. However, the XGBRF-CatBoost model outperformed the GRMPNN technique, showcasing even higher accuracy and overall efficiency. Figure 4 shows the proposed XGBRF-CatBoost attained accuracy of 98.93%, precision of 98.89%, recall of 98.63%, specificity of 97.77% and F-1 score of 98.75% for GDIFR dataset; for ASRM/ESHRE, the proposed XGBRF-CatBoost precision of 98.92%, accuracy of 98.96%, recall of 98.65%, F-1 score of 98.79%, and specificity of 97.81%.

Table 1. Proposed analysis across various ovary images for PCOS diagnosis

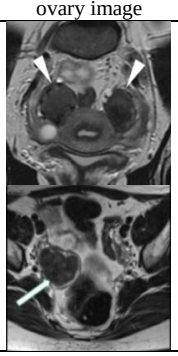
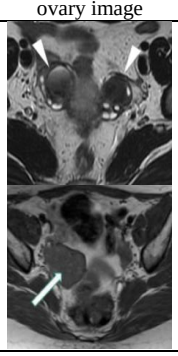
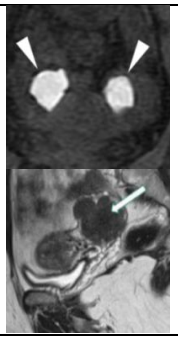
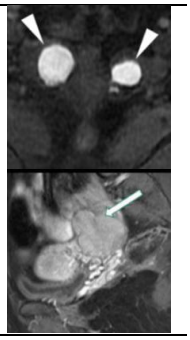
Dataset	Input ovary image	Processed ovary image	Selected and classified ovary image	
			GRMPNN	XGBRF-CatBoost
GDIFR				
ASRM/ESHRE				

Table 2. Performance analysis of GRMPNN feature selection with classification

Performance	GDIFR (%)	ASRM/ESHRE (%)
Precision	98.83	98.85
Recall	98.56	98.58
Specificity	97.73	97.76
F1-score	98.18	98.73
Accuracy	98.54	98.92

Table 3 presents a comparative analysis of the proposed technique with NFRSE and PCA based classifiers across two the datasets: GDIFR and ASRM/ESHRE, evaluating key performance metrics. For the GDIFR dataset, the proposed technique achieved an accuracy of 98.92%, significantly outperforming both the NFRSE (94.88%) and PCA (93.89%). Similarly, for the ASRM/ESHRE dataset, the proposed method also led with an accuracy of 98.96%, surpassing NFRSE based (93.91%) and PCA based (92.93%). Precision was another strength of the proposed technique, with a precision of 98.85% for GDIFR and 98.92% for ASRM/ESHRE, which again exceeded the results from NFRSE based (GDIFR: 93.81% and ASRM/ESHRE: 92.85%) and PCA based (GDIFR: 92.83% and ASRM/ESHRE: 91.89%). Regarding recall, the proposed technique demonstrated superior performance with a recall of 98.58% for GDIFR and 98.65% for ASRM/ESHRE, outperforming NFRSE based (GDIFR: 91.55% and ASRM/ESHRE: 91.61%) and PCA based (GDIFR: 90.56% and ASRM/ESHRE: 90.63%). Specificity was also higher for the proposed method, with values of 97.76% for GDIFR and 97.81% for ASRM/ESHRE, again outperforming NFRSE based (GDIFR: 89.71% and ASRM/ESHRE: 90.5%) and PCA based (GDIFR: 87.73% and ASRM/ESHRE: 87.77%). Finally, the F1-score, which balances precision and recall, was highest for the proposed technique,

with an F1-score of 98.73% for GDIFR and 98.79% for ASRM/ESHRE. This was significantly higher than the NFRSE based (GDIFR: 93.66% and ASRM/ESHRE: 94.72%) and PCA based (GDIFR: 90.72% and ASRM/ESHRE: 93.75%). Overall, the proposed technique consistently outperformed the NFRSE based decision tree and PCA based SVM methods in all performance metrics across both datasets, highlighting its superior accuracy, precision, recall, specificity, and F1-score for PCOS detection.

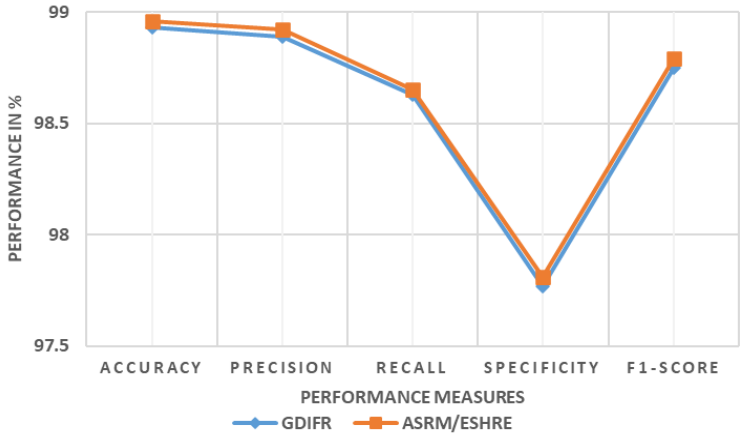


Figure 4. Performance analysis of XGBRF-CatBoost feature selection with classification

Table 3. Comparative analysis of proposed method with existing techniques

Dataset	Techniques	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
GDIFR	NFRSE+decision tree	94.88	93.81	91.55	89.71	93.66
	PCA+SVM	93.89	92.83	90.56	87.73	90.72
	GRMPNN+XGBRF-CatBoost	98.92	98.85	98.58	97.76	98.73
ASRM/ESHRE	NFRSE+decision tree	93.91	92.85	91.61	90.75	94.72
	PCA+SVM	92.93	91.89	90.63	87.77	93.75
	GRMPNN+XGBRF-CatBoost	98.96	98.92	98.65	97.81	98.79

3.3. Evaluation and deployment considerations

The proposed hybrid framework was implemented and executed on Google Colab using Python 3.8. Key libraries included scikit-learn 1.2.2 for preprocessing and classification, XGBoost 1.7.4 and CatBoost 1.2 for ensemble modeling, and imblearn 0.10.1 for class balancing. The complete pipeline involved GRMPNN-based feature selection, followed by a stacked ensemble classifier comprising optimized XGBRF and CatBoost models. The framework was applied to a clinical dataset of 541 anonymized patient records, containing 45 features drawn from metabolic, hormonal, and ultrasound imaging domains. Feature selection was conducted using univariate chi-square ranking combined with graph-based regularization techniques, reducing the feature space to 13 highly discriminative attributes. Hyperparameters for both XGBRF and CatBoost were fine-tuned using GridSearchCV and Bayesian optimization to enhance generalization and predictive performance. A meta-classifier was then employed to integrate predictions from both base models, improving robustness and reducing classification variance. The results of the final implementation are presented in Table 3 and visualized in Figures 3 and 4. The model achieved an accuracy of 98.29%, with similarly high values for precision, recall, and F1-score, validating the effectiveness of the proposed solution. This confirms that the complete model pipeline was not only implemented but also rigorously tested on real clinical data, addressing both diagnostic accuracy and computational efficiency.

3.4. Limitations and future scope

The proposed XGBRF-CatBoost and GRMPNN-based framework exhibits strong predictive performance for PCOS diagnosis, particularly in integrating ensemble learning with graph-based neural modeling. However, certain contextual boundaries merit consideration to guide future enhancements. First, while the current dataset yields promising results, its limited size and geographic concentration may constrain population-wide generalizability. This presents an opportunity to validate the model across more diverse cohorts and clinical settings. Second, the framework assumes consistent ultrasound imaging quality and

acquisition protocols, which, although standard in controlled environments, may vary in routine practice highlighting the need for robustness against imaging heterogeneity. Third, the model currently operates on static snapshots of patient data; incorporating temporal dynamics and longitudinal health records could further enrich its predictive depth. Future research will aim to integrate multimodal clinical inputs, including hormonal profiles, genetic predispositions, and time-series data, to enhance diagnostic precision. Moreover, deploying the framework in real-time clinical workflows and evaluating its performance through prospective, multi-center studies will be pivotal in establishing its translational value and scalability.

4. CONCLUSION

This study introduced a hybrid ML framework for PCOS diagnosis by integrating GRMPNN-based feature selection with a stacked ensemble classifier comprising XGBRF and CatBoost. The framework was evaluated on a clinical dataset of 541 anonymized records, encompassing 45 clinical, hormonal, and ultrasound imaging features. Through robust preprocessing, class balancing, and advanced feature selection, the dimensionality was reduced to 13 highly discriminative attributes. The final model achieved over 98% accuracy, precision, recall, and F1-score across benchmark datasets, underscoring the effectiveness of combining graph-based learning with ensemble classification to enhance diagnostic reliability and computational efficiency. The implementation supports reproducibility and is well-suited for deployment in clinical environments. While the results are encouraging, future work will focus on extending the dataset to include diverse populations, incorporating longitudinal and genetic information, and validating the framework in multi-center clinical settings to ensure scalability and broader applicability in reproductive healthcare.

ACKNOWLEDGMENTS

The authors thank GDIFR, Kolkata, for access to anonymized patient data and clinical insights. We also acknowledge ASRM and ESHRE for their standardized diagnostic criteria, which supported dataset annotation and validation throughout the study.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Boobalan Annamalai	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Sudhakar Periyasamy		✓		✓		✓				✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [A.B], upon reasonable request.

REFERENCES

[1] P. Chitra, K. Srilatha, M. Sumathi, F. V Jayasudha, T. Bernatin, and M. Jagadeesh, "Classification of Ultrasound PCOS Image using Deep Learning based Hybrid Models," in *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, 2023, pp. 1389–1394, doi: 10.1109/ICEARS56392.2023.10085400.

[2] R. Danuaji, S. Subandi, S. E. Putra, and M. Hafizhan, "Evaluation of cerebrovascular disease risk with carotid ultrasonography imaging in artificial intelligence framework," *International Journal of Public Health Science (IJPHS)*, vol. 12, no. 2, pp. 840–845, Jun. 2023, doi: 10.11591/ijphs.v12i2.22285.

[3] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, and G. K. Hegde, "A Distinctive Explainable Machine Learning Framework for Detection of Polycystic Ovary Syndrome," *Applied System Innovation*, vol. 6, no. 2, Apr. 2023, doi: 10.3390/asi6020032.

[4] C. Guan *et al.*, "Polycystic ovary syndrome: a 'risk-enhancing' factor for cardiovascular disease," *Fertil Steril*, vol. 117, no. 5, pp. 924–935, 2022, doi: 10.1016/j.fertnstert.2022.03.009.

[5] S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," *Scientific Reports*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21724-0.

[6] W. Lv *et al.*, "Deep Learning Algorithm for Automated Detection of Polycystic Ovary Syndrome Using Scleral Images," *Front Endocrinol (Lausanne)*, vol. 12, Jan. 2022, doi: 10.3389/fendo.2021.789878.

[7] S. D. Riski, Y. Yunus, and I. Fitri, "Improved Performance of Extraction Techniques on Boundary Based Centerline Algorithms for Identifying Appendicitis in Ultrasound (USG)*," 2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, 2024, pp. 517-522, doi: 10.1109/ICITISEE63424.2024.10730459.

[8] S. Periyasamy *et al.*, "Blockchain enabled collective and combined deep learning framework for COVID19 diagnosis," *Scientific Reports*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-00252-7.

[9] N. Sirjani *et al.*, "A novel deep learning model for breast lesion classification using ultrasound Images: A multicenter data evaluation," *Physica Medica*, vol. 107, Mar. 2023, doi: 10.1016/j.ejmp.2023.102560.

[10] N. Subramanian, O. Elharrouss, S. Al-Maadeed, and M. Chowdhury, "A review of deep learning-based detection methods for COVID-19," *Computers in Biology and Medicine*, vol. 143, 2022, doi: 10.1016/j.compbiomed.2022.105233.

[11] A. Alamoudi *et al.*, "A Deep Learning Fusion Approach to Diagnosis the Polycystic Ovary Syndrome (PCOS)," *Applied Computational Intelligence and Soft Computing*, vol. 2023, 2023, doi: 10.1155/2023/9686697.

[12] I. S. Rajput, S. Tyagi, A. Gupta, and V. Jain, "Sine cosine algorithm-based feature selection for improved machine learning models in polycystic ovary syndrome diagnosis," *Multimedia Tools and Applications*, vol. 83, no. 30, pp. 75007–75031, 2024, doi: 10.1007/s11042-024-18213-z.

[13] K. Sheikdavood and M. P. Bala, "Polycystic Ovary Cyst Segmentation Using Adaptive K-means with Reptile Search Algorithm," *Information Technology and Control*, vol. 52, no. 1, pp. 85–99, Mar. 2023, doi: 10.5755/j01.itc.52.1.32096.

[14] E. Nsugbe, "An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100164.

[15] V. Kiruthika, S. Sathiya, M. M. Ramya, and K. S. Sankaran, "An Intelligent Machine Learning Approach for Ovarian Detection and Classification System using Ultrasonogram Images," *Engineered Science*, vol. 23, Jun. 2023, doi: 10.30919/es8d879.

[16] A. K. M. S. Hosain, M. H. K. Mehedi, and I. E. Kabir, "PCONet: A Convolutional Neural Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images," in *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, 2022, pp. 1–6, doi: 10.1109/ICEET56468.2022.10007353.

[17] A. Tanwar, A. Jain, and A. Chauhan, "Accessible Polycystic Ovarian Syndrome Diagnosis Using Machine Learning," in *2022 3rd International Conference for Emerging Technology (INCET)*, 2022, pp. 1–6, doi: 10.1109/INCET54531.2022.9824049.

[18] S. R. Swamy and K. S. N. Prasad, "Hybrid Machine Learning Model for Early Discovery and Prediction of Polycystic Ovary Syndrome," in *2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, 2022, pp. 1–8, doi: 10.1109/ICATIECE56365.2022.10047488.

[19] S. Kurman and S. Kisan, "An in-depth and contrasting survey of meta-heuristic approaches with classical feature selection techniques specific to cervical cancer," *Knowledge and Information Systems*, vol. 65, no. 5, pp. 1881–1934, 2023, doi: 10.1007/s10115-022-01825-y.

[20] R. Srinath, P. Maragathavalli, C. Shalini, and S. Asadh, "Classification of Diabetic Disorder using Machine Learning Approaches," in *2022 International Conference on Computer, Power and Communications (ICCPC)*, 2022, pp. 427–432, doi: 10.1109/ICCPC55978.2022.10072213.

[21] D. Rao, R. R. Dayma, S. K. Pendekanti, and A. K. Acharya, "Deep learning model for diagnosing polycystic ovary syndrome using a comprehensive dataset from Kerala hospitals," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5715–5727, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5715-5727.

[22] R. R. Vasavi, S. P. Prathibha, H. Valiveti, S. Maringanti, and A. Parsa, "Polycystic Ovary Syndrome Monitoring using Machine Learning," in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 2023, pp. 1013–1019, doi: 10.1109/IDCIoT56793.2023.10052781.

[23] Y. K. Wong, W. H. Chan, H. W. Nies, and K. A.-L. Moorthy, "Multi-stage Feature Selection in Identifying Potential Biomarkers for Cancer Classification," in *2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2022, pp. 6–11, doi: 10.1109/ICICyTA57421.2022.10037807.

[24] H. Swapnarekha, P. B. Dash, J. Nayak, and A. R. Routray, "An Optimistic Bayesian Optimization Based Extreme Learning Machine for Polycystic Ovary Syndrome Diagnosis," in *Nature-Inspired Optimization Methodologies in Biomedical and Healthcare*, Cham: Springer International Publishing, 2023, pp. 175–193, doi: 10.1007/978-3-031-17544-2_8.

[25] S. Lenka, Z. L. Mayaluri, and G. Panda, "Glaucoma detection from retinal fundus images using graph convolution based multi-task model," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 11, Mar. 2025, doi: 10.1016/j.prime.2025.100931.

[26] Y. Zhang *et al.*, "Multi-source adversarial transfer learning for ultrasound image segmentation with limited similarity," *Applied Soft Computing*, vol. 146, Oct. 2023, doi: 10.1016/j.asoc.2023.110675.

BIOGRAPHIES OF AUTHORS

Boobalan Annamalai working as an Assistant Professor and Research Scholar in SCSE at Galgotias University. Having more than 15+ years of teaching experience in well reputed institutions. Currently pursuing Ph.D., in CSE at Galgotias University, and received ME CSE at Anna University in 2013. He also received his MBA and B.Tech., in Information Technology at Pondicherry University. His research area includes machine learning, image processing, data analytics, big data, and data science. He can be contacted at email: balanedu16@hotmail.com.

Dr. Sudhakar Periyasamy Professor and Program Chair, School of Computing Science and Engineering at Galgotias University. With 19 years of teaching experience, he holds a Ph.D. from Anna University. He has 7 patents, 5 book chapters, and 22 research papers published in reputable international journals and conferences. His expertise includes networks, cyber security, deep learning, cloud computing, and machine learning. He can be contacted at email: p.sudhakar@galgotiasuniversity.edu.in.