

MACHINE LEARNING

A Review on Machine Learning, Deep Learning, and Ensemble Transfer Learning Approaches with Hyperparameter Optimization for Polycystic Ovary Syndrome (PCOS) Detection from Ultrasound Sonography Images

Rutuja Nalawade, B. Gupta

Department of Information Technology, Siddhant College of Engineering, Sudumbare, Pune, Maharashtra, India

Published October 2026

Abstract

Polycystic ovary syndrome (PCOS) is one of the most prevalent endocrine diseases among women of reproductive age. The problem of diagnosing this condition on the basis of ultrasound scanning pictures is complicated by various factors, including inter-patient variability, image noise, and follicle morphology irregularities. Researchers have been gradually shifting from conventional machine learning approaches to deep learning algorithms, transfer learning, and ensembled transfer learning models fine-tuned with the help of genetic algorithms, Bayesian Optimization, Optuna, and other metaheuristic techniques in order to increase the accuracy and generalization ability of PCOS diagnostics. The present paper provides a comprehensive overview of the latest research efforts aimed at creating automated PCOS detection systems. The main findings and performance of machine learning-based, convolutional neural network (CNN)-based, and transfer learning-based approaches are discussed in detail. Moreover, the paper highlights the efficiency of hybrid approaches, which combine optimized CNNs (EfficientNetB7, DenseNet201) and a Fuzzy Inference System applied through an image enhancement module, and advanced optimization methods, such as Genetic Algorithm. Particular attention is paid to the limitations of existing PCOS detection systems, including the lack of suitable data sets, and opportunities for future research are identified.

Keywords: polycystic ovary syndrome; ultrasound imaging; machine learning; deep learning; transfer learning

1 Introduction

Polycystic Ovary Syndrome is a common hormonal disorder that involves irregular menstruation, high levels of androgen, and multiple ovarian cysts. Without proper treatment, PCOS can lead to infertility, obesity, cardiovascular diseases, and psychological problems. Transvaginal and transabdominal ultrasound scanning is the standard diagnostic procedure for evaluating ovarian morphology in PCOS. However, it is a tedious procedure, requiring the physician's experience and attention. This fact explains why numerous researchers try to develop an automatic PCOS screening system using artificial intelligence. The first studies used classical machine learning algorithms for classifying PCOS patients. However, the recent trend is to employ convolutional neural networks and transfer learning techniques in order to extract meaningful patterns from ultrasound images. Furthermore, some authors propose to combine several CNNs architectures using different pre-

trained weights. Also, instead of using traditional hyperparameter optimization procedures, researchers try to apply genetic algorithms to find optimal CNN settings.

The purpose of this study is to present up-to-date research in PCOS screening using machine learning methods, analyze trends, and identify gaps. In other words, all the relevant articles were grouped together to create a common understanding of the current state-of-the-art. Moreover, each article provided a brief description of methods used and results achieved. Finally, based on the observed information, some of the critical points were identified that could help future researchers to improve the results of PCOS screening using machine learning methods.

2 Motivation and scope of the review

The rationale behind this review arises from the fact that although many individual studies report extremely high accuracy in PCOS detection, these are obtained using widely varying databases, pre-processing pipelines, and evaluation procedures, thus raising questions about comparability. Additionally, while ensemble and genetic-algorithm-optimized transfer learning solutions reporting accuracy figures in excess of 99% (e.g., EffiDenseGenOp, which combines EfficientNetB7 and DenseNet201, utilizes Genetic Algorithm-optimized hyperparameters, and incorporates a Fuzzy Inference System for noise-aware image enhancement) have been proposed, a comprehensive comparison of individual contributions of various techniques (convolutional networks, traditional machine learning, ensemble learning, etc.) has yet to be made. This review, therefore, (i) provides an overview of the state-of-the-art in machine learning, deep learning, transfer learning, and ensemble approaches to PCOS detection; (ii) summarizes the literature on hyperparameter optimization, including the Genetic Algorithm and its application beyond the PCOS use-case; (iii) compiles a list of relevant performance figures; and (iv) identifies areas for future research.

3 Literature review

3.1 Machine learning-based approaches

A significant number of prior studies utilized classic supervised learning algorithms trained on the conventional variables rather than ultrasound images for PCOS diagnosis. The application of Random Over-Sampling and SMOTE techniques for class imbalance mitigation in conjunction with Random Forest, K-NN, Decision Tree, and XGBoost classifiers were implemented, where the latter algorithm demonstrated almost 96% accuracy on clinical data. The PCOS/PCOD records underwent similar bagging, boosting, stacking, and Random Forest procedures. In particular, Gradient Boosting achieved the best performance, reaching 91.7% when operating on clinical measurements. CS-PCOS, a method employing the chi-squared statistic for feature selection, and ten-fold cross-validation were applied alongside ten different hyperparameter-tuned ML algorithms. The simplest Gaussian Naive Bayes showed the best performance, reaching 100% accuracy on clinical data with extremely low computational complexity, while prolactin, thyroid-stimulating hormone, and blood pressure were detected as the key discriminative features. Finally, another approach using the CatBoost classifier was compared to the Bernoulli Naive Bayes, Random Forest, Decision Tree, and Logistic Regression algorithms, demonstrating nearly 93.9% accuracy. An Extreme Learning Machine was tuned using Bayesian optimization in combination with a Genetic Algorithm and showed better performance than traditional SVM and Multi-Layer Perceptron, achieving 99.31% accuracy. Although all these methods provided

high classification performance while being computationally efficient, they required proper engineering of input features to achieve high accuracy.

3.2 Deep learning and convolutional neural network approaches

In order to directly benefit from the visual information contained in ovarian ultrasound images, several studies utilized convolutional neural networks (CNNs). The first CNN-based follicle-segmentation method managed to classify PCOS and non-PCOS ovaries with nearly 83% accuracy. Another deep learning architecture called PCOSNet was introduced that achieved around 98.1% accuracy in classifying normal vs abnormal ovaries. The wavelet-domain convolutional network (PCOS-WaveConvNet) utilized a Haar wavelet 2D Discrete Wavelet Transform to decompose images into four subbands, followed by a CNN to classify PCOS and non-PCOS ovaries with close to 99.7% accuracy. Another sequential CNN model that combined a customized CNN and VGGNet16 achieved about 99.4% accuracy by harnessing the key characteristics of ovarian images. Comparing the results of self-defined CNNs and popular networks like VGG16 and ResNet50, along with their support vector machine and artificial neural network counterparts, showed that common architectures usually yield better results. In other words, transfer learning (i.e. utilizing existing networks) appears to be more beneficial than training a CNN from scratch, especially considering the limited amount of data available for training.

3.3 Transfer learning-based approaches

Due to the availability of reasonably large publicly available PCOS ultrasound datasets, transfer learning with ImageNet pretrained backbones seems to be the dominant approach. A custom ResNet model showed significantly better performance than the VGG16, ResNet50, and CNN baselines, with an accuracy of 96.7% vs. approximately 58% for the others in the same comparison. Using a pre-trained VGGNet16 for feature extraction followed by the stacking ensemble with an XGBoost meta-learner resulted in a high accuracy of 99.89% on a smaller ovarian ultrasound dataset. The same approach with fine-tuning VGG16 demonstrated strong performance for the PCOS classification task. Finally, the same accuracy of around 95% was achieved by fusing the features from VGG-16 and ResNet-50 via wavelet decomposition. Apart from the VGG- and ResNet-based architectures, one study used hybrid features with support vector machines and DenseNet to reach 97.68% accuracy, also highlighting the benefits of transfer learning via dense networks over traditional handcrafted features in this domain.

3.4 Ensemble and hybrid architectures

Because individual transfer learning backbones often focus on learning different types of visual patterns, such as efficient global context modeling and dense feature reuse, several works have explored ensembling multiple visual analysis pipelines. A hybrid model, which combines the results of VGG-16 and ResNet-50 backbones, outperformed individual architectures, which suggests that architectural diversity within the ensemble drives performance gains due to complementary discriminative pattern discovery. Attention-based hybrid visual analysis pipelines have also been explored; in particular, one work chained together an adaptive bilateral filter for noise reduction followed by an AResUNet for 2D and multi-modal image processing. This combination reportedly improved accuracy by approximately 2% compared to the baseline. The most comprehensive empirical evaluation of visual backbone ensembling comes in the form of the EffiDenseGenOp framework. It benchmarks 27 possible pairwise ensemble combinations of 8 transfer learning backbones (ResNet152V2, Inception-

V3, XceptionV2, DenseNet201, VGG19, MobileNetV2, EfficientNetB7, and NASNetMobile) and reports that the EfficientNetB7 + DenseNet201 combination achieves the best performance due to the complementary nature of their respective feature reuse mechanisms.

3.5 Hyperparameter optimization and genetic algorithm-based tuning

Hyperparameter configuration, including such variables as learning rate, dropout, dense-layer width, batch size, or optimizer, significantly impacts the performance of deep and ensembled models. Grid and random search approaches demonstrate limited efficacy in the context of high-dimensional hyperparameter spaces, prompting the adoption of metaheuristic and Bayesian alternatives in the literature. The application of Bayesian optimization to the Extreme Learning Machine within the PCOS scope resulted in an improved model after the Genetic Algorithm addition. In turn, Genetic Algorithms have been applied to a variety of tasks in medical imaging and prediction. Specifically, Many Layer Transfer Learning Genetic Algorithm (MLTLGA) has demonstrated its efficacy in terms of selecting layers for InceptionV3 fine-tuning for pneumonia classification with around 97% of accuracy. Another approach utilizing a Genetic Algorithm and an adapted Xception architecture achieved 97.62% of accuracy in COVID-19 detection from X-ray images. A similar algorithm was employed in conjunction with a heterogeneous stacking ensemble (K-NN, SVM, Random Forest, Gaussian Naive Bayes) for Alzheimer's disease detection, improving the accuracy from 70.6% to approximately 96.7%. Finally, the ResNet-50V2-based hybrid model with a Genetic Algorithm utilization achieved better performance than a Random Search or Bayesian Optimization-based counterpart in terms of acute lymphoblastic leukemia classification, with the accuracy of 98.46%. Overall, the effectiveness of Optuna and TPOT-based hyperparameter optimization in the context of cardiovascular and heart disease prediction has also been demonstrated. Thus, the efficiency of the Genetic Algorithm utilization in particular, and evolutionary computation in general, has been proved, as the accuracy was improved for all three tasks. Notably, such optimization techniques allow deep neural networks to explore lower-dimensional, non-convex, and complex optimization spaces, unlike grid search, random search, or manual search. Moreover, the combination of ensemble members' weights with their hyperparameters during optimization might be the reason why they outperformed traditional approaches. Similar conclusions were drawn by the EffiDenseGenOp study in terms of genetic optimization efficacy for densely connected networks.

3.6 Image enhancement and explainability

A common theme in recent research on PCOS detection using ultrasound scans is that image quality, noise, and attenuation are significant factors in obtaining accurate results. The EffiDenseGenOp study, for instance, introduces a Fuzzy Inference System that enhances images depending on the degree of pixel noise, estimating it using a Gaussian-Triangular membership function. This approach was chosen by the authors after testing various options on images with different levels of artificially generated noise, which were evaluated in terms of PSNR.

Meanwhile, similar research into ultrasound-specific attenuation and intensity normalization can potentially be applied to other medical fields, addressing shadowing issues in similar ways. In terms of explainability, multiple articles report on the application of both LIME and SHAP to disentangle how PCOS prediction models incorporate factors such as insulin levels and follicle numbers – in one case resulting in a 97.42% accuracy score across all samples (after combining the SMOTE-Tomek algorithm with Grey Wolf Optimization and PODBoost classifier). Additionally, visualization-based techniques allow medical professionals to see exactly which parts of an image contribute most to a

model's conclusions – for instance, using Class Activation Maps or Grad-CAM to emphasize follicles that the AI used to determine a PCOS diagnosis.

4 Comparative analysis of representative studies

Table 1 presents the representative results of the reviewed articles in this survey. The columns identify the main approach investigated by the authors, the data source used in each study, and the best result (in terms of accuracy) found by the authors over the selected benchmark. It is important to note that accuracy values are not directly comparable between the studies, since different authors used different training/test splits, different class distributions in their datasets, and different approaches to cross-validation. However, these numbers help us identify general trends in the research, such as the fact that most articles report the highest accuracy when using ensembled approaches with extensive hyperparameter optimization on top of transfer learning architectures.

Table 1. Representative studies on PCOS detection.

Ref.	Approach / technique	Data source	Reported performance
[1]	VGGNet16 feature extractor with stacking ensemble and XGBoost meta-learner	594 ovary USG images	99.89% accuracy
[2]	PCOSNet – custom CNN classifier	3856 ovary USG images	98.12% accuracy
[3]	Self-defined CNN vs. VGG16/ResNet50 and their SVM/ANN hybrids	277 ovary USG images	75%–93% accuracy range
[5]	Hybrid VGG16 + ResNet50 with 2D-DWT feature fusion	3856 ovary USG images	95% accuracy
[6]	PCOS-WaveConvNet (wavelet-based ConvNet)	3856 ovary USG images	99.7% accuracy
[7]	Custom ResNet vs. VGG16, ResNet50, CNN	Ovarian USG dataset	96.7% (custom ResNet)
[9]	CNN + VGGNet16 sequential model	1924 ovary USG images	99.4% accuracy
[26]	ELM with Bayesian-optimization and Genetic Algorithm	Clinical/physical PCOS dataset	99.31% accuracy
[28]	Random Forest, KNN, Decision Tree, XGBoost with SMOTE/ROS balancing	Clinical PCOS dataset	Up to 96% (XGBoost)
[29]	Bagging, Random Forest, AdaBoost, Gradient Boosting	541 instances, 45 features	91.7% (Gradient Boosting)
[35]	Chi-squared feature selection (CS-PCOS) with 10 ML models	Clinical PCOS dataset	100% (Gaussian Naive Bayes)
[37]	CatBoost vs. Bernoulli NB, Random Forest, Decision Tree, Logistic Regression	541 PCOS instances	93.90% (CatBoost)
[41]	Fine-tuned VGG16 transfer-learning model	Ovary USG images	Reported strong classification performance
[44]	SVM combined with DenseNet features	Ovary USG images	97.68% accuracy
[45]	Adaptive bilateral filtering with Attention Residual U-Net (AResUNet)	2D and multimodal USG images	≈2% improvement over baselines
[49]	Gaussian NB, kNN, SVM, Random Forest, DNN with LIME/SHAP explainability	Clinical PCOS dataset	97% accuracy

Ref.	Approach / technique	Data source	Reported performance
[50]	SMOTE-Tomek + Grey Wolf Optimization + PODBoost + LIME	Demographic/clinical PCOS dataset	97.42% accuracy
Base study	Ensembled EfficientNetB7 + DenseNet201 with Genetic-Algorithm hyperparameter tuning and Fuzzy Inference System	Kaggle ovary USG dataset (upsampled to 8000 images)	99.58% test accuracy, 98.97% AUC-ROC

5 Research gaps and challenges

- **Dataset scarcity and imbalance:** most prior works used only a handful of publicly available or hospital-specific ultrasound datasets, requiring heavy finetuning and upsampling for external validity, limiting generalizability across imaging devices and populations.
- **Inconsistent evaluation protocols:** differences in train/test splitting, lack of standardized benchmark datasets, and varying preprocessing pipelines between works impede meaningful comparison of performance accuracy across studies.
- **High computational cost of ensemble and Genetic Algorithm search:** while ensembling of multiple transfer networks with genetic algorithm hyperparameter search offers significant gains in accuracy, the resource costs of such an approach greatly surpass standard practice, diminishing returns on clinical utility due to complexity and computational requirements.
- **Limited external and multi-center validation:** very few prior works validated performance on ultrasound images from multiple medical institutions, scanners, or populations, as would be required for clinical deployment.
- **Interpretability methods not embedded in model design:** while some recent works have applied feature visualization and explanation frameworks like CAM, Grad-CAM, LIME, SHAP to PCOS detection networks, these are typically used only in post-hoc analysis rather than built into the design of the models themselves.
- **Dependence on image quality:** PCOS detection accuracy is often strongly tied to image noise levels or attenuation, and while some works have applied Fuzzy Inference Systems or denoising autoencoders to this task, methods for handling these difficulties in a real-world clinical setting remain an open question.

6 Future directions

- Multimodal fusion utilizing ultrasound imaging along with clinical, hormonal, and demographic data for increased diagnostic accuracy beyond visual inspection of imaging alone.
- Use of more efficient ensemble methods, possibly incorporating knowledge distillation, to preserve diagnostic performance of the ensemble while reducing computational burden at inference and training time for practical clinical use.
- Exploration of a wider range of optimization techniques beyond genetic algorithms for architecture and hyperparameter search, including but not limited to Grey Wolf, Particle Swarm, jSO, and other similar metaheuristics.
- Encouraging the use of explainable AI throughout the training process, not just as an interpretability tool but as part of the training objective itself, to further increase model reliability and clinical utility.

- Generation and publication of larger, more diverse, multi-center datasets with standardized annotations to enable better comparisons between methods trained on different data.
- Investigating the use of learned methods for noise reduction and image quality enhancement beyond fuzzy logic and attenuation correction to make models more resistant to variations in imaging quality seen in practice.

7 Conclusion

This review's analysis allows presenting the progress in the field of PCOS detection using computational methods. First, it started with classic machine learning classification methods applied to tabular data extracted from clinical questionnaires. Next, the convolutional neural networks and single transfer learning image classifiers appeared, and the current trend is to design ensembled transfer learning models optimized with genetic algorithms and advanced by fuzzy-logic image preprocessing methodologies. However, regardless of the complexity of the considered solutions, most of the researchers agreed that the best performance in PCOS detection was achieved by the combination of the efficient computer vision architecture (EfficientNetB7 or DenseNet201) and the genetic algorithm with fuzzy inference system for hyperparameter optimization (see Table 1). Moreover, while some of the latest studies tried to address the issues of limited data availability, the necessity of using a particular model for each patient, the complexity of interpreting AI-based results, and other questions, there are still some problems to be solved in order to make AI-based PCOS screening accessible and reliable for patients all over the world.

References

- [1] Brindha PG, Rajalaxmi RR. Comparative study of CNN and transfer learning techniques in the classification of PCO ultrasound images. In: *Proc. Int. Conf. Netw. Commun. (ICNWC)*. 2023. p. 1–6.
- [2] Tiwari S, Maheshwari P. PCOS-WaveConvNet: A wavelet convolutional neural network for polycystic ovary syndrome detection using ultrasound images. In: *Proc. 9th Int. Conf. Inf. Technol. Trends (ITT)*. 2023. p. 117–122.
- [3] Nagodavithana B, Ullah A. Diagnosis of polycystic ovarian syndrome (PCOS) using deep learning. In: *Proc. Int. Conf. Inf. Technol. Appl., Singapore: Springer*. 2023. p. 47–61.
- [4] Prasher S, Nelson L. Follicle prediction for polycystic ovary syndrome diagnosis from ovarian ultrasound images using CNN. In: *Proc. 10th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*. 2023. p. 789–793.
- [5] Harish KP, Dhivyanchali MN, Devi KN, Krishnamoorthy N, Sree RD, Dharanidharan R. Smart diagnostic system for early detection and prediction of polycystic ovary syndrome. In: *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*. 2023. p. 1–6.
- [6] Lakshmi MJ, Spandana DS, Raj H, Niharika G, Kodipalli A, Kamal S, et al. Prediction of PCOS and PCOD in women using ML algorithms. In: *Proc. Int. Conf. Inf. Commun. Technol. Intell. Syst., Singapore: Springer*. 2023. p. 97–115.
- [7] Srivastav S, Guleria K, Sharma S. A transfer learning-based fine tuned VGG16 model for PCOS classification. In: *Proc. 2nd Int. Conf. Intell. Data Commun. Technol. Internet Things (IDCIoT)*. 2024. p. 1074–1079.
- [8] Silambarasan E, Nirmala G, Mishra I. *Polycystic ovary syndrome detection using optimized SVM and DenseNet*. *Int. J. Inf. Technol.* ;17:1039–1047.
- [9] Bedi P, Goyal SB, Rajawat AS, Kumar M. An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting polycystic ovary syndrome. *Decis. Analytics J.* 2024;10:100366.
- [10] Jain P, Mishra RK, Deep A, Jain NK. Xplainable AI for deep learning model on PCOD analysis. In: *XAI Based Intelligent Systems for Society 5.0*. Amsterdam, The Netherlands: Elsevier. 2024. p. 131–152.
- [11] Moral P, Mustafi D, Sahana SK. PODBoost: An explainable AI model for polycystic ovarian syndrome detection using grey wolf-based feature selection approach. *Neural Comput. Appl.* 2024;36(30):18627–18644.

-
- [12] Ghosh A, Srinivasan K. EffiDenseGenOp: Ensemble transfer learning with hyperparameter tuning using genetic algorithm optimization for PCOS detection from ultrasound sonography images. *IEEE Access*. 2025;13:54285–54312.