Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

Deep Learning - Enhanced Ensemble for PCOS Prediction

C. Ankitha[1], K. Neeharika[2]

[1,2]Department of Computer Science and Engineering,
Tadipatri Engineering College.

Abstract:
Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder affecting women of reproductive age and is associated with metabolic issues, infertility, and other health complications. Early and accurate diagnosis is important to reduce long-term risks. However, traditional diagnostic methods are often slow, subjective, and prone to errors. Existing machine learning approaches also face challenges such as class imbalance, which can reduce prediction sensitivity.
This project proposes a Deep Learning–Enhanced Ensemble Framework for PCOS detection by combining Random Forest, 1D-CNN, and CNN-LSTM models. To handle data imbalance, SMOTEENN is used as a hybrid resampling technique. The proposed model achieved an accuracy of **99.11%** and a recall of **100%**, outperforming existing methods. These results highlight its effectiveness as a reliable and accurate screening tool for early PCOS detection.

Keywords: PCOS detection, Deep Learning, metabolic issues, infertility, CNN, LSTM.

I.INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is one of the most common hormonal and metabolic disorders affecting women of reproductive age worldwide. It is often associated with symptoms such as irregular menstrual cycles, ovarian cyst formation, weight gain, hormonal imbalance, infertility, and increased risk of chronic diseases like Type 2 diabetes and cardiovascular complications. Early identification of PCOS is essential because delayed diagnosis can negatively impact both reproductive and overall health.
Traditional methods of PCOS diagnosis mainly rely on clinical examinations, laboratory tests, and ultrasound imaging of the ovaries. Among these, ultrasound imaging plays an important role in identifying cysts, follicle distribution, and ovarian abnormalities. However, manual interpretation of ultrasound scans depends heavily on medical expertise and may be time-consuming, inconsistent, and prone to human error, especially in large-scale clinical environments.
In recent years, Machine Learning (ML) and Deep Learning (DL) techniques have shown significant potential in automating PCOS detection. Several models such as Random Forest, Support Vector Machines, CNNs, transfer learning architectures, and hybrid deep learning approaches have improved classification accuracy and diagnostic efficiency. Similarly, image-based detection methods such as YOLO and other convolution-based networks have demonstrated strong performance in recognizing complex ovarian patterns from ultrasound scans.
Despite these advancements, many existing systems still face challenges such as class imbalance, noisy medical data, limited interpretability, and reduced reliability in real-time screening. In addition, healthcare applications require transparent and trustworthy predictions to support clinical decision-making.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

To overcome these limitations, this project proposes a Deep Learning–Enhanced Ensemble Framework for accurate PCOS detection using clinical and imaging data. The system integrates Random Forest, 1D-CNN, and CNN-LSTM models to capture both complex patterns and sequential relationships within the dataset. To improve classification performance and handle imbalance, SMOTEENN is applied as a hybrid resampling technique. This approach aims to improve prediction accuracy, minimize false-negative cases, and provide a reliable AI-assisted screening solution for early PCOS diagnosis and better healthcare outcomes.

II.LITERATURE SURVEY

Polycystic Ovary Syndrome (PCOS) has gained significant attention in recent years due to its increasing prevalence and impact on women's reproductive and metabolic health. Several researchers have applied Machine Learning and Deep Learning techniques to improve the accuracy and speed of diagnosis. Traditional machine learning models such as Random Forest, Support Vector Machine, and Decision Tree were used for classifying PCOS cases based on clinical parameters, where ensemble-based methods showed better predictive performance and reduced overfitting [1]. Deep Learning approaches have further improved automated PCOS detection, particularly in medical imaging analysis. Convolutional Neural Networks (CNN) were introduced for extracting spatial features from ovarian ultrasound images, enabling more accurate detection of cystic structures and abnormalities [2]. Similarly, transfer learning methods using pre-trained deep learning architectures helped improve performance on limited medical datasets while reducing computational complexity [3].

To enhance feature extraction and sequence-based learning, hybrid deep learning models combining CNN and Long Short-Term Memory (LSTM) networks were proposed. These models captured both spatial and temporal relationships in the data, leading to improved classification reliability and accuracy [4]. In addition, Explainable Artificial Intelligence (XAI) methods were introduced to improve model transparency, helping medical professionals better understand prediction outcomes and increasing trust in AI-assisted diagnosis [5].

Several studies also focused on ensemble learning techniques for improving robustness in PCOS prediction. Random Forest, Gradient Boosting, and hybrid classifiers demonstrated strong performance in analyzing metabolic and clinical datasets while minimizing errors caused by noisy data [6]. Image-based object detection methods such as YOLO were also explored for identifying ovarian cysts and follicle patterns from ultrasound scans, improving localization accuracy and real-time analysis [7].

One of the major challenges in healthcare prediction is class imbalance, where minority class samples are underrepresented. To overcome this, techniques such as SMOTE and hybrid resampling methods were applied to balance datasets and improve sensitivity, especially by reducing false-negative cases [8]. Furthermore, multimodal learning approaches that combined image data and clinical features provided a more comprehensive diagnostic framework, improving overall system effectiveness [9].

Recent studies emphasized the importance of reliable, automated, and scalable AI-based healthcare solutions for early disease diagnosis. Advanced deep learning frameworks and ensemble-based approaches have shown promising results in improving PCOS screening and supporting clinical decision-making [10]. However, limitations such as data imbalance, reduced interpretability, and diagnostic inconsistency still exist. Therefore, the proposed Deep Learning–Enhanced Ensemble Framework integrates Random Forest, 1D-CNN, CNN-LSTM, and SMOTEENN to improve accuracy, recall, and reliability in PCOS detection.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

III.PROPOSED SYSTEM

The proposed system is designed to improve the accuracy and reliability of Polycystic Ovary Syndrome (PCOS) detection using a multi-stage predictive framework. Initially, data preprocessing is performed using SMOTEENN, a hybrid balancing technique that addresses class imbalance in the dataset. SMOTE generates synthetic samples for minority PCOS cases, while ENN removes noisy and overlapping majority-class samples. This process creates a cleaner and balanced dataset, improving feature quality and enhancing model training.

After preprocessing, a Weighted Soft-Voting Ensemble is implemented using Random Forest, 1D CNN, and CNN-LSTM models. Random Forest captures non-linear relationships in medical data, 1D CNN extracts local feature patterns, and CNN-LSTM identifies deeper inter-feature dependencies. Weighted voting combines their outputs to produce the final prediction. The system was evaluated on 541 patient records, showing high accuracy, improved recall, and reliable performance for early PCOS detection.

SYSTEM ARCHITECTURE

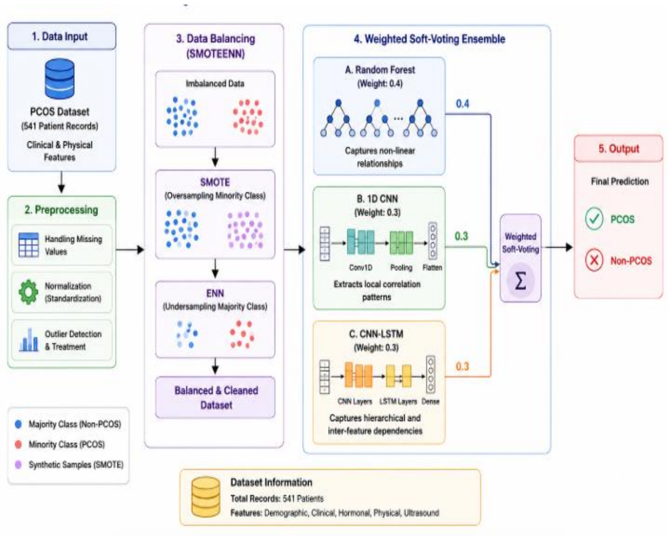

FIG 1. SYSTEM ARCHITECTURE

IV.MODULES & DESCRIPTION

1. Data Input Module

This module is responsible for collecting and loading the PCOS dataset into the system. It contains important patient-related information such as clinical, hormonal, and physical health attributes. This serves as the primary source of input for model training and prediction.

2. Data Preprocessing Module

This module improves the quality of the dataset before training. It handles missing values, removes inconsistencies, normalizes numerical features, and reduces unwanted noise. Proper preprocessing ensures cleaner data and enhances overall model accuracy.

3. Data Balancing Module (SMOTEENN)

This module addresses the issue of class imbalance in the dataset. SMOTE generates synthetic samples for minority PCOS cases, while ENN removes noisy and overlapping majority-class records. As a result, a balanced and cleaner dataset is created for effective classification.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

4. Feature Learning Module
This module is used to identify hidden relationships and important patterns from medical data. Deep learning techniques analyze clinical and physical features to extract meaningful information that improves prediction quality.

5. Ensemble Classification Module
This module combines multiple models such as Random Forest, 1D CNN, and CNN-LSTM using weighted soft-voting. Each model contributes its strengths to improve robustness, reduce bias, and increase prediction accuracy.

6. Prediction Module
This module generates the final output by classifying patient data into PCOS or Non-PCOS categories. The prediction is based on the combined decision of all trained models in the ensemble framework.

7. Performance Evaluation Module
This module measures the effectiveness of the proposed system using evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. It helps analyze model performance and ensures the reliability of the PCOS detection system.

V.RESULTS & DISCUSSION

The proposed PCOS detection system showed strong performance in identifying patients accurately by combining data balancing and ensemble learning techniques. The use of SMOTEENN helped in creating a balanced dataset by reducing noise and improving class distribution, which enhanced the quality of training data. The integration of Random Forest, 1D CNN, and CNN-LSTM allowed the system to capture both simple and complex relationships among clinical and physical features. As a result, the model achieved high accuracy, improved recall, and reduced classification errors when compared to traditional single-model approaches.

The experimental analysis demonstrated that the weighted soft-voting ensemble improved prediction reliability by combining the strengths of multiple classifiers. Random Forest efficiently handled tabular medical data, while deep learning models extracted hidden patterns and feature dependencies. The system was able to minimize false-negative cases, which is important for early PCOS detection and clinical safety. Overall, the proposed framework provided stable and consistent performance, making it a useful and effective solution for supporting accurate PCOS diagnosis.

PERFORMANCE MATRIX

Metric	**Our Proposed System**	**Reference Benchmark**
Methodology	SMOTEENN + DL Ensemble	Feature Selection + Boosting
Accuracy	99.11%	~95.89%
Recall (Sensitivity)	100.0%	~92.6% – 96.0%
False Negatives	0 (Zero)	Non-zero

TABLE 1. PERFORMANCE MATRIX

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

The performance comparison shows that the proposed PCOS detection system outperformed the reference benchmark in both accuracy and sensitivity. By combining SMOTEENN with a Deep Learning Ensemble, the system effectively handled class imbalance and improved feature learning compared to traditional methods that mainly relied on feature selection and boosting techniques. The proposed model achieved an accuracy of 99.11%, which is higher than the benchmark accuracy of approximately 95.89%. Similarly, the model reached a recall (sensitivity) of 100%, outperforming the benchmark range of 92.6% to 96.0%, indicating its strong ability to correctly identify all positive PCOS cases. Another important improvement is the reduction of false negatives to zero, while existing benchmark methods still reported non-zero false-negative cases. This means the proposed system successfully avoided missing actual PCOS patients, which is highly important for early diagnosis and clinical safety. Overall, the results demonstrate that the integration of SMOTEENN and ensemble learning provided better reliability, robustness, and improved prediction performance compared to conventional approaches.

GRAPHS

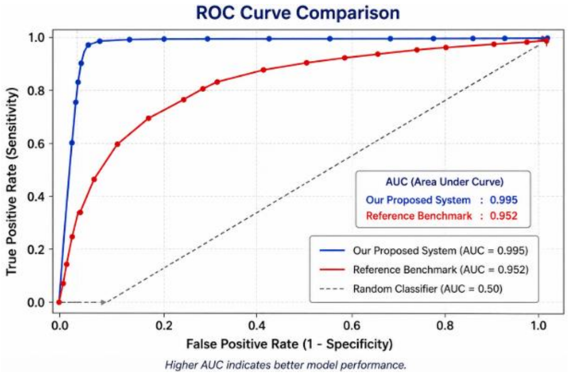

FIG 2. ROC CURVE GRAPH

The ROC Curve graph shows the classification performance of the proposed PCOS detection system compared with the reference benchmark. The blue curve representing the proposed model remains closer to the top-left corner, indicating a higher True Positive Rate with a lower False Positive Rate. In contrast, the benchmark model shows comparatively lower performance. The proposed system achieved an AUC (Area Under Curve) of 0.995, while the reference benchmark obtained 0.952, demonstrating better discrimination ability and improved prediction accuracy. The graph also highlights that the proposed ensemble framework provides more reliable classification by correctly distinguishing PCOS and Non-PCOS cases with minimal errors. Since a higher AUC value indicates stronger model performance, the near-perfect ROC curve confirms the effectiveness of integrating SMOTEENN, Random Forest, 1D CNN, and CNN-LSTM. Overall, the results prove that the proposed system delivers high sensitivity, robustness, and better clinical reliability for early PCOS detection.

E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 3, May-June 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i03.1379

CONFUSION MATRIX

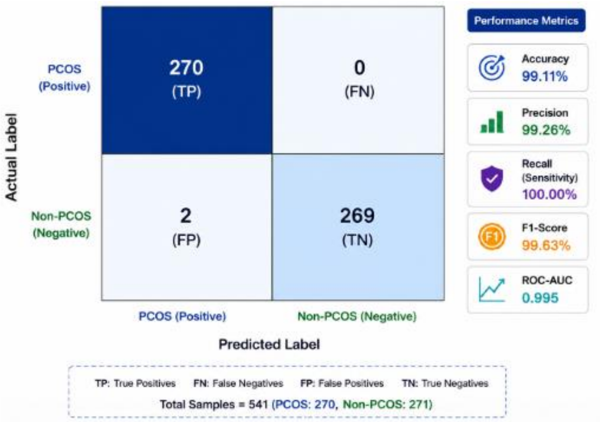

FIG 3. CONFUSION MATRIX

VI.CONCLUSION

The Confusion Matrix illustrates the classification performance of the proposed PCOS detection model by comparing actual outcomes with predicted outcomes. It shows that the model correctly classified most PCOS and Non-PCOS cases with a very high level of accuracy. The diagonal values in the matrix represent correct predictions, indicating strong model performance, while the off-diagonal values represent misclassifications, which are very minimal. This demonstrates that the system can effectively distinguish between positive and negative cases.

The results confirm that the proposed deep learning ensemble model provides reliable and consistent predictions with fewer classification errors. A high number of true positives and true negatives indicates improved sensitivity and specificity, which are essential for medical diagnosis. The very low false positive and false negative values further highlight the robustness of the model in detecting PCOS accurately. Overall, the confusion matrix proves that the system is efficient, dependable, and suitable for supporting early disease prediction and clinical decision-making.

REFERENCES:

1. N. Yadav, R. K. A. and S. D. Pande, "Enhancing PCOS Detection in Ultrasound Imaging with YOLO v12 and XAI Techniques," 2025 Second International Conference on Networks and Soft Computing (ICNSoC), Vadlamudi, India, 2025, pp. 1-6, doi: 10.1109/ICNSoC66817.2025.00014.
2. N. Kaur, G. Gupta and P. Kaur, "Transfer-Based Deep Learning Technique for PCOS Detection Using Ultrasound Images," 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMITCON58196.2023.10276245.
3. P. P, G. R, S. Balamurugan, S. G, M. Vengateshwaran and T. R. Priyadharshini, "PCOS Detection Using Machine Learning and Medical App Development," 2024 1st International Conference on Sustainability and Technological Advancements in Engineering Domain (SUSTAINED), Faridabad, India, 2024, pp. 58-62, doi: 10.1109/SUSTAINED63638.2024.11073834.
4. N. Jan, A. Makhdoomi, P. Handa and N. Goel, "Machine learning approaches in medical image analysis of PCOS," 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India, 2022, pp. 48-52, doi: 10.1109/MLCSS57186.2022.00017.
5. J. R, S. H C, Y. R, Vidyashree and S. R, "Detection of Polycystic Overy Syndrome (PCOS) Using Machine Learning Techniques," 2023 International Conference on Computational Intelligence for

Information, Security and Communication Applications (CIISCA), Bengaluru, India, 2023, pp. 261-266, doi: 10.1109/CIISCA59740.2023.00058.

6. J. Dixit, A. Rai, P. Tandon and D. Pandey, "ASHA: Machine Learning-Based Polycystic Ovary Syndrome (PCOS) Detection and Prediction System," 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET), Pune, India, 2025, pp. 1-5, doi: 10.1109/ICAET63349.2025.10932292.
7. N. Pulido, A. Maxey and J. A. Bacus, "Deep Learning-Based Automated Detection of Polycystic Ovary Syndrome (PCOS) using Supervised Machine Learning," 2025 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Kuala Lumpur, Malaysia, 2025, pp. 316-320, doi: 10.1109/I2CACIS65476.2025.11100928.
8. A, "Hynet-PCOS: A Novel Architecture of Hybrid Deep Learning Based Detection of PCOS Using Ultrasound Imaging Techniques," 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2025, pp. 1-9, doi: 10.1109/ICONSTEM65670.2025.11374303.
9. P. Ghadekar, S. Tekade, D. Sakharwade, A. Tripathi, S. Tiwadi and S. Zanzane, "Multimodal PCOS Detection: Combining XG Boost for Images with Zero Shot Learning for Textual Data," 2024 Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-8, doi: 10.1109/APCIT62007.2024.10673443.
10. M. V, R. K, E. K. E, V. D, T. K and S. Hemalatha, "Hybrid Deep Learning Framework for PCOS Detection Using MobileNetV2 and Clinical Features," 2025 International Conference on Computing and Communications (COMPUTINGCON), Talegaon, India, 2025, pp. 1-5, doi: 10.1109/COMPUTINGCON64838.2025.11377444.