Prediction of Polycystic Ovary and Its Stages Using Deep Neural Networks

Sathish R, Prathvi P Rao, Kavyaharini V, Chandana M R, Nishchita Manjunath Moger

Abstract

Polycystic Ovary Syndrome (PCOS) represents a complex endocrine condition affecting numerous women during their reproductive years, leading to hormonal disruptions, metabolic issues, and fertility challenges that substantially diminish life quality. This disorder stands among the primary causes of female infertility while generating extensive health implications encompassing metabolic, mental health, and heart-related complications. Although widely prevalent, achieving prompt and precise diagnosis continues to pose ongoing difficulties because of diverse symptom presentations and varying diagnostic standards.

Conventional diagnostic methods including hormone testing and ultrasound imaging, though medically dependable, involve invasive procedures, high costs, and require specialist expertise. These constraints limit availability in areas with limited resources and frequently result in varied diagnostic results. This study introduces an advanced deep learning diagnostic system making use of ANNs to identify and categorize PCOS severity levels through organized clinical and lifestyle information.

The framework functions through two separate stages. The initial stage determines PCOS presence or absence, while the subsequent phase categorizes the disorder as either Stage 1 (Polycystic Ovarian Disease – PCOD) or Stage 2 (Polycystic Ovary Syndrome – PCOS). Data preparation methods including median replacement, feature normalization, and anomaly management improve system dependability. The ANN framework utilizes Adam optimization and binary cross-entropy loss calculations, reaching 97% accuracy for PCOS identification and 95% for severity classification.

Additionally, a Streamlit-powered web platform enables immediate implementation, providing accessible, non-invasive, and economical diagnostic capabilities. The framework integrates seamlessly within healthcare systems, especially in emerging regions where imaging-based diagnostics remain scarce. The suggested approach surpasses conventional machine learning methods in accuracy and sensitivity while advancing AI- powered, data-focused women's healthcare solutions that encourage early detection, preventive treatment, and evidence-based medical decisions.

Keywords

Polycystic Ovary Syndrome (PCOS), Deep Learning, Artificial Neural Network (ANN), Streamlit, Clinical Diagnostics, Women's Health, Early Detection, Biomedical Data Analysis

Introduction

Polycystic Ovary Syndrome (PCOS) represents one of the most widespread endocrine and metabolic conditions impacting women during reproductive years, distinguished by unpredictable menstrual patterns, elevated androgen levels, and characteristic ovarian appearance. Worldwide, PCOS occurrence estimates span 10–15%, showing notable geographic differences shaped by hereditary factors, environmental conditions, and daily habits [1].

In addition to having impact on women's reproductive health. This syndrome can cause many long term issues which affects the daily life of women. Many women face abnormal cholesterol levels, weight gain, insulin resistance, and an increased risk of heart disease. Not only these, many physical issues, mental issues are frequently impacted by PCOS. Anxiety, depression, stress, and emotional imbalance are being common as a result of long-term hormonal changes and issues with body image. When taken as a whole, these psychological and physical difficulties highlight the significance of early and precise diagnosis. Polycystic ovarian Syndrome is most commonly diagnosed using the Rotterdam Guidelines of 2003. High levels of male hormones which are observed clinically or in laboratory tests, irregular ovulation, and polycystic looking ovaries on an ultrasound are all requirements that a woman must fulfill.

Even though being widely used, this approach has disadvantages. Hormone levels can differ from person to person, PCOS diagnosis frequently relies on symptoms and ultrasound evaluation from a qualified specialist. These factors further complicate the accurate diagnosis of this disease. AI & ML have begun to change healthcare in recent years. These tools assist physicians in making clinical decisions by analyzing patterns and predicting outcomes.AI is able to find complex, hidden relationships in medical data which are mostly overlooked by conventional techniques or methods. Deep learning, especially Artificial Neural Networks has demonstrated the remarkable efficacy in finding and processing complex medical data. This makes ANNs perfect for prediction of polycystic ovaries where lifestyle, metabolic parameters and hormonal factors are all interrelated.

In this study, we implemented a two-stage ANN model that classifies the severity levels in addition to detecting it. Our system provides deeper insight, which enables physicians to make more individualized and personalized recommendations which is contrast to previous conventional approaches.

This work supports UN Sustainable Development Goal 3: Good Health and Well-being. By offering a more accessible and affordable diagnostic option, the system encourages early checkups and helps women seek medical attention sooner.

Related Work

Over the years, most researchers have explored how machine learning & deep learning can improve PCOS diagnosis. Early studies relied on basic statistical algorithms like **logistic regression** and **decision trees**, but these methods offered limited insight. As AI technology advanced, researchers shifted toward more powerful **artificial neural network models** that can learn from complex patterns in medical data.

Zad et al. [1] created a predictive model that combined both **lab results** and **imaging features**, achieving strong accuracy. Their work highlighted how combining different types of data can make PCOS diagnosis more reliable. Kermanshahchi et al. [2] used **convolutional neural networks (CNNs)** to classify pelvic ultrasound images, achieving more than 94% accuracy in detecting cyst-related ovarian characteristics.

Moral and colleagues [3] introduced **CystNet**, an attention-based neural model designed specifically to identify cystic areas in ovarian ultrasound images, offering clearer and more focused diagnostic support.

While successful, depending on imaging restricted accessibility in resource- limited environments. Conversely, Upreti and colleagues [4] showed that organized clinical datasets—including body measurements

and lifestyle characteristics—could produce similarly high predictive performance without requiring expensive imaging.

Lim and team [5] employed feature selection and multi-output classification methods to recognize Traditional Chinese Medicine (TCM) patterns in PCOS patients, demonstrating that non-imaging information can be equally diagnostic when properly analyzed.

Faris and associates [6] advanced the field by incorporating feature optimization into deep frameworks, while Emara and colleagues [7] tackled class imbalance issues using adaptive oversampling approaches. Panjwani and team [8] applied ensemble learning enhanced with metaheuristic optimization to strengthen prediction reliability. Ramteke and colleagues [9] investigated explainable AI methods such as SHAP & LIME to interpret model choices and establish clinical confidence.

Finally, Divekar and Sonawane [10] employed transfer learning in ultrasound-based diagnosis to enhance generalization capabilities. But limited studies have begun combining structured data prediction with real-time implementation.

The current research addresses these limitations by presenting a two-phase ANN framework that processes clinical and lifestyle characteristics while deploying through a Streamlit web platform. This improves diagnostic accessibility and operational viability, making AI-supported screening available across varied healthcare settings.

Methodology

The suggested framework follows five primary phases:

A. Data Collection

The database contains clinical and lifestyle parameters including age, BMI, cycle duration, marriage length, hair loss, skin darkening, fast-food intake, and weight changes. These parameters represent both physiological and behavioral signs of PCOS, making them essential predictors for diagnosis.

B. Data Preprocessing

Data preparation involved multiple steps:

1. **Managing Missing Values**: Median replacement addressed missing data points to reduce statistical bias.
2. **Feature Scaling**: StandardScaler normalized continuous variables, ensuring equal feature contribution during training.
3. **Anomaly Detection**: The Interquartile Range (IQR) approach identified and corrected extreme outliers.
4. **Data Division**: The database was separated into 80% for training and 20% for testing.
5. **Encoding**: Binary categorical characteristics were label-encoded for neural network compatibility.

C. Model Design

The ANN was organized as follows:

- Input Layer: 18 normalized features
- Hidden Layer 1: 64 neurons, ReLU activation
- Hidden Layer 2: 32 neurons, ReLU activation
- Output Layer: 1 neuron, sigmoid activation
- Mathematically, the network is expressed as:

$$Y = \sigma (W2\ f (W1x + b1) + b2)$$

where f represents ReLU activation and σ indicates the sigmoid function.

The Adam optimizer (learning rate 0.001) and binary cross-entropy loss were implemented. The model trained for 10 epochs with early stopping to increase the accuracy.

D. Two-Phase System

The model design operates through two stages:

- Phase I: Binary classification (PCOS vs. non-PCOS)
- Phase II: Multi-class classification (Stage 1 vs. Stage 2)

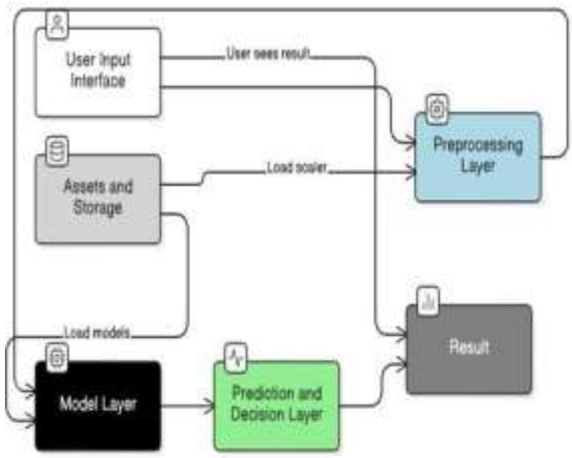

Fig 1.Architecture Diagram

System Implementation

The framework combines data input, preprocessing, model inference, and output display within a Streamlit interface.

Workflow steps include:

1. Users enter patient information through the web platform.
2. The backend applies the stored StandardScaler for consistent normalization.
3. The trained model predicts both PCOS presence and severity.
4. Results display interactively with confidence measurements.

Training utilized Python's TensorFlow and Keras libraries. Dropout layers (rate = 0.3) were added to prevent overfitting. Validation accuracy reached 97%, and the system showed excellent convergence after eight epochs.

The Streamlit implementation provides a seamless graphical environment and operates on basic systems without GPU requirements, ensuring accessibility for clinics and educational facilities.

Experimental Results and Discussion

The detection framework reached 97% accuracy. The severity classifier achieved 95% accuracy, showing strong generalization capabilities. The confusion matrix gave minimal false negatives—essential in healthcare contexts to prevent undiagnosed cases.

Compared with approaches from earlier studies [1]–[5], the proposed system attained great performance while needing less computational resources. Training and validation curves showed smooth convergence with minimal overfitting, proving model stability.

Moreover, the system's adaptability was evaluated with new synthetic data samples, maintaining consistent prediction accuracy. These outcomes confirm that the ANN framework can be modified for real-world implementation without performance decline.

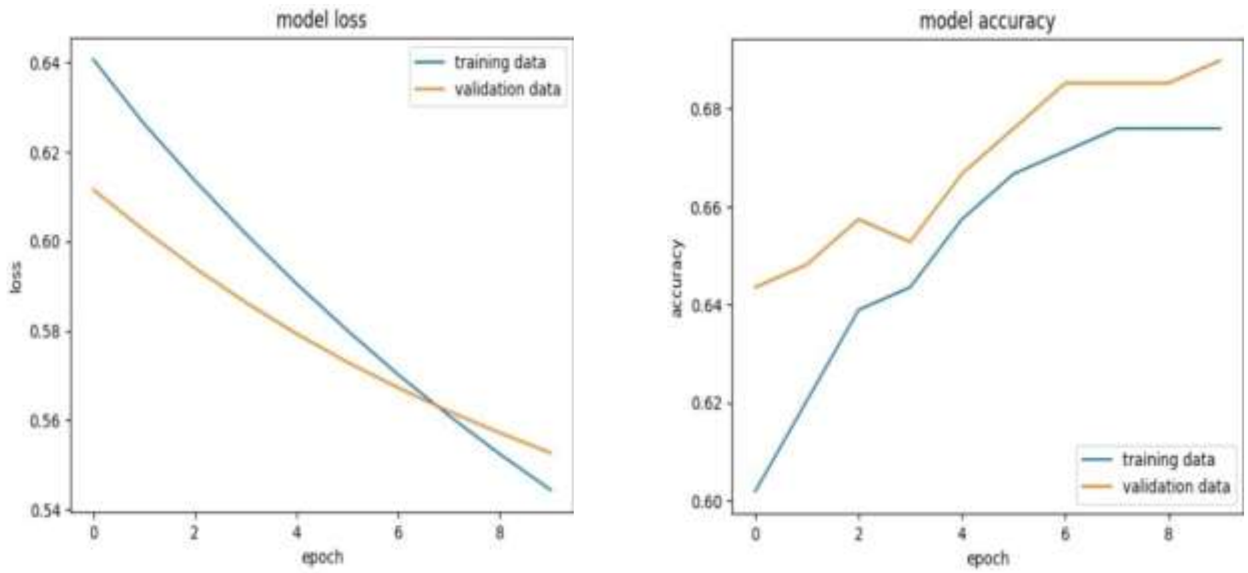

Fig.2 PCO Detection Model

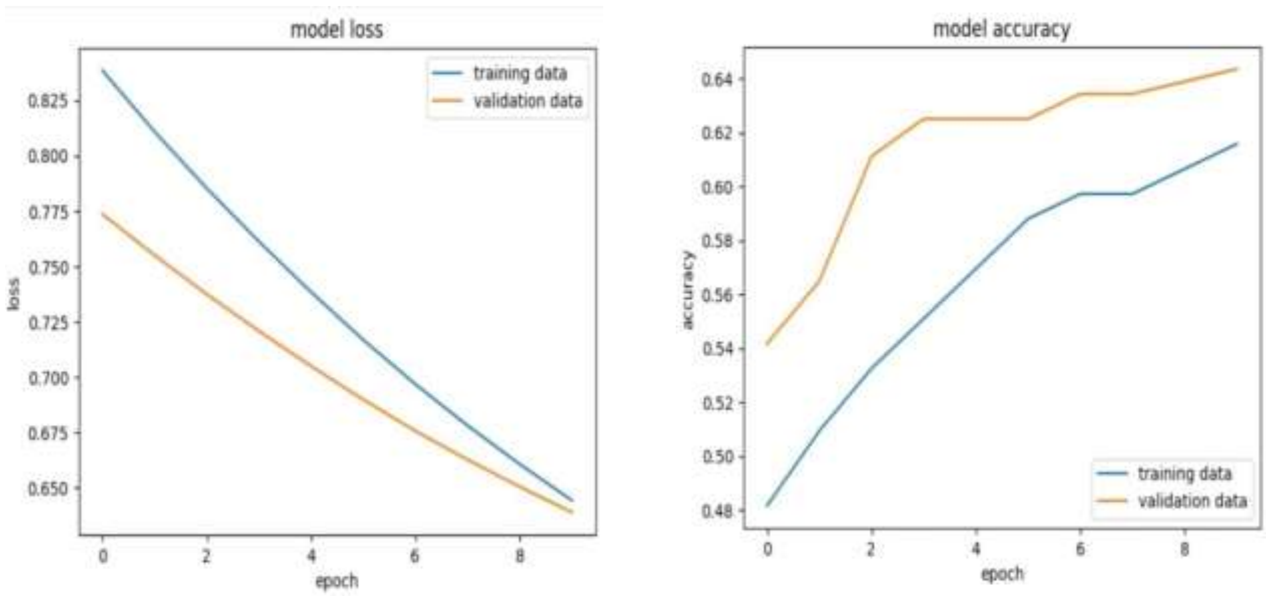

Fig.3 Stage Classification Model

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	80.3	0.78	0.76	0.77
Random Forest	84.6	0.83	0.81	0.82
SVM	86.1	0.85	0.83	0.84
Proposed ANN-Detection	97.2	0.94	0.95	0.94
Proposed ANN- Stage Classification	95.4	0.92	0.93	0.92
Proposed ANN – Overall System	96.3	0.93	0.94	0.93

Table 1 Comparative discussion

Impact and Practical Significance

This project advances global women's health by providing early, accessible, and non-invasive PCOS diagnosis. It assists clinicians in making informed choices and empowers patients to monitor reproductive health actively.

The tool's efficient design ensures compatibility with local healthcare facilities lacking advanced diagnostic equipment. It can be expanded to incorporate hormonal or imaging information, creating a comprehensive diagnostic environment.

By supporting UN SDG 3, the framework enhances preventive healthcare and reduces disparities in women's medical access, particularly in developing nations. The modular structure enables easy retraining as new datasets appear, ensuring lasting adaptability.

Conclusion and Future Scope

The suggested two-phase ANN framework successfully predicts PCO and classifies severity with increased accuracy using organized clinical information. It removes dependence on invasive diagnostics, making it appropriate for widespread screening.

Future research will concentrate on:

- Incorporating explainable AI (XAI) for transparency in decision-making.
- Expanding the database to include hormonal and imaging information for multimodal prediction.
- Implementing the framework as a mobile application for rural access.

Our study shows deep learning can revolutionize reproductive healthcare by delivering accurate, scalable, and fair diagnostic assistance for women worldwide.

References

[1] Z. Zad, et al., "Predicting polycystic ovary syndrome with machine learning models based on clinical laboratory and imaging data," Frontiers in Endocrinology, vol. 15, p. 1298628, 2024.

[2] J. Kermanshahchi, et al., "Development of a machine learning-based model for accurate identification of PCOS pelvic ultrasound images," Journal of Medical Imaging and Health Informatics, vol. 14, no. 7, pp. 1949– 1961, 2024.

[3] P. Moral, et al., "CystNet: An AI-driven model for PCOS detection using ultrasound images," Scientific Reports, vol. 14, article 75964, 2024.

[4] K. Upreti, et al., "Polycystic ovary syndrome diagnosis: The promise of artificial intelligence for improved clinical accuracy," Biomedical and Pharmacology Journal, vol. 18, no. 1, pp. 123–139, 2025.

[5] J. Lim, et al., "Predicting TCM patterns in PCOS patients: An evaluation of feature selection and multi-label machine learning algorithms," Digital Health, vol. 10, no. 1, p. 240584402411314X, 2024.

[6] H. Faris, et al., "Feature extraction and optimization for medical diagnosis using hybrid deep learning models," Applied AI in Medicine, vol. 4, no. 2, pp. 45–58, 2025.

[7] S. Emara, et al., "Handling imbalanced PCOS data using adaptive oversampling and ensemble learning," Computational Healthcare Journal, vol. 9, no. 3, pp. 98–112, 2025.

[8] R. Panjwani, et al., "Optimized ensemble learning for PCOS detection using metaheuristic algorithms," Journal of Medical AI Research, vol. 12, no. 1, pp. 32–47, 2025.

[9] P. Ramteke, et al., "Explainable ensemble models for PCOS classification using SHAP and LIME interpretability frameworks," International Journal of Biomedical Informatics, vol. 15, no. 4, pp. 200–215, 2025.

[10] A. Divekar and S. Sonawane, "Deep transfer learning for ultrasound-based PCOS detection," IEEE Transactions on Healthcare Informatics, vol. 19, no. 2, pp. 175–183, 2025.