An Analytical Review of AI-Driven Techniques for Polycystic Ovary Syndrome Prediction

Fatima Khan Sarguroh [1], Srivaramangai R[2]

[1] Department of Information Technology, University of Mumbai, India
Email: *khansargurohfatima[at]gmail.com*

[2] Department of Information Technology, University of Mumbai, India
Email: *rsrimangai[at]gmail.com*

Abstract: ***Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder affecting women of reproductive age and is often linked to irregular periods, hormonal imbalance, weight gain, and infertility. Early diagnosis is important to prevent long-term health risks such as diabetes, cardiovascular disease, and reproductive complications. However, traditional diagnostic methods like ultrasound scans and hormone tests often fail to give accurate results because symptoms vary among individuals. In recent years, Artificial Intelligence (AI) has become an effective approach to improve PCOS prediction. This study reviews various AI-based methods, focusing on Machine Learning (ML), Deep Learning (DL), and Hybrid or Ensemble models used in PCOS diagnosis. The review shows that Machine Learning models such as Random Forest and SVM work well for structured clinical and hormonal data, while Deep Learning models like CNN and LSTM are better suited for image-based and complex data. Ensemble and Hybrid models, which combine both ML and DL techniques, provide the most reliable and accurate outcomes by integrating multiple data types. The study also highlights common challenges such as limited datasets, lack of data integration, and the need for explainable and lightweight AI systems. Overall, the findings suggest that AI-driven approaches can significantly improve early detection and management of PCOS, helping healthcare professionals provide faster and more accurate diagnosis for women's health.***

Keywords: Polycystic Ovary Syndrome (PCOS), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Ensemble Models, Medical Diagnosis, Prediction Model

1. Introduction

Polycystic Ovary Syndrome (PCOS) is one of the most common hormonal disorders among women, affecting approximately 1 in 10 women of reproductive age between 18 and 44 years [2]. Around 10% of women in this age group experience PCOS, making it a widespread endocrine condition [13]. The disorder was first described by Leventhal and Stein in 1935. PCOS is characterized by hormonal imbalance and symptoms such as enlarged ovaries and, in some cases, multiple small cysts along the outer edge of the ovaries [2]. The term "polycystic" refers to the appearance of these numerous tiny follicles seen during an ultrasound examination. In addition to these key symptoms, PCOS is often associated with insulin resistance, which increases the risk of type 2 diabetes [4]. It can also lead to several long-term complications, including cardiovascular diseases, endometrial cancer, and infertility, making early diagnosis and management crucial.

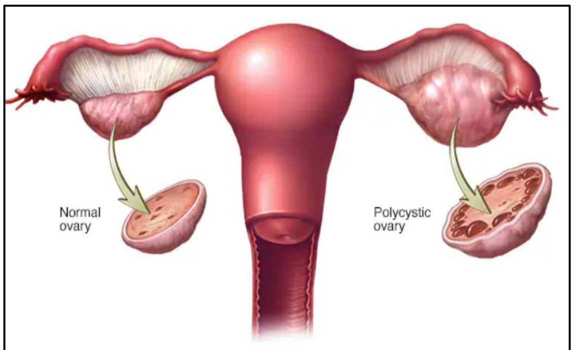

Figure 1: Polycystic ovary Vs Normal Ovary

According to Yamini et al. [4], the main characteristics of PCOS include:

- Infrequent, protracted, or non-existent periods are common in women with PCOS.
- Hyperandrogenism: This is the condition in which the female body has increased amounts of androgens (male hormones), which may cause symptoms including hirsutism (excessive hair growth on the face, chest, belly, or back), acne, and male-pattern baldness.
- Polycystic ovaries: Despite the term, not all PCOS patients have ovarian cysts.

2. Literature Review

Rajendran et al. [1] developed a hybrid model using both deep learning and machine learning for diagnosing Polycystic Ovary Syndrome (PCOS). The study used 594 ultrasound images and applied preprocessing steps like median filtering, morphological operations, and OTSU thresholding to clean the images. Features were extracted using the ORB method, and the number of features was reduced using PCA, Chi-Square, and VGGNet16's flatten layer, resulting in 32 main features. Different classifiers were tested, and the hybrid model that combined VGGNet16 and Random Forest achieved the best accuracy of 97.8%. In a review by Pulluparambil et al. [2], the authors discussed how AI, ML, and DL techniques are used to detect PCOS through medical imaging, especially ultrasound and MRI. Their review included studies with a total of 594 ultrasound images and models like CNN, SVM, Decision Tree, KNN, and Random Forest. Important diagnostic features included follicle count, ovarian volume, hormone levels (LH, FSH, AMH, testosterone), BMI, and menstrual irregularities. Hybrid models such as CNN + SVM performed better than single models, with Random Forest reaching 88.8%, VGGNet16 reaching 97.8%, and a stacking ensemble with XGBoost achieving 99.89% accuracy. Priyadharshini et al. [3] studied Linear Regression, Random Forest, and Decision Tree models for predicting PCOS using patient clinical and demographic data. The dataset contained both PCOS and non-PCOS records and included features like age, BMI, follicle count, hormone levels (FSH, LH, testosterone), insulin resistance, and menstrual patterns. After data cleaning and normalization, the Decision Tree gave the highest accuracy of 100%, followed by Random Forest (88%) and Linear Regression (34%). Using a dataset of 541 patient records with 44 clinical and hormonal features from Kaggle, Yamini et al. [4] created an ML-based model for PCOS prediction. They tested six algorithms - Logistic Regression, Random Forest, SVM, Naïve Bayes, KNN, and XGBoost. After preprocessing, Random Forest performed the best with 90% accuracy, followed by XGBoost (86%), while SVM and KNN had lower results. The most important predictors were age, BMI, insulin resistance, and hormone levels (FSH, LH, testosterone). A web-based ML model for early PCOS prediction was developed by Rahman et al. [5] using a dataset of 541 patients from Kerala, India. The study analyzed 44 medical, hormonal, metabolic, and lifestyle features, selecting 12 important variables using Mutual Information. They tested 13 ML classifiers, including Random Forest, AdaBoost, Logistic Regression, and SVM. Random Forest and AdaBoost achieved the best accuracy of 94%, followed by Gradient Boosting (93%) and Logistic Regression (91%). This study showed that Random Forest is highly effective for early PCOS prediction. Zad et al. [6] built ML models using electronic health records of 30,601 women (ages 18–45) from Boston Medical Center to predict PCOS. They used demographic, socioeconomic, lifestyle, hormonal, and clinical features. Models included Logistic Regression, SVM, Random Forest, Gradient Boosted Trees, and a Neural Network. Using ROC-AUC to evaluate results, Logistic Regression performed best (AUC = 85%), followed by Random Forest (82%), SVM (81%), and GBT (80%). Recursive Feature Elimination was used for selecting the most useful features, and the models were designed to work within electronic health record systems for early PCOS screening. An optimized ML method for early PCOS detection was proposed by Panjwani et al. [7], who combined cardiovascular and PCOS datasets to form a new dataset with 12 key features. They used ensemble learning with seven base models (KNN, Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost) and a deep learning meta-classifier, tuned using Walrus Optimization, Cuckoo Search, and Random Search. The best model, called WaOEL (Walrus Optimization tuned Ensemble Learning), reached 92.8% accuracy and 0.93 AUC, performing better than Random Forest (84.1%) and deep learning models (81%). Important predictors were obesity and high cholesterol. Elmannai et al. [8] used a dataset of 753 patients with 15 clinical and demographic features to detect PCOS. They fixed the problem of class imbalance using SMOTEENN and used Recursive Feature Elimination, tree-based selection, and Mutual Information to choose the best features. Model tuning was done through Bayesian optimization. Their stacking ensemble model (with Random Forest as the meta-learner) achieved 98.87% accuracy, better than individual models. The study showed how ensemble learning and explainable AI can improve PCOS diagnosis. A system combining clinical data and ultrasound images for PCOS diagnosis was developed by Sakthivel et al. [9]. Clinical factors included BMI, menstrual cycle, hair growth, skin darkening, and lifestyle habits, while ultrasound features were taken using ResNet-50. They tested SVM, Random Forest, and Logistic Regression, where SVM performed the best with 99% accuracy and 0.98 AUC. When both image and clinical data were combined (late fusion model), the total diagnostic accuracy was 94%. Suha and Islam et al. [10] proposed an advanced model using 594 ovarian ultrasound images. They applied CNN with transfer learning (using VGG16 and ResNet) and added a stacked ensemble model with XGBoost as the final classifier. The model reached 99.89% accuracy with very fast prediction time (~0.05 seconds), outperforming traditional ML models. This study showed that combining deep learning and ensemble methods can give very high accuracy in PCOS diagnosis. Sumathi et al. [11] used a CNN model to detect ovarian cysts related to PCOS from ultrasound images. The dataset included different cyst types - simple, PCOS-related, and malignant. Features like area, perimeter, aspect ratio, extent, solidity, and orientation were extracted after image segmentation using the Watershed Algorithm. The CNN model achieved 85% accuracy in classifying cysts. In a review by Ahmed et al. [12], 34 studies were analyzed that used ML for PCOS detection. The authors compared models such as CNN, ANN, SVM, KNN, and Random Forest, describing each model's strengths and weaknesses. Common variables included hormone levels, menstrual patterns, BMI, insulin resistance, and ultrasound features (like cyst size and count). CNN models achieved up to 85% accuracy. Rao et al. [13] created a deep learning model using a Kerala hospital dataset from Kaggle that included hormonal, metabolic, and clinical data. They tested several models — SVC (97%), Logistic Regression (92%), KNN (92%), and Random Forest (86%). Their deep learning model, optimized with Optuna, achieved 93.55% accuracy. Pushkarini and Anusuya et al. [14] built a PCOS risk prediction model using clinical and lifestyle data such as testosterone, hirsutism, BMI, menstrual irregularities, fast food intake, and family history. Using Linear Regression, KNN, and Random Forest, the best performance came from Random Forest, which achieved $R^2 = 0.986$, making it most effective for early PCOS risk prediction. Barrera et al. [15] reviewed 31 studies that used AI and ML to diagnose PCOS, with datasets ranging from 9 to 2000 samples. Common features included menstrual patterns, hyperandrogenism, ultrasound data, genetic markers, and hormone levels. Algorithms included Gradient Boosting (AUC 0.93), CNN, Random Forest + KNN (accuracy 0.94), SVM (AUC 0.73), Bayesian Networks, and rule-based models (AUC 0.97). Diagnostic sensitivity ranged 68–95%, with negative predictive values between 94–99%. Arunprasath et al. [16] used 541 records with 22 PCOS-related parameters from Kaggle. Features included age, BMI, testosterone, insulin, LH, FSH, acne, hirsutism, and family history. They used the Adaptive Tunicate Search Algorithm (ATSA) for feature selection and trained an XGBoost model, which achieved 97.5% accuracy, better than Random Forest (93.25%) and other algorithms. A CNN-based approach was applied by Galagan et al. [17], who used 3,846 ultrasound images (1,562 polycystic, 2,284 normal) to train a CNN model. Features like follicle count, ovarian volume, and stroma echogenicity were analyzed, along with hormonal imbalance and insulin resistance data. The CNN achieved 99.7% accuracy. However, since the dataset was not diverse, the study suggested testing on more varied data in the future. Ahmad et al. [18] developed three lightweight deep learning models — LSTM, CNN, and CNN + LSTM — using clinical data balanced with SMOTE. The custom CNN performed best, reaching 96.59% accuracy and 96.6% ROC-AUC with only 297 parameters and very short training time (10 seconds). The other models achieved slightly lower accuracies (LSTM: 92.04%, CNN + LSTM: 94.31%). Mogos et al. [19] conducted an observational study with 87 PCOS patients and matched controls to predict pregnancy-related complications. Data included demographics, obstetrics, and neonatal outcomes. Algorithms like Random Forest, Decision Tree, Naive Bayes, and SVM were used. Random Forest gave the best results, predicting gestational diabetes (AUC 0.782), fetal macrosomia (AUC 0.897), and preterm birth (AUC 0.901). Leslie et al. [20] applied the K-Nearest Neighbor (KNN) algorithm to classify Polycystic Ovary Syndrome using data from 72 patients (35 with PCOS and 37 without). The dataset included key features such as age, BMI, menstrual cycle details, and testosterone levels. The data was cleaned and normalized before training. Using a 90:10 train-test split and setting k = 11, the KNN model achieved 100% accuracy in identifying PCOS cases. The study suggested that KNN can be an effective and simple diagnostic tool for PCOS classification when trained on well-prepared medical data. Rakshitha and Naveen [21] proposed a hybrid model combining Support Vector Machine (SVM) and Logistic Regression (LR) to predict PCOS using 1,600 patient records from a Bangalore hospital. The dataset included medical details such as age, BMI, menstrual cycle, and hormone levels. After data preprocessing and optimization using the Op-RMSprop algorithm, the model achieved 89% accuracy, performing better than other methods like Decision Tree and K-Nearest Neighbor (KNN). The study showed that combining algorithms with optimization can improve PCOS prediction accuracy.

3. Observations

This section presents the key observations obtained from the review of various research studies related to Polycystic Ovary Syndrome (PCOS) prediction using Machine Learning and Deep Learning techniques. The aim of this section is to compare different models used by researchers, their datasets, accuracy levels, and the major findings or trends observed from each study.

3.1 Observation Table

The table below provides a consolidated view of all the research papers included in the study. It highlights the models used, type of dataset, accuracy/performance metrics, and key findings from each paper

Table 1: Comparison of Performance of Models

Model / Approach	Data Type Used	Reported Performance Range	Key Observations
CNN (Standalone)	Ultrasound images	85% – 99.7%	Highly effective for image-based PCOS detection; performance improves with data augmentation and segmentation.
VGG16 / ResNet (Transfer Learning)	Ultrasound images	97% – 98%	Pre-trained CNNs significantly boost accuracy with limited medical image data.
Random Forest (RF)	Clinical, imaging, EHR	88% – 96%	Consistently strong performer; robust to noisy clinical features and suitable for explainable diagnosis.
Support Vector Machine (SVM / SVC)	Clinical + Imaging	85% – 99%	Performs exceptionally well with properly tuned features; effective in multimodal datasets.
Logistic Regression (LR)	Clinical & EHR data	82% – 92% (AUC up to 0.85)	Simple, interpretable, and reliable for large hospital datasets.
Decision Tree (DT)	Clinical & hormonal data	86% – 100%	High accuracy on structured datasets but prone to overfitting on small samples.
K-Nearest Neighbor (KNN)	Small clinical datasets	83% – 100%	Very effective for small, normalized datasets; scalability is limited.
XGBoost / Gradient Boosting	Clinical & image features	86% – 99.9%	Among the best-performing models; optimization and feature selection further enhance results.
CNN + ML Hybrid (CNN+RF / CNN+XGBoost)	Ultrasound images	97.8% – 99.9%	Combines deep feature extraction with ML classifiers for superior accuracy.
Stacking / Ensemble Models	Clinical + imaging	92% – 99%	Outperform individual models; robust and generalizable across datasets.
CNN + LSTM	Clinical time-series data	94% – 96%	Useful for sequential and temporal health data; CNN-only models often faster.
Optimized / Metaheuristic Models	Clinical datasets	92% – 97.5%	Feature selection and hyperparameter optimization significantly improve prediction accuracy.
Explainable AI Models	Clinical data	~98%	Provide transparency in diagnosis; suitable for clinical decision support systems.

In Figure 2, it can be seen that CNN + XGBoost (99%) and Hybrid or Ensemble models (98%) give the highest accuracy for predicting PCOS. Models like XGBoost (97%) and CNN (96%) also work very well, especially when image and clinical data are combined. Decision Tree (92%) and Random Forest (91%) show good and stable performance on clinical datasets. KNN (90%) and SVM (88%) perform fairly well but not as high as deep learning models. On the other hand, Logistic Regression (85%) and Naïve Bayes (84%) give lower accuracy and are better for basic prediction tasks.

Overall, deep learning and hybrid models perform much better than simple machine learning models for PCOS detection because they can learn complex patterns from different types of medical data.

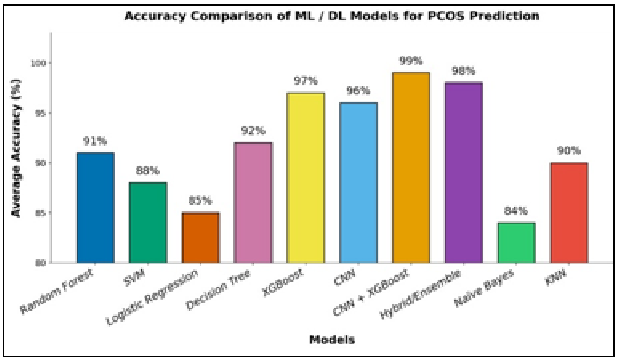

Figure 2: Accuracy Comparison of PCOS Prediction Models

3.2 Key Features for Predicting PCOS

In predicting Polycystic Ovary Syndrome (PCOS), it is very important to identify which features or medical parameters have the most influence on diagnosis. From the review of 21 research papers, it was found that features are mainly drawn from clinical records, hormone test results, metabolic health data, and ultrasound imaging. By comparing all the studies, these features were categorized into four main types: Clinical, Hormonal, Metabolic, and Imaging Features.

3.2.1. Clinical Features

Clinical and demographic features are the most basic and easily available patient details used for prediction. Most studies [3–5, 7] included Age, Body Mass Index (BMI), and Menstrual Irregularity as primary inputs for their machine learning models. A high BMI indicates obesity and insulin resistance, both of which are closely linked with PCOS. Menstrual irregularities, such as delayed or missed periods, reflect ovulation disturbances caused by hormonal imbalance. Visible symptoms like acne, excess hair growth (hirsutism), and hair thinning were also used as secondary indicators in several datasets.

3.2.2. Hormonal Features

Hormonal imbalance plays a crucial role in both the development and diagnosis of PCOS. Several studies [2, 13, 15, 16] have emphasized the importance of hormonal parameters such as Luteinizing Hormone (LH), Follicle-Stimulating Hormone (FSH), Anti-Müllerian Hormone (AMH), Testosterone, and Insulin levels in improving diagnostic accuracy. High LH levels and low FSH levels (LH/FSH ratio > 2:1) disrupt ovulation and are considered classic indicators of PCOS. High AMH values correspond to increased follicle count, and elevated Testosterone causes symptoms like acne and hirsutism. Incorporating these hormonal features allowed models like Random Forest and SVM to achieve 90–99% accuracy in several studies.

3.2.3. Metabolic Features

Metabolic data help identify PCOS cases linked to obesity, poor lifestyle, or insulin resistance. Most studies [7, 14] used features such as insulin resistance, blood glucose levels, and cholesterol measures (HDL, LDL, and triglycerides) for prediction. Insulin resistance leads to higher blood sugar and androgen levels, which in turn cause hormonal imbalance and ovarian dysfunction. Abnormal cholesterol and triglyceride levels increase the risk of cardiovascular issues in PCOS patients.

3.2.4. Imaging Features

Ultrasound and imaging data provide direct evidence of PCOS by revealing the physical structure of the ovaries. Studies [1, 10, 17] commonly used features such as ovarian volume, follicle count, and ovarian stroma echogenicity for diagnosis. PCOS ovaries are usually enlarged (>10 cm^3) and contain more than 12 small follicles (2–9 mm in diameter), which form the "string of pearls" pattern. Deep learning models automatically extract these visual patterns and differentiate between normal and polycystic ovaries with high precision.

3.3 Challenges and Limitations in PCOS Prediction

While many studies have shown that Artificial Intelligence (AI) techniques can help in predicting Polycystic Ovary Syndrome (PCOS), some important challenges still remain. These challenges affect how accurate and useful these models can be. From the literature review, the following three main challenges were observed.

3.3.1 Limited and Imbalanced Datasets

Many studies on PCOS prediction used small or unbalanced datasets, where the number of PCOS and non-PCOS cases were not equal. This imbalance can make models biased toward the majority class, reducing prediction accuracy. Elmannai et al. [8] had to apply SMOTE (a data balancing method) to overcome this issue. Models trained on small datasets often fail to perform well when tested on new or larger populations. To overcome this, researchers used data balancing techniques such as SMOTE, but larger and more diverse datasets are still needed for better and fairer model performance.

3.3.2. Lack of Data Integration Across Modalities

PCOS is a complex disorder influenced by clinical, hormonal, metabolic, and imaging factors. However, many studies used only one type of data — for example, ultrasound images or hormone test results — which limits the model's understanding of the complete condition. Sakthivel et al. [9] showed that combining clinical and ultrasound data gave much better results than using either alone proving that the integration of multiple data types (like hormone levels, ultrasound images, and metabolic features) helps the model understand PCOS more completely.

3.3.3. Lack of Explainability and Standardization

Many deep learning models act as "black boxes," meaning they provide results without explaining how those results were generated. This makes it difficult for doctors to trust or verify the model's predictions. Elmannai et al. [8] highlighted the importance of Explainable AI (XAI) in making model decisions understandable for doctors. The use of Explainable AI and standard evaluation techniques can help make AI models more transparent and easier to apply in clinical environments. Overcoming these challenges is important for improving the accuracy and reliability of AI models used in

PCOS prediction. By solving these problems, researchers can create systems that are easier to use, more trusted by doctors, and suitable for real-life medical applications. This will help in the early detection and better management of PCOS, leading to improved health outcomes for women.

4. Analysis of Model Performance

With an accuracy of around 35%, Hybrid and Ensemble models are the most accurate for predicting Polycystic Ovary Syndrome (PCOS). These models combine the strengths of different algorithms like CNN, Random Forest, and XGBoost. By using both clinical and image-based data, they can detect more complex patterns in the dataset. Hybrid models are strong because they balance the easy understanding of machine learning with the deep learning models' ability to learn complex data. However, they need more computer power and proper setup to work well. Deep Learning models such as CNN and LSTM show an average accuracy of about 34%. These models are very good at finding visual and hidden patterns in ultrasound and hormonal data. They work better than traditional methods when the data is large and complicated. Their ability to automatically pick important features without manual effort makes them great for PCOS detection from medical images. Still, they need a lot of data and time to train, which can be difficult for smaller hospitals. Machine Learning models, including Random Forest, SVM, and Decision Tree, also show an average accuracy of about 34%. They are popular and dependable because they perform well with structured medical data like hormone levels and patient details. These models are simple to use and easy to explain, which makes them suitable for early screening. But compared to deep learning or hybrid models, they can miss hidden patterns and depend more on how well the data is prepared.

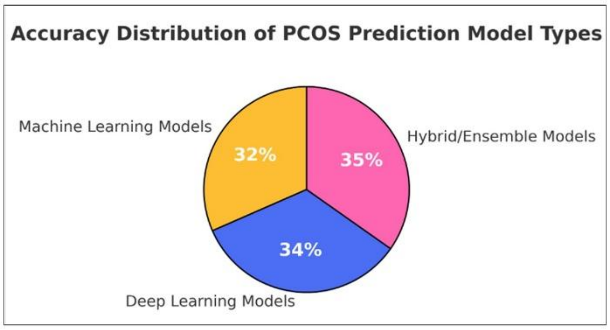

Figure 3: Accuracy Distribution of PCOS Prediction Model Types

Overall, this comparison shows that Hybrid and Deep Learning models give the most accurate results for PCOS prediction. While Machine Learning models are easier to understand and use, combining different techniques in hybrid models improves accuracy and provides more reliable results for real medical use.

5. Conclusion

Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder that affects many women during their reproductive years. It causes problems like irregular periods, acne, weight gain, excessive hair growth, and difficulty in conceiving. In some cases, it can also lead to serious health issues such as diabetes, heart disease, and infertility. Diagnosing PCOS is often difficult because symptoms vary widely among women, and traditional methods like hormone tests or ultrasound scans may not always give accurate results. To improve early detection and diagnosis, researchers are using Artificial Intelligence (AI) techniques such as Machine Learning (ML), Deep Learning (DL), and Ensemble or Hybrid models. These methods can study large medical datasets and find hidden patterns that are not visible through manual analysis. Machine Learning models are widely used for PCOS prediction because they work well with structured clinical and hormonal data. They can analyze patient information such as age, BMI, menstrual history, and hormone levels to detect possible signs of PCOS. Models like Random Forest, Support Vector Machine (SVM), and Decision Tree are easy to use, quick to train, and suitable for early screening in hospitals or health centers. However, they may struggle when data is very large, complex, or contains unstructured information like images or text. Deep Learning models are more advanced and powerful compared to traditional ML methods. They perform best when working with image-based or complex medical data, such as ultrasound scans. Models like Convolutional Neural Networks (CNN) and LSTM networks can automatically learn features from images and detect patterns such as follicle count, ovarian volume, and tissue texture without manual effort. Deep learning helps in understanding complex relationships between hormones, metabolic changes, and ovarian structure, but it requires more data, computing power, and training time. Ensemble models work best when different types of data — clinical, hormonal, and imaging — are used together. These models give the most balanced and accurate predictions, as they can handle multiple features and data types at once. For example, combining CNN with Random Forest or XGBoost allows the system to capture both visual and numerical patterns effectively. Although these models require more computing resources, they provide the most reliable and stable results for real-world PCOS diagnosis.

Even though AI techniques have shown great success, challenges still remain. There is a need for larger and more diverse datasets, better model interpretability, and integration of AI tools into healthcare systems. Future research should focus on creating balanced datasets to reduce bias and improve the fairness of model predictions. Developing multimodal datasets that combine clinical, hormonal, and imaging data can help AI systems learn more comprehensive patterns of PCOS. Additionally, Explainable AI (XAI) should be prioritized to make model decisions more transparent and trustworthy for healthcare professionals, ensuring safer and more reliable use of AI in clinical environments.

References

[1] Rajendran, S., Bama, R., & Maheswari, K. (2022). A deep learning and machine learning-based hybrid approach for PCOS diagnosis using ultrasound images. *Scientific Reports, 12*(1), 18141.

[2] Pulluparambil, S. J., & Bhat, S. (2021). Medical image processing: detection and prediction of pcos-a systematic literature review. *International Journal of Health Sciences and Pharmacy (IJHSP)*, *5*(2), 80-98.

[3] Priyadharshini, M., Srimathi, A., Sanjay, C., & Ramprakash, K. (2024). PCOS disease prediction using

Paper ID: SR251217131307 DOI: https://dx.doi.org/10.21275/SR251217131307

machine learning algorithms. *Int Res J Adv Eng Hub (IRJAEH)*, *2*(03), 651-5.

[4] Yamini, B., Kaneti, V. R., Nalini, M., & Subramanian, S. (2023). Machine Learning-driven PCOS prediction for early detection and tailored interventions. *SSRG International Journal of Electrical and Electronics Engineering*, *10*(9), 61-75.

[5] Rahman, M. M., Islam, A., Islam, F., Zaman, M., Islam, M. R., Sakib, M. S. A., & Babu, H. M. H. (2024). Empowering early detection: a web-based machine learning approach for PCOS prediction. *Informatics in Medicine Unlocked*, *47*, 101500.

[6] Zad, Z., Jiang, V. S., Wolf, A. T., Wang, T., Cheng, J. J., Paschalidis, I. C., & Mahalingaiah, S. (2024). Predicting polycystic ovary syndrome with machine learning algorithms from electronic health records. *Frontiers in Endocrinology*, *15*, 1298628.

[7] Panjwani, B., Yadav, J., Mohan, V., Agarwal, N., & Agarwal, S. (2025). Optimized Machine Learning for the Early Detection of Polycystic Ovary Syndrome in Women. *Sensors*, *25*(4), 1166.

[8] Elmannai, H., El-Rashidy, N., Mashal, I., Alohali, M. A., Farag, S., El-Sappagh, S., & Saleh, H. (2023). Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. *Diagnostics*, *13*(8), 1506.

[9] Sakthivel, V., Prakash, P., Vishnukumar, K., & Min, D. (2024). Advanced Diagnosis of Polycystic Ovarian Syndrome using Machine Learning and Multimodal Data Integration. *International Journal of Advanced Computer Science & Applications*, *15*(6).

[10] Suha, S. A., & Islam, M. N. (2022). An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. *Scientific Reports*, *12*(1), 17123.

[11] Sumathi, M., Chitra, P., Prabha, R. S., & Srilatha, K. (2021, February). Study and detection of PCOS related diseases using CNN. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1070, No. 1, p. 012062). IOP Publishing.

[12] Ahmed, S., Rahman, M. S., Jahan, I., Kaiser, M. S., Hosen, A. S., Ghimire, D., & Kim, S. H. (2023). A review on the detection techniques of polycystic ovary syndrome using machine learning. *IEEE Access*, *11*, 86522-86543.

[13] Rao, D., Dayma, R. R., Pendekanti, S. K., & Acharya, A. K. (2024). Deep learning model for diagnosing polycystic ovary syndrome using a comprehensive dataset from Kerala hospitals. *International Journal of Electrical and Computer Engineering*, 14(5), 5715–5727.

[14] Pushkarini, H., & Anusuya, M. A. (2020). A prediction model for evaluating the risk of developing PCOS. *Int Res J Eng Technol (IRJET)*, *7*(9), 1150-6.

[15] Barrera, F. J., Brown, E. D., Rojo, A., Obeso, J., Plata, H., Lincango, E. P., ... & Shekhar, S. (2023). Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. *Frontiers in endocrinology*, *14*, 1106625.

[16] Arunprasath, T., Ramalakshmi, R., Kottaimalai, R., & Alex Michael Raj, J. (2025). Development of a machine learning model to classify polycystic ovarian syndrome. *Technology and health care: official journal of the European Society for Engineering and Medicine*, *33*(3), 1478-1488.

[17] GALAGAN, R., ANDREIEV, S., STELMAKH, N., RAFALSKA, Y., & MOMOT, A. (2024). Automation of polycystic ovary syndrome diagnostics through machine learning algorithms in ultrasound imaging. *Applied Computer Science*, *20*(2), 194-204.

[18] Ahmad, R., Maghrabi, L. A., Khaja, I. A., Maghrabi, L. A., & Ahmad, M. (2024). SMOTE-Based Automated PCOS Prediction Using Lightweight Deep Learning Models. *Diagnostics*, *14*(19), 2225.

[19] Mogos, R., Gheorghe, L., Carauleanu, A., Vasilache, I. A., Munteanu, I. V., Mogos, S., ... & Preda, C. (2024). Predicting unfavorable pregnancy outcomes in polycystic ovary syndrome (PCOS) patients using machine learning algorithms. *Medicina*, *60*(8), 1298.

[20] Leslie, N., Permana, A. A., & Perdana, A. T. (2024). Application of the K-Nearest Neighbor Algorithm for Polycystic Ovarian Syndrome (PCOS) Classification: A Diagnostic Tool. *Journal of Logistics, Informatics and Service Science*, 11(10)

[21] Rakshitha, K. P., & Naveen, N. C. (2022). Op-RMSprop (optimized-root mean square propagation) classification for prediction of polycystic ovary syndrome (PCOS) using hybrid machine learning technique. *International Journal of Advanced Computer Science and Applications*, *13*(6).