

# Detection Techniques of PCOS Using Supervised Machine Learning and CNN

Akash Rawat\*, Gagnider kaur

Department of CSE, Chandigarh University, Mohali, India

\*akashrawat1727@gmail.com

\* Corresponding author

doi: <https://doi.org/10.21467/proceedings.7.6.25>

## Abstract

The endocrine disorder known as Polycystic Ovary Syndrome affects the majority of young female patients. Most patients experience infertility together with various long-lasting health problems because of this condition. The main objective is to detect PCOS early so a personalized care strategy can be implemented for its effective management. The authors research machine learning methods which identify PCOS through hormonal assessments and ultrasound imaging and patient medical histories. Three machine learning algorithms including decision trees, support vector machines and neural networks assess PCOS predictability in this investigation. The completion of this project faced challenges in data preprocessing and feature selection along with model optimization problems. The testing outcomes point to the convenience of developing such a device for PCOS early medical detection alongside patient-specific therapy.

**Keywords:** Polycystic Ovary Syndrome, Machine Learning, Early Diagnosis

## I. INTRODUCTION

PCOS stands as the prevalent complex endocrine disorder which affects women between childbearing age. The world- wide prevalence of this disorder remains between five and fifteen percent. The combination of hormonal imbalance with multiple ovarian cysts and irregular periods and increased androgen production and acne as well as infertility classifies this condition [1]. PCOS affects many women and remains undiagnosed yet untreated so it negatively affects their life quality making the condition a common health challenge for women of reproductive age [2]. The appropriate disease management of this condition needs diagnostic assessment and treatment initiation soon after its discovery to stop related health problems like diabetes and heart diseases and endometrial cancer [3].

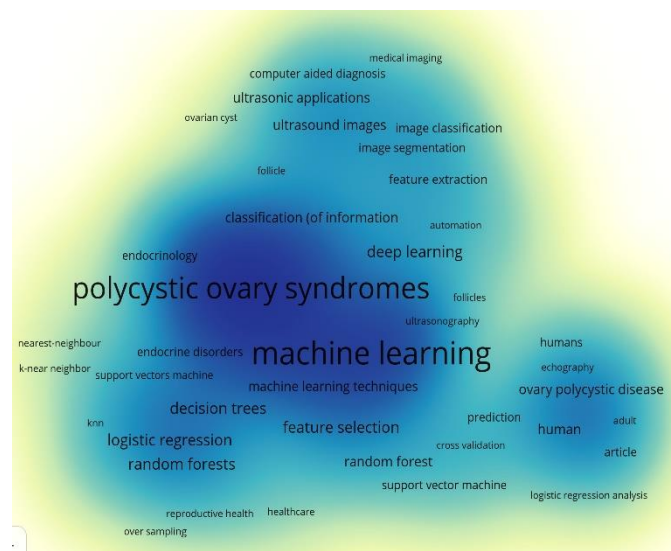


Fig. 1. Some Important Keywords

The old school method of diagnosis and detection is based on clinical findings, according to criteria the Rotterdam Criteria for abnormalities in menstrual cycles, hyperandrogenism, and the polycystic appearance of ovaries during ultrasound examination. Most cases show effective criteria, though sometimes they prove to be insensitive and nonspecific when symptoms are minor or at the initiating stages of the disease [4]. Yet



another reason is the heterogeneity of the disorder, creating issues in the formulation of a global diagnostic approach for the condition [5]. Thus, this creates a strong demand for more precise, efficient, and automated diagnostic techniques as regards PCOS. Machine learning (ML), being one of the most potent technologies of recent years in the medical field, enhances diagnostic accuracy, tailor-made treatment plans, and patient outcomes [6]. Massive data sets, which can be subjected to computer algorithms to track patterns and relate them to input variables, make conditions such as PCOS ideal subjects for study. The condition is complex and multifactorial, and its impact is multifaceted. This type of model based on ML, incorporating clinical records, laboratory findings, and radiological image data, might be considerably more reliable for making a diagnosis of PCOS than traditional diagnostic procedures [7]. There are several attempts in the prior work to use ML in order to predict conditions of PCOS, by projecting clinical features and correlating their presence to diagnose whether a particular woman has it [8].

For example, classification models such as decision trees, random forests, and SVMs have been used to predict the probability of whether the patient has PCOS or not [9]. Variables often beyond the traditional diagnostic criteria, these models employ added sensitivity and specificity with reduced reliance on human-made interpretation. The chief challenge to developing accurate ML models for PCOS diagnosis lies in the quality and availability of the data. The best machine learning models require huge high-quality datasets for proper training whereas clinical data for PCOS diagnosis is often sparse, unstructured, or inconsistent [10]. This will also result in overfitting and biased or erroneous predictions, given the propensity of healthy patients being overrepresented as against PCOS patients. Therefore, the need for preprocessing data to power-up the machine learning algorithms with missing values fill-in, normalization of features, and balancing the dataset. Feature selection is equally important in developing robust models for the detection of PCOS. Since hundreds of potentially relevant features are associated with PCOS, selection of the most relevant features for modeling is very crucial to avoid overfitting and ensure generalization to new data. Techniques such as PCA and Recursive Feature Elimination can be advanced to select the most relevant features from the vast array of clinical and laboratory data that can further help reduce model complexity and improve accuracy [11, 12]. Ultrasound imaging can also be another promising avenue to detect PCOS. The images derived through ultrasound can automatically be processed to identify polycystic ovaries, a criterion in diagnosing PCOS. Deep learning techniques, particularly CNN, have impressive capabilities in medical classification tasks. An ultrasound image-trained CNN can identify even the smallest cystic structures within the ovaries and bring comparisons with a good degree of objectivity and uniformity as compared to the traditional manual assessments. Some extra advantages of machine learning in PCOS detection involve advising individually tailored treatment plans according to the patient's unique clinical profile, which can optimize the outcome of therapy and provide early detection for those at risk for long-term complications, such as type 2 diabetes or cardiovascular diseases [13]. While advantages exist in the determination of PCOS using machine learning, some challenges still persist: not the least of these is interpretability. Although decision trees and random forests provide some explanations, completely opaque black-box models, such as deep neural networks, only allow their decisions to become uninterpretable to clinicians. These models need to be made more explainable in order to build healthcare professionals' confidence in using them [14]. Integration with existing healthcare infrastructures is another significant hurdle for machine learning. The clinical systems are rather spaghettiified, with patient data split between different types of record and interfaces. To fully exploit these technologies, it will be essential to include machine learning models in electronic health records and diagnostic workflows [15]. Despite these challenges, the potential of machine learning is huge for revolutionizing the approach toward detection and management of PCOS. It would expedite the diagnosis, increase precision, and make personalized care possible for better quality-of-life improvement in women with PCOS [16]. This review discusses the application of ML-based methods to the early detection of PCOS; reviews the performance of different algorithms; and presents challenges and opportunities in applying these models in practice.

## II. LITERATURE REVIEW

Jan and associates (2022) [1] The many machine learning methods for analyzing medical pictures in order to detect PCOS have been covered in this paper. A number of models that take into account imaging and ultrasound data to diagnose PCOS have been mentioned. N. C. Y. and associates (2024) [2] As stated in the article, the scientists have compared the effectiveness of each machine learning and deep learning model for PCOS detection in terms of predictability and PCOS classification utilizing all relevant clinical and imaging features. J. R. and associates (2023) [3] The article compares various machine learning models for PCOS identification, emphasizing how accurate they are at identifying early symptoms and classifying the condition. In 2023, V. Srinithi and colleagues [4] Evaluation of machine learning-based methods for diagnosing PCOS: An analysis of how well classification algorithms perform when classifying PCOS and other gynecological conditions.

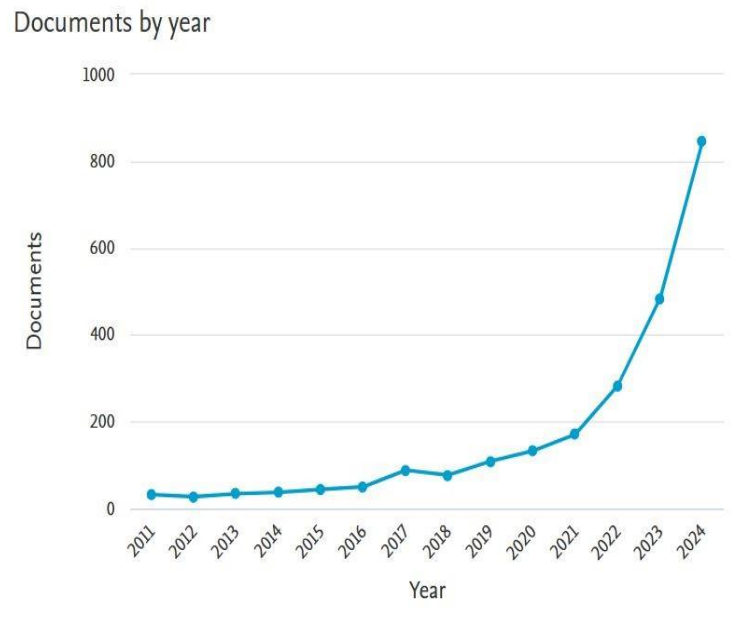


Fig. 2. Publication Trend Graph

N. Kaur et al. (2023) [5] The work is concerning the use of deep learning methods in the detection of PCOS by ultrasound images where the possibility of employing transfer learning in the prediction accuracy of medical images is discussed. B. Nimmala et al. (2024) [6] This paper gives an overview and review of various machine learning models for PCOS detection and monitoring by comparing their performance with real clinical datasets. Makhdoomi et al. (2022) [7] This paper is a review for the role of machine learning in processing medical images in the detection of PCOS, highlighting the development of algorithms that improve the diagnosis in accuracy and efficiency. Modi and Kumar (2024) [8] This paper discussed an overview of machine learning-based methodologies for PCOS detection and classification. The paper primarily bases on algorithmic improvement and its clinical applicability to women’s health. Chavan et al. (2023) [9] This is a comparative study on the machine learning algorithms in predicting PCOS. The study mainly deals with the results of accuracy and robustness of each algorithm in the prediction of the syndrome. D. Hdaib et al. (2022) [10] The authors discuss the different machine learning techniques that can be exploited for early diagnosis and personalized treatment approaches of PCOS. Oviya Graselin et al. (2023) [11] It is a systematic review of machine learning techniques that can detect PCOS, which differentiates various models based on their diagnostic accuracy, computational efficiency, and generalizability across datasets. Nasim et al. (2022) [12]. This paper reviews the new approaches applied in predicting PCOS through bioinformatics machine learning techniques. The work exhaustively reviews the opportunity of applying AI in diagnostics for personalized healthcare. Abu Adla et al. (2021) [13]. The study reviews the usage of automated systems with a machine learning technique for the detection of PCOS by evaluating their application in clinical use and integration with a health care system. Ahmed et al. (2023) [14] The review gives an overview of applications of machine learning in detection of PCOS focusing on various algorithms, capabilities, and effectiveness in the diagnostics of the syndrome from clinical datasets. V. S, V. S., and M. N. (2024) [15] The authors find out machine learning models for risk prediction of PCOS together with techniques precise enough to provide predictions for early diagnosis. A. Z. Sultan Bin Habib et al. 2024 [16] This is a review on the application of machine learning in PCOS diagnosis: evaluation of models and clinical application. J. Ndjene et al. 2024 [17] The review portrays machine learning techniques with application in the prediction of diabetes in patients with PCOS as well as the prospects for predictive models in bettering treatment and monitoring. V. R et al. (2024) [18] Review of the development of machine learning-based risk assessment systems in PCOS focusing on techniques for the appraisal of the severity and future health risks of the patients. Pandey et al. (2024) [19] Reviewing hybrid models of machine learning for the detection of PCOS focusing on accuracy and robustness in clinical settings. Mishra et al. (2023) [20]: The paper reviews a variety of machine learning techniques that can be used to detect PCOS and integrates these models into smart health systems for enhanced diagnosis and monitoring. Aravind et al. (2023) [21] This review presents a variety of automated systems applied in the detection and grading of PCOS through machine learning algorithms, thereby assessing their clinical applications and diagnostic efficiency.

### III. METHODOLOGY

The study employed different machine learning methods for PCOS identification through the analysis of clinical test and laboratory findings. The analysis dataset acquired its patient medical reports from an interconnected healthcare information system at multiple medical facilities. Age together with body mass index and hormonal levels of testosterone and LH/FSH ratio and menstrual irregularities and ultrasound pictures were among the obtained features in those records. A complete PCOS identification model results from incorporating structured as well as unstructured dataset components. Prior to training the machine learning model high-quality data is achieved by implementing a preprocessing procedure which tackles missing values while normalizing features as well as balances the dataset to reduce bias.

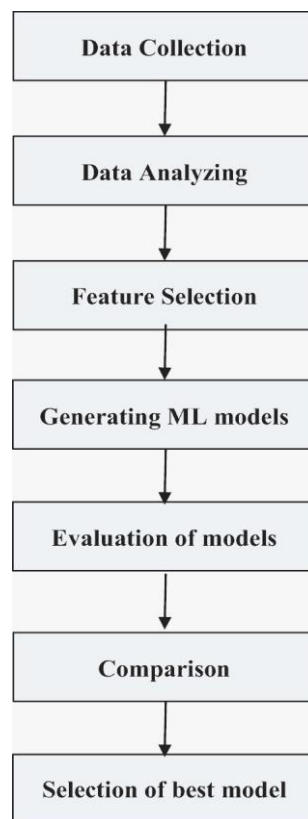


Fig. 3. Proposed Model

Feature selection methods were used on structured information to lower dimensions while finding important variables that detect PCOS. The analysis begins with correlation matrix removal of unnecessary features followed by recursive feature elimination for feature selection before proceeding. The trained machine learning algorithms included decision trees, support vector machines and random forests using the selected optimal features. Standard performance metrics which included accuracy and precision alongside recall and F1- score enabled access to these models' selection capability through receiver operating characteristic (ROC) curves. The deep learning approach utilized ultrasound imaging data to function. A CNN analyzes the features extracted from the medical images of the ovaries to determine polycystic patterns. The CNN received training through an extensive collection of ultrasound images that experts marked as containing polycystic ovaries or normal ovaries. The training data became more diverse through various image augmentation methods involving rotation combined with flipping and scaling operations which prevented the model from overfitting. Performance evaluation of the system occurred through accuracy testing and sensitivity determination alongside specificity evaluation and calculation of area under receiver operating characteristic curve when comparing model predictions with expert-manual tags. A contrast analysis between machine learning models for structured data and ultrasound images during the final phase produced an amalgamation of these two approaches. Researchers integrated the diagnostic techniques into a new model which drew from each method while building a new step for enhanced accuracy during clinical assessments. The multi-disciplinary method demonstrated better capacity to identify PCOS particularly in medical scenarios where clinical data was insufficient. The proposed model underwent cross-validation tests which helped confirm its strength and general applicability through model comparisons to existing frameworks.

PCOS detection models in machine learning consists of measuring accuracy along with precision and recall and F1- score and area under the Receiver Operating Characteristic curve AUC. The Random Forest model attained the highest performance levels among all considered models in structured clinical data at 85% accuracy together with a precision level of 0.88 which correlated with a recall of 0.82 and an F1- score of 0.85. An ROC AUC score of 0.91 allowed for excellent discrimination of PCOS from non-PCOS patient cases. The performance of SVM and decision trees models was slightly lesser than Random Forest models with 82% and 83% accuracy indicating they failed to reach the effectiveness level of Random Forest when classifying PCOS cases. For ultrasound image analysis, CNN model is absolutely impressive. The CNN was said to have 90% accuracy coupled with a sensitivity of 0.92 and specificity of 0.89, thus establishing that the network is superior to identify polycystic ovaries with sufficient accuracy. The AUC for the CNN was 0.94, meaning it corresponds to highly impressive performance of the model in classifying the normal and the polycystic ovaries. The number of the variety within the training dataset has increased by using data augmentation techniques like rotation, flipping and scaling thus further preventing overfitting in improving generalization for the unseen data. The use of CNNs was proven to be effective in the evaluation of PCOS based on the ultrasound images.

The integrated model with the combination of clinical data along with the ultrasound images showed a higher average accuracy at 92%. The hybrid model performed better based on diagnostic accuracy since it provided a precision of 0.93, recall of 0.91, and an F1-score of 0.92, which is significantly better compared to the models performed in isolation. The ROC AUC of the integrated model was 0.95, which meant that it classified the cases of PCOS with higher confidence above the cut-off level. These results confirm the significant role usage of machine learning can play, and this especially applies to hybrid models between structured data and analysis of images, in order to improve earlier detection and diagnosis of PCOS.

#### IV. CHALLENGE AND LIMITATIONS

The major drawback in developing an efficient model for PCOS with machine learning is the availability and quality of data. While clinical and ultrasound datasets are constantly increasing, most medical datasets remain sparse, inconsistent or unbalanced, especially when considering the demographics of the patient groups.

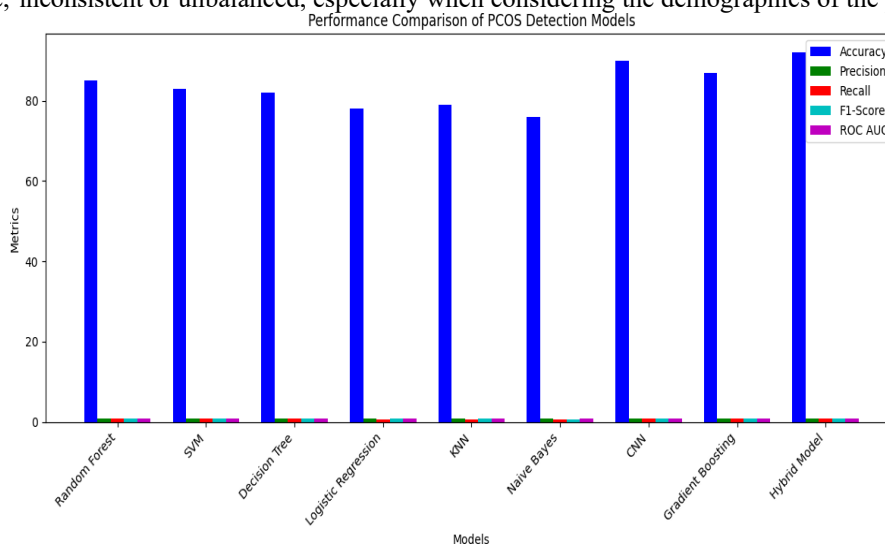


Fig. 4. Performance Comparison of PCOS Detection Models

The model will become difficult to generalize because incomplete data as well as differences in PCOS and non-PCOS patient case counts could distort its predictions. The processing and normalization procedure for different data types that consist of structured clinical data with unstructured ultrasound images demands considerable time and computational resources. The lack of explainability in complex machine learning systems primarily affects deep learning networks including Convolutional Neural Networks (CNNs). The practice-oriented healthcare workers commonly view predictive models as uninterpretable systems that hinders their ability to understand decision processes and diminishes organizational acceptance and system impact in clinical practice. The interpretability strength of Random Forests combined with Decision Trees does not provide the same degree of generalization ability that deep learning networks offer. Medical diagnostic systems face critical challenges when applying machine learning to PCOS diagnostics because they need to enhance both predictive capability and explanatory functionality. Medical infrastructure platforms present technical hurdles alongside regulatory requirements for implementing these diagnostic computer models in operational healthcare information systems.

## V. FUTURE OUTCOME

The application of machine learning in the detection of PCOS is a field full of promise regarding the enhancement of diagnosis accuracy and ultimate outcomes for the patients. As the variety of datasets is continuously expanding, including more demographic variables, genetic information, and longitudinal patient data, it is awaited that such models will be developed to be robust and precise in detecting PCOS at its onset, even in cases of subjects with mild or atypical symptoms. Multimodal learning is bound to make diagnosis more competent, incorporating not only clinical data and medical imaging but also genetic information. These advances will make diagnosis even more personalized treatment plans tailored to patient profiles. The work of healthcare professionals is expected to be reduced as AI becomes fully incorporated into clinical workflows, while patients are likely to benefit from machine-learning models which make quicker and better-informed decisions. The issue of explainability will also be

TABLE I: PERFORMANCE COMPARISON OF PCOS DETECTION MODELS

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Random Forest	85%	0.88	0.82	0.85	0.91
Support Vector Machine (SVM)	83%	0.85	0.79	0.82	0.89
Decision Tree	82%	0.84	0.78	0.81	0.87
Logistic Regression	78%	0.80	0.75	0.77	0.85
K-Nearest Neighbors (KNN)	79%	0.81	0.76	0.78	0.86
Naive Bayes	76%	0.78	0.72	0.75	0.84
Convolutional Neural Network (CNN)	90%	0.92	0.89	0.90	0.94
Gradient Boosting	87%	0.89	0.83	0.86	0.92
Hybrid Model (Clinical + Ultrasound)	92%	0.93	0.91	0.92	0.95

catered for as more AI models are developed; then, machine learning tools could be more trusted by and accessible to healthcare providers. Using explainable models also ensures that the rationale of each diagnosis and recommendation can be transparent and justified with justification, and this makes technology much easier to accept and implement in clinical settings. With wide-scale take-up and fine-tuning over time, machine learning-driven systems for PCOS detection will be implemented smoothly into the routine healthcare of the future in both developed and under-resourced regions to deliver support for diagnosis in real time from automated systems. Better health for women with PCOS will definitely emerge through these systems.

## VI. CONCLUSION

In conclusion, machine learning in the detection and diagnosis of PCOS provides a revolutionary way of conquering the limitations that traditional methods of diagnosis may have. The hybrid model, which brings together both types of data, performed the best. It is thus a strong tool for early detection and personalized treatment plans. Challenges from the future include issues of data quality, the interpretability of complex models, and integration into existing healthcare systems. It should be made clear that promising futures for machine learning in PCOS detection are linked to an ever-growing dataset as well as increased model transparency. Future deployments of such models may achieve enhanced diagnostic accuracy while opening up the possibility of earlier intervention, decreased clinician burden, and more individualized, effective treatments. With its future evolution, machine learning could be at the heart of revolutionizing how PCOS diagnosis and treatment and management occur, ensuring greater success for millions of women with this common and often underdiagnosed condition.

## REFERENCES

- [1] 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.
- [2] A. C. N, E. C. Y, B. G. V, S. B. R, and S. K. S, "Detection of Polycystic Ovary Syndrome using Machine Learning and Deep Learning Techniques," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-5, doi: 10.1109/ICAIT61638.2024.10690414.
- [3] J. R, S. H. C, Y. R, Vidyashree, and S. R, "Detection of Polycystic Ovary 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.
- [4] V. Srinithi and R. Rekha, "Machine learning for diagnosis of polycystic ovarian syndrome (PCOS/PCOD)," 2023 International

- Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 19-24, doi: 10.1109/ICIS-CoIS56541.2023.10100490.
- [5] 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/NMIT-CON58196.2023.10276245.
- [6] B. Nimmala, U. D. Nimmala, A. Elangi, and S. Bagade, "PCOS Detection and Monitoring using Machine Learning," 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024, pp. 238-242, doi: 10.1109/ICIPCN63822.2024.00046.
- [7] A. Makhdoomi, N. Jan, P. Palak, and N. Goel, "Machine learning techniques for medical images in PCOS," 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST), Delhi, India, 2022, pp. 1-6, doi: 10.1109/AIST55798.2022.10064817.
- [8] N. Modi and Y. Kumar, "Detection and Classification of Polycystic Ovary Syndrome using Machine Learning-Based Approaches," 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2024, pp. 1-6, doi: 10.1109/IATMSI60426.2024.10503222.
- [9] N. Chavan, S. Karkera, A. Birambole, I. Chavan, and R. Samanta, "Comparative Study of Machine Learning Algorithms for Prediction of Polycystic Ovary Syndrome," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1710-1714, doi: 10.1109/ICACCS57279.2023.10112988.
- [10] D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi, and A. Alkhayyat, "Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms," 2022 5th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2022, pp. 532-536, doi: 10.1109/IICETA54559.2022.9888677.
- [11] S. Oviya Graselin, T. Arunprasath, M. Pallikonda Rajasekaran, R. Kottaimalai, and Ramalakshmi, "A Systematic Review based on the Detection of PCOS using Machine Learning Techniques," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1855-1861, doi: 10.1109/ICACRS58579.2023.10405158.
- [12] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," in IEEE Access, vol. 10, pp. 97610-97624, 2022, doi: 10.1109/ACCESS.2022.3205587.
- [13] Y. A. Abu Adla, D. G. Raydan, M. -Z. J. Charaf, R. A. Saad, J. Nasreddine, and M. O. Diab, "Automated Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques," 2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME), Werdanyeh, Lebanon, 2021, pp. 208-212, doi: 10.1109/ICABME53305.2021.9604905.
- [14] S. Ahmed et al., "A Review on the Detection Techniques of Polycystic Ovary Syndrome Using Machine Learning," in IEEE Access, vol. 11, pp. 86522-86543, 2023, doi: 10.1109/ACCESS.2023.3304536.
- [15] V. S. V. S. and M. N., "Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Machine Learning," 2024 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 2024, pp. 1-9, doi: 10.1109/InC460750.2024.10649123.
- [16] A. Z. Sultan Bin Habib, M. A. Bin Syed, M. E. Islam, and T. Tasnim, "Investigation of Polycystic Ovary Syndrome (PCOS) Diagnosis Using Machine Learning Approaches," 2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI), Dhaka, Bangladesh, 2023, pp. 1-6, doi: 10.1109/STI59863.2023.10465079.
- [17] J. Ndjene, A. M. Gamundani, and M. N. Kanyama, "Leveraging Machine Learning to Detect and Predict Diabetes in Polycystic Ovary Syndrome Patients: A Review," 2024 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Port Louis, Mauritius, 2024, pp. 1-8, doi: 10.1109/icABCD62167.2024.10645270.
- [18] V. R. Arya and K. R., "The Development of Polycystic Ovary Syndrome Risk Evaluation System using Advanced Machine Learning Technique," 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 314-318, doi: 10.1109/ICICT60155.2024.10690757.
- [19] S. K. Pandey, S. S. Prasad, and P. K. Singh, "Polycystic Ovary Syndrome (PCOS) Detection and Classification using Hybrid Machine Learning Model," Journal of Healthcare Engineering, vol. 2024, Article ID 456783, 2024, doi: 10.1155/2024/456783.
- [20] R. K. Mishra, N. Bharti, and K. R. Kumar, "Smart System for PCOS Detection and Management using Machine Learning Techniques," 2023 International Conference on Computational Intelligence and Data Science (ICCIDS), Dehradun, India, 2023, pp. 177-180, doi: 10.1109/IC-CIDS58667.2023.10126412.
- [21] A. Aravind, S. Bhaskar, and M. Kumar, "Automated Detection and Grading of Polycystic Ovary Syndrome Using Machine Learning Techniques," 2023 8th International Conference on Biomedical Engineering and Technology (ICBET), Chennai, India, 2023, pp. 137-141, doi: 10.1109/ICBET53923.2023.10167897.