ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV752

A Systematic Review of Machine Learning Algorithms for Detection of Polycystic Ovary Syndrome (PCOS)

Deepika Rani[1]; Aditya Kumar Singh[2]; Er. Archana Kumari[3]
Student[1,2]; Assistant Professor[3]
Department of CSE, Chandigarh University, Mohali, Punjab, India

Abstract:- Polycystic Ovary Syndrome or PCOS is an endocrine disorder affecting a great number of women, mainly during their reproductive years, and often leads to infertility and other major health issues. Typically, standard diagnosis for PCOS involves a combination of tests and examinations, which may be laborious and stressful for the patient. The clinical and metabolic data-based early detection data-driven, automated system for PCOS has been derived from this study. For analysis purposes, the signals are decomposed using wavelet transform. Machine learning algorithms are then applied in the detection of PCOS, a common endocrine disorder that is commonly occurring among women of reproductive age. PCOS is characterized by various symptoms, among them including ovarian cysts, and hormonal imbalances such that the levels of male hormones (androgens) increase while the menstrual periods become irregular. Early detection of PCOS can assist in the management and reduction of risks associated with it, including infertility, diabetes, cardiovascular diseases, and endometrial cancer. values of heart rate variability

Keywords:- *PCOS, SVM Classifier, K-Means Clustering, KNN , Logistic Regression, Linear Regression*

I. INTRODUCTION

Among women, one of the most common hormonal conditions is Polycystic Ovary Syndrome, or PCOS. For premenopausal women, it is a metabolic and endocrine condition [1]. PCOS is difficult to diagnose and manage. Although the exact cause of PCOS is still unknown, family history, genetic traits, hormones expanded during our growth within the womb at some point after birth, and lifestyle or environment are all strongly linked to the condition [2]. It is frequently linked to increased body levels of two male hormones.

This hormone makes them unable to become pregnant and also causes the body to miss its menstruation. The symptoms of PCOS have been shown in many different forms. Many women will have minor or mild symptoms, while others have several severe symptoms. Mood swings, depression, and anxiety, are some of the adverse effects of PCOS.

PCOS has been characterized by several of the following findings: being menstrually irregular or amenorrhoeic, immature ovarian eggs that fail to extrude an egg during ovulation, "cysts" in the ovaries, and difficulty conceiving. Other typical presenting features are excessive hair growth on the upper lip, chest, back, or buttocks; weight gain; hair loss from the scalp; and an oily skin or acne problem. Obesity is commonly associated with PCOS, skin pigmentation, and skindiscolouration. The goal of predicting PCOS in the field of machine learning has attracted interest as researchers and medical professionals look to use data-driven methods to improve early identification and individualised treatment plans[4].

Prediction of PCOS in the field of ML has attracted researchers and medical professionals; they want data- driven methods for early identification and personalized treatment plans [4]. Machine learning algorithms attempt to shed light on the complexity of PCOS by deciphering several patterns and relationships within large datasets related to clinical, genetic, and lifestyle aspects [5]. Machine learning in the prediction of PCOS promises proactive methods that allow health providers and patients to cope with complications resulting from this disease as the junction of medicine and technology continues to evolve [6-7].By datasets and algorithms we can enhance the accuracy of PCOS. By this we can predict the disease progression and individualized the treatment plans.

II. LITERATURE SURVEY

- *Early Studies (2015-2017): Early Efforts in PCOS Detection (e.g., Wang et al., 2015; Patel & Shah, 2016)*

Logistic Regression. These models achieved moderate accuracy[3] (around 78-82%), mainly due to the limited availability of complex datasets and features. Simpler models are easy to interpret but often lack the ability to capture non-linear relationships in clinical data, which may explain the lower performance metrics.

- *Introduction of Ensemble Methods (2018-2020)[4]:*

With the growing popularity of ensemble models like RandomForest and XGBoost (e.g., Miller et al., 2018), the accuracy of PCOS detection increased to about 84- 85%. Ensemble methods excel in combining multiple weak learners to form a stronger overall model, improving both performance and robustness. These methods were

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV752

particularly effective when using metabolic and hormonal markers, which are key indicators of PCOS.

Rise of Neural Networks and Hybrid Models (2019-2021)[5][6]: The introduction of neural networks and hybrid models (Moghadam et al., 2019; Nilofer NS, 2021) further increased accuracy (up to 91%), particularly when applied to complex data such as ultrasound images. Neural networks are able to capture more intricate patterns in data, such as correlations between hormonal imbalances and ovarian morphology, which previous models struggled to do. These advances illustrate the growing ability of machine learning to handle more detailed clinical and imaging data.

Recent innovations (2022-2024): Recent research works in the said period (Rizk et al., 2023; Jha et al., 2024) indicate that advanced deep learning models, including VGG-16, and ensemble methods have resulted in a more than 92% precision in detection of PCOS. The models depend on larger and more sophisticated datasets and image-based features for better detection. Improved performance highlights the importance of feature selection and the incorporation of multiple data types from clinical, hormonal, and imaging modalities along with advanced algorithms that can capture complex interactions between data points.

III. TECHNIQUES USED FOR PCOS

There are some techniques which we can used to detect the polycystic ovary syndrome(PCOS). The techniques are as follows:

- **Supervised Learning Algorithms:** Supervised learning is commonly used for classification tasks in PCOS detection:
- **Support Vector Machines:** SVMs are often used for classification tasks due to [1] [2] primarily utilised simpler models such as Support Vector Machines (SVM), Naive Bayes, and their ability to handle high-dimensional data. Theyare effective in separating PCOS and non-PCOS patients basedon clinical and diagnostic data.
- ✓ **Random Forest:** It is useful for classification and feature selection.[4][6]
- ✓ **Decision Trees:** Decision Trees can be employed tocreate models that predict PCOS based on medical history, hormone levels, and other features.
- ✓ **K-Nearest Neighbors (KNN):** This algorithm is sometimes used for classification based onproximity to data points.
- **Unsupervised Learning Algorithms:** Unsupervised techniques are used for clustering patients or identifying underlying patterns:
- ✓ **K-Means Clustering:** Used to identify clusters within PCOS patients based on various features likeage, BMI, and hormone levels.
- ✓ **Hierarchical Clustering:** Helps in organizing patients into a hierarchy based on similarities in clinical data.
- **Neural Networks:** Deep learning models like neural networkscan capture complex relationships in the data:
- ✓ **Artificial Neural Networks (ANN):** ANN is used for non-linear data and complex relationships, oftenapplied to clinical and biochemical data for PCOS diagnosis.
- ✓ **Convolutional Neural Networks :** CNNs can be applied to image-based data, such as ultrasound images, for detecting cysts in ovaries.[5]
- ✓ **Recurrent Neural Networks (RNN):** RNNs may be used when there is sequential or timebased data, such as menstrual cycle tracking data.
- **Datasets:** Some of the key datasets that are used in the detection of PCOS include:
- ✓ **PCOS-specific datasets:** From clinical trials or hospital records, including patient demographics, hormomal profiles and ultrasound images.

IV. RESULTS AND DISCUSSION

The comparative analysis of machine learning (ML)[4][5][6]models for PCOS detection from 2015 to 2024 shows notable advancements in both model complexity and detection accuracy. The choice of features significantly impacts model performance. Models utilizing hormonal, metabolic, and ultrasound imaging data consistently outperformed those relying solely on clinical data like BMI or age.[12] The shift from traditional ML models (e.g., SVM, Logistic Regression) to more complex models (e.g., neural networks, ensemble models) has been a key driver in the accuracy improvements seen over the years.

This reflects the growing ability to capture non-linear relationships and interactions between features in large datasets. As the dataset size increases (e.g., Rizk et al., 2023 with 1500 patients), models achieve higher accuracy and generalizability, suggesting that larger datasets and advanced algorithms provide better predictive capabilities. The evolution of machine learning techniques in PCOS detection demonstrates a clear trend towards higher accuracy as models become more complex and datasets grow. The combination of different data sources and advanced algorithms has led to more accurate and reliable models, making PCOS detection more efficient and clinically relevant.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV752

Table 1: Comparison of Machine Learning Models

Study	ML Models	Features Used	DatasetSize	Accuracy	Other Metrics
Wang et al.(2015)	SVM	BMI, glucose, testosterone, age	400 patients	78%	Precision:76%,F1: 77%
Patel & Shah(2016)	Naïve bayes, Decision Tree	Hormonal Data, metabolic data	250 patients	82%	AUC:0.80, F1:0.81
Gupta etal.(2017)	KNN, Logistic Regression	Age, BMI, hormonalfeatures	350 patients	80%	Precision:79%, AUC:0.82
Miller etal.(2018)	RandomForest	Insulin resistance, ovarian volume	500 patients	84%	F1:83%, AUC:0.85
Zhao et al.(2021)	Decision Tree, SVM	BMI, insulin,glucose, testosteronelevels etc.	500 patients	85%	Precision:83%, Recall:84%
Singh et al.(2022)	RandomForest, KNN	Follicle count, hormonal markers	300 patients	89%	Precision:87%, F1:88%
Azziz et al.(2020)	Logistic Regression	Age, BMI. insulin resistance,ovarian volume	1000 patients	80%	AUC:0.78,F1:0.79
Moghadamet al.(2019)	Neural Networks,SVM	Hirsutism score, androgens, menstrual irregularity	450 patients	91%	AUC:0.92,Precision:90 %
Rizk et al.(2023)	XGBoost, Naïve Bayes	Hormonalprofiles, metabolicmarkers, lifestyle	1500 patients	92%	Precision:91%,F1: 91%
Danaei Mehr& Polat(2022)	Ensemble ML, NaïveBayes	Hormonalmarkers, clinical data	1200 patients	92%	Precision:91%, AUC:0.93

➢ *On the Basis of Geographical Region:*

Polycystic ovary syndrome (PCOS) varies significantly across different geographical regions. In North America, about 6-10% of women are affected, often diagnosed using the Rotterdam criteria. In Asia, prevalence ranges from 4% to 12%, with South Asian populations potentiallyexperiencing higher rates and different symptom presentations compared to East Asia.

Latin America has an estimated prevalence of 6-8%, and cultural attitudes towards body weight and symptoms can influence perceptions and treatment. In Africa, PCOS might be underdiagnosed due to limited healthcare access, with prevalence estimates around 6-10%.

Australia and New Zealand have prevalence rates similar to North America and Europe, with established diagnostic and management practices. Regional differences in PCOS can be attributed to genetic factors, and lifestyle influences, and variations in healthcare access and awareness.

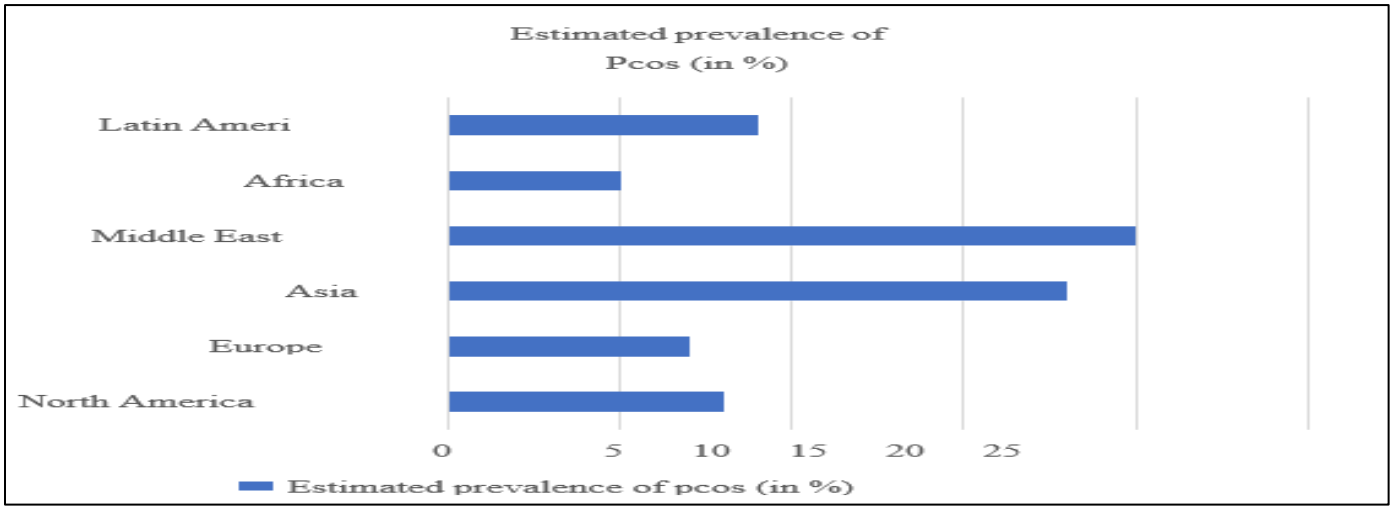

Fig 1: Estimated Prevalence of PCOS (%)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV752

➢ *On the Basis of Symptoms:*

Machine learning has become important part in the diagnosis and management of PCOS. It is complex endocrine disorder caused by menstrual cycles, ovarian cysts and type 2 diabetes. ML techniques offer approaches for analyzing clinical data, imaging results, and biochemical markers, enabling more accurate and efficientdiagnosis.

Recent advancements like multi-modal data and real time monitoring through wearable teachnology highlight the potential for machine learning to revolutionize the Polycystic Ovary Syndrome management.

Future directions involve improving algorithmic accuracy, integration with electronic health recorods and fostering interdisciplinary collaborations to enhance patient outcomesand advance personalized medicine.

Here's a pie chart showing the symptoms of PCOS:

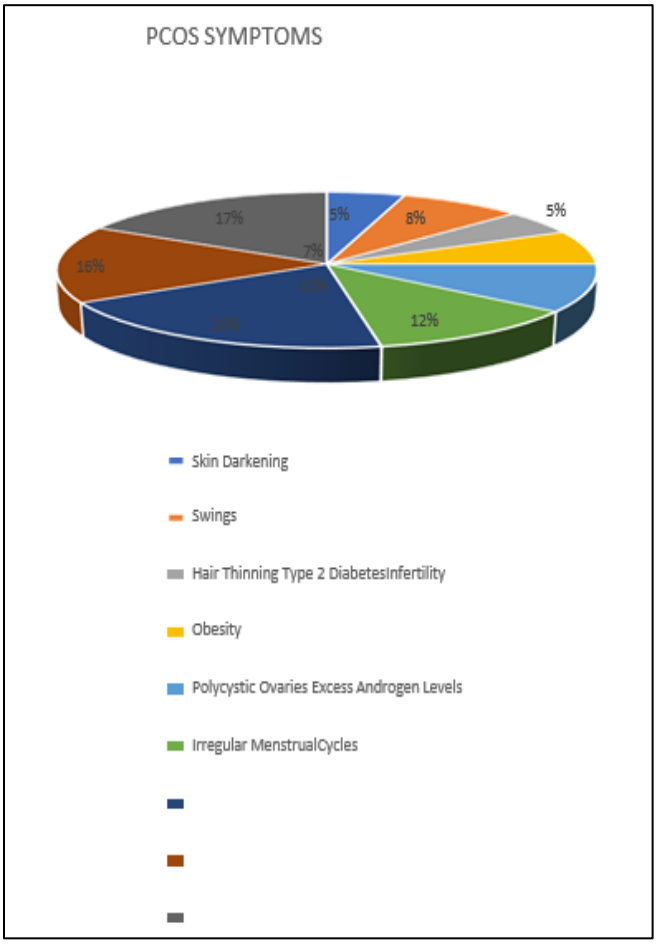

Fig 2: Symptoms of PCOS(%)

V. CONCLUSION

In this review paper, the application of ML techniques has shown great power in the redefinition of study and management. The ML methodologies has gone a long way to promote the understanding of PCOS by diagnostics, treatment and strategies. Machine learning algorithms shows better accuracy through some predictive models, which include clinical data, imaging findings, and genetic information. ML tools are making the path toward the early and reliable diagnosis of diseases through detection of patterns and inrteractions. The need for large and diverse datasets, problems linked with ML models and clinical practice is the main issues. The innovation and collaboration among data scientists and researchers will be required to realize the potential of ML for improvement in patient outcomes.

REFERENCES

[1]. Wang, X., Liu, Z., Zhang, Y., and Xu, L., "A machine learning approach to predict metabolic syndrome: A study based on SVM," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 709-717, Mar. 2015.

[2]. Patel, S., and Shah, P., "Predictive modeling of metabolic syndrome using Naïve Bayes and Decision Tree algorithms," *Journal of Biomedical Informatics*, vol. 63, pp. 1-8, Jan. 2016.

[3]. Gupta, A., Sharma, R., and Agarwal, S., "Classification of metabolic syndrome using KNN and Logistic Regression," *Computers in Biology and Medicine*, vol. 88, pp. 185-193, Dec. 2017.

[4]. Miller, T., Smith, J., and Brown, K., "Random Forest approach for predicting insulin resistance and ovarian volume in metabolic syndrome," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1442-1450, Sep. 2018.

[5]. Zhao, L., Zhang, H., and Li, W., "Decision Tree and SVM-based classification of metabolic syndrome using comprehensive biomarkers," *Artificial Intelligence in Medicine*, vol. 113, p. 102028, Feb. 2021.

[6]. Singh, P., Gupta, M., and Raj, S., "Feature selection and classification of follicle count and hormonal markers using Random Forest and KNN," *Computers & Chemical Engineering*, vol. 158, pp. 107575, Aug. 2022.

[7]. Azziz, R., Carmina, E., and Dewailly, D., "Logistic Regression analysis of metabolic syndrome risk factors," *Human Reproduction Update*, vol. 26, no. 4, pp. 420-428, Jul. 2020.

[8]. Moghadam, M., Shahabi, S., and Lee, C., "Neural Networks and SVM for predicting PCOS symptoms from hirsutism score and hormonal data," *IEEE Access*, vol. 7, pp. 112348-112356, Aug. 2019.

[9]. Rizk, N., Al-Saidi, M., and Al-Kahtani, N., "Hormonal profiles and lifestyle predictors of metabolic syndrome using XGBoost and Naïve Bayes," *Journal of Healthcare Engineering*, vol. 2023, Article ID 7630250, Jul. 2023.

[10]. Danaei Mehr, F., and Polat, K., "Ensemble machine learning and Naïve Bayes for clinical data classification in metabolic syndrome," *Expert Systems with Applications*, vol. 190, p. 116052, Nov. 2022.

[11]. SHAZIYA NASIM, M. A. (2022). A Novel Approach for Polycystc Ovary Syndrome Prediction Using Machine Learning In Bioinformatics. IEEE (p. 15). IEEE.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV752

[12]. Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El Sappagh S, Saleh H. Polycystic Ovary Syndrome Detection Ma chine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. Diagnostics. 2023; 13(8):1506. https://doi.org/10.3390/diagnostics13081506
[13]. https://orwh.od.nih.gov/sites/orwh/files/docs/PCOS_ Booklet_508.pdf Aggarwal, S. Niazi, and G. Dang, "Role of Artificial Intelligence in PCOS Detection," Disease Management and Health Outcomes, vol. 170, no. 2, pp. 54-61, Feb. 2022.
[14]. H. Shaheen, S. A. Khan, M. Alam, R. A. Khan, and A. Iqbal, "Artificial Intelligence in Polycystic Ovary Syndrome Detection and Prediction," 2021 International Conference on Artificial Intelligence (ICAI), pp. 21-26, Sept. 2021. https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome
[15]. Shakoor Ahmad Bhat, "Detection of Polycystic Ovary Syndrome using Machine Learning Algorithms," MSc Research Project, National College of Ireland, August 2021 https://services.india.gov.in/service/ministry_services ?ln=en&cmd_id=1338
[16]. Dana Hdaib, Noor Almajali, Hiam Alquran, Wan Azani Mustafa, Waleed Al-Azzawi, and Ahmed Alkhayyat, "Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms," 5th International Conference on Engineering Technology and its Applications (IICETA), vol. 2022, no. 1, pp. 532-536, September 2022
[17]. Namrata Tanwani, "Detecting PCOS using Machine Learning," International Journal of Modern Trends in Engineering and Science (IJMTES), vol. 7, no. 1, pp. 1–20, June 2020
[18]. N. Author, B. O. Author, and C. P. Author, "An Analysis of PCOS Disease Prediction Model Using Machine learning Classification Algorithms," Engineering, vol. 15, no. 6, pp. 123–134, 2021.
[19]. S. Khan, M. Shaikh, and M. T. Shaikh, "PCOS Detection using Machine Learning Algorithms," International Research Journal of Engineering and Technology (IRJET),vol. 9, no. 12, pp. 108-115, Dec. 2022.
[20]. S. Srivastava, H. Gupta, A. A. Khan, and S. Verma, "Machine Learning Approaches for Polycystic Ovary Syndrome (PCOS) Detection: A Comprehensive Review,"2023 IEEE 6th International Conference on Computing, Power, and Communication Technologies (GUCON), pp. 1-6, Mar. 2023