ICEST-2K24 [International Conference on Engineering, Science and Technology]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Review on Polycystic Ovary Syndrome Detection Using Machine Learning

Supriya H.Kale[1], Dr.D.B.Hanchate[2], Dr.S.S.Bere[3]

[1]P.G.Scholar, Department of Computer Engineering, DGOIET, Swami Chincholi Daund,Pune-413110, Maharashtra, India

[2]Head Department of Computer Engineering, DGOIET, Swami Chincholi Daund,Pune-413110, Maharashtra, India

[3]Assistant Professor, Department of Computer Engineering, DGOIET, Swami Chincholi Daund,Pune-413110, Maharashtra, India

ABSTRACT

Polycystic ovary syndrome among women of tine or above 16year age has been steadily increasing, necessitating accu rate and early prediction methods. Machine Learning techniques have shown great potential in this area due to their capability to analyze huge of data and identify complex datasets. This research paper aims to review existing studies on the PCOS prediction with the help of Machine Learning(ML). PCOS, a prevalent endocrine disorder in women during their reproductive years, is marked by hormonal imbalances, including elevated levels of male hormones like androgen.

Index Terms—PCOS, ML,SVM, Hormonal Fluctuations ,an drogen,Supervised,Data Sources

I. INTRODUCTION

Polycystic ovary syndrome is a multifaceted condition that impacts numerous females globally.It is by hormonal fluctua tions and the increased level in male hormone and androgen, irregular period's cycle, and the tiny cysts on the ovaries. The exact cause of PCOS is not fully understood, and its symptoms can vary widely from person to person [1].

The complexity of polycystic ovary syndrome extends be yond its physiological manifestations, delving into the psycho logical and emotional well-being of affected individuals [2]. The condition not only disrupts normal hormone levels and menstrual cycles but also there may be chance of abortion in pregnancy due to PCOS and In future there may be chance of gynecological cancer. PCOS growing in many females at an early edges. Moreover, the heterogeneity of PCOS symptoms poses a challenge in accurately diagnosing and managing the condition [3]

In the realm of machine learning, the quest to predict PCOS has garnered attention as researchers and healthcare profes sionals seek to leverage data-driven approaches to enhance early detection and personalized treatment strategies [4]. By uncovering patterns and correlations within extensive datasets encompassing clinical, genetic, and lifestyle factors, machine learning algorithms aim to offer insights into the multifaceted nature of PCOS [5]. As the intersection of medicine and technology continues to evolve [6], the ability of machine learning in PCOS

fore casting holds promise for empowering healthcare providers and individuals with proactive measures toward addressing the complexities of this condition [7].

II. LITRETURE REVIEW

Several studies have great attention on the use of machine learning techniques for the prediction and diagnosis of PCOS. These studies have shown promise in analyzing patterns and markers that can cooperate in the early detection as well as management of PCOS. By analyzing large datasets of patient clinical data, machine learning algorithms can help identify risk factors and early indicators of PCOS, leading to more proactive and personalized care for affected individuals [8]– [11].

For instance, an article by Xie et al. utilized a machine learning approach called Random Forest to predict PCOS based on clinical and biochemical features [12]The research attained an 85 percent accuracy rate in forecasting PCOS, showcasing the promise of machine learning in this field. In a separate investigation conducted by Chen et al., a Support Vector Machine algorithm was utilized to anticipate PCOS by integrating clinical, metabolic, and genetic indicators [17] The research reported an accuracy of 89 percentin PCOS prediction, further reinforcing the potential of machine learn ing for improved diagnosis and management. Furthermore, machine learning has also been utilized in predicting specific manifestations and associated complications of PCOS (Chen, 2019) [17].For instance, a study by Huang et al. utilized a deep learning model to predict insulin resistance in women with PCOS.

III.METHODOLOGY

The methodologies employed in machine learning for PCOS prediction vary across studies, with some utilizing supervised learning techniques such as logistic regression, decision trees, and support vector machines. Others have explored the po tential of unsupervised learning methods like clustering and dimensionality reduction to uncover hidden patterns within diverse datasets.

Furthermore, the integration of multiomics data, including genomics, metabolomics, and proteomics, has emerged as a key area of interest in advancing the accuracy and depth of PCOS prediction models. Leveraging these comprehensive molecular datasets alongside incorporating clinical and demo graphic data enables a comprehensive grasp of the fundamental factors influencing the onset and advancement of PCOS.

IV.CURRENT CHALLENGES AND FUTURE DIRECTIONS

Despite the promising outcomes of machine learning in PCOS prediction, several challenges persist. The heterogeneity of PCOS symptoms and manifestations necessitates the need for more robust and adaptable algorithms that can accommo date the diverse clinical presentations of the condition. Further more, challenges concerning data quality, standardization, and interoperability among healthcare systems impede the efficient utilization of machine learning for predicting PCOS.

In future research directions may involve the mixture of real-time or longitudinal data to capture dynamic changes in PCOS-related parameters, enabling proactive interventions and personalized monitoring. Additionally, the development of interpretable machine learning models that provide transparent insights into the reasoning behind predictions will be crucial in gaining the trust of healthcare providers and facilitating the

translation of research findings into medical practice.In conclusion, the intersection of machine learning and PCOS prediction represents a compelling frontier in addressing the complexities of this multifaceted condition. By overcoming current challenges and embracing future advancements, the field holds great potential in revolutionizing the early detection, management, and individualized care of individuals affected by PCOS.

V. POLYCYSTIC OVARY SYNDROME

A. Introduction

PCOS is gynecological condition occurs due to hormonal fluctuations .Hormonal Fluctuations may cause in ovaries generating multiple cysts. Ovaries are unable to fertilize eggs or sometimes occurs critical complications in pregnancy.

The main symptoms of PCOS are androgen level increased, disturbed periodic cycle and ovaries with cysts. There may be chance of abortion in pregnancy due to PCOS and In future there may be chance of gynecological cancer. PCOS growing in many females at an early edges. There are 69 to 70 percent women's of the recent report. As per 2023 reports 12.8 percent females of America,8 Percent womens of Spain and 31.3 womens of Asia is diagnosed PCOS (SHAZIYA NASIM, 2022) [1]

To treat PCOS need to change daily lifestyle ,need to reduce weight take diet as per prescribed. Daily exercise has great outcomes in minimizing androgen level and hy per androgenism. Machine learning, a branch of artificial intelligence, has shown promise in predicting and diagnosing PCOS using various data sources such as hormone levels, ultrasound images, and clinical symptoms. In this review, we will explore the current state of machine learning applications for PCOS prediction, including the approaches, challenges, and anticipation. Through this review, we aim to provide a

Fig. 1. Architecture of PCOS

Sr no.	Feature	Non-Null Count	Data Type	Sr no.	Feature	Non-Null Count	Data Type
1	Sl. No	541	int64	22	TSH (mIU/L)	541	float64
2	Patient File No.	541	int64	23	AMH(ng/mL)	540	float64
3	PCOS (Y/N)	541	int64	24	PRL(ng/mL)	541	float64
4	Age (yrs)	541	int64	25	Vit D3 (ng/mL)	541	float64
5	Weight (Kg)	541	float64	26	PRG(ng/mL)	541	float64
6	Height (Cm)	541	float64	27	RBS(mg/dl)	541	float64
7	BMI	541	float64	28	Weight gain(Y/N)	541	int64
8	Blood Group	541	int64	29	hair growth(Y/N)	541	int64
9	Pulse rate(bpm)	541	int64	30	Skin darkening (Y/N)	541	int64
10	RR (breaths/min)	541	int64	31	Hair loss(Y/N)	541	int64
11	Cycle(R/I)	541	int64	32	Pimples(Y/N)	541	int64
12	Cycle length(days)	541	int64	33	Fast food (Y/N)	540	float64
13	Marriage Status (Yrs)	540	float64	34	Reg. Exercise(Y/N)	541	int64
14	Pregnant(Y/N)	541	int64	35	BP _Systolic (mmHg)	541	int64
15	No. of abortions	541	int64	36	BP _Diastolic (mmHg)	541	int64
16	FSH(mIU/mL)	541	float64	37	Follicle No. (L)	541	int64
17	LH(mIU/mL)	541	float64	38	Follicle No. (R)	541	int64
18	FSH/LH	541	float64	39	Avg. F size (L) (mm)	541	float64
19	Hip (inch)	541	int64	40	Avg. F size (R) (mm)	541	float64
20	Waist (inch)	541	int64	41	Endometrium (mm)	541	float64
21	Waist- Hip Ratio	541	float64				

Fig. 2. Dataset exhausted understanding of the potential of machine learning in the analyzing and management of PCOS.

B. Machine Learning-based Approaches for PCOS Prediction

In our swiftly evolving world, the importance of precise predictions and early detection for diverse health conditions cannot be emphasized enough. Machine learning has surfaced as a valuable asset in healthcare studies, facilitating the cre ation of predictive models that aid in timely disease diagnosis and management.

C. Data Sources for PCOS Prediction

To predict and diagnose PCOS using machine learning, various data sources are utilized, including hormone levels, ovarian morphology from ultrasound images, and clinical symptoms. Hormone levels such as luteinizing hormone, follicle-stimulating hormone, and androgens play a crucial role in the diagnosis of PCOS. Machine learning models can analyze these hormone levels along with other biomarkers to identify patterns and correlations indicative of PCOS.

Moreover, ultrasound scans of the ovaries offer valuable insights for predicting PCOS. Ovarian morphology, indicated by the presence of numerous small follicles, is a pivotal feature of PCOS. Machine learning algorithms can be instructed to scrutinize these ultrasound images and recognize patterns linked with polycystic ovaries. Additionally, clinical manifes tations like irregular menstrual cycles, acne, and excessive hair growth serve as significant markers for PCOS.By incorporating these clinical symptoms into the predictive models, machine learning can provide a more comprehensive approach to PCOS prediction.Specific methodologies and challenges associated with utilizing these data sources for PCOS prediction using machine learning techniques are:svm,RF,GNB.

D. Machine Learning Algorithms for PCOS Prediction

Machine learning algorithms have shown promise in pre dicting and diagnosing polycystic ovary syndrome. These algorithms can analyze and interpret various data sources to identify patterns and make accurate predictions. Various machine learning algorithms have been employed for PCOS prediction, including decision trees, random forests, support vector machines, logistic regression, artificial neural networks, and ensemble methods. Decision trees are popular for PCOS prediction due to their capability to handle both numerical and categorical data.

E. Importance of PCOS Prediction

Polycystic Ovarian Syndrome affects millions of women worldwide and can have significant impacts on their health and well-being. The ability to accurately predict and diagnose PCOS is crucial in providing timely and effective medical interventions to reduce risk for affected individuals. With the advancement of machine learning algorithms, there is potential to enhance the fidelity of PCOS prediction, leading to earlier detection and better management of this condition. (Polycystic ovary syndrome, 2023) [3]

F. Understanding the Complexity of PCOS

PCOS is a complex internal secretion disorder characterized by hormonal fluctuation, irregular menstrual cycles, and the presence of ovarian cysts. Additionally, PCOS is associated with an increased risk of abortion, obesity, type 2 diabetes, and cardiovascular disease. By delving deeper into the in tricate interplay of various biological markers and clinical manifestations, machine learning algorithms can offer a more comprehensive understanding of PCOS, potentially uncovering subtle patterns and associations that may not be apparent through old diagnostic approaches. [12]–[16]

G. Leveraging Machine Learning for PCOS Prediction

Machine learning has the potential to revolutionize PCOS prediction by utilizing large amounts of data to analyze patterns and relationships that can aid in fast detection and accurate diagnosis. With the help of ML algorithms, healthcare professionals can develop models that integrate diverse bio logical markers and clinical variables to create a more holistic approach to PCOS prediction.

H. The Promise of Personalized Medicine

Furthermore, machine learning techniques can enable the development of personalized predictive models that take into account individual variations in the presentation of PCOS. This tailored approach holds the promise of

optimizing treatment strategies and interventions based on the specific characteris tics and needs of each patient, ultimately improving results and wellbeing health for individuals with PCOS.

MACHINE LEARNING TECHNIQUES FOR PCOS PREDICTION

Various machine learning methods have been explored in light of this PCOS prediction.These include but are not limited to:

1) Support Vector Machines
2) Random Forest
3) Neural Networks
4) Logistic Regression
5) Naive Bayes Classifier
6) Gradient Boosting

Each of these techniques has its own strengths and weak nesses, and their suitability for PCOS prediction may vary depending on the specific dataset and features being consid ered. Research studies have demonstrated the effectiveness of these techniques in accurately classifying and predicting PCOS based on various clinical and biological markers.

VI. INTEGRATION OF MULTIMODAL DATA FOR HOLISTIC PREDICTION

In addition to traditional clinical and genetic data, the mixure of multimodal data, like imaging and omics data, has shown promise in enhancing the predictive capabilities of machine learning models for PCOS. Integrating information from multiple modalities can provide a more extensive view of the basic physiological and molecular characteristics asso ciated with PCOS, leading to improved prediction accuracy and personalized healthcare strategies.

CHALLENGES AND LIMITATIONS

Despite the potential advantages, employing machine learn ing for PCOS prediction presents certain challenges and limi tations. Although the prospect of utilizing machine learning in PCOS prediction is promising, there are hurdles that must be tackled. These challenges encompass the necessity for stan dardized data collection and management, the comprehension of intricate model outcomes, and the ethical deliberations concerning the utilization of confidential health information. Addressing these challenges will be crucial in realizing the full potential of machine learning in PCOS prediction and man agement (Che, 2023) [20]. Moving forward, further research efforts should focus on strain existing prediction models, integrating diverse data sources, and validating the clinical applicability of machine learning-based approaches. Collabo ration between researchers, clinicians, and data scientists will be essential in advancing the field of PCOS prediction and improving healthcare outcomes for individuals affected by this syndrome.

a) *1. Overcoming Data Heterogeneity::* One of the pri mary challenges in leveraging machine learning for PCOS pre diction is the heterogeneity of data sources. Biological markers and clinical variables related to PCOS are often captured using various measurement techniques and instruments, leading to a diverse range of data formats.
b) *2. Leveraging Data for Improved Prediction Models::* The availability of large datasets encompassing clinical, ge netic, and biochemical information allows for the growth of robust machine learning models for PCOS prediction. These models can leverage diverse data sources to identify key features and risk factors associated with PCOS, facilitating more accurate predictions and personalized interventions. Ad

ditionally, the integration of advanced imaging techniques and molecular profiling further expands the scope of data-driven approaches, enabling a multidimensional assessment of PCOS related factors.

By delving deeper into the complexities of PCOS and hitch up the power of machine learning, researchers and healthcare professionals can strive towards more effective prediction and management of this condition, ultimately making a positive impact on the lives of individuals affected by PCOS.

c) *3. Uncovering Subtle Patterns in PCOS:* Uncovering Subtle Patterns in PCOS: PCOS is a multifaceted condi tion with intricate connections between hormonal imbalances, metabolic irregularities, and reproductive health. Machine learning algorithms offer a unique opportunity to uncover subtle patterns and associations within this complex web of in terconnected factors. By analyzing large and diverse datasets, these algorithms can identify hidden relationships that may not be evident through traditional diagnostic methods. This deeper understanding may lead to more precise and personalized interventions for individuals with PCOS, ultimately improving their overall well-being.

d) *4. Personalized Medicine and Tailored Interventions:* The promise of personalized predictive models lies in their ability to account for individual variations in the presentation of PCOS. By taking into consideration a patient's specific biological, clinical, and geneticharacteristics, these models can tailor treatment strategies and interventions to address the unique needs of each individual. This approach holds great potential in optimizing outcomes and quality of life for those with PCOS, fostering a more individualized and effective approach to managing the condition

e) *Expanding the Scope of Data-Driven Approaches:* In addition to clinical, genetic, and biochemical information, the integration of advanced imaging techniques and molecular profiling further enriches the dataset used for PCOS prediction. These multidimensional data-driven approaches enable a more holistic assessment of PCOS-related factors, paving the way for enhanced prediction accuracy and personalized healthcare strategies.

As researchers and healthcare professionals continue to delve deeper into the complexities of PCOS and harness the power of machine learning, the potential for more effec tive prediction and management of this condition becomes increasingly evident. The collaboration between researchers, clinicians, and data scientists will be essential in advancing the field of PCOS prediction, ultimately improving healthcare outcomes for individuals affected by this syndrome.

VII. CONCLUSION

Through leveraging the capabilities of machine learning, there exists the opportunity to enhance the early identification and treatment of PCOS, consequently resulting in improved health outcomes for impacted women. This review endeav ors to offer a thorough examination of the present state of machine learning applications in predicting PCOS, aiming to underscore avenues for future research and advancement.

VIII. REFERENCES

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

[2]. 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

[3]. S. Deodikar, A. Shaikh, S. Jadhav, A. Joshi and S. Mutakekar. ”PCOS Detection: A Study of the Literature”. International Journal for Research in Applied Science and Engineering Technology. vol. 11. no. 2. pp. 544-549. Feb. 2023. https://doi.org/10.22214/ijraset.2023.49079.

[4]. E. Nsugbe. ”An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome”. Healthcare analytics. vol. 3. pp. 100164-100164. Nov. 2023. https://doi.org/10.1016/j.health.2023.100164.

[5]. N. Thomas and A. Kavitha, ”PREDICTION OF POLYCYSTIC OVAR IAN SYNDROME WITH CLINICAL DATASET USING A NOVEL HYBRID DATA MINING CLASSIFICATION TECHNIQUE”.

[6]. D. Anda, ”Comparative Analysis of Artificial Intelligence in the Diag nosis of Polycystic Ovary Syndrome”.

[7]. V. Thakre. ”PCOcare: PCOS Detection and Prediction using Machine Learning Algorithms”. Bioscience biotechnology research communications. vol. 13. no. 14. pp. 240-244. Dec. 2020. https://doi.org/10.21786/bbrc/13.14/56.

[8]. T. M. Barber and S. Franks. ”Obesity and polycystic ovary syn drome”. Clinical Endocrinology. vol. 95. no. 4. pp. 531-541. Jan. 2021. https://doi.org/10.1111/cen.14421.

[9]. S. H.Kale, ”Polycystic Ovary Syndrome Prediction using Machine Learning”.

[10]. J. Lim et al.. ”Machine learning classification of polycystic ovary syndrome based on radial pulse wave analysis”. BMC Comple mentary Medicine and Therapies. vol. 23. no. 1. Nov. 2023. https://doi.org/10.1186/s12906-023-04249-5.

[11]. S. A. Suha and M. N. Islam. ”An extended machine learning technique for polycystic ovary syndrome detection using ovary ul trasound image”. Scientific Reports. vol. 12. no. 1. Oct. 2022. https://doi.org/10.1038/s41598-022-21724-0.

[12]. S. Kharb and A. Joshi. ”Multi-omics and machine learning for the prevention and management of female reproductive health”. Frontiers in Endocrinology. vol. 14. Feb. 2023. 10.3389/fendo.2023.1081667.

[13]. N. Xie, F. Wang, J. Zhou, C. Liu and F. Qu. ”Establishment and Analysis of a Combined Diagnostic Model of Polycystic Ovary Syndrome with Random Forest and Artificial Neural Network”. BioMed Research International. vol. 2020. pp. 1-13. Aug. 2020. 10.1155/2020/2613091.

[14]. R. Miotto, F. Wang, S. Wang, X. Jiang and J. T. Dudley. ”Deep learning for healthcare: review, opportunities and challenges”. Brief ings in Bioinformatics. vol. 19. no. 6. pp. 1236-1246. May. 2017. 10.1093/bib/bbx044.

[15]. P. Soni and S. Vashisht. ”Exploration on Polycystic Ovarian Syndrome and Data Mining Techniques”. Oct. 2018. 10.1109/cesys.2018.8724087. [16] W. Kim. ”Knowledge-based diagnosis and prediction using big data and deep learning in precision medicine”. Investigative and Clinical Urology. vol. 59. no. 2. pp. 69-69. Jan. 2018. 10.4111/icu.2018.59.2.69. [17] Chen, B. (2019). The role of MiRNA in polycystic ovary syndrome(PCOs).Gene,Volume.706,2019,Pages,91-96,ISSN0.378,1119 ,www.sciencedirect.com/science/article/pii/S037811191930

[16]. Polycystic ovary syndrome. (2023, JUNE 28). Retrieved JUNE 28, 2023, from www.who.int: https://www.who.int/news-room/fact sheets/detail/polycystic-ovary-syndrome

[17]. S. Bharati, P. Podder and M. R. H. Mondal. ”Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms”. 2020 IEEE Region 10 Symposium (TENSYMP). Jan. 2020. https://doi.org/10.1109/tensymp50017.2020.9230932.

[18]. Che, Y. (2023). Polycystic Ovary Syndrome: Challenges and Possible Solutions. MDPI.