A REVIEW ON ENHANCED DIAGNOSIS AND MONITORING OF POLYCYSTIC OVARY SYNDROME (PCOS) USING MACHINE LEARNING

Rubeena Rustum
Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous)Hyderabad, Telangana,India.
rubeena.rustum@gmail.com

T. Amulya
Department of CSE
Gokaraju Rangaraju Institute of Engineering & Technology. Telangana,India.
thadaamulyareddy@gmail.com

M. Harshitha
Department of CSE
Gokaraju Rangaraju Institute of Engineering & Technology. Telangana,India.
mharshitha122@gmail.com

B. Shilpi
Department of CSE Gokaraju Rangaraju Institute of Engineering & Technology. Telangana,India.
bukyashilpi2@gmail.com .

Shaik Karishma
Department of CSE Gokaraju Rangaraju Institute of Engineering & Technology. Telangana,India.
Shaikkarishma1764@gmail.com

ABSTRACT: Polycystic Ovary Syndrome (PCOS) represents a substantial health concern for women, underscoring the need for early detection to mitigate associated complications. This review critically evaluates recent advancements in machine learning (ML) and ensemble learning methodologies aimed at constructing transparent and dependable diagnostic models for PCOS. The progress in ML and ensemble learning offers promising prospects in the domain of medical diagnostics. Our research endeavors to formulate a model for PCOS diagnosis, prioritizing transparency and trust through both local and global explanations.To achieve this objective, we employ diverse feature selection methods and various ML models, including logistic regression (LR), random forest (RF), decision tree (DT), naive Bayes (NB), support vector machine (SVM), k-nearest neighbor (KNN), xgboost, and Adaboost algorithms. These models are utilized to identify optimal features and construct the most effective diagnostic model. Stacking ML models, which leverage the strengths of the best base ML models combined with a meta-learner, are proposed to further enhance overall model performance. Bayesian optimization is utilized for fine-tuning the ML models.Addressing the issue of class imbalance is crucial, and we address this concern by incorporating Synthetic Minority Oversampling Techniques (SMOTE) and Edited Nearest Neighbor (ENN). Experimental trials are conducted using a benchmark PCOS data set, divided into two ratios: 70:30 and 80:20. The results indicate that the Stacking ML approach with RF feature selection achieved the highest accuracy, reaching 100%, surpassing other models. This research contributes significantly to the early and accurate diagnosis of PCOS, potentially leading to improved healthcare outcomes for affected individuals.

Keywords-Machine Learning,PCOS,Stacking ML models, Feature Selection,Model performance.

I. INTRODUCTION

The synergy between technology and human efforts holds great potential for advancing healthcare and services.

Machine learning, a subset of artificial intelligence, empowers systems to learn and improve autonomously, without explicit programming. It focuses on developing algorithms that can leverage provided datasets for learning purposes.

Applications of machine learning have ushered in significant transformations in the healthcare industry, encompassing areas such as detection, data prediction, and image recognition. One pressing healthcare concern is Polycystic Ovary Syndrome (PCOS), a prevalent hormonal disorder among women of childbearing age. PCOS is a complex condition associated with infertility, anovulation, cardiovascular disease, type 2 diabetes, and obesity. Shockingly, it often goes undiagnosed, with up to 70% of affected women remaining unaware of their condition.

The treatment of PCOS typically involves medication, lifestyle adjustments, and medical procedures. Treatment options include birth control pills, diabetes management, fertility treatments, anti-androgen medications, and ultrasound scans. In cases where these interventions prove ineffective, surgical procedures like ovarian drilling may be considered to improve ovulation by reducing male hormone levels. The underlying causes of PCOS involve insulin resistance and hyperparathyroidism. Clinically, PCOS manifests with reproductive, metabolic, and psychological symptoms, posing a significant health burden to women. Diagnosis typically relies on clinical, biochemical, and radiological tests. Unfortunately, due to the complexity of its pathophysiology, PCOS is often diagnosed through the exclusion of other conditions with similar symptoms. This leads to a reliance on numerous clinical tests and unnecessary imaging procedures, driving up healthcare costs and causing distress for patients.

Early and accurate PCOS detection through minimal testing and imaging is crucial, as the condition directly affects ovarian function and carries risks of miscarriage, infertility, gynecological cancer, and emotional distress. PCOS is characterized by elevated androgen levels, menstrual irregularities, and the presence of cysts on the ovaries. Skin changes, such as increased body hair and acne, along with metabolic issues, may also occur.

Machine learning's potential in PCOS detection remains relatively unexplored. The system proposed in this paper aims to help individuals understand their PCOS status, whether they are asymptomatic or have the condition. Importantly, it could enable PCOS determination by analyzing medical records alone, eliminating the need for costly and sometimes inaccessible physical examinations, such as ultrasounds. This approach has the potential to make PCOS diagnosis more accessible and efficient for a broader population..

II. LITERATURE SURVEY

Chauhan et al. employed a decision tree classifier and achieved an accuracy of 81% for PCOS diagnosis. Their study suggests the potential for further improving accuracy by tracking symptoms over time and utilizing patient data. Prapty and Shitu evaluated the performance of K-nearest neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, and Random Forest classifiers. While Random Forest performed well, it's worth noting that no feature engineering was conducted in the machine learning process. Abu Adla et al. used a hybrid feature selection approach, combining filters and wrappers, along with SVM to achieve an accuracy of 91.6%. They also employed K-Fold Cross Validation (KCV) to ensure unbiased data splitting, contributing to robust results.

Nabi et al. conducted research on PCOS detection in Bangladeshi women. They collected a dataset, used various classifiers, and achieved impressive results, with SVM, XGB Classifiers, and Gaussian Naive Bayes all achieving high accuracy. Khan Inan et al. proposed the use of Extreme Gradient Boosting (XGBoost) and resampling techniques, including Edited Nearest Neighbour (ENN) and Synthetic Minority Oversampling Techniques (SMOTE). Their study demonstrated the effectiveness of XGBoost in PCOS detection. Bharati et al. ranked features and identified the significant feature as the ratio of Follicle-stimulating hormone (FSH) to Luteinizing hormone (LH). They achieved a testing accuracy of 91.01% using a hybrid random forest and logistic regression approach. Escobar et al. suggests that treatment should be symptom-oriented, long term, dynamic and adapted to the changing circumstances, personal needs and expectations of the individual patient. Oza and Bokhare implemented logistic regression for predicting disease.

IV.METHODOLOGY

1. Machine Learning Algorithms in PCOS Diagnosis:

1.1. Feature Selection and Extraction:

Machine learning algorithms often begin with feature selection and extraction to identify the most informative variables for PCOS diagnosis. Various physiological, hormonal, and clinical parameters can be integrated to improve the accuracy of PCOS detection. Algorithms such as Random Forest, Gradient Boosting, and Recursive Feature Elimination have been employed effectively in this context.

1.2. Predictive Models:

Predictive models for PCOS use supervised learning algorithms (SVM, Logistic Regression, Neural Networks) trained on labeled datasets. SVM establishes a hyperplane for classifying PCOS, Logistic Regression estimates probabilities, and Neural Networks learn complex patterns. Input features include hormonal levels and demographic data. Models adjust weights during training to minimize prediction differences. Evaluation metrics, like accuracy and precision, gauge model performance. Cross-validation ensures robustness. Interpretability and data complexity guide algorithm selection. The models contribute to accurate PCOS diagnosis by learning patterns in input features.

1.3. Clustering Techniques:

Clustering techniques, such as k-Means and hierarchical clustering, employ unsupervised learning to discern subgroups among PCOS patients with shared characteristics. These methods enable the identification of clusters based on common features, aiding in the development of personalized treatment plans. By categorizing patients into homogeneous groups, clustering enhances our comprehension of PCOS's heterogeneity. This approach is valuable for tailoring interventions to specific subgroup needs, potentially improving the effectiveness of treatments. Unsupervised learning methods play a crucial role in unraveling the complexity of PCOS, allowing healthcare practitioners to target interventions more precisely.

2. Data Sources for PCOS Diagnosis:

2.1. Medical Imaging:

Medical imaging techniques, notably ultrasound and magnetic resonance imaging (MRI), have been seamlessly incorporated with machine learning (ML) algorithms for the identification of ovarian cysts and evaluation of ovarian morphology. These applications play a pivotal role in the diagnostic criteria for Polycystic Ovary Syndrome (PCOS). ML algorithms enhance the accuracy and efficiency of ovarian cyst detection, providing a valuable tool for clinicians. The integration of ultrasound and MRI with ML enables automated analysis, expediting the diagnosis process. By leveraging advanced image processing and pattern recognition, ML contributes to more precise assessments of ovarian structures. This intersection of medical imaging and ML holds promise for improving the early detection and understanding of PCOS through non-invasive and technology-driven approaches.

2.2. Hormone Profiling:

Hormonal imbalances, a defining feature of PCOS, have prompted the development of machine learning (ML) models for comprehensive hormone profiling. These models specifically analyze levels of key hormones like follicle-stimulating hormone (FSH), luteinizing hormone (LH), and anti-Müllerian hormone (AMH). By leveraging ML algorithms, these models enable a more nuanced understanding of hormonal patterns associated with PCOS. Automated analysis of hormone levels aids in the early and accurate diagnosis of PCOS, contributing to more targeted treatment strategies. ML algorithms enhance the efficiency of interpreting complex hormone profiles, providing clinicians with valuable insights. The integration of technology into hormone profiling not only streamlines diagnostic processes but also supports ongoing research into the intricate endocrine aspects of PCOS. This intersection of ML and hormone profiling holds promise for advancing personalized and data-driven approaches to PCOS diagnosis and management.

2.3. Electronic Health Records (EHRs):

Electronic Health Records (EHRs) serve as rich repositories of clinical data, containing vital information on patients, including those with Polycystic Ovary Syndrome (PCOS). Machine learning (ML) algorithms are adept at mining EHRs to discern patterns related to PCOS. By analyzing medical history, symptoms, and comorbidities, ML aids in the identification and diagnosis of PCOS patients. This technology-driven approach enhances the efficiency of evaluating diverse patient data within EHRs. ML algorithms contribute to a more nuanced understanding of PCOS, leveraging comprehensive information embedded in electronic records. The integration of EHRs and ML aligns with a data-driven paradigm, offering a valuable tool for healthcare practitioners in the diagnostic process. This intersection holds potential for improved patient outcomes and personalized healthcare strategies in the realm of PCOS.

.

3. Challenges and Future Directions:

3.1. Diagnostic Inaccuracy:

Diagnostic accuracy in Polycystic Ovary Syndrome (PCOS) is frequently compromised due to existing methods, resulting in misdiagnosis or delayed diagnosis. The intricate nature of PCOS symptoms, characterized by hormonal imbalances, irregular menstrual cycles, and ovarian cysts, adds complexity to the diagnostic process. The overlap of PCOS symptoms with those of other conditions further contributes to diagnostic uncertainty. Clinicians face challenges in distinguishing PCOS from related disorders, impacting the precision of diagnostic outcomes. The absence of definitive biomarkers and reliance on subjective criteria heightens the potential for misinterpretation. Such diagnostic challenges hinder timely intervention and personalized treatment strategies for individuals with PCOS. There is a critical need for more accurate and nuanced diagnostic tools to address the complexities inherent in PCOS symptomatology. Advances in machine learning and comprehensive data analysis hold promise for improving the precision of PCOS diagnosis in the future.

3.2 Heterogeneous Presentation:

Polycystic Ovary Syndrome (PCOS) presents a heterogeneous clinical landscape, encompassing a broad spectrum of manifestations. The diverse range of symptoms, including irregular menstrual cycles, hormonal imbalances, and ovarian cysts, complicates the development of a standardized diagnostic protocol. The variability in symptom severity and presentation among individuals with PCOS further challenges diagnostic standardization. Clinicians encounter difficulties in establishing a one-size-fits-all approach due to this clinical heterogeneity. The absence of universally agreed-upon diagnostic criteria adds to the complexity of devising a standardized protocol. PCOS's multifaceted nature necessitates a personalized diagnostic approach, considering individual variations in symptoms and severity. This clinical diversity underscores the importance of adopting flexible and nuanced diagnostic strategies tailored to the unique characteristics of each patient. Emerging research and technological advancements hold promise in addressing the challenges posed by PCOS's heterogeneous presentation, aiming for more accurate and personalized diagnostic methodologies.

3.3 High Cost and Time Burden:

Traditional diagnostic methods for Polycystic Ovary Syndrome (PCOS), including hormonal tests, ultrasounds, and clinical examinations, pose significant financial and time burdens for patients. These procedures are often expensive, requiring multiple tests and consultations, contributing to the overall cost of diagnosis. The financial strain and time commitment may deter some individuals from seeking timely assessment for PCOS symptoms. Long waiting times for appointments and test results can further delay the diagnostic process, hindering early intervention. The cumulative expenses associated with these diagnostic procedures may impact accessibility, especially for individuals with limited financial resources. The high cost and time burden associated with traditional diagnostic approaches highlight the need for more streamlined and cost-effective methods for PCOS assessment. Emerging technologies and innovative diagnostic strategies, such as machine learning applications and electronic health records, hold potential in alleviating these challenges, making diagnosis more accessible and efficient.

3.4 Limited Accessibility:

Limited accessibility to specialized healthcare facilities and diagnostic tests for Polycystic Ovary Syndrome (PCOS) is a significant challenge, exacerbated by geographical, economic, and infrastructural factors. Geographical disparities may result in unequal access, especially for individuals residing in remote or underserved areas. Economic constraints can impede individuals with financial limitations from seeking specialized care and diagnostic procedures. In regions with inadequate healthcare infrastructure, the availability of PCOS-specific diagnostic facilities may be limited, contributing to delays in diagnosis. Transportation challenges and long travel distances can further hinder patients from reaching specialized healthcare centers. The cumulative impact of these barriers results in delayed or restricted access to crucial diagnostic services for PCOS. Efforts to improve accessibility through telehealth initiatives, community outreach programs, and mobile diagnostic units may mitigate these challenges, ensuring a more equitable distribution of PCOS diagnostic resources.

3.5 Lack of Predictive Insights:

Monitoring the progression of Polycystic Ovary Syndrome (PCOS) and assessing treatment responses is challenging due to the absence of effective tools for predicting disease evolution. The inherent unpredictability of PCOS complicates the development of reliable prognostic indicators. Clinicians face difficulties in anticipating how the syndrome may evolve over time for individual patients, impeding the formulation of personalized treatment strategies. The lack of precise predictive insights limits the ability to tailor interventions to the specific needs of each patient. Without reliable prognostic tools, healthcare providers may struggle to preemptively address emerging complications or optimize therapeutic approaches. Research efforts to identify robust predictive markers and the integration of advanced technologies, such as machine learning, hold promise for enhancing predictive insights in the monitoring and management of PCOS. Improved understanding of the long-term trajectory of PCOS can inform more targeted and personalized healthcare interventions.

3.6. Transparency and Interpretability:

Current diagnostic methods for Polycystic Ovary Syndrome (PCOS) face challenges related to transparency and interpretability, presenting difficulties for both patients and medical professionals. The intricacies of traditional diagnostic procedures, such as hormonal tests and ultrasound examinations, may not be readily transparent to individuals undergoing these tests, leading to confusion and anxiety. The rationale behind diagnostic outcomes is often not clearly communicated, impeding patients' comprehension of their conditions. Medical professionals, too, may face challenges in interpreting complex diagnostic results, hindering effective communication with patients. The lack of transparency in diagnostic processes may contribute to miscommunication and patient dissatisfaction. Efforts to enhance transparency through clear communication, patient education, and accessible information platforms can empower individuals in understanding and participating in their diagnostic journey. Innovations in diagnostic technologies, such as explainable artificial intelligence, aim to improve the interpretability of diagnostic outcomes, fostering a more informed and collaborative approach to PCOS diagnosis.

3.7 Integration of Multi-Modal Data:

The complexity of PCOS diagnosis and monitoring demands the integration of diverse data modalities, including clinical, hormonal, and imaging data. These multi-modal datasets offer a comprehensive view of the syndrome's heterogeneous nature. Advanced analytical techniques, such as machine learning and data fusion, are crucial for extracting meaningful insights from these varied sources. Integrating clinical data, which includes symptoms and medical history, with hormonal profiles enhances the diagnostic precision of PCOS. Hormonal data, encompassing levels of FSH, LH, and AMH, provides valuable endocrine insights. The incorporation of imaging data, such as ultrasound and MRI results, aids in visualizing ovarian morphology. The synergy of these data modalities allows for a more nuanced understanding of PCOS, facilitating personalized and targeted treatment strategies. The challenges lie in developing sophisticated analytical frameworks capable of handling the complexity and heterogeneity inherent in multi-modal PCOS data. Advances in these techniques hold promise for refining diagnostic accuracy and treatment efficacy.

V.CONCLUSION

Machine Learning (ML) is poised to revolutionize the landscape of Polycystic Ovary Syndrome (PCOS) diagnosis, offering the potential to elevate accuracy, efficiency, and personalization in healthcare. By adeptly harnessing diverse data sources, including clinical records, hormonal profiles, and imaging results, ML algorithms demonstrate the capability to facilitate early detection and tailored management of PCOS. The intricate and multifaceted nature of PCOS symptoms can be better navigated through the sophisticated analytical prowess of ML, allowing healthcare professionals to decipher intricate patterns and correlations within patient data.

However, the integration of ML into clinical practice necessitates a cautious approach, acknowledging and mitigating ethical, privacy, and standardization challenges. The responsible use of patient data is paramount to uphold confidentiality and prevent unauthorized access. Furthermore, ensuring standardization across ML applications is essential for consistent and reliable results. The ethical implications of algorithmic decision-making, potential biases, and the protection of patient privacy should be meticulously addressed through robust regulatory frameworks and ethical guidelines.

Despite these challenges, the transformative potential of ML in PCOS diagnosis remains significant. The amalgamation of advanced technology and responsible practices promises to usher in a new era in women's health. As research in this field continues to evolve, the anticipation is for ML to not only enhance diagnostic accuracy and personalized management but also contribute to a broader understanding of PCOS pathophysiology. The careful integration of ML into clinical workflows has the potential to redefine healthcare practices, ultimately leading to improved outcomes and a brighter future for individuals managing PCOS.

.

VI. REFERENCES:

[1] https://www.mdpi.com/2571-5577/6/2/32
[2] https://ieeexplore.ieee.org/document/8929674
[3] https://www.ijres.org/papers/Volume-11/Issue-1/1101165170.pdf
[4] https://ieeexplore.ieee.org/abstract/document/9230932
[5] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9885199
[6]

https://git.org.in/GIT_JET/Papers/Regular%20Edition/8_GIT_JET_Regular_Edition_2022.pdf
[7] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10137609/
[8] https://www.frontiersin.org/articles/10.3389/fpubh.2021.789569/full
[9] https://www.medrxiv.org/content/10.1101/2023.07.27.23293255v1
[10] https://www.researchgate.net/publication/348627784_PCOcare_PCOS_Detection_and_Prediction_using_Machine_Learning_Algorithms
[11] https://www.researchgate.net/publication/345804076_Diagnosis_of_Polycystic_Ovary_Syndrome_Using_Machine_Learning_Algorithms
[12] https://www.hindawi.com/journals/acisc/2023/9686697/
[13] https://www.semanticscholar.org/paper/i-HOPE%3A-Detection-And-Prediction-System-For-Ovary-Denny-Raj/1d9ee28205fff2565fb37e4f4256263294786f9e
[14] https://www.semanticscholar.org/paper/Hot-Deck-Methodsfor-Imputing-Missing-Data-The-of-Joenssen-Bankhofer/853253faf9d7ee66a4ebd749659c463cdc475f7c
[15] https://ieeexplore.ieee.org/document/9604905

Authors:

1. Rubeena Rustum
Assistant Professor
Computer Science and Engineering
Gokaraju Rangaraju institute of Engineering and Technology
rubeena.rustum@gmail.com
2. Thada Amulya Reddy
Student
Computer Science and Engineering
Gokaraju Rangaraju institute of Engineering and Technology
thadaamulyareddy@gmail.com
3. Macha Harshitha
Student
Computer Science and Engineering
Gokaraju Rangaraju institute of Engineering and Technology
Mharshitha122@gmail.com

4. Bukya Shilpi
Student
Computer Science and Engineering
Gokaraju Rangaraju institute of Engineering and Technology
bukyashilpi@gmail.com
5. Karishma Shaik
Student
Computer Science and Engineering
Gokaraju Rangaraju institute of Engineering and Technology
Shaikkarishma1764@gmail.com