JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction and Detection of PCOS using Machine learning Approach

[1] Mr.Meghraj Walkikar, [2]Ms. Amruta Nevare, [3]Ms. Shweta Bansode,[4] Ms. Aasawari Mohite,
Prof. Dr. Geetha Chillarge

[1]Department Of Computer Engineering,
[1]Marathwada Mitra Mandal's College of Engineering, Pune, India

***Abstract*—Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age, characterized by hormonal imbalances, irregular menstruation, and ovarian cysts. Timely identification of PCOS is crucial for effective management and mitigating long-term complications. This study proposes an innovative machine learning approach for PCOS detection, utilizing a dataset containing clinical and biochemical features. The dataset encompasses demographic details, medical history, and hormonal profiles of both PCOS-diagnosed patients and healthy individuals. Diverse machine learning algorithms, including logistic regression, random forest, and support vector machines, are deployed to formulate predictive models for PCOS detection. Feature selection techniques are applied to discern the most relevant features for model training. The experimental findings underscore the efficacy of the proposed approach in accurately identifying PCOS. The developed models exhibit notable accuracy, sensitivity, and specificity, suggesting their potential utility as diagnostic tools for PCOS. This research significantly contributes to advancing PCOS diagnosis by harnessing machine learning techniques to enhance early detection and management of this prevalent condition.**

Index Terms - **Polycystic Ovary Syndrome, Menstrual irregularity, Polycystic ovaries, Clinical features**

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder that affects individuals with ovaries, particularly women of reproductive age. It is estimated to affect 6% to 12% of women worldwide, making it one of the most common hormonal disorders among this population. PCOS is characterized by a combination of symptoms, including irregular menstrual cycles, excess androgen levels (male hormones), and polycystic ovaries (ovaries with multiple small cysts). The exact cause of PCOS is not fully understood, but it is believed to involve a combination of genetic and environmental factors. PCOS can have a significant impact on a woman's health and quality of life. In addition to irregular periods and fertility issues, PCOS is associated with an increased risk of other health problems, including type 2 diabetes, high blood pressure, and heart disease. It can also cause emotional and psychological symptoms, such as anxiety and depression, due to its effects on hormone levels and appearance. Diagnosing PCOS can be challenging, as there is no single test that can definitively diagnose the condition. Instead, healthcare providers rely on a combination of symptoms, physical exams, and laboratory tests to make a diagnosis. Treatment for PCOS focuses on managing symptoms and addressing the underlying hormonal imbalances. This may include lifestyle changes (such as diet and exercise), medications to regulate menstrual cycles and reduce androgen levels, and fertility treatments for those trying to conceive. In recent years, there has been growing interest in using machine learning techniques for the detection and diagnosis of various medical conditions, including PCOS. Machine learning offers the potential to analyze large and complex datasets to identify patterns and relationships that may not be apparent to human observers. By leveraging machine learning, it may be possible to develop more accurate and efficient methods for PCOS detection, which could improve patient outcomes and reduce healthcare costs.

This study aims to explore the use of machine learning approaches for the detection of PCOS using a dataset of clinical and biochemical features. Various machine learning algorithms will be applied to develop predictive models for PCOS detection, with the goal of improving the accuracy and efficiency of PCOS diagnosis. The findings of this study could have significant implications for the early detection and management of PCOS, ultimately improving the quality of life for women affected by this condition.

Following are the major contributions of this research:

The dataset incorporates 44 features, and through rigorous analysis, we pinpointed 19 features deemed most crucial for PCOS identification. To enhance the informativeness of the data for model training, a series of pre-processing techniques were applied exclusively to the PCOS data, excluding the infertility dataset. Additionally, we present a comprehensive comparison of various machine learning algorithms specifically tailored for the task of detecting PCOS.

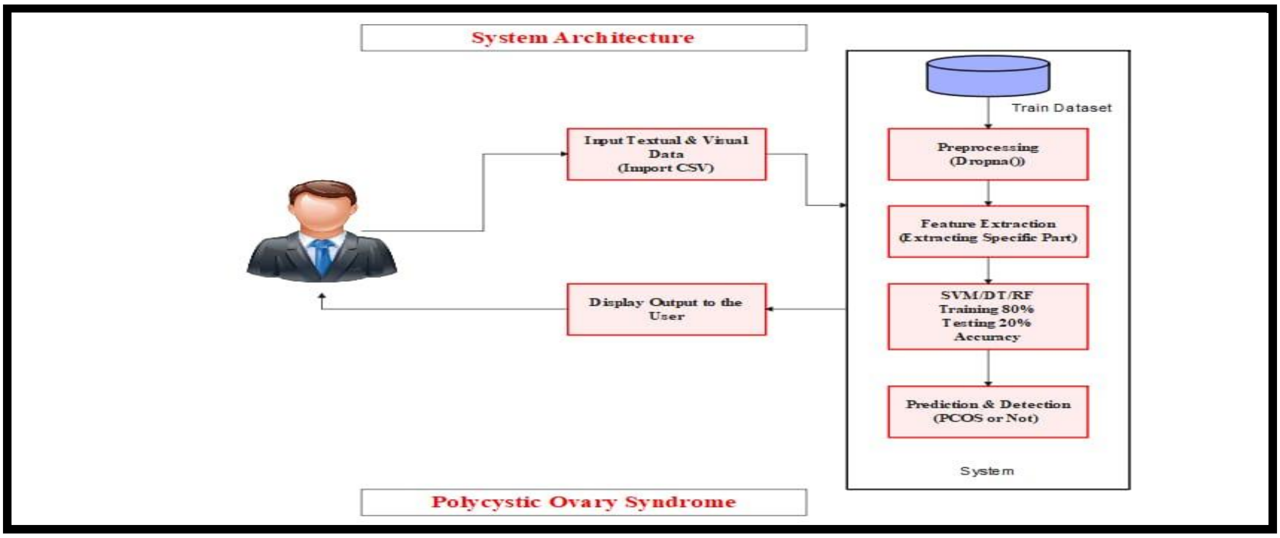
System Architecture
Input Textual & Visual Data (Import CSV)
Display Output to the User
Train Dataset
Preprocessing (Dropna())
Feature Extraction (Extracting Specific Part)
SVM/DT/RF Training 80% Testing 20% Accuracy
Prediction & Detection (PCOS or Not)
System
Polycystic Ovary Syndrome

Figure 1: System Architecture

A. Problem Statement

Design an efficient model for the diagnosis and analysis of polycystic Ovary Syndrome(PCOS). Compare and Evaluate the performance of the SVM model using a different machine learning approach on PCOS data to identify the most accurate and reliable classification technique for PCOS diagnosis.

II. MOTIVATION AND BACKGROUND

PCOS is frequently underdiagnosed or diagnosed late, resulting in delayed treatment and potential health complications. This intricate and often misunderstood medical condition significantly impacts individuals, particularly women of reproductive age. The drive behind our PCOS project stems from a deep concern for the well-being of those affected, a commitment to bridging gaps in knowledge and awareness, and a dedication to enhancing diagnosis, treatment, and the overall quality of life for individuals living with PCOS.

Polycystic ovary syndrome is widespread, affecting women globally, including African and white women (5%-8%), individuals in Spain (6.8%-13%), and showing a higher prevalence in Asian countries. Common symptoms experienced by women with PCOS include hair fall, unusual blood pressure, excessive weight, irregular menstrual cycles, and potential complications during pregnancy. Detection methods, such as 3D ultrasounds and pulsed-Doppler ultrasounds, have shown efficiency. While previous studies utilized Logistic Regression and Bayesian classifiers for automated screening, our results demonstrate increased satisfaction. Another study focused on PCOS identification using ultrasonographic, clinical, and endocrine factors, such as LH/FSH, ovarian volume, BMI range, and primary infertility, but the limited dataset of 60 patients restricts the reliability of valuable decisions. In our research, a morphological image processing filter with a watershed algorithm was applied to ultrasound images for PCOS detection.

The primary aim of our study is to identify PCOS presence in patients using diverse machine learning methods, including SVM, KNN, and Random Forest Algorithm, striving for improved accuracy compared to the previously used Logistic Regression and Naive-Bayes classification. This research seeks to contribute to the advancement of PCOS diagnosis, providing more accurate and reliable detection methods to positively impact the lives of individuals with PCOS.

III. PROPOSED SYSTEM

- **Dataset Collection and Preparation**: Focus on collecting a comprehensive dataset that includes relevant features such as hormonal levels, clinical symptoms, and demographic information. Ensure the dataset is well-curated and balanced to avoid bias in model training.
- **Feature Selection and Engineering**:Examine a range of feature selection techniques to uncover the most crucial attributes for predicting PCOS. Additionally, explore the realm of feature engineering to devise novel features or representations adept at capturing key patterns within the dataset.
- **Model Selection and Evaluation**: Assess an array of machine learning models, encompassing both conventional classifiers and deep learning models, to identify the optimal approach for predicting PCOS. Employ cross-validation and diverse evaluation metrics to gauge the performance of the models.
- **Data Imbalance Handling**: Address class imbalance in the dataset using techniques such as oversampling, under sampling, or synthetic data generation. This is important to ensure that the model can effectively learn from both positive and negative instances of PCOS.
- **Interpretability and Explainability**: Prioritize the interpretability of the machine learning models to understand the underlying factors contributing to PCOS prediction. This can help in building trust with clinicians and patients.
- **Integration of Multi-Modal Data**: Explore the possibility of incorporating diverse data modalities, including genetic, imaging, and clinical data, to enhance the precision of PCOS prediction models. Implement techniques for multi-modal data fusion to effectively amalgamate information derived from various sources.
- **Real-Time Monitoring and Personalization**: Explore the possibility of real-time monitoring for PCOS prediction using wearable devices and continuous data streams. Develop personalized prediction models that consider individual patient characteristics.

- **Clinical Validation and Deployment**: Validate the performance of the machine learning models in clinical settings to ensure their effectiveness in real-world applications. Collaborate with healthcare providers to integrate the models into clinical practice.
- **Ethical and Privacy Considerations**: Address ethical and privacy concerns related to the use of machine learning for PCOS prediction. Ensure that data security, informed consent, and bias mitigation strategies are in place.
- **Collaboration and Knowledge Sharing**: Collaborate with other researchers and healthcare professionals working on PCOS to share knowledge and insights. This can help in advancing the field and improving patient outcomes.

TABLE 1
PERFORMANCE COMPARISON AMONG 4 METHOD

Method	**Accuracy %**	**Precision %**	**Recall %**	**F1 Score %**
KNN	75	65	66	64
SVM	90	82	82	81
NAIVE	93	80	80	80
RANDOM FOREST	93.5	84	84	84

KNN does not give much satisfactory result where the performance of SVM is good enough to detect PCOS correctly.

Classification algorithms

SVM

The Support Vector Machine (SVM) stands out as a potent machine learning algorithm, particularly utilized for binary classification tasks, with the primary goal of minimizing generalization errors by explicitly constructing decision boundaries. SVM's effectiveness in handling complex datasets surpasses that of many other machine learning models. It excels in scenarios where the data can be linearly separated, adjusting the decision boundary through a hyperplane strategically positioned between the nearest points of each categorical class.

In the context of PCOS diagnosis, SVM is employed for binary classification, assigning a numerical target factor of 1 to denote patients suffering from PCOS and 0 for those not affected. The remaining columns, encompassing variables like follicle-stimulating hormone, luteinizing hormone, thyroid hormone, anti-mullerian hormone level, prolactin level, etc., act as predictors. These predictors play a crucial role in training the SVM model, facilitating the establishment of an optimized decision boundary. This enables SVM to effectively discern between patients with and without PCOS based on the provided features. The application of SVM in this medical context underscores its proficiency in handling intricate datasets and discerning nuanced patterns within them, contributing to its widespread utility in various classification tasks.

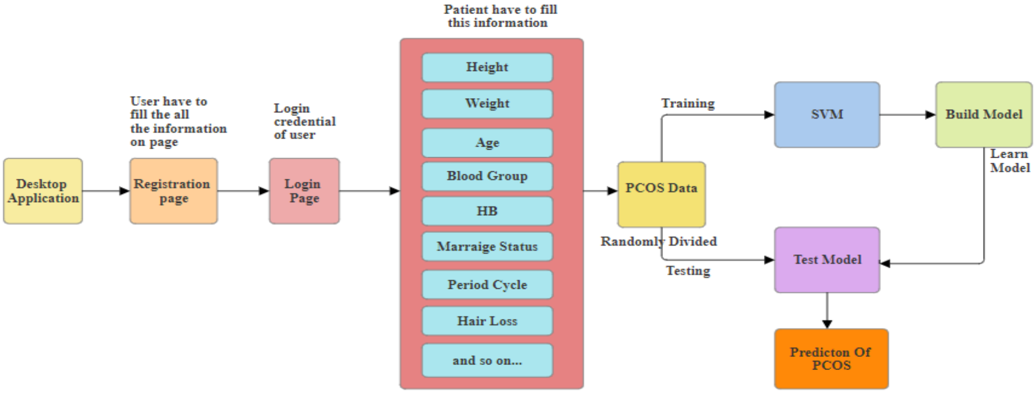

Figure 2: SVM Working

IV. SCOPE

In the forthcoming stages of our research, we envision the incorporation of Convolutional Neural Networks (CNN) in an optimized fashion to elevate the efficacy of our model. Furthermore, our aspirations include an in-depth exploration of hyperparameter tuning for machine learning algorithms and the implementation of advanced feature selection techniques, all aimed at achieving superior performance. Looking ahead, the future scope of PCOS research could encompass an expansive investigation into the genetic, environmental, and hormonal factors that contribute to the condition. This trajectory may also involve delving into the development of highly personalized treatment approaches, integrating the capabilities of AI or machine learning to not only predict but effectively manage symptoms. Moreover, our project's future initiatives might concentrate on amplifying awareness and educational outreach regarding PCOS, particularly in underserved communities. Collaborative efforts with healthcare professionals are also envisaged to enhance the diagnostic processes and establish more effective long-term management strategies. These prospective endeavors underscore our commitment to advancing PCOS research and making a meaningful impact on both understanding and addressing this complex medical condition.

V. CHALLENGES AND OPPORTUNITIES

In future ,we wish to include CNN, in an optimized form, to improve our model . Furthermore , we would like to perform more extensive hyperparameter tuning of machine learning algorithm and more improved feature selection for better performance. The future scope of a PCOS could involve expanding research to better understand the genetic, environmental, and hormonal factors contributing to the condition. It might also explore the development of more personalized treatment approaches, incorporating AI or machine learning to predict and manage symptoms. Additionally, the project could focus on increasing awareness and education about PCOS in underserved communities, and collaborating with healthcare professionals to improve diagnosis and long-term management strategies.

VI. CONCLUSION

This paper presents the application of SVM for PCOS detection and represents a significant stride towards automating and enhancing the diagnostic process. The SVM-based system showcased promising results, demonstrating a high degree of accuracy, precision, and recall in identifying PCOS. This technological advancement holds immense potential in revolutionizing PCOS diagnosis, offering a rapid, objective, and reliable method for healthcare professionals. We selected statistically significant and discriminating features that best describe the PCOS condition, followed by providing the selected features into an Extreme Gradient Boosting model. Our detailed experiments, conducted on a benchmark dataset shows the massive potential of this integrated solution. In future, we wish to include CNN, in an optimized form, to improve our model. Furthermore, we would like to perform more extensive hyperparameter tuning of machine learning algorithms and more improved feature selection for better performance.

VII. REFERENCES

[1] Bharati, S., Podder, P., & Mondal, M. R. H. (2020, June). Diagnosis of polycystic ovary syndrome using machine learning algorithms. In 2020 IEEE region 10 symposium (TENSYMP) (pp. 1486-1489). IEEE.

[2] Deshpande, S. S., & Wakankar, A. (2014, May). Automated detection of polycystic ovarian syndrome using follicle recognition. In 2014 IEEE international conference on advanced communications, control and computing technologies (pp. 1341-1346). IEEE.

[3] Soni, P., & Vashisht, S. (2018, October). Exploration on polycystic ovarian syndrome and data mining techniques. In 2018 3rd International Conference on Communication and Electronics Systems (ICCES) (pp. 816-820). IEEE.

[4] Prapty, A. S., & Shitu, T. T. (2020, December). An efficient decision tree establishment and performance analysis with different machine learning approaches on polycystic ovary syndrome. In 2020 23rd International conference on computer and information technology (ICCIT) (pp. 1-5). IEEE.

[5] Inan, M. S. K., Ulfath, R. E., Alam, F. I., Bappee, F. K., & Hasan, R. (2021, January). Improved sampling and feature selection to support extreme gradient boosting for PCOS diagnosis. In 2021 IEEE 11th annual computing and communication workshop and conference (CCWC) (pp. 1046-1050). IEEE.

[6] Dewi, R. M., Adiwijaya, Wisesty, U. N., & Jondri. (2018, March). Classification of polycystic ovary based on ultrasound images using competitive neural network. In Journal of Physics: Conference Series (Vol. 971, p. 012005). IOP Publishing.

[7] Thufailah, I. F., & Wisesty, U. N. (2018, March). An implementation of Elman neural network for polycystic ovary classification based on ultrasound images. In Journal of Physics: Conference Series (Vol. 971, No. 1, p. 012016). IOP Publishing.

[8] Deng, Y., Wang, Y., & Chen, P. (2008, August). Automated detection of polycystic ovary syndrome from ultrasound images. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4772-4775). IEEE.

[9] Potočnik, B., & Zazula, D. (2002). Automated analysis of a sequence of ovarian ultrasound images. Part I: segmentation of single 2D images. Image and vision computing, 20(3), 217-225.

[10] Rethinavalli, S., & Manimekalai, M. (2016). A Hypothesis Analysis on the Proposed Methodology for Prediction of Polycystic Ovarian Syndrome. International Journal of Science, Engineering and Computer Technology, 6(11), 396.