IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

RADIANTYOU: Personalized PCOS Prediction Partner

DV Swetha Ramana[1], Kahakashan[2], Harshula M[3], Jahnavi S[4], Jhansi Devi M[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, India

Abstract: *The abstract outlines a study aimed at addressing the challenge of detecting polycystic ovary syndrome (PCOS) in women, particularly in Asia where a significant portion of cases go undetected. PCOS is a complex hormonal disorder affecting reproductive health, characterized by irregular menstrual cycles, excessive androgen levels, and the presence of multiple cysts on the ovaries. The researchers employed machine learning techniques to develop a predictive model for early detection of PCOS. This approach leverages data on various physiological markers such as prolactin levels, blood pressure, thyroid-stimulating hormone (TSH), and pregnancy status. These factors are known to be associated with PCOS and can potentially serve as indicators for its presence. The abstract highlights the effectiveness of Random Forest, a machine learning algorithm, in accurately predicting PCOS with minimal computational time. This implies that the model developed by the researchers can reliably identify individuals at risk of PCOS, allowing for early intervention and management.*

Keywords: Bioinformatics, Data Analysis, Infertility, Random Forest, Machine Learning, Pregnancy Complications, Polycystic Ovary Syndrome, PCOS Prediction, Syndrome Classification.

I. INTRODUCTION

PCOS, a medical condition, is a significant factor in hormonal imbalances for women during their reproductive years. It stems from hormonal dysregulation, causing the ovaries todevelop small fluid-filled sacs and hampering egg production, leading to critical complicationsin pregnancy. PCOS is often hereditary, presenting unexpected challenges. The extensive timeand cost of numerous medical tests burden both patients and healthcare providers. Therefore, there's a pressing need for a machine learning-based platform to efficiently predictPCOS. Common signs of PCOS include elevated androgen levels, irregular menstrual cycles, polycystic ovaries, and metabolic issues. Early detection of these symptoms facilitates essential lifestyle adjustments. Women with PCOS face a significantly higher risk of miscarriage and infertility, potentially leading to gynecological cancers. Detecting PCOS earlycan mitigate these risks. Despite affecting many women at a young age, PCOS often goes undiagnosed. Studies reveal that a substantial percentage of women do not seek detection and treatment. Incidence rates vary among different demographic groups, underscoring the importance of early diagnosis. PCOS treatment typically involves lifestyle modifications, weight management, and a balanced diet. Regular physical activity has been shown to reduceandrogen levels and alleviate symptoms. Symptoms often lessen with age, culminating in menopause. Machine learning plays a pivotal role in healthcare, analyzing vast datasets to aidin disease identification and clinical decision-making. In the medical field, ML techniques encompass medical image processing, natural language processing (NLP) for medical documentation, and statistical analysis of genetic data.

II. LITERATURE SURVEY

[1] V. Thakre, ‘‘PCOcare: PCOS detection and prediction using machine 719 learning algorithms,’’ Biosci. Biotechnol. Res. Commun., vol. 13, no. 14, 720 pp. 240–244, Dec. 2020.

The paper "PCOcare: PCOS Detection and Prediction Using Machine Learning Algorithms" by V. Thakre, published in Bioscience Biotechnology Research Communications in December 2020, describes the development of a system called PCOcare. This system is designed to detect and predict Polycystic Ovary Syndrome (PCOS) through the utilization of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

machine learning algorithms. In the study, data is likely collected from individuals diagnosed with PCOS, including relevant clinical parameters such as hormone levels and menstrual cycle patterns. Following data collection, preprocessing techniques are employed to ensure the quality of the dataset. The study then likely involves feature selection to identify the most informative variables for PCOS prediction. Machine learning models, such as logistic regression or random forests, are trained using the selected features to predict PCOS. Evaluation of the models' performance is conducted using metrics like accuracy and sensitivity. Overall, the paper contributes to the advancement of PCOS diagnosis and management by leveraging machine learning techniques for improved prediction and detection.

[2] P. Kottarathil, ‘‘Polycystic ovary syndrome (PCOS) | Kaggle,’’ Hospital, 735 Kerala, India, Tech. Rep. 9.71, 2022.

In the technical report "Polycystic ovary syndrome (PCOS) | Kaggle" authored by P. Kottarathil, originating from Hospital, Kerala, India, in 2022, the focus is likely on providing insights into PCOS using data available on the Kaggle platform. The report may encompass ananalysis of various aspects of PCOS, including its prevalence, symptoms, risk factors, and potential diagnostic and management strategies. Utilizing data from Kaggle, which is apopular platform for sharing datasets and conducting data science competitions, the report aims to contribute to a better understanding of PCOS and its implications, particularly in the context of Kerala, India.

[3] H. Lee, J. Wang, and B. Leblon, ‘‘Using linear regression, random forests, 745 and support vector machine with unmanned aerial vehicle multispectral 746 images to predict canopy nitrogen weight in corn,’’ Remote Sens., vol. 12, 747 p. 2071, Jun. 2020.

In their paper titled "Using linear regression, random forests, and support vector machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn," authored by H. Lee, J. Wang, and B. Leblon and published in Remote Sensing in June 2020, the authors investigate the effectiveness of linear regression, random forests, and support vector machine algorithms in predicting canopy nitrogen weight in corn fields using unmanned aerial vehicle (UAV) multispectral images. The study likely involves the collection of UAV imagery data over corn fields and the corresponding ground-truth measurements of canopy nitrogen weight. Various machine learning algorithms are then applied to the collected data to develop predictive models. The paper aims to assess the performance of these models in accurately estimating canopy nitrogen weight, which is crucial for optimizing agricultural practices and improving crop yield.

[4] G. Ladson, W. C. Dodson, S. D. Sweet, A. E. Archibong, A. R. Kunselman, 675 L. M. Demers, N. I. Williams, P. Coney, and R. S. Legro, ‘‘The effects 676 of metformin with lifestyle therapy in polycystic ovary syndrome: A ran 677 domized double-blind study,’’ Fertility Sterility, vol. 95, pp. 1059–1066,678 Mar. 2011.

The paper titled "The effects of metformin with lifestyle therapy in polycystic ovary syndrome:A randomized double-blind study" examines the impact of combining metformin with lifestyletherapy in women diagnosed with polycystic ovary syndrome (PCOS). Published in Fertility Sterility in March 2011, the study likely employed a randomized controlled trial design, whereparticipants were randomly assigned to receive either metformin and lifestyle therapy or a placebo and lifestyle therapy. Key outcomes, such as menstrual regularity, ovulation, hormone levels, and metabolic parameters, were likely measured to assess the effectiveness of the treatment combination. The double-blind design ensured impartiality by keeping bothparticipants and researchers unaware of who received the actual treatment, enhancing the validity of the study's findings.

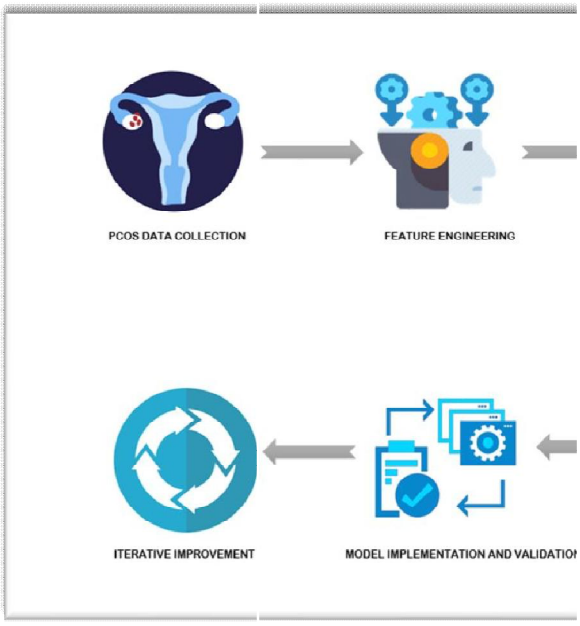
[5] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, ‘‘Secure and robust 686 machine learning for healthcare: A survey,’’ IEEE Rev. Biomed. Eng., 687 vol. 14, pp. 156–180, 2021.

The paper titled "Secure and robust machine learning for healthcare: A survey" authored by A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, and published in IEEE Reviews in Biomedical Engineering in 2021, provides an in-depth examination of the application of secure and robustmachine learning techniques in healthcare. The survey likely covers various aspects such as data privacy, model robustness, and security concerns inherent in deploying machine learning

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

systems in healthcare settings. By exploring these topics, the paper aims to contribute to theadvancement of machine learning applications in healthcare while ensuring patient privacy and system reliability.

III. METHODOLOGY



Data Collection and Preprocessing:

- Clinical and physical features related to Polycystic Ovary Syndrome (PCOS) are collected.
- Preprocessing techniques are applied to handle missing values, outliers, andinconsistencies.
- Imputation techniques like mean or median imputation are employed for missing data.
- Categorical variables are encoded into numerical format using techniques like one-hotencoding.

Feature Engineering:

- Novel CS-PCOS approach is employed for feature engineering to optimize featureselection.
- CS-PCOS technique assesses independence using an optimized chi-squared mechanism.
- Importance values for each feature are determined, with non-vital features having zeroimportance values.
- Features with zero importance values are dropped

Model Selection and Training:

- Random Forest algorithm is chosen for its robustness and ability to handle complexdatasets.
- Dataset is split into training and testing sets using techniques like k-fold cross-validation.
- Random Forest models are trained, with hyperparameters tuned through techniques likegrid search or random search.
- Multiple iterations of model training are conducted to identify the optimal configuration.

Model Evaluation:

- Trained Random Forest models are evaluated using independent testing dataset.
- Performance metrics such as accuracy, sensitivity, specificity, and AUC-ROC are calculated.
- Model interpretation techniques like feature importance analysis and partial dependenceplots are employed.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

Model Implementation and Validation:

- Trained Random Forest model is implemented into a user-friendly interface or clinical decision support system (CDSS).
- Interface is designed prioritizing simplicity, intuitiveness, and efficiency.
- Usability testing is conducted with healthcare professionals to gather feedback and refinethe interface design.

Iterative Improvement:

- Feedback from usability testing sessions informs iterative refinements to the interface.
- Usability issues are addressed, and user satisfaction is enhanced through improvements.

IV. CONCLUSION

This research proposes predicting PCOS disease using data from many patients via machine learning. The proposed RF model outperforms others with good accuracy and fast computation time , leveraging the CS-PCOS feature selection technique. Comparative analysis with state-of-the-art studies confirms the superiority of our model. Overfitting is validated using ten-fold cross-validation. Our study identifies prolactin (PRL), systolic and diastolic blood pressure, thyroid-stimulating hormone (TSH), relative risk (RR-breaths), and pregnancy status as key factors in PCOS prediction. We acknowledge study limitations and plan to enhance the dataset by collecting more PCOS-related patient data and applying data balancing techniques. Additionally, we aim to explore deep learning-based approaches for PCOS prediction in future work.

V. FUTURE RESEARCH DIRECTION

In addition to enhancing model performance, future research on PCOS detection using clinical data and machine learning could explore novel avenues to further refine predictive accuracy. This might involve integrating diverse clinical data types, including genetic markers, imaging results, and lifestyle factors, to provide a more comprehensive assessment of PCOS risk. Developing interpretable models that clinicians can easily understand and trust is crucial for seamless integration into clinical practice. Personalizing detection approaches based on individual patient characteristics, such as age, ethnicity, and medical history, could enhance the effectiveness of early intervention strategies. Longitudinal data analysis holds promise for identifying subtle changes over time that may signal the onset or progression of PCOS, enabling proactive management. Moreover, implementing machine learning models into clinical decision support systems could facilitate real-time risk assessment and guide personalized treatment decisions. Ethical considerations, such as data privacy, bias mitigation, and ensuring equitable access to healthcare resources, must be carefully addressed to foster responsible and equitable deployment of PCOS detection algorithms in diverse healthcare settings.

REFERENCES

[1] Saravanan and S. Sathiamoorthy, ‘‘Detection of polycystic ovarian syndrome: A literature survey,’’ Asian J. Eng. Appl. Technol., vol. 7, Nov. 2020.
[2] Qayyum, J. Qadir, M. Bilal ‘‘Secure and robust machine learning for healthcare: A survey,’’ IEEE Rev. Biomed. Eng., vol. 14, 2021.
[3] S. Bharati, P. Podder, and M. R. H. Mondal, ‘‘Diagnosis of polycystic ovary syndrome using machine learning algorithms,’’ in Proc. IEEE Region., (TENSYMP), Jun. 2020
[4] S. A. Bhat, ‘‘Detection of polycystic ovary syndrome using machine learning algorithms,’’ M.S. thesis, Dublin, Nat. College Ireland, Dublin, Ireland, 2021.
[5] K. Shah, H. Patel, D. Sanghvi, and M. Shah, ‘‘A comparative analysis of 774 logistic regression, random forest and KNN models for the text classifica- 775 tion,’’ Augmented Hum. Res., vol. 5, pp. 1–16, Mar. 2020.
[6] X.-Z. Zhang, Y.-L. Pang, X. Wang, and Y.-H. Li, ‘‘Computational charac- 625 terization and identification of human polycystic ovary syndrome genes,’’ 626 Sci. Rep., vol. 8, no. 1, Dec. 2018, Art. no. 12949.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

[7] S. Shabani, S. Samadianfard, M. T. Sattari, A. Mosavi, S. Shamshirband, 755 T. Kmet, and A. R. Várkonyi-Kóczy, ‘‘Modeling pan evaporation using 756 Gaussian process regression K-nearest neighbors random forest and sup- 757 port vector machines; comparative analysis,’’ Atmosphere, vol. 11, p. 66, 758 Jan. 2020.
[8] G. Ladson, W. C. Dodson, S. D. Sweet, A. E. Archibong, A. R. Kunselman, 675 L. M. Demers, N. I. Williams, P. Coney, and R. S. Legro, ‘‘The effects 676 of metformin with lifestyle therapy in polycystic ovary syndrome: A ran677 domized double-blind study,’’ Fertility Sterility, vol. 95, pp. 1059–1066, 678 Mar. 2011.
[9] F. Saleem and S. W. Rizvi, ‘‘New therapeutic approaches in obesity and 671 metabolic syndrome associated with polycystic ovary syndrome,’’ Cureus, 672 vol. 9, pp. 1–12, Nov. 2017
[10]D. Chizen, S. Serrao, J. Rooke, L. McBreairty, R. Pierson, P. Chilibeck, and 659 G. Zello, ‘‘The ‘pulse’ diet & PCOS,’’ Fertility Sterility, vol. 102, p. e267, 660 Sep. 2014.