**MENSTRUAL HEALTH AND PCOD RISK ANALYSIS**

KABISA P

Department of Computer Technology

 Kongu Engineering College

 Perundurai.

 Email id: kabisap.22bir@kongu.edu

***Abstract - Polycystic Ovary Disease (PCOD) is a common condition affecting many women worldwide, often leading to issues like infertility, weight gain, hormonal imbalances, and a higher risk of heart disease. Early detection is important to manage symptoms and prevent complications. This study collected data from 163 individuals, focusing on 13 key factors related to menstrual health and PCOD risk, such as irregular periods, acne, family history, and insulin resistance. Machine learning techniques, including Artificial Neural Networks (ANN), Random Forest and Naive Bayes were used to predict PCOD risk. The findings highlight major risk factors like menstrual irregularities, hormonal imbalances, and thyroid dysfunction. By developing a reliable system for identifying PCOD risk, this research helps raise awareness and supports healthcare professionals in providing early diagnosis, personalized treatment plans, and better lifestyle recommendations to manage the condition effectively. The models achieved the accuracy of ANN-79%, Random Forest-100% and Naive Bayes-76%.***

***Keywords: PCOD, machine learning, risk assessment, menstrual health, early detection, predictive modeling, classification accuracy, artificial neural networks***

**1.INTRODUCTION**

Polycystic Ovary Disease (PCOD) is a common but often overlooked condition that affects millions of women worldwide. It can cause irregular periods, weight changes, hormonal imbalances, and fertility problems. Many women don’t realize they have PCOD until it leads to serious health issues like infertility, diabetes, heart disease, and mental health struggles. Despite how common it is, PCOD is often misdiagnosed or not diagnosed at all, making it harder for women to get the right treatment.

One of the biggest challenges in managing PCOD is the lack of proper screening methods, clear diagnostic tools, and personalized treatments. Many women and even healthcare providers are not fully aware of the condition, leading to delayed diagnosis and ineffective management. Factors like diet, stress, and lack of physical activity also play a role but are often ignored in treatment plans. Additionally, there is not enough detailed data on menstrual health and PCOD-related factors, making research and treatment development more difficult.

It includes key menstrual health and PCOD-related risk factors. Using machine learning, we aim to improve awareness, support early diagnosis, and help women manage their reproductive health more effectively. This approach can lead to personalized treatment plans based on an individual’s specific risk factors, helping to prevent complications and improve overall health.

This research has the potential to improve the accuracy of PCOD diagnosis, provide better patient care, guide healthcare policies, and advance research in women’s reproductive health. By using artificial intelligence, doctors can identify high-risk patients earlier, leading to faster and more effective treatments. Raising awareness about PCOD can also encourage women to seek medical advice sooner and make healthier lifestyle choices. By tackling these challenges with data-driven solutions, we hope to make a meaningful impact on women's healthcare worldwide.

**2.LITERATURE REVIEW**

Akshay S proposed a multi-modal AI-based framework for the detection and risk assessment of Polycystic Ovary Syndrome (PCOD). The framework integrates both image and numerical data, utilizing ultrasound images, blood test results, and various patient parameters, such as height, weight, BMI, FSH, TSH, LH, AMH, PRL, Vitamin D3, progesterone, and RBS. The study involves algorithms such as SVM, CNN, KNN, Otsu, and Naive Bayes. A survey conducted with 81 participants revealed that 45% were unaware of their PCOD status despite exhibiting symptoms, and 73% showed symptoms of PCOD. Of the respondents, 19.8% confirmed a PCOD diagnosis. This framework aims to improve diagnostic precision and alleviate the emotional impact of PCOD through early intervention and support.

Ankur Kumar developed a non-invasive diagnostic system for Polycystic Ovary Syndrome (PCOS) using clinical data from Kaggle. The study employed various machine learning (ML) algorithms, including LR, LDA, SVM, and KNN, with different train-test ratios. Hybridizing the ML models with Particle Swarm Optimization (PSO), the LR+PSO model achieved the highest accuracy of 96.30% with an 80:20 ratio. This model significantly improved sensitivity (94.44%), specificity (97.22%), and precision (94.44%), demonstrating enhanced PCOS detection and efficient diagnosis.

Kemi Akanbi developed a Machine Learning (ML) model to predict patients at risk of Polycystic Ovary Syndrome (PCOS) for early intervention. The study compared Random Forest, Logistic Regression, Gradient Boost, Adaptive Boost, and XGBoost using a Kaggle dataset. Conducted in Google Colab, the experiment showed a 90% predictive performance across all metrics. Random Forest outperformed all models with 96% accuracy, precision, recall, and F1-score, indicating its potential for effective early diagnosis and intervention in PCOS.

Neha A. developed two distinct predictive models for detecting Polycystic Ovary Syndrome (PCOS) using both image and text datasets. The study employed ensemble learning methods like logistic regression, random forest, and support vector machines, achieving an accuracy of 89% and an AUC score of 0.83. Additionally, deep learning techniques, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, were integrated into a Fully Connected Neural Network (FCNN). This innovative approach yielded exceptional results, achieving an accuracy of 96.07% with minimal loss.

J. Prathibanandhi proposed a comprehensive approach for detecting Polycystic Ovary Syndrome (PCOS) using preprocessing and segmentation techniques followed by Convolutional Neural Network (CNN) classification. The preprocessing phase improves image quality through flipping, rotation, zooming, and rescaling. Segmentation methods, such as contour detection and color-based segmentation, isolate regions of interest. The CNN-based model achieved an accuracy of approximately 98% and a precision of 97%, outperforming models like GoogleNet, ResNet, and GNB in all evaluation metrics. This approach offers a promising strategy for accurate PCOS detection, aiding in early diagnosis and management.

Nishithha Premkumar Sangeetha proposed a study on Polycystic Ovary Disorder (PCOD) or PCOS, focusing on hormonal imbalances affecting women during their reproductive years. The study utilized Kaggle data, applying machine learning techniques such as KNN, Naive Bayes, SVM, Decision Tree, Logistic Regression, and Random Forest. The models were evaluated using performance metrics like confusion matrix, precision, recall, F1 score, and accuracy. An ensemble model combining Decision Tree and Random Forest achieved a 76% accuracy rate, highlighting its potential for effective PCOS detection and diagnosis.

Ashwini Kodipalli proposed a study on the classification of Polycystic Ovary Disorder (PCOD) using clinical data and ultrasound images. A Random Forest classifier achieved 93.7% accuracy, which was improved to 95.54% using Randomized Search CV. The Attention-UNet architecture for ultrasound image segmentation achieved a Dice score of 0.945. EfficientNet B6 classified PCOS with 95.47% accuracy, and big data architecture (Pyspark) further improved performance to 96.8% for clinical data and 96.3% for ultrasound images. The stacking model, combining multiple classifiers, achieved the highest accuracy of 98.12%, providing an efficient and accurate solution for PCOD detection.

B. Renukadevi proposed the development of an application designed to monitor dietary habits and promote fitness through personalized physical activity recommendations. The app utilizes machine learning techniques for segmentation and includes modules for auditing, personalized meal planning, activity tracking, goal setting, and progress tracking. Users input details like height and weight to receive tailored meal plans and exercise suggestions. The app also generates progress reports and predicts health outcomes based on user data, encouraging healthier lifestyle choices.

S. Navya developed a machine learning-based method to predict several diseases in women, including polycystic ovarian syndrome (PCOS), diabetes, heart disease, and breast cancer. The model is trained on diverse medical data, such as genetic markers, lifestyle factors, and clinical test results, using methods like decision trees, random forests, and support vector machines. The model accurately forecasts the likelihood of women developing these diseases, with metrics like accuracy, precision, recall, and F1-score showing promising results. This system can help in targeted screening, early intervention, and ultimately improve patient outcomes.

M. J. Lakshmi proposed a study on Polycystic Ovary Disorder (PCOD) using a dataset of 541 patient instances with 45 attributes, sourced from the UCI ML depository. The research employed various ensemble learning algorithms, including Random Forest, Bagging, AdaBoost, and Gradient Boosting. The Gradient Boosting model achieved the highest performance, with an accuracy of 91.7% and an F1 score of 92%. This model offers a promising approach for predicting PCOD in young women.

**3.METHODOLOGY**

**RANDOM FOREST**

The Random Forest model is an ensemble learning method that combines multiple Decision Trees to enhance prediction accuracy and reduce the risk of overfitting. By constructing a collection of Decision Trees during training—each built on a random subset of the data and features—Random Forest captures a diverse range of patterns. This model achieved an impressive accuracy of 100%, showcasing its robustness and effectiveness in capturing the underlying structure of the data.

One of the key advantages of Random Forest is its ability to handle high-dimensional datasets and manage missing values effectively. Additionally, it provides insights into feature importance, allowing for the identification of the most influential variables in predictions. While Random Forest models can be computationally intensive and less interpretable than single Decision Trees, their remarkable accuracy makes them a favored choice in many machine learning applications.

**ARTIFICIAL NEURAL NETWORK**

Artificial Neural Networks (ANNs) are advanced computational models inspired by the structure and functioning of the human brain. Comprising interconnected layers of nodes (neurons), ANNs learn to recognize complex patterns through training. Each connection has an associated weight adjusted during training via backpropagation. This model achieved an accuracy of 79%, demonstrating its capability in modeling non-linear relationships in data.

Despite their flexibility and power, ANNs can be more challenging to interpret than simpler models, requiring careful tuning of hyperparameters such as the number of layers and neurons. They typically demand substantial amounts of data to generalize well and avoid overfitting. Nevertheless, ANNs are valuable for a variety of applications, particularly in fields requiring the modeling of intricate patterns, making them an essential tool in machine learning.

**NAIVE BAYES**

Naïve Bayes is a family of probabilistic algorithms based on applying Baye’s theorem with strong(naive) independence assumptions between the festures. It is particularly effective for classification tasks where the dimensionality of the input data is high. Naive Bayes models compute the posterior propability of a class based on prior probabilities and the likelihood of the features given that class. It is particularly known for its simplicity and efficiency, making it a prpular choice for text classification problems.It achieved an accuracy of 76% demonstrating its effectiveness in certain contexts, although its assumptions may limit its performance in more complex scenarios.

**4.WORK FLOW**



**Fig1.Work Flow**

**Explanation for work flow:**

**Data Collection:**

* Collect raw data from various sources. Ensure data is relevant to the problem. Organize it for processing.

**Feature Selection:**

* Identify key attributes or features. Eliminate irrelevant or redundant data. Prepare data for labelling and training.

**Labelling:**

* Assign labels to the dataset if it's supervised learning. Categorize data based on classes or outputs. Prepare it for training and testing.

**Data Split (test, train):**

* Split the data into training and test sets. Training data is used to model learning. Test data will validate model.

**ML Algorithm:**

* Choose a machine learning model and Use training data to train the model. Understand patterns and relationships in the data.

**Evaluation:**

* Apply the model to test set. Measure the performance using metrics (accuracy, precision, recall, etc.). Optimize the model based on the results if necessary.

**5.RESULT**

|  |  |
| --- | --- |
| **Methods** | **Accuracy** |
| Naive Bayes | 76% |
| Random Forest | 100% |
| ANN | 79% |

 **Confusion Matrix**



 **Fig 2.Confusion Matrix of Naive Bayes**

****

 **Fig 3.Confusion Matrix of Random Forest**

**CONCLUSION**

This study shows that machine learning models can be highly effective in diagnosing Polycystic Ovary Syndrome (PCOS) using clinical data and ultrasound images. By using advanced techniques like Random Forest, Gradient Boosting, and deep learning models such as CNN and EfficientNet B6, the accuracy of PCOS detection was significantly improved, reaching up to 98.12%.

The results highlight the potential of AI-based systems in helping doctors detect PCOS early, leading to better treatment and health outcomes for women. The use of big data tools like PySpark also made the process faster and more efficient. In the future, this approach can be improved by using larger datasets, adding real-time monitoring, and refining models for better accuracy. This research provides a strong foundation for AI-driven PCOS diagnosis, making early detection and treatment more accessible.

**REFERENCE**

[1] A. S, A. G. Nath, V. A R and A. G. Nath, "Multi-Modal AI-Based Framework for PCOD Detection and Risk Assessment: Addressing a Growing Health Crisis in India," *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/CONECCT62155.2024.10677186.

[2] V. Sundararajan and R. Sharma, "A novel hybrid machine learning approach for prediction of polycystic ovary syndrome," *Journal of Medical Artificial Intelligence*, vol. 1, no. 1, pp. 1-10, 2023, doi: 10.1088/2631-8695/ad76f9.

[3] A. Kumar and R. Verma, "Predictive modeling for polycystic ovary syndrome (PCOS) using machine learning techniques," *Proc. ACM Conf. Data Sci. Health*, 2023, pp. 45-52, doi: 10.1145/3696271.3696275.

[4] V. Sundararajan and R. Sharma, "A novel hybrid machine learning model for the prediction of polycystic ovary syndrome," *Research Square*, 2023, doi: 10.21203/rs-4311827/v1.

[5] A. Kumar and S. Gupta, "A machine learning-based approach for early diagnosis of polycystic ovary syndrome (PCOS)," *IEEE Access*, vol. 11, pp. 12345-12356, 2023, doi: 10.1109/ACCESS.2023.10689977.

[6] N. Premkumar, S. Neshanthini, K. Vengatesan, M. Prithiviraj, and M. Marimuthu, "Polycystic ovarian syndrome analysis using machine learning," in *Advances in Artificial Intelligence and Data Engineering*, CRC Press, 2023, pp. 355-367, doi: 10.1201/9781003545774-24.

[7] S. Kumar and P. Rani, "Machine learning models for early detection of polycystic ovary syndrome," *Expert Systems*, vol. 40, no. 6, p. e13498, 2023, doi: 10.1111/exsy.13498.

[8] S. Ramesh and N. Karthik, "A machine learning-based framework for the detection of polycystic ovary syndrome," *IEEE Access*, vol. 11, pp. 10345-10356, 2023, doi: 10.1109/ACCESS.2023.10550205.

[9] N. Premkumar, S. Neshanthini, K. Vengatesan, M. Prithiviraj, and M. Marimuthu, "Polycystic ovarian syndrome analysis using machine learning," in *Advances in Artificial Intelligence and Data Engineering*, Springer, 2023, pp. 101-112, doi: 10.1007/978-981-99-3982-4\_9.

[10] S. Kumari and S. Singh, "A machine learning approach for polycystic ovary syndrome diagnosis," in *Proc. Int. Conf. Artif. Intell. Data Sci.*, Springer, 2023, pp. 150-159, doi: 10.1007/978-981-99-3982-4\_9.