

Title: Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Traditional Machine Learning and Deep Learning Algorithms

Author: Aunik Hasan Mridul, Nowreen Ahsan, Syeda Sadia Alam, Sonia Afrose, Zakia Sultana, Md. Tanvir Mahmud kafi

Abstract: In recent years, there has been a noticeable rise in the prevalence of Polycystic Ovary Syndrome (PCOS), a complex endocrine disorder that affects a significant portion of the population, particularly women of reproductive age. PCOS is characterized by hormonal imbalances, irregular menstrual cycles, and the presence of multiple small cysts on the ovaries. Beyond its reproductive implications, PCOS is associated with various metabolic disturbances, including insulin resistance, obesity, dyslipidemia, and increased risk for type 2 diabetes and cardiovascular disease. To gain a comprehensive understanding of the disease's severity and its multifaceted impact on women's health, distinguishing between standard and affected diagnostic reports is imperative. In this study, we propose the application of algorithmic models to enable early detection and raise awareness of potential health risks associated with PCOS. Our approach is straightforward and well-suited for the prediction of uncomplicated cases of PCOS in real-world scenarios. Our dataset, sourced from various medical databases and clinical records, served as the foundation for our research. We employed a wide array of classifiers, including Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Convolution Neural Network (CNN), Long Short-Term Memory Network (LSTM), Bi-Directional Long Short-Term Memory Network (BLSTM), Random Forest (RF), Logistic Regression (LR), Gradient Boosting (GB), K-Nearest Neighbor (KNN), Adaboost Classifier (ABC), Decision Tree (DT), Support Vector Machine (SVM), Quadratic Discriminant Analysis (QDA), Ridge Classifier (RC), Passive Aggressive (PA), Gaussian Naïve Bayes (GNB), and ensemble techniques, to comprehensively explore and evaluate the predictive capabilities of each model in identifying PCOS and its associated complications. The results yielded notable success, with the Boosted Random Forest (RF) and Support Vector Classifier (SVC) classifier emerging as the most accurate, boasting an impressive accuracy rate of 98.278%. Furthermore, the Stacking Classifier RDAS exhibited an accuracy of 99.32%. Our optimization efforts, which included hyperparameter tuning, further enhanced the performance of each classifier. Based on extensive experimentation and a review of contemporary research, our findings unequivocally endorse the Random Forest (RF) and Support Vector Classifier (SVC) boosting classifier as exceptionally proficient, demonstrating a remarkable accuracy rate of 99.32% in the precise prediction of PCOS disease.

Keywords: PCOS; Bagging; Boosting; Ensemble; Machine learning; Deep learning

<https://cspub-ijcisim.org/index.php/ijcisim/article/view/712/597>