# Sakhi - Predictive Modeling for Early Detection of Women's Health Conditions using Machine Learning

**Dr. Nupur Giri*, Attreyee Mukherjee*, Amogh Inamdar*, Saumya Tripathi*, Yashodhan Sharma***
VES Institute of Technology Mumbai, Maharashtra,
India nupur.giri@ves.ac.in, 2021.attreyee.mukherjee@ves.ac.in, 2021.amogh.inamdar@ves.ac.in,
2021.saumya.tripathi@ves.ac.in
2021.yashodhan.sharma@ves.ac.in

**Article Info**

***Article type:***

*Research*

***Article History:***

*Received: 2024-03-22*

*Revised: 2024-05-18*

*Accepted: 2024-06-05*

***Keywords:***

Polycystic Ovary Syndrome (PCOS), En- dometriosis, Urinary Tract Infection (UTI), Machine Learning.

**A B S T R A C T**

For those who are designated as female at birth, health is essential to general well-being; nevertheless, conditions such as PCOS, endometriosis, and UTIs pose serious obstacles to reproductive health. Sakhi is a unique application that uses machine learning (ML) techniques to detect disorders proactively and solve them. Sakhi uses a customized questionnaire to gather user data. Machine learning methods such as Decision Tree Classifier, Logistic Regression, Random Forest Classifier, Support Vector Machines, and XGBoost classifier are then used to analyze trends in the user's health history and symptoms. Sakhi is a non-invasive platform for users to monitor and manage their health health, offering individualized insights and suggestions, by spotting deviations from typical swings in hormone levels. This creative strategy encourages access to healthcare services and understanding of health, enabling people to mitigate the effects of health disorders through proactive healthcare management.

## INTRODUCTION

In India, women facing conditions such as Polycystic Ovary Syndrome (PCOS), Endometriosis, and Urinary Tract Infec- tions (UTIs) often encounter unique challenges due to various factors including societal norms, limited access to healthcare, and cultural stigmas surrounding women's health. These are significant health concerns affecting millions of individuals worldwide and in India. Early and accurate diagnosis of these conditions is much needed for effective treatment and manage- ment. PCOS is a prevalent health disorder that affects women who are fertile and poses a serious threat to public health. 8–13 percent of women who are in the age for reproduction have the illness, and up to 70 percent of cases go untreated[1].Five to ten percent of females in their reproductive years suffer from this condition, which predominantly impacts women's fertility (15–45) [2] Endometriosis affects 1 in 10 women and those classified as females of reproductive age at birth. 176 million women globally, or 10% of all women, suffer with endometrio- sis. Infertile women may have an endometriosis prevalence of up to 30 to 50 percent.1.5 million women have endometriosis, and the number of people with diabetes who are born with a female gender assignment is likewise comparable[3]. Ten to fifteen percent of reproductive-age women and thirty to fifty percent of infertile or pelvic pain-affected women are affected by this condition. It should be mentioned, nonetheless, that endometriosis may also occur in young women and that some people develop it after menopause. In the majority of in- stances seen, women experience Endometriosis between their menarche and menopausal ages and when they go through the peak sickness when they are between 25 years to 45 years. Seven percent of the cases are also hereditary. [4] Urinary tract infections (UTIs) are seen mostly in adult women with a prevalence of 50-60%. Approximately 20% of women over 65 have this condition, compared to approximately 11% of the general population. Adult women will get a UTI at least once between 50 and 60 percent of cases [5]. With these issues increasing at an alarming rate in women, an application called Sakhi is being created as a part of this research, which will pave the way towards early detection of these issues, creating awareness about these problems in society, aiding in its diagnosis, and helping women lead a better life.

## LITERATURE REVIEW

Machine Learning (ML) has emerged as a valuable tool in healthcare, offering the potential to enhance diagnostic accuracy and patient outcomes. This literature survey explores the use of ML in predicting PCOS, Endometriosis, and UTI, focusing on research papers that have employed ML models as classifiers. By examining various approaches and method- ologies, this survey aims to identify trends, challenges, and opportunities in the application of ML for these conditions. Paper [6] introduces an enhanced machine-learning method

for PCOS detection via ovary ultrasound images. It evaluates different machine learning algorithms, like Random Forest, and deep learning approaches, to assess their performance. The results show that the Random Forest classifier outperforms other models, achieving an accuracy of 88.8%. Paper [7] dis- cusses the use of the XGBoost classifier using 19 variables to achieve optimal performance in predicting health imbalances in girls. Paper [8] utilizes the Random Forest classifier as part of an ensemble machine-learning approach to detect Polycystic Ovary Syndrome (PCOS) based on patient symptoms. It shows that, with 95.7% accuracy, the model using the 'Gradient Boosting' classifier as the meta learner surpasses the others in classifying patients with and without PCOS. There are many models that aim to enhance the classification accuracy as introduced in Paper [9] a new intelligence system, K-M- SVM, which combines k-means with LS-SVM for classify- ing PCOS. Paper [10] discusses the application of machine learning techniques for recognizing endometriosis through the Analogy of Endometriosis Recognition. It utilizes 38 RNA-seq datasets and 80 enrichment-based DNA methylation (MBD- seq) datasets for analysis.Paper [11] employs machine learning methods to analyze data. It applies Random Forest, Support Vector Machines, and Partial Least Squares Discriminant Analysis (PLSDA) to classify endometriosis using samples trained on both transcriptomics and methylomics data.

## NOVELTY OF WORK

Data was methodically collected from nearby hospitals in the localities. The collection was done from a balanced crowd which included women mostly in the age groups from 18 to 60. Their demographics and eco-social backgrounds were evenly distributed and consisted of women in all types of occupa- tions ranging from domestic workers, housewives, students, teachers, and public/private sector employees. 37 parameters were recorded for PCOS,33 parameters were recorded for Endometriosis and 52 parameters were recorded for UTI using a questionnaire. Separation was done to all the data into three different datasets about the three other issues faced. The training was performed on these datasets with a Random Forest Classifier, Decision Tree Classifier, Support Vector Ma- chine Classifier, Logistic Regression, and XGBoost Classifier thereby classifying and predicting whether a person is facing either of the above-mentioned issues. The evaluation pro- cess involved calculating the Accuracy, Precision, Recall, F-1 Score, and Cross-validation to draw comparisons between all of the models used and to understand the model's performance.

## METHODOLOGY

### *A. **Parameter selection***

The parameters chosen for each health condition aren't medical terms that need sophisticated medical tests. These were picked with the guidance of Gynaecologists who are experts in this field. These parameters were carefully chosen and are fully observable symptoms by the patients. It's been observed that many women experience these symptoms when they have the mentioned health issues.

### B. *Data Collection*

Data collection was done using an online form created by us, which was in both Hindi and English. This form was shared with female peers, teachers, relatives, and other females. Hospitals were also collaborated with to gather information from patients directly. They were chatted with, questions were asked and their answers were noted. The questions were based on the parameters that were decided upon with the help of Gynecologists and these were completely non- medical. It was made sure that the questions were simple to understand. This approach helped to gather lots of different kinds of information, which made the research more helpful and reliable.

Table 1 summarizes the data collected, showing 191 total responses. Among them, there are 25 positive cases each for PCOS, Endometriosis, and UTI.

| **Response Type** | **Number** |
|---|---|
| Total Responses | 191 |
| Positive Cases for PCOS | 25 |
| Positive Cases for Endometriosis | 25 |
| Positive Cases for UTI | 25 |

TABLE I
TABLE 1: SUMMARY OF DATA COLLECTED

### C. *Data Preprocessing and Cleaning*

In the research, preprocessing techniques were employed to enhance the quality of the dataset and address inherent challenges associated with dimensionality and class imbalance. Initially, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the feature space while pre- serving the required information. This technique facilitated the extraction of meaningful features from the dataset and ensured that the analysis captured the most important aspects of the data. Through PCA, the original feature set was brought into a smaller set of principal components, thereby mitigating the risk of overfitting and improving the efficiency of subsequent machine-learning algorithms.

Furthermore, to remove class imbalance, a common issue in the dataset where positive cases were underrepresented, an oversampling technique was employed. The Synthetic Minor- ity Over-sampling Technique (SMOTE) was utilized to in- crease the number of instances in the minority class, ensuring a more balanced distribution of classes in the training data. By oversampling the minority class, the aim was to alleviate the biases introduced by class imbalance, enabling the models to learn from a more representative dataset. This preprocessing step was instrumental in enhancing the capabilities of the machine learning models, contributing to the reliability and effectiveness of the findings. Data before and after SMOTE technique - Before the SMOTE analysis, the distribution of the data was 87.4% for people without the health condition and 12.6% for people with the health condition. After the SMOTE analysis, the distribution of the oversampled data became 50% for people without the health condition and 50% for people with the health condition.

### D. *Data split*

Selecting the test and train protocol is the most important part of any research. The dataset for the research was divided into two parts - testing and training.

- 70% training data and 30% testing data
- 75% training data and 25% testing data
- 80% training data and 20% testing data

Using appropriate training and testing data splits, such as 70:30, 75:25, or 80:20, is essential in predictive modeling for diseases like PCOS, Endometriosis, and UTIs, particularly when dealing with small datasets like that in the research. These splits help in evaluating model performance, preventing overfitting, ensuring generalization to new cases, and obtaining statistically reliable results.

### *E. **Correlation Matrices***

A correlation heatmap visually represents the strength and direction of relationships between multiple variables through color-coded cells, with darker hues indicating stronger cor- relations. Figure 1 illustrates the correlation heatmap for the PCOS dataset, highlighting the top 10 parameters' relation- ships. Notable findings include a moderate positive correlation between the top 10 parameters. The parameters shown here are

- Mental Health Score, Personal Hygiene, Average Meals, Job satisfaction, First Period Age, and Sanitary Product Frequency. Figure 2 shows a correlation heatmap for the Endometriosis dataset's parameters. Positive connections have been observed

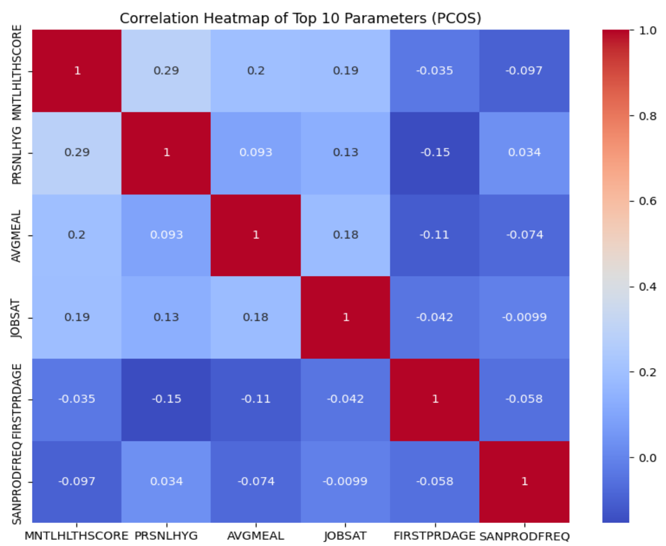


Fig. 1. Correlation heatmap for PCOS data and parameters

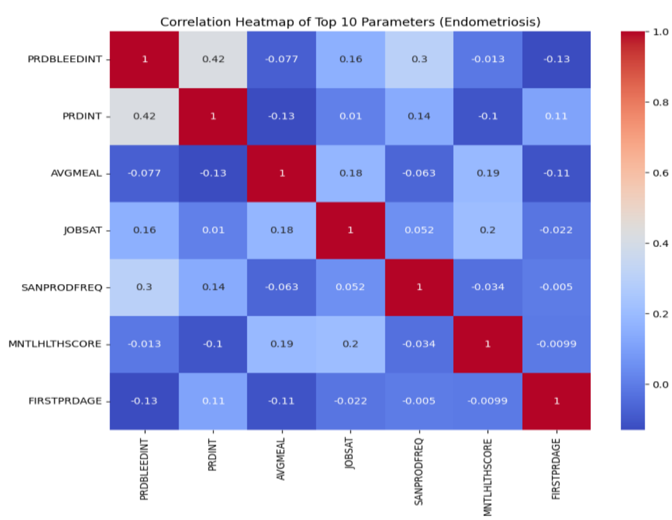


Fig. 2. Correlation heatmap for Endometriosis data and parameters

between a few of the top 10 endometriosis parameters. Below are the parameters for the Heatmap of Endometriosis: Probable Bleeding Intensity, Period Intensity, Average Meal, Job Satisfaction, Sanitary Product Frequency, Mental Health Score, First Period Age. Figure 3 shows the correlation heatmap for the UTI dataset's parameters, there are weak positive correlations between some of the top 10 parameters for UTI (Urinary Tract Infection).

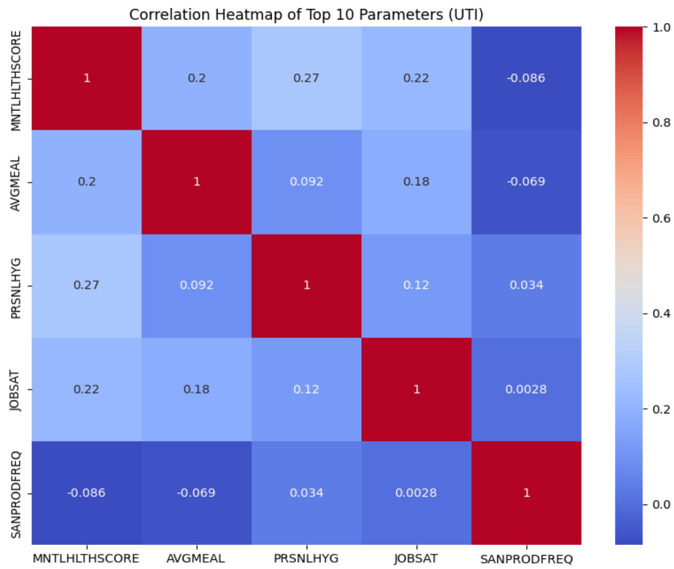


Fig. 3. Correlation heatmap for UTI data and parameters

These are the parameters of the Heatmap for UTI:
Mental Health Score, Average Meals, Personal Hygiene, Job Satisfaction, Sanitary Product Frequency, First Period Age.

## MACHINE LEARNING MODELS USED

### A. *Random Forest Classifier*

Random Forest Classifier is a common ensemble learning algorithm for classification and regression tasks. During train- ing, this type of supervised learning approach creates a large number of decision trees and outputs the mean prediction (regression) or the mode of the classes (classification) for each tree. The study assessed the effectiveness of the Random Forest Classifier on endometriosis, PCOS, and urinary tract infections (UTIs) using the cleaned datasets, maintaining a random state of 46 and an n-estimator size of 100. After maintaining the minimum sample leaf at three and the max- imum tree depth size at five for each of the three data split ratios (80:20, 75:25, and 70:30), the performance and analysis for the PCOS dataset were completed. In order to get better performance in terms of accuracy, the Endometriosis dataset's maximum depth was also maintained at five and the minimum sample leaf at three, respectively, for the 80:20 split; for the 75:25 and 70:30 splits, while for the UTI dataset, it was maintained at None and one.

### B. *Support Vector Machines*

The Support Vector Machine (SVM) is a strong supervised learning technique used in classification and regression tasks. It is most well-known for its effectiveness in classification tasks. The reason Support Vector Machines were used in the research analysis was because of the fact that it can effectively classify data that is not separable linearly by transforming the input space into a higher-dimensional space using a kernel. Its effect on the research analysis was measured using different performance metrics. The prediction was done keeping the kernel as a Radial Basis Function and the regularization parameter as 1.0 for PCOS and UTI prediction for all the three data splitting ratios

(80:20, 75:25, and 70:30) while for En- dometriosis the kernel was set as Linear and the regularization parameter as 0.1 to get better performance metrics.

### C. *Decision Tree Classifier*

A decision tree classifier is a supervised learning algorithm that mostly performs classification tasks. Through the use of basic decision rules deduced from the characteristics of the data, it forecasts the value of a target variable. The research carried out a thorough evaluation of a decision tree classifier's effectiveness on the cleaned medical datasets, on endometriosis, PCOS, and urinary tract infections (UTIs). The performance measures of the Decision Tree Classifier on the datasets were concluded after putting no definite limits to maximum tree depth and keeping the minimum sample leaf as 1. These parameters along with criterion as gini were set for all three different dataset splits (80:20, 75:25, and 70:30).

*Logistic Regression*

Logistic regression is a supervised learning approach that is used for binary classification tasks with a categorical output variable and only two classes (e.g., Yes/No, True/False, 0/1). Logistic Regression proved to be a good model for evaluating the performance measures in the analysis. Keeping the regu- larization term as l2, the inverse of regularization strength as 1.0, and the maximum iterations for lbfgs solver as 100, good results were obtained in prediction using this machine learning model. These parameters were maintained the same for all the different ratios of data splitting used.

### D. *XGBoost Binary Logistic Classifier*

XGBoost's efficiency stems from its optimization tech- niques like regularization, parallel processing, and handling missing values, contributing to its fast execution and high predictive accuracy. It incorporates advanced features such as tree pruning and cross-validation, which help prevent over- fitting and enhance the model's generalization capabilities. Additionally, its implementation of weighted quantile sketch for approximate tree learning ensures that even large-scale datasets can be processed quickly without sacrificing accuracy, benefiting medical datasets with large volumes of complex data. By tuning the hyperparameters—such as a learning rate of 0.3, a maximum depth of 6, and 100 iterations—researchers can achieve optimal performance. The subsample ratio close to 1 and the column sample ratio of 1 ensure that almost all data points and features are considered, capturing un- derlying relationships in the dataset. These hyperparameters collectively contribute to the model's robustness and accuracy, demonstrating that XGBoost can be effectively used in medical research. The great performance metrics obtained with these settings underscore XGBoost's capability to provide reliable and insightful results, making it a valuable tool for medical data analysis and research, ultimately aiding in better decision- making and improved healthcare outcomes.

## RESULTS

A comprehensive analysis was conducted of Receiver Operating Characteristic curves and prediction probabilities from the discussed machine learning models on the cleaned datasets. Figure 4 shows the combined graph of the Receiver Operating Characteristic (ROC) curves resulting from the analysis. The ROC curves and the AUC values were found only for the 80:20 data split since it showed the best results and Table 2 contains information about the performance measures concluded in our research using different models on different health conditions. In the research work, different Machine Learning models were used to predict health issues like PCOS, Endometriosis, and UTI. The Machine Learning models used are a Decision Tree Classifier, Random Forest Classifier, Logistic Regression, Support Vector Machine(SVM), and xgBoost. The findings of analyzing several machine learning classifiers for the prediction were remarkable. Across all three health issues, Random Forest, Decision Tree, and XGBoost models consistently demonstrate

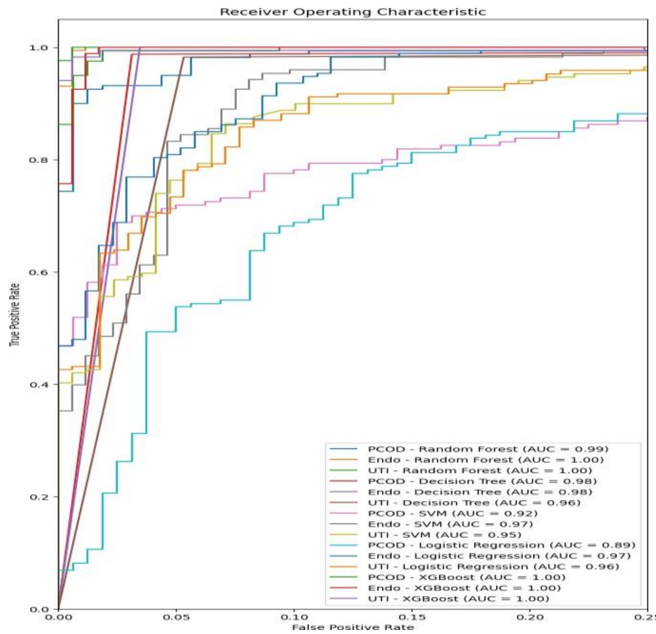

Fig. 4. Receiver Operating Characteristic curves resulted from the research analysis.

strong performance, achieving high Area Under the Curve (AUC) values, with some reaching perfect classification (AUC = 1). Support Vector Machine (SVM) and Logistic Regression models also perform well, even though they showed slightly lower AUC values compared to the previous methods. The models' great prediction capacity and powerful discriminating ability in differentiating between positive and negative cases of the corresponding health conditions are indicated by their high AUC scores. The performance measures in Table 2 include Accuracy, Precision, Recall, F1 Score, and Cross-Validation. The analysis is performed on different splits between the test and train data- 70% training data and 30% testing data (70:30), 75% training data and 25% testing data (75:25), 80% training data and 20% testing data (80:20). The best results were seen in the prediction of PCOS using the xgBoost Classifier for 80:20 data split. While Random Forest Classifier showed the best results for Endometriosis, xgBoost yet again showed great conclusions for the same dataset, in all the different data split ratios. On the other hand, the Random Forest Classifier proved to be the best model for predicting UTI with all three different data split ratios. While xgBoost also performed very well in predicting UTI for the 70:30 data split ratio, Decision Tree Classifier and XGBoost showed their power in predicting Endometriosis with the same data split.

Such remarkable results highlight the potential of using machine learning algorithms for early detection and prediction of these important women's health concerns. These results provide fresh possibilities for the development of accurate and reliable diagnostic tools for improving healthcare delivery. The remarkable predictive accuracy and discriminative power exhibited by these models signify a pivotal advancement in early detection and prognosis, heralding a new era of precision medicine in women's healthcare. These findings not only underscore the transformative potential of Machine Learning in healthcare but also pave the way for the development of precise and reliable diagnostic tools. Healthcare professionals can optimize patient care pathways, facilitate timely interventions, and significantly elevate healthcare outcomes for women worldwide. Therefore, these findings open up new possibilities for multidisciplinary cooperation, development, and innovation in the fight for accessible and equitable healthcare delivery. Thus the research introduces a shift towards proactive healthcare strategies, fostering a brighter future for women's health globally. By leveraging cutting- edge technology, this approach promises to reduce the burden of disease, improve quality of life, and ensure that women receive the most effective treatments tailored to their specific needs.

## STATEMENT OF CONFLICT OF INTEREST

The authors declare no competing interests, financial or non-financial, that could influence the content of this article. They have no relevant affiliations, funding, or relationships that could be perceived as a conflict of interest. The authors are independent and unbiased in their research and presentation of the manuscript's content.

| ML Models Used | Hormonal Issue | 80:20 split | | | | | 75:25 split | | | | | 70 : 30 split | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F1 Score | Cross Validation | Accuracy | Precision | Recall | F1 Score | Cross Validation | Accuracy | Precision | Recall | F1 Score | Cross Validation |
| Random Forest Classifier | PCOS | 82.81% | 0.8125 | 0.8387 | 0.8254 | [0.884, 0.835, 0.800] | 85.00% | 0.8571 | 0.8571 | 0.8571 | [0.888, 0.875, 0.850] | 89.58% | 0.898 | 0.898 | 0.898 | [0.827, 0.880, 0.865] |
| | Endometriosis | 97.14% | 0.9333 | 1 | 0.9655 | [0.935, 0.957, 0.946] | 96.55% | 0.925 | 1 | 0.961 | [0.931, 0.965, 0.930] | 97.12% | 0.9423 | 1 | 0.9703 | [0.889, 0.963, 0.913] |
| | UTI | 98.53% | 0.9615 | 1 | 0.9804 | [0.933, 0.900, 0.967] | 97.65% | 0.9444 | 1 | 0.9714 | [0.894, 0.881, 0.964] | 97.06% | 0.9333 | 1 | 0.9655 | [0.924, 0.798, 0.897] |
| Decision Tree Classifier | PCOS | 85.00% | 0.8163 | 0.9302 | 0.8696 | [0.763, 0.800, 0.775] | 86.25% | 0.8333 | 0.9302 | 0.8791 | [0.713, 0.800, 0.775] | 86.46% | 0.8571 | 0.875 | 0.866 | [0.653, 0.720, 0.770] |
| | Endometriosis | 93.10% | 0.8571 | 1 | 0.9231 | [0.908, 0.919, 0.907] | 93.10% | 0.8571 | 1 | 0.9231 | [0.897, 0.930, 0.907] | 92.31% | 0.8462 | 1 | 0.9167 | [0.889, 0.926, 0.925] |
| | UTI | 90.59% | 0.8974 | 0.8974 | 0.8974 | [0.776, 0.798, 0.869] | 87.06% | 0.8182 | 0.9231 | 0.8675 | [0.765, 0.774, 0.881] | 85.29% | 0.7679 | 0.9556 | 0.8515 | [0.861, 0.747, 0.885] |
| Support Vector Classifier | PCOS | 83.75% | 0.8571 | 0.8372 | 0.8471 | [0.788, 0.725, 0.800] | 83.75% | 0.8571 | 0.8372 | 0.8471 | [0.788, 0.725, 0.800] | 81.25% | 0.8125 | 0.8125 | 0.8125 | [0.800, 0.787, 0.757] |
| | Endometriosis | 87.14% | 0.7714 | 0.9643 | 0.8571 | [0.870, 0.935, 0.891] | 87.36% | 0.7955 | 0.9459 | 0.8642 | [0.885, 0.942, 0.884] | 86.54% | 0.7966 | 0.9592 | 0.8704 | [0.864, 0.938, 0.875] |
| | UTI | 89.41 | 0.875 | 0.897 | 0.8861 | [0.871, 0.833, | 89.41 | 0.875 | 0.897 | 0.886 | [0.871, 0.833, | 82.35% | 0.7755 | 0.844 | 0.8085 | [0.899, 0.835, |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % | | 4 | | 0.869] | % | | 4 | 1 | 0.869] | | | 4 | | 0.872] |
| **Logistic Regression** | **PCOS** | **76.56%** | **0.7353** | **0.8065** | **0.7692** | **[0.814, 0.835, 0.800]** | **75.00%** | **0.7619** | **0.7619** | **0.7619** | **[0.812, 0.838, 0.787]** | **76.04%** | **0.7600** | **0.7755** | **0.7677** | **[0.787, 0.813, 0.770]** |
| | **Endometriosis** | **85.06%** | **0.7347** | **1** | **0.8471** | **[0.897, 0.919, 0.919]** | **85.06%** | **0.7347** | **1** | **0.8471** | **[0.897, 0.919, 0.919]** | **85.58%** | **0.7544** | **0.9773** | **0.8515** | **[0.901, 0.926, 0.900]** |
| | **UTI** | **87.06%** | **0.8333** | **0.8974** | **0.8642** | **[0.871, 0.857, 0.845]** | **87.06%** | **0.8333** | **0.8974** | **0.8642** | **[0.871, 0.857, 0.845]** | **83.33%** | **0.7800** | **0.8667** | **0.8211** | **[0.899, 0.810, 0.846]** |
| **XGBoost Classifier** | **PCOS** | **93.75%** | **0.9130** | **0.9767** | **0.9438** | **[0.775, 0.825, 0.862]** | **93.75%** | **0.9130** | **0.9767** | **0.9438** | **[0.775, 0.825, 0.862]** | **92.00%** | **0.8846** | **0.9583** | **0.9200** | **[0.773, 0.787, 0.865]** |
| | **Endometriosis** | **95.40%** | **0.9211** | **0.9722** | **0.9459** | **[0.943, 0.977, 0.954]** | **95.40%** | **0.9211** | **0.9722** | **0.9459** | **[0.943, 0.977, 0.954]** | **94.00%** | **0.8800** | **1** | **0.9362** | **[0.914, 1.000, 0.912]** |
| | **UTI** | **91.76%** | **0.8636** | **0.9744** | **0.9157** | **[0.941, 0.905, 0.893]** | **91.76%** | **0.8636** | **0.9744** | **0.9157** | **[0.941, 0.905, 0.893]** | **93.00%** | **0.8800** | **0.9778** | **0.9263** | **[0.937, 0.873, 0.833]** |

TABLE II
TABLE 2: COMPARISON OF ALL MODELS

## REFERENCES

[1] World Health Organization: WHO & World Health Organization: WHO. (2023, June 28). Polycystic ovary syndrome. https://www.who.int/news- room/fact-sheets/detail/polycystic-ovary-syndrome

[2] Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El- Sappagh S, Saleh H. Polycystic Ovary Syndrome Detection Ma- chine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. Diagnostics (Basel). 2023 Apr 21;13(8):1506. doi: 10.3390/diagnostics13081506. PMID: 37189606; PMCID: PMC10137609.

[3] Endometriosis facts and figures — Endometriosis UK. (n.d.). https://www.endometriosis-uk.org/endometriosis-facts-and-figures

[4] Smolarz B, Szyłło K, Romanowicz H. Endometriosis: Epidemiol- ogy, Classification, Pathogenesis, Treatment and Genetics (Review of Literature). Int J Mol Sci. 2021 Sep 29;22(19):10554. doi: 10.3390/ijms221910554. PMID: 34638893; PMCID: PMC8508982.

[5] Medina M, Castillo-Pino E. An introduction to the epidemiol- ogy and burden of urinary tract infections. Ther Adv Urol. 2019 May 2;11:1756287219832172. doi: 10.1177/1756287219832172. PMID: 31105774; PMCID: PMC6502976.

[6] Suha, S. A., & Islam, M. N. (2022, October 12). An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Scientific Reports. https://doi.org/10.1038/s41598- 022-21724-0

[7] Pan, L., Liu, G., Mao, X., Li, H., Zhang, J., Liu, L., & Liang, H. (2018, January 1). Prediction Models for Girls with Suspected Central Precocious Puberty Using Machine Learning Algorithms. Social Science Research Network. https://doi.org/10.2139/ssrn.3215549
[8] Ding DC, Chen W, Wang JH, Lin SZ. Association between poly- cystic ovarian syndrome and endometrial, ovarian, and breast cancer: A population-based cohort study in Taiwan. Medicine (Baltimore). 2018 Sep;97(39):e12608. doi: 10.1097/MD.0000000000012608. PMID: 30278576; PMCID: PMC6181615.
[9] Faris, N. N., & Miften, F. S. (2022, June 22). An intelligence model for detection of PCOS based on k-means coupled with LS-SVM. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.7139
[10] Arches, P. M. C., Austero, A. R., Diaz, A. J. A., Taer, H. J. C., & Fabregas, A. C. (2021, December 22). Detecting Urinary Tract Infection (UTI) thru Analytes level using Convolutional Neural Network and Support Vector Machine. https://doi.org/10.1145/3512576.3512591
[11] H.T. Lin and C.J. Lin. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report, Department of Computer Science, National Taiwan University, 2003.
[12] Akter, S., Xu, D., Nagel, S. C., Bromfield, J. J., Pelch, K. E., Wilshire, G. B., & Joshi, T. (2019, September 4). Machine Learning Classifiers for Endometriosis Using Transcriptomics and Methylomics Data. Frontiers in Genetics. https://doi.org/10.3389/fgene.2019.00766