Predictive Modeling for Early Detection of Women's Health Conditions Using Machine Learning

Submitted in partial fulfillment of the requirements of the degree

BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By

Amogh Inamdar/ 18

Attreyee Mukherjee/ 33

Yashodhan Sharma/55

Saumya Tripathi/ 61

Name of the Mentor

Dr. Nupur Giri

Vivekanand Education Society's Institute of Technology,

An Autonomous Institute affiliated to University of Mumbai

HAMC, Collector's Colony, Chembur,

Mumbai-400074

(AY 2023-24)

CERTIFICATE

This is to certify that the Mini Project entitled **"Predictive Modeling for Early Detection of Women's Health Conditions using Machine Learning"** is a bonafide work of **Amogh Inamdar(18), Attreyee Mukherjee(33), Yashodhan Sharma(56), Saumya Tripathi(61)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Computer Engineering".**

(Dr. Nupur Giri)
Mentor

(Dr. Nupur Giri)
Head of Department

(Dr. Mrs. J. M. Nair)
Principal

Mini Project Approval

This Mini Project entitled "**Detection of Issues Caused by Abnormal Hormonal Changes in Menstruators Using Machine Learning"** by **Amogh Inamdar(18), Attreyee Mukherjee(33), Yashodhan Sharma(56), Saumya Tripathi(61)** are approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1……
(Internal Examiner Name & Sign)

2……
(External Examiner name & Sign)

Date: 13.04.2024

Place: Mumbai, India.

Contents

ABSTRACT

From 1 in 10 women suffering from PCOS worldwide to currently 3-4 in 10 women, PCOS is now exponentially increasing among women due to an unhealthy lifestyle. The literature says that 1 in every 5 women in India suffers from PCOS.

In 2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths globally. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer.

Endometriosis affects roughly 10% (190 million) of reproductive-age women and girls globally.

Menopause is one of the most significant events in a woman's life and brings in several physiological changes that affect the life of a woman permanently. There have been a lot of speculations about the symptoms that appear before, during, and after the onset of menopause.

Since women face all these issues, detecting these after a significant period can affect the patient mentally, emotionally, and physically.

Hence, we aim to create a project that will make it easier for women to find out whether they might have these issues, be aware of it, and get cautious for the future using our app to predict these at an early stage looking at various symptoms.

ACKNOWLEDGEMENT

We extend our sincere gratitude to the individuals and institutions that have been instrumental in the successful completion of our project, "Detection of Issues Caused by Abnormal Hormonal Changes in Menstruators Using Machine Learning." This project has been a collaborative effort, and we appreciate the support and contributions from the following:

Our esteemed project mentor, Dr. Nupur Giri, provided invaluable guidance and expertise throughout the research process, ensuring the clinical relevance of our work.

The healthcare professionals, doctors, and hospitals shared their expertise and provided critical parameters and insights for our project. Their involvement and collaboration were essential in ensuring the accuracy and relevance of our research.

The Department of Computer Engineering for their constant and invaluable guidance. Their collective wisdom and dedication to the field have been instrumental in shaping our knowledge and skills. Their support has been an essential pillar of our academic and professional growth, and we are thankful for their continuous encouragement.

CHAPTER 1: INTRODUCTION

1.1 Introduction

Our project will help women to be able to detect the possibility of having issues due to abnormal hormonal changes by answering questions that would be the parameters for detecting these issues. If any of the issues are predicted, our app will give a warning to the user to visit the doctor and take medical steps.

The parameters we are choosing will be common for all the issues and by clustering certain parameters, we will be able to predict the corresponding problem faced by the user.

For this, along with independent parameters, we also would need a correlation between the hormonal issues.

Correlation between PCOS, Endometriosis, and Menopause-

Characteristics and consequences of PCOS have been previously associated with both increased and decreased risk of breast cancer. For example, infertility due to an ovulatory disorder has been shown to decrease breast cancer risk while obesity increases breast cancer risk among postmenopausal women and decreases risk among premenopausal women.

An association between PCOS and endometrial cancer was first suggested over fifty years ago by a series of small case reports, but these studies were limited by a lack of control groups and small numbers. Three meta-analyses, with overlapping studies, have reported a significantly increased risk of endometrial cancer among women with PCOS. However, they each included estimates from analyses that did not take into account BMI which is a common characteristic of PCOS and a strong and well-established risk factor for endometrial cancer.

Women with PCOS tend to reach menopause about two years later than women without PCOS. PCOS doesn't go away with menopause, so you can continue to have symptoms. Some symptoms of PCOS are similar to those of perimenopause.

1.2 Motivation

The motivation behind our project is rooted in a deep commitment to women's health and well-being. We recognize that hormonal issues can often go unnoticed or undiagnosed, leading to unnecessary suffering and complications. By harnessing the power of machine learning and data analysis, we aim to empower women with a proactive tool that allows them to monitor their hormonal health and detect potential issues early. Our goal is to create a user-friendly application that serves as a guardian of women's health, providing timely

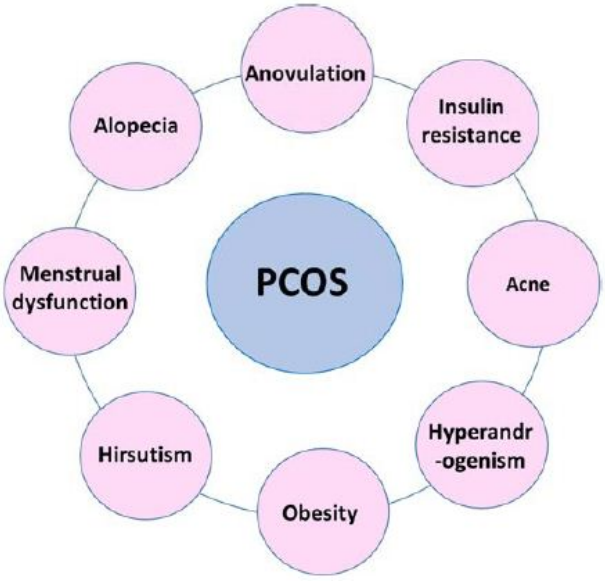
warnings and guidance when anomalies are detected in their responses to carefully chosen parameters. We believe that knowledge is the first step to prevention and early intervention, and this project seeks to provide women with the knowledge and tools they need to make informed decisions about their health. Through this endeavor, we aspire to improve the quality of life for women by promoting early detection and medical intervention, ultimately contributing to a healthier and happier community of menstruators.

1.3 Problem Statement

PCOS/PCOD-

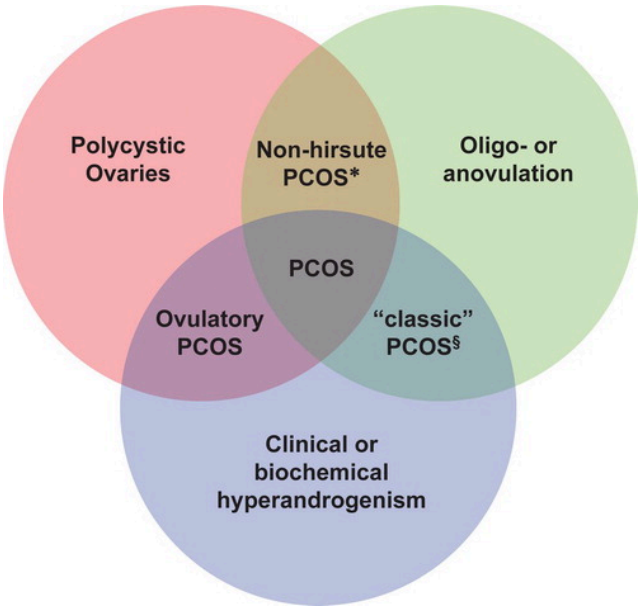
Polycystic Ovary Syndrome (PCOS) is a medical condition that causes hormonal disorders in women in their childbearing years. The hormonal imbalance leads to a delayed or even absent menstrual cycle. Women with PCOS majorly suffer from excessive weight gain, facial hair growth, acne, hair loss, skin darkening, and irregular periods leading to infertility in rare cases.

Polycystic Ovary Disease (PCOD) is a common disorder among women with no exact cure known to date. It shows various symptoms and may even contribute to long-term health problems. It is largely ignored, due to a lack of awareness and usually detected when women try for conception. The existing methodologies and treatments are insufficient for early-stage detection and prediction. To confirm the diagnosis various hormonal blood tests along with an ultrasound scan are required which leads to multiple trips to well-equipped hospitals in urban areas.



Some Causes of PCOS

Figure 1.1



Different Types of PCOS

Figure 1.2

ENDOMETRIOSIS-

Endometriosis is a chronic gynecological condition that affects 5–10% of women of reproductive age. Women with endometriosis have endometrial-type tissue outside of the uterus. In exceptional cases, endometriosis lesions may reach organs distant from the pelvis such as the membranes of the lungs, heart, limbs, and brain. As a result, and in response to the substances that this tissue produces, the immune system is activated, and a chronic inflammatory process is triggered, leading to the formation of adhesions, scars, and cysts between the pelvic and abdominal organs. Endometriosis tissue can also penetrate various organs in the body, including the digestive and urinary systems, and attach to nerves.

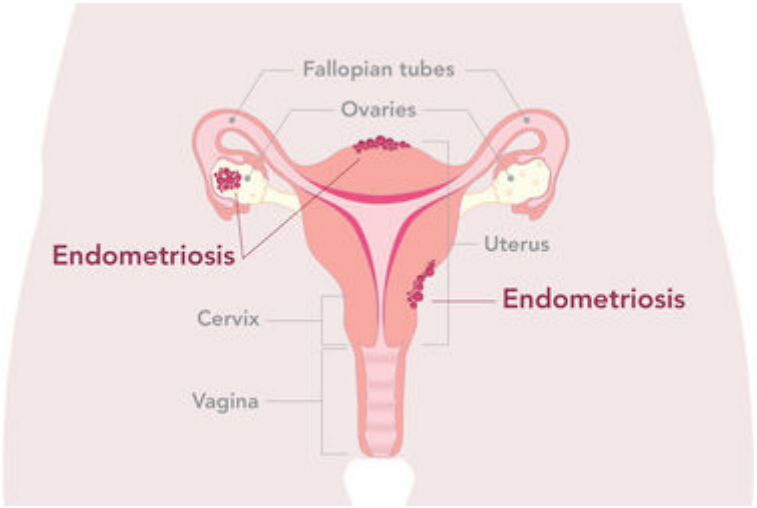


Figure 1.3

POST-MENOPAUSAL ISSUES-

Menopause is one of the most significant events in a woman's life and brings in a number of physiological changes that affect the life of a woman permanently. There have been a lot of speculations about the symptoms that appear before, during, and after the onset of menopause.

Menopause is the permanent cessation of menstruation resulting in the loss of ovarian follicle development.

The following are some of the issues faced by women during and post-menopause:-

Vasomotor symptoms- affect up to 75% of peri-menopausal women. Symptoms last for 1–2 years after menopause in most women, but may continue for up to 10 years or longer in others.

Urogenital atrophy - Urogenital atrophy results in vaginal dryness and pruritus, dyspareunia, dysuria, and urinary urgency.

Osteoporosis - Musculoskeletal symptoms characterized by backache, fractures on minimal trauma, decreased height, and mobility are common due to osteoporosis.

Depression- Although most women transition to menopause without experiencing psychiatric problems, an estimated 20% have depression at some point during menopause.

- **Hormonal changes**
- **Life stressors**
- **Psychological or social conditions**

Cognitive functions - Memory problems are common complaints in perimenopausal and recent postmenopausal women. The increased frequency of cognitive complaints in menopausal women suggests that memory problems in this population are related to the menopause transition rather than to the aging process.

Problems with sleep - Insomnia occurs in 40–50% of women during the menopausal transition, and problems with sleep may or may not be connected to mood disorders.

Psychiatric disorders-

- **Schizophrenia**
- **Bipolar disorder**
- **Panic disorder**
- **Obsessive-compulsive disorder**

Along with these, we also **propose** to predict the approximate age of menopause for the user provided they haven't hit their menopause yet. This we can do by looking at certain symptoms/parameters given by the user.

Time For Menopause-The medical definition of menopause is no menstrual bleeding for a year. Most women experience menopause between age 40 and 58, and the average age at menopause is 51, according to the North American Menopause Society. Many women are surprised when they go through menopause in their forties because they think they're too young, but it's not unusual. Women with shorter menstrual cycles, defined as less than 25 days, are more likely to reach menopause early than women with normal-length cycles (26 to 34 days), according to a study of 634 women published in *Menopause*Researchers also found that the women with short menstrual cycles had a higher frequency of total menopause symptoms, and were more likely to have certain menopause symptoms, including midlife sleep problems, heart discomfort, and depressive symptoms.

CHAPTER 2: LITERATURE SURVEY

2.1 Survey of Existing System

Name	Takeaways
1] Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimised Feature Selection and Explainable Artificial Intelligence-	Understanding how to detect PCOS using Machine Learning Models. Understanding how to use Optimized Feature Selection and Explainable Artificial Intelligence

H.Elmannai, N.El-Rashidy, I.Mashal	
2] Association between Polycystic Ovarian Syndrome and Endometrial, Ovarian, and Breast Cancer- D.Ding, W.Chen, J.Wang, S.Lin	Understanding the relation of PCOS with Endometrial Cancer. Learning that there is no direct relationship between PCOS, Ovarian and Breast Cancer
3] Menopausal Symptoms- K.Ruddy, A.Partridge	Understanding Menopause from a medical perspective. Understanding parameters for determining the age of Menopause, Perimenopause, and Post-Menopausal Symptoms
4] Risk of Endometrial, Ovarian and Breast Cancer in Women with Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis- J.Barry, M.Aziza, P.Hardiman	Understanding the relations using Meta-Analysis

Table 2.1

2.2 Research Gap

1. Limited Focus on Early Detection: Many existing studies may primarily focus on diagnosing hormonal disorders after symptoms have become severe. A research gap may exist in developing early detection methods that can identify issues in their early stages, enabling timely intervention and prevention.

2. Diversity and Representation: Research in women's health and hormonal changes should address issues of diversity and representation. A research gap may exist in creating models that are inclusive and applicable to a wide range of individuals, accounting for various ethnicities, ages, and socioeconomic backgrounds.

3. Data Privacy and Ethical Concerns: The ethical use of personal health data is a significant concern. Research should address how to ensure privacy and data security while developing machine learning models for hormonal issue detection.

4. User-Centered Design: Existing applications and tools might not fully consider the user experience and preferences. There may be a research gap in developing user-centered applications that encourage user engagement and provide information in a clear and accessible manner.

5. Interconnected Hormonal Issues: Many hormonal disorders and issues are interconnected. A research gap could involve developing models that consider the complex interplay between different hormonal conditions and provide comprehensive assessments.

6. Clinical Validation and Integration: While machine learning models may show promise in early detection, a research gap exists in conducting clinical validation studies to confirm the effectiveness of these models and integrating them into healthcare systems for practical use.

7. Educational Components: A research gap may involve the development of educational components within the application to enhance users' understanding of hormonal health, its importance, and the potential issues they may encounter.

8. Feedback Mechanisms: The project may benefit from research on how to provide users with constructive feedback and recommendations that are actionable and encourage proactive health-seeking behaviors.

9. Long-term Monitoring: Focusing on long-term hormonal health monitoring may address a research gap. Many issues may manifest over time, and continuous monitoring is essential for early detection.

10. Cross-Disciplinary Collaboration: Building bridges between machine learning experts and medical professionals is essential. Research can explore ways to facilitate collaboration and communication between these fields to develop more accurate and practical models.

2.3 Mini Project Contribution

Our mini-project contribution to this project was to develop a machine-learning model to predict PCOS/PCOD, Endometriosis, and UTI based on a patient's symptoms and medical history. We trained the model on a dataset for which, the data collection was done using an online form created by us, which was in both Hindi and English. We shared this form with female peers, teachers, relatives, and other females we knew. We also worked with hospitals

to gather information from patients directly. We chatted with them, asking questions and noting their answers. This approach helped us gather lots of different kinds of information, which made our research more helpful and reliable.

We are also developing a Flutter app that allows users to input their symptoms and medical history and receive a prediction of their risk of having PCOS, Endometriosis, and UTI. The app also provides users with information and the different treatment options available.

CHAPTER 3: PROPOSED SYSTEM

3.1 Introduction

Abnormal hormonal fluctuations within the menstrual cycle can often signify underlying health issues, necessitating timely and accurate detection for effective intervention. However, the identification of such irregularities remains a complex challenge, requiring a nuanced understanding of hormonal patterns and their potential implications.

In recent years, advancements in machine learning techniques have revolutionized the healthcare sector, offering promising solutions for the early detection and diagnosis of various health conditions. Leveraging the power of data-driven insights, the proposed system aims to utilize machine learning algorithms to discern patterns indicative of irregular hormonal changes in menstruators. By integrating data from diverse sources, including physiological indicators, lifestyle patterns, and symptomatology, this system seeks to provide a comprehensive framework for the timely identification of potential health issues related to abnormal hormonal fluctuations.

This report outlines the proposed system's key objectives, methodologies, and anticipated outcomes, highlighting the potential implications for both medical practitioners and individuals seeking proactive menstrual health management. Through the effective application of machine learning techniques, this system endeavors to contribute to a proactive and personalized approach to menstrual healthcare, fostering early intervention and improved health outcomes for menstruating individuals.

3.2 Architectural Framework / Conceptual Design

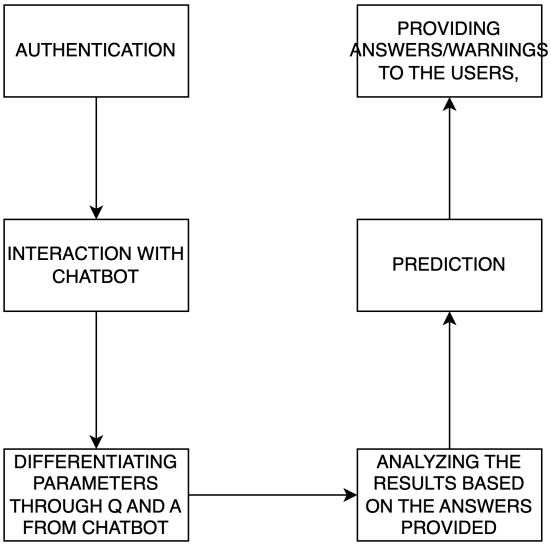


Figure 2.1

3.3 Algorithm and Process Design

Algorithm Selection:

We tried different algorithms to see the best fit for our dataset. The algorithms that we tried were:

- Random Forest Classifier:

  Random Forest is a common ensemble learning algorithm for classification and regression tasks. It is a form of supervised learning technique that generates a large number of decision trees during training and outputs either the mode of the classes (classification) or the mean prediction (regression)of each tree. After using the Random Forest is a common ensemble learning algorithm for classification and regression tasks. It is a form of supervised learning technique that generates a large number of decision trees during training and outputs either the mode of the classes (classification) or the mean prediction (regression) of each tree. The research evaluated the performance of the Random Forest Classifier our cleaned datasets, on endometriosis, PCOS, and urinary tract infections (UTIs). The investigation attempted to determine how accurate these classifiers were at making predictions by using Receiver Operating Characteristic (ROC) curves. An Area Under the Curve (AUC) value of 0.47 for PCOS indicates that the classifier performed only moderately well, indicating difficulties in correctly recognizing this condition based on the attributes of the supplied data. On the other hand, the classifier performed

exceptionally well for endometriosis, with an AUC value of 0.93, indicating its strong predictive ability to identify this particular medical disease. With an AUC score of 0.90, the classifier demonstrated exceptional accuracy on UTI data as well.

- Support Vector Machine:

  The Support Vector Machine (SVM) is a strong supervised learning technique used for classification and regression tasks. However, it is most well-known for its effectiveness in classification tasks. Using Receiver Operating Characteristic (ROC) curves, the study assessed the performance of Support Vector Machine (SVM) on our cleaned datasets on endometriosis, urinary tract infections (UTI), and PCOS. With an Area Under the Curve (AUC) score of 0.98 for PCOS, the model performed exceptionally well, demonstrating its excellent accuracy in diagnosing this condition. In the same way, the model performed quite well for endometriosis (AUC = 0.99), indicating that it is strong in identifying this medical disease. Additionally, it performed very well on UTI data, producing an AUC value of 0.93, highlighting its ability to correctly identify urinary tract infections.

- 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 a our cleaned medical datasets, on endometriosis, PCOS, and urinary tract infections (UTIs). The analysis's main metrics, Receiver Operating Characteristic (ROC) curves, were used to evaluate the classifier's predictive accuracy for each condition. The results showed that the classifier performed exceptionally well in detecting PCOS patients, as demonstrated by the outstanding Area Under the Curve (AUC) value of 0.98. With an AUC score of 0.93, the classifier also performed well in identifying cases of endometriosis, demonstrating its excellence in managing this particular medical problem. Additionally, the classifier performed well with UTI data, producing an AUC value of 0.88, indicating its ability to reasonably identify instances of urinary tract infections.

- 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).The study examined the value of the Area Under the Curve

(AUC) in assessing the ability of models to predict the conditions using our cleaned medical datasets, on endometriosis, PCOS, and urinary tract infections (UTIs). AUC values are an essential parameter; a perfect score of 100 percent signifies that the model is completely capable of differentiating between positive and negative examples. In particular, the perfect AUC of 100\% indicates that the models have excellent discriminatory power in correctly identifying instances associated with PCOS and endometriosis datasets. An AUC of 99 percent true positive rate (TPR) in the case of UTI indicates that the model performed exceptionally well in properly detecting positive occurrences.

- XGBoost:

  The gradient boosting algorithm is implemented in an efficient and streamlined manner by XGBoost, which stands for eXtreme Gradient Boosting. For supervised learning tasks like regression and classification, it is particularly prevalent. In the case of PCOS, a perfect score indicates that the model can accurately identify 100\% of genuine positive and true negative situations. More specifically, the AUC score measures the degree of separability between positive and negative classes. In the case of endometriosis, the research showed that the model obtained an 100\% AUC value, indicating that it is very accurate in differentiating between positive and negative cases. Similar to the UTI example, the model's efficacy in identifying urinary tract infections was highlighted by the fact that, despite its shortcomings, the AUC score accurately classified 99\% of true positive and true negative cases, indicating great performance.

Feature Selection and Engineering:

In our project, we performed an extensive feature selection and engineering process to determine which parameters are most relevant to detecting hormonal issues. This step is crucial in improving the model's accuracy. Features that are highly correlated with hormonal conditions and symptoms would have been selected and used as input variables for all the models that we used. We also found the correlation heatmap for the same.

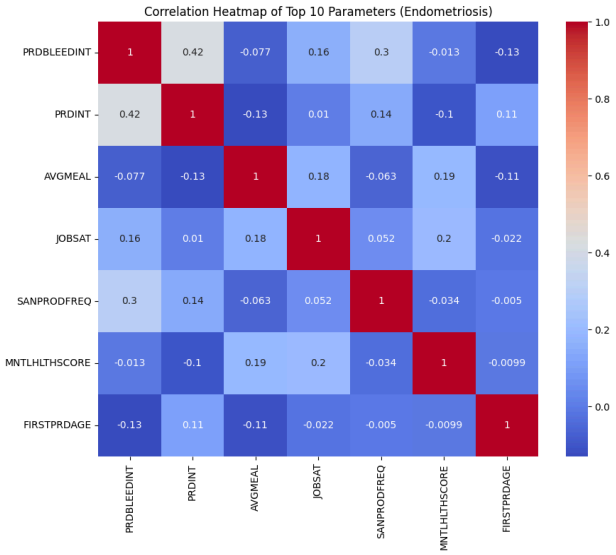


Figure 3.3.1 For Endometriosis

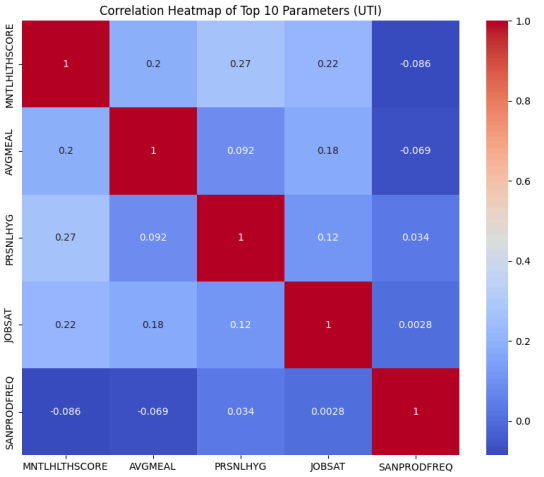


Figure 3.3.2: For UTI

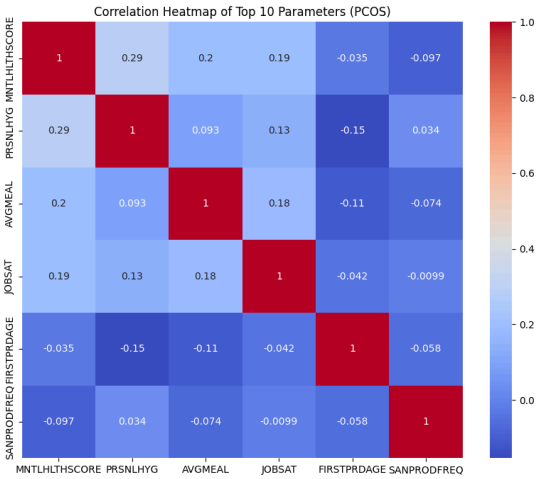


Figure 3.3.3 For PCOD/PCOS

3.4 Methodology Applied

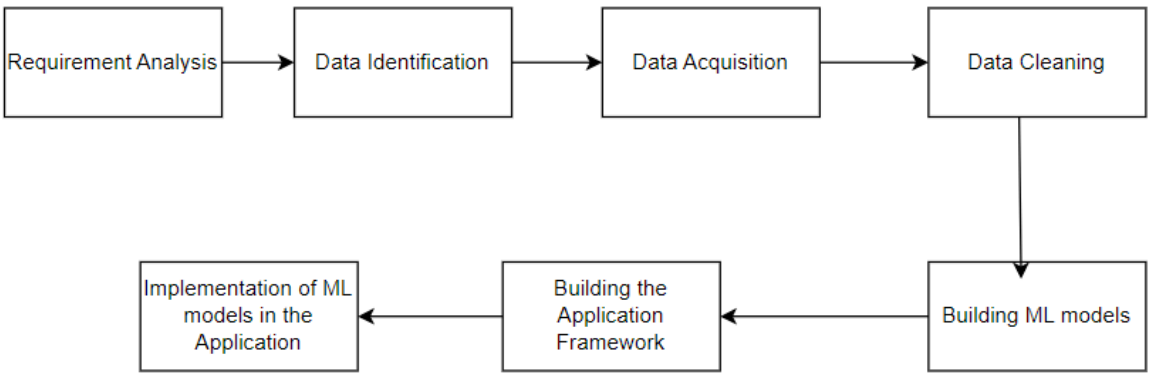


Figure 3.4.1

3.5 Hardware & Software Specifications

Hardware Tools:

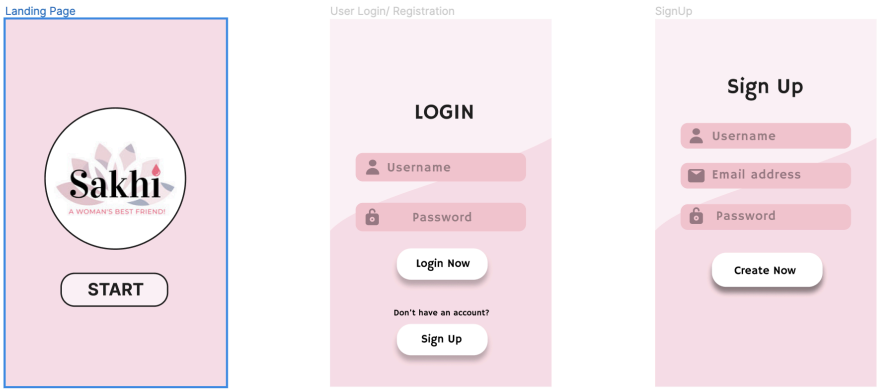
- Computer System with 8 GB RAM

Software Tools:

- Python
- Android/Flutter SDK
- Jupyter/Colab Notebook

3.6 Experiment and Results for Validation and Verification

App design prototype



Accuracy of all the models:

ML Model Used	Hormonal Issue	80:20 Split					75:25 Split					70:30 Split				
		A	P	R	F1	CV	A	P	R	F1	CV	A	P	R	F1	CV
Random Forest Classifier	PCOS	98.44%	0.9688	1	0.9841	[0.965, 0.965, 0.941]	97.5	97.62	97.62	0.9762	[0.975, 0.9625, 0.975]	97.92 %	0.9796	0.9796	0.9796	[0.9733, 0.9733, 0.9729]
	Endo	97.14%	0.9333	1	0.9655	[0.924, 0.891, 0.957]	98.85	97.37	1	0.9867	[0.954, 0.9884, 1.0]	98.08 %	0.9608	1	0.98	[0.9383, 0.9753, 1.0]
	UTI	98.53%	0.9615	1	0.9804	[0.944, 0.989, 1.000]	98.82	97.14	1	0.9855	[0.9765, 0.9881, 0.9643]	96.08 %	0.913	1	0.9545	[0.9873, 0.9494, 0.9359]
Decision Tree Classifier	PCOS/D	97.50%	1	0.9535	0.9762	[0.888, 0.850, 0.938]	96.25	97.62	95.35	0.9647	[0.8875, 0.85, 0.9375]	97.50 %	1	0.9535	0.9762	[0.925, 0.8375, 0.925]
	Endo	96.55%	0.9459	0.9722	0.9589	[0.920, 0.977, 0.965]	96.55	92.31	1	0.96	[0.8966, 0.9884, 0.9419]	95.40 %	0.9211	0.9722	0.9459	[0.908, 0.9767, 0.9302]
	UTI	89.41%	0.8571	0.9231	0.8889	[0.918, 0.857, 0.917]	89.41	84.09	0.9487	0.8916	[0.9059, 0.881, 0.9167]	89.41 %	0.8409	0.9487	0.8916	[0.9176, 0.881, 0.9167]
Support Vector Machine	PCOS/D	97.50%	1	0.9535	0.9762	[0.975, 0.875, 1.000]	97.5	1	95.35	0.9762	[0.975, 0.875, 1.0]	97.50 %	1	0.9535	0.9762	[0.975, 0.875, 1.0]
	Endo	98.57%	0.9655	1	0.9825	[1.000, 1.000, 0.989]	98.85	97.37	1	0.9867	[1.0, 1.0, 0.9884]	99.04 %	0.98	1	0.9899	[1.0, 1.0, 0.9875]
	UTI	96.47%	0.950	0.9744	0.962	[0.988, 0.988, 0.988]	96.47	0.95	0.9744	0.962	[0.9882, 0.9881, 0.9881]	96.47 %	0.95	0.9744	0.962	[0.9882, 0.9881, 0.9881]
Logistic Regression	PCOS/D	98.44%	0.9688	1	0.9841	[1.000, 1.000, 0.988]	97.5	95.45	1	0.9767	[1.0, 1.0, 0.9875]	97.92 %	0.9608	1	0.98	[1.0, 1.0, 0.9865]
	Endo	98.85%	0.973	1	0.9863	[0.989, 0.988, 0.988]	98.85	97.3	1	0.9863	[0.9885, 0.9884, 0.9884]	98.85 %	0.973	1	0.9863	[0.9885, 0.9884, 0.9884]
	UTI	95.29%	0.9268	0.9744	0.95	[0.988, 0.976, 0.940]	95.29	0.9268	0.9744	0.95	[0.9882, 0.9762, 0.9405]	95.29 %	0.9268	0.9744	0.95	[0.9882, 0.9762, 0.9405]
XGBoost	PCOS/D	97.50%	1	0.9535	0.9762	[0.925, 0.887, 0.938]	97.5	1	95.35	0.9762	[0.925, 0.8875, 0.9375]	98%	1	0.9535	0.9762	[0.925, 0.8875, 0.9375]

Classifier	Endo	97.70%	0.9722	0.9722	0.9722	[0.966, 1.000, 0.977]	97.7	97.22	0.9722	0.9722	[0.9655, 1.0, 0.9767]	98%	0.9722	0.9722	0.9722	[0.9655, 1.0, 0.9767]
	UTI	95.29%	0.9268	0.9744	0.95	[0.941, 0.929, 0.976]	95.29	0.9268	0.9744	0.95	[0.9412, 0.9286, 0.9762]	95%	0.9268	0.9744	0.95	[0.9412, 0.9286, 0.9762]

Table 3.6.1

Here the the following terms represent:

- P-Precision
- R-Recall
- F1-F1 Score
- CV-Cross Validation

The Table 3.6.1 contains information about the performance measures concluded by us using different models on the different hormonal conditions. These performance measure includes 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 shaded cells depict the best accuracies.

Chatbot Questions

1. What is your Age?
2. What is your Height?
3. What is your weight?
4. What is your Blood Group?
5. Do you have irregular sleep cycles?
6. Cycle Length(in days)
7. Do you experience excruciating pain during your periods?
8. Do you have any white discharges frequently?
9. Do you have an absence of periods for some months?
10. Do you have pain during intercourse?
11. Nausea and vomiting
12. Pain in lower abdomen
13. On an Average, how many meals do you have in a day?
14. Are you Pregnant(Y/N)
15. Hair growth(Y/N)
16. Facial Hair(Y/n)
17. Skin darkening(Y/N)
18. Hair Loss(Y/N)
19. Do you have any Pimples/Acne(Y/N)
20. Have Fast Food(Y/N)
21. Regular exercise(Y/N)
22. Do you consume excessive caffeine?
23. Have you used ORAL contraceptive pills?

Figure 3..6.1

3.7 Result Analysis and Discussion

1. Significance of the Project:

Our project holds great significance as it empowers women to proactively assess their health risks related to PCOS, endometrial cancer, menopausal symptoms, and UTIs through a user-friendly chatbot and machine learning. By enabling early detection, the project may lead to more effective treatments and better health outcomes. It respects user privacy and provides accessible, data-driven insights, educating women about health issues and encouraging them to seek timely medical attention. Furthermore, it has the potential to efficiently allocate healthcare resources, scale to a broader audience, support research, and foster collaboration with medical professionals, ultimately contributing to improved women's health and well-being.

2. Data Collection and Features:

Our project incorporates several key features that make it a valuable tool for women's health. The chatbot interface allows users to easily input their health information, symptoms, and concerns in a private and user-friendly manner. The integration of machine learning, specifically logistic regression, enables personalized risk assessments for PCOS, endometrial cancer, menopausal symptoms, and UTIs, providing users with data-driven insights into their health status. Furthermore, the project emphasizes user education by providing information on risk factors and preventive measures, empowering women with knowledge to make informed decisions about their health. Privacy and data security are prioritized to ensure the confidentiality of user information. This project's scalability, research potential, and ability to collaborate with healthcare professionals make it a promising and versatile tool for women's health, with the capacity to positively impact a broad audience and contribute to early detection and improved well-being.

Feature Importance:
The selection and importance of features in your machine learning model for risk assessment in women's health are pivotal to the model's performance and interpretability. Each feature represents a piece of information that contributes to the model's ability to make accurate predictions. Features related to symptoms and health data are at the core of your

risk assessment, and their significance varies depending on their correlation with the specific health condition being assessed.

- Symptoms like abnormal vaginal bleeding and age at menopause can be highly influential in assessing endometrial cancer risk.
- Age is a common and essential feature, serving as a strong predictor for conditions like menopausal symptoms or endometrial cancer.
- Medical history, family history, and demographic factors can also play a critical role in the risk assessment. These features provide valuable context and contribute to the model's ability to tailor risk assessments to individual users.
- Lifestyle and behavioral factors, such as diet, exercise, and BMI, as well as hormonal and reproductive factors, like menstrual irregularity and hormone use, are of paramount importance for conditions like PCOS and menopausal symptoms.
- In sum, the choice and importance of features are fundamental in ensuring the accuracy and effectiveness of your model in assessing health risks for women, and they help provide tailored and personalized recommendations for risk reduction and early detection.

3.8 Conclusion and Future work.

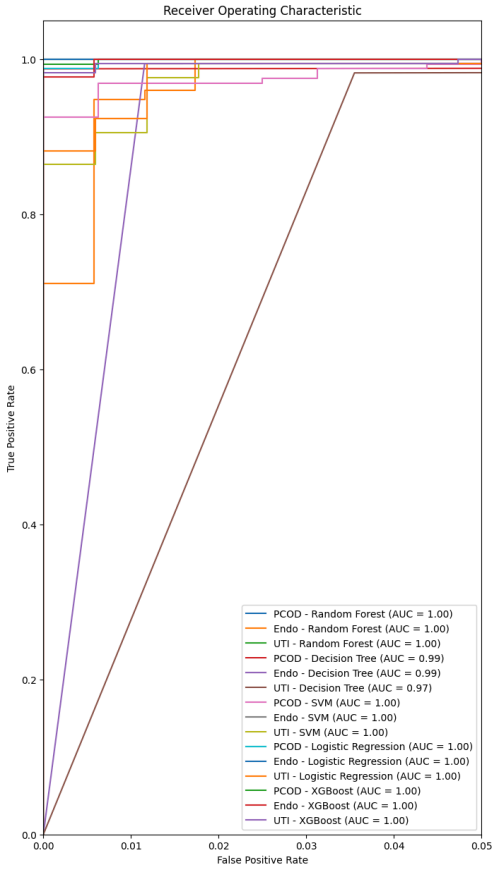


Figure 3.8.1

We have successfully understood the different types of hormonal issues faced by women and developed a correlation between the symptoms of the disorders and their possibility to occur . In our research work, we have used different Machine Learning models to predict hormonal issues like PCOD/PCOS, Endometriosis and UTI. The Machine Learning models used by us are - Decision Tree Classifier, Random Forest Classifier, Logistic Regression, Support Vector Machine(SVM) and xgBoost. We conducted a comprehensive analysis of Receiver Operating Characteristic curves and prediction probabilities from various machine learning models on our cleaned datasets. Figure 3.8.1 shows the combined graph of the Receiver Operating Characteristic curves we got from our analysis.

REFERENCES

Knowledge Gathering

[1] Do PCOD and PCOS mean the same thing or are they different | UNICEF India

[2] cs.huji.ac.il/~shais/UnderstandingMachineLearning/copy.html

[3] https://www.medicalnewstoday.com/articles/319363#symptoms-and-diagnosis

[3]https://www.everydayhealth.com/menopause/at-what-age-will-you-enter-menopause.aspx

Journal Papers

[1] H. Elmannai et al., “Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence,” Diagnostics, vol. 13, no. 8, p. 1506, Apr. 2023, doi: https://doi.org/10.3390/diagnostics13081506.

[2]D.-C. Ding, W. Chen, J.-H. Wang, and S.-Z. Lin, “Association between polycystic ovarian syndrome and endometrial, ovarian, and breast cancer,” Medicine, vol. 97, no. 39, p. e12608, Sep. 2018, doi: https://doi.org/10.1097/md.0000000000012608.

[3]K. J. Ruddy et al., “Menopausal symptoms and fertility concerns in premenopausal breast cancer survivors,” Menopause, vol. 18, no. 1, pp. 105–108, Jan. 2011, doi: https://doi.org/10.1097/gme.0b013e3181ef39f8.

[4]M. R. Ataollahi, J. Sharifi, M. R. Paknahad, and A. Paknahad, “Breast cancer and associated factors: a review,” Journal of medicine and life, vol. 8, no. Spec Iss 4, pp. 6–11, 2015.

[5]J. A. Barry, M. M. Azizia, and P. J. Hardiman, “Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis,” Human Reproduction Update, vol. 20, no. 5, pp. 748–758, Mar. 2014, doi: https://doi.org/10.1093/humupd/dmu012.